

Thermal unfolding dynamics of proteins probed by nonlinear infrared spectroscopy

by

Hoi Sung Chung

M. S. Seoul National University, 2000

B. S. Seoul National University, 1998

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

February 2007

© Massachusetts Institute of Technology 2007. All rights reserved.

Signature of Author

Department of Chemistry

January 10, 2007

Certified by

Andrei Tokmakoff

Associate Professor of Chemistry

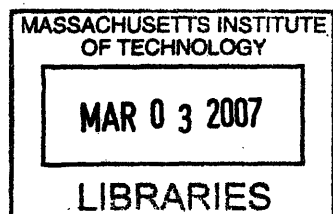
Thesis Supervisor

Accepted by

Robert W. Field

Haslam And Dewey Professor Of Chemistry

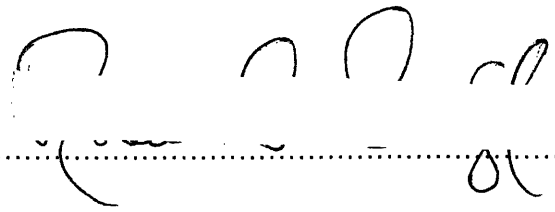
Chairman, Departmental Committee on Graduate Students



ARCHIVES

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

Professor Robert W. Field
Chair

Professor Robert G. Griffin


Professor Andrei Tokmakoff
Thesis Supervisor

Thermal unfolding dynamics of proteins probed by nonlinear infrared spectroscopy

by

Hoi Sung Chung

Submitted to the Department of Chemistry
on January 10, 2007 in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis presents spectroscopic approaches to study the thermal unfolding dynamics of proteins. The spectroscopic tool is nonlinear infrared (IR) spectroscopy of the protein amide I band. Among various nonlinear IR techniques, two-dimensional infrared (2D IR) spectroscopy, which is an IR analogue of 2D NMR, is the most informative. A 2D IR spectrum is obtained from a double Fourier transform of the heterodyned third-order nonlinear signal, which is generated by three consecutive interactions between femtosecond IR pulses and the vibrations of the system. This technique is sensitive to the presence of β -sheet structure in proteins through the formation of cross peaks between the two characteristic vibrational modes of β -sheets. In this work, 2D IR spectroscopy is used to measure equilibrium thermal unfolding of ribonuclease A and ubiquitin. For transient unfolding studies, the temperature of the solution is rapidly raised by a nanosecond temperature jump (T-jump) laser, which is followed by probing structural changes of proteins with dispersed vibrational echo (DVE) spectroscopy or 2D IR spectroscopy. DVE spectroscopy is a homodyne measurement of the third-order signal, in which the spectrum is related to a projection of a complex 2D IR spectrum onto one of the frequency axes (ω_3). In spite of its reduced dimension measurement, DVE spectroscopy is sensitive enough to be utilized in transient probing with less experimental challenges. From transient thermal unfolding studies of ubiquitin probed by DVE spectroscopy, complicated non-exponential relaxations are observed on the microsecond timescale, which are followed by ms unfolding. Non-exponential relaxation is interpreted as downhill unfolding of a transient species populated around the top of a barrier (transition state) due to the barrier shift caused by a rapid T-jump. Variations in the unfolding transition state of ubiquitin are further investigated with temperature-dependent T-jump experiments and mutation studies. Experimental conclusions are supported by calculations of the unfolding free energy surfaces using statistical mechanical modeling. T-jump 2D IR spectroscopy is also performed to remove ambiguities in the projected domain of DVE spectroscopy and provide a new spectroscopic measure for transient unfolding through line-broadening analysis.

Thesis Supervisor: Andrei Tokmakoff

Title: Associate Professor of Chemistry

To Mother, Father, and Kyung-sun

Acknowledgements

Perhaps, one of the biggest changes in my life was to start my PhD study at Boston where I had never been before. Five and half years later, now, I just feel like waking up from a long dream. Sometimes the dream seemed scary and bitter, but I feel that it ends up with a sweet one. To finish my study, I have been beholden to many people around me. First of all, I would like to thank my advisor, Professor Andrei Tokmakoff for his solid supports and guidance. He is positive-minded and has always encouraged me whenever I was in trouble in experiments, projects, and everything. Also, his generosity, probably coming from his tallness, covered all embarrassing moments brought by my silliness.

Probably, it was my best luck to join the Tokmakoff group and meet a bunch of nice people. There are two subgroups: the protein group and so called “Team Water.” I joined the protein group and started to work closely with Nuri Demirdöven and Munira Khalil, an amazing research duo. It was very hard to follow their pace at the first moment but I learned a lot of things that I should know in the lab through working with them. Especially, I worked with Munira for about six months before she left. Her last project, which was my first project, was building a temperature jump instrument. I can say that her wise decisions at several crucial stages of the development speeded up the realization of the tough experiment. I owe many works in this thesis to her preceding help. Dr. Chris Cheatum was also a member of the protein subgroup but his knowledge covers both projects and I could learn many sciences from him.

There were three guys in the water project: Chris Fecko, Joel Eaves, and Joe Loparo. Although topics were different, their scientific inputs into my research are valuable. Particularly, Joe was my year but he joined the group much earlier than I, and had already known many things in advance. Sometimes, this made him take more burdens, which should have been shared by me. I would like to thank him for this.

My graduate life became more fruitful by other members who joined the group afterwards. Dr. Matt DeCamp’s advices and comments based on his deep understanding of science were really helpful in polishing my work. I cannot omit all discussions with members of protein group, Adam Smith, Ziad Ganim, and Kevin Jones. Their assistances and comments in experiments and theoretical works spread over my accomplishments. I have also enjoyed Kevin’s ice cream. Every moment spent with Lauren DeFlores, Sean Roberts, Rebecca Nicodemus, Poul Petersen, and Josh Lessing were also beneficial to me.

For all administrative help, I would like to thank Anne Hudson. Her professional and kind assistances have made my studies easier and more efficient. For advices and academic supports, I would also like to thank Professor Robert W. Field and Professor Robert G. Griffin.

I owe many things to my friends outside school. I thank my old friends (and roommates), Bum Suk Zhao and Hong Myung Lee for sharing many funs and difficulties. Delicious food cooked by Taeho Shin nourished my spoiled stomach. Most of all, I cannot imagine my life here for five and half years without advices and guidance of Inhee Chung as a former settler in the Boston metro area. Her positive thoughts, encouragements and cheers have made favorable changes in my graduate life as well. I am really thankful for the great friendship with her.

Finally, I would like to thank my parents Se-Yang Chung and Myo-Sun Shin and my sister Kyung-Sun. Their endless supports, love, and prayers have made me as I am. I dedicate my small achievements here to them.

Contents

List of Figures

List of Tables

1	Introduction	16
1.1	Protein folding problem: Folding paradigms	16
1.2	Experimental approaches	18
1.2.1	Two-state folder: Fast folding and unfolding dynamics	18
1.2.2	Folding and unfolding dynamics	19
1.2.3	Biological relevance of in vitro folding	20
1.3	Initiation and probe of fast unfolding dynamics of proteins	20
1.3.1	Initiation method: Laser temperature jump	20
1.3.2	Probing method: Infrared spectroscopy	21
1.3.3	Nonlinear infrared spectroscopy	22
1.3.3.1	New structure-sensitive probe	22
1.3.3.2	Amide I vibrational spectroscopy	26
1.3.3.3	Application to the anti-parallel β sheet	27
1.4	Thesis outline	28
	References	29
2	Visualization and characterization of the IR active amide I vibrations of proteins	34
2.1	Introduction	34
2.2	Methods	36
2.2.1	Calculation of amide I eigenstates and IR spectrum	36
2.2.2	Doorway state calculation	38
2.2.3	Visualization of amide I vibrations	41
2.3	Results and Discussion	41
2.3.1	Characterization of the vibrations of secondary structure	41
2.3.2	Spatial Correlation Functions	43
2.3.3	IR spectra	47
2.3.4	β -Protein: Concanavalin A	49

2.3.5	α -Protein: Myoglobin	52
2.3.6	α/β -Protein: Ubiquitin	54
2.4	Conclusions	56
	References	59
3	Projection relationships in Nonlinear infrared spectroscopy	61
3.1	Third-order response function of 2D IR spectroscopy	62
3.2	Projection relationships	64
	References	67
4	Experimental setups and data analyses	68
4.1	Generation of infrared pulses	68
4.2	Time delay control and generation of the third order signal	70
4.3	2D IR and DVE spectroscopy	72
4.4	T-jump laser	73
4.5	Control of the delay between T-jump and probe pulses	74
4.6	Collection of transient data	76
4.7	Temperature controlled Sample cell	78
4.8	Calculation of the temperature relaxation	78
4.9	Balanced detection	80
4.10	Future directions	82
4.11	Singular value decomposition	83
	4.11.1 Equilibrium experiments	84
	4.11.2 Transient experiments	84
	References	85
4-A	Appendix: Drawing of brass sample cell	86
4-B	Appendix: Solving heat diffusion equation	87
5	Nonlinear IR spectroscopy of conformational change of RNase A during thermal unfolding	91
5.1	Introduction	91
5.2	Experimental	93
5.3	Results and discussion	94
	5.3.1 2D IR spectroscopy of proteins with antiparallel β -sheets	94

5.3.2	FTIR spectra	96
5.3.3	2D IR spectra	98
5.3.4	Dispersed nonlinear signals: Pump-probe and vibrational echoes	101
5.3.5	SVD analysis of spectral changes	103
5.4	Conclusions	107
	References	109
5-A	Appendix: Thermodynamics of two-state folding from SVD analysis	111
6	Conformational changes during the nanosecond to millisecond unfolding of ubiquitin	113
6.1	Introduction	113
6.2	Methods	114
6.3	Results	115
6.3.1	Equilibrium thermal unfolding	116
6.3.2	Transient spectral change	117
6.4	Discussion	119
	References	126
7	Temperature dependent unfolding dynamics of wildtype ubiquitin	128
7.1	Introduction	128
7.2	Methods	130
7.3	Results	131
7.3.1	Thermodynamic analysis	131
7.3.2	Ultrafast responses of the solvated region	132
7.3.3	Transient spectral changes of ubiquitin unfolding	135
7.3.4	Analysis of unfolding in ν_{\perp} region	136
7.3.5	Unfolding monitored in other frequency regions	142
7.4	Discussion	144
7.4.1	Population change during temperature relaxation	144
7.4.2	Deviation from the two-state kinetics	147
7.5	Conclusion	150
	References	152

8	Calculation of free energy surface of ubiquitin unfolding	154
8.1	Introduction	154
8.2	Calculation methods	156
8.2.1	The Munõz-Eaton model	156
8.2.2	Beyond the ME model: Consideration of partially folded states	162
8.3	Results and discussion	167
8.3.1	Determination of the empirical parameters	167
8.3.2	Unfolding pathways	169
8.3.3	Free energy changes by T-jump	171
8.3.4	Orthogonal coordinates	174
8.4	Limitation of the ME model and its improvement	176
	References	177
8-A	Appendix: Exact solution of the ME model	179
8-B	Appendix: Calculation of free energy surface	181
8-C	Appendix: Calculation of folding probability on FE	184
9	Transition state of ubiquitin unfolding explored by mutation studies	189
9.1	Introduction	189
9.1.1	Background: Φ value analysis	190
9.2	Experimental	191
9.2.1	Overview of the mutants	192
9.3	Results	194
9.3.1	Equilibrium thermal unfolding	194
9.3.2	Transient thermal unfolding	195
9.3.2.1	Overview	195
9.3.2.2	Mutant <i>l</i>	196
9.3.2.3	Mutant <i>l</i>	199
9.3.2.4	Mutant <i>g</i>	201
9.4	Discussion	202
9.4.1	Transition state	202
9.4.2	Fast unfolding phase of mutant <i>l</i>	205
9.5	Conclusion	206
	References	208

10	Transient 2D IR spectroscopy of ubiquitin unfolding	209
10.1	Introduction	209
10.2	Experimental	211
10.3	Results and discussion	214
10.3.1	Equilibrium differences	214
10.3.2	Ultrafast responses	215
10.3.3	Transient thermal unfolding of ubiquitin	217
10.3.4	Comparison with DVE results	220
10.3.5	Dynamics monitored by diagonal and off-diagonal regions	222
10.3.6	Homogeneous broadening during thermal unfolding of protein	224
10.3.7	Polarization effect	227
10.4	Concluding Remarks	229
	References	230

List of Figures

1-1	Free energy curve and populations of folded and unfolded species before and after temperature jump.....	19
1-2	2D IR spectroscopy	23
1-3	Lineshapes in 2D IR spectrum	25
1-4	2D IR spectra of proteins with different β -sheet contents.....	27
2-1	Transformation from site basis to doorway states.....	39
2-2	Projection of the antiparallel β sheet on a plane.....	42
2-3	Amide I oscillator correlation functions and their Fourier transforms	44
2-4	Experimental and calculated infrared spectra of proteins	48
2-5	Visualization and characterization of the ν_{\perp} and ν_{\parallel} doorway modes for concanavalin A.....	50
2-6	Visualization and characterization of three doorway modes for myoglobin.....	53
2-7	Phase correlation of vibrations of ubiquitin	56
3-1	Projection relationships between the 2D IR spectrum, pump-probe, and DVE spectra....	66
4-1	Experimental layout. Layout of the laser components, detector, controller, and flowchart of the frequency division are shown.....	69
4-2	Typical characteristics of a femtosecond IR pulse	70
4-3	Enlarged illustration of the sample cell region.....	71
4-4	Control of the time delay.....	75
4-5	Detection scheme of transient data.....	77
4-6	Calculation and measurement of the temperature relaxation in the sample cell	79
4-7	Balanced heterodyne detection.....	81
5-1	The linear FTIR and 2D IR correlation spectra of poly-L-lysine and RNase A.....	94
5-2	Temperature-dependent amide I FTIR spectra of the thermal denaturing of RNase A.....	97
5-3	Temperature-dependent 2D IR correlation spectra \tilde{S}'_C of RNase A obtained in the crossed-polarization geometry	98
5-4	2D IR difference spectrum	99

5-5	Absolute value non-rephasing spectra of RNase A for the native and thermally denatured states obtained in the crossed-polarization geometry	101
5-6	Temperature-dependent dispersed pump-probe and dispersed vibrational echo spectra	102
5-7	Difference spectra between the highest and lowest temperature and the second SVD component spectra of FTIR, DVE, and DPP	105
5-8	Thermal melting curves obtained from the scaled second components of the FTIR, DVE, DPP, and 2D-IR spectra	106
6-1	Temperature-dependent IR spectroscopy of ubiquitin unfolding	115
6-2	DVE difference spectra normalized to the peak signal at several time delays from the T-jump	118
6-3	Two-dimensional projection of the β -sheet and calculation of IR spectrum	120
6-4	An illustration of a reduced free energy surface	122
7-1	Temperature dependent T-jump experiment	129
7-2	Equilibrium thermal unfolding of ubiquitin	132
7-3	Comparison of transient changes of NMA and ubiquitin	134
7-4	Temperature dependent transient thermal unfolding of ubiquitin	136
7-5	Comparison of the equilibrium DVE difference spectra (blue) and the first SVD component spectra (red) of the transient data	138
7-6	Extracting relaxation parameters	139
7-7	Spectral relaxations of the $\nu_{ }$ and random coil region	143
7-8	Folded population changes after a T-jump	146
8-1	Possible states in the ME model. Each residue is allowed only to have a folded (F) or unfolded (U) state	156
8-2	1D free energy curve of the ME model	158
8-3	Comparison of the approximated and exact calculation by 2-D free energy surfaces at 58°C	162
8-4	Linear sequence, structure and the β -sheet registry of ubiquitin	163
8-5	Addition of native interactions between different native stretches	165
8-6	Comparison of the hybrid model	167
8-7	Determination of empirical parameters by comparing with experimental data	168

8-8	Folding probability of the residues in the β sheet at several locations on the free energy surface	170
8-9	Changes in free energy surface and populations	172
8-10	Projection of the unfolding trajectories of MD simulations onto the free energy surface.....	174
8-11	Free energy surfaces projected onto the number of folded residues in the α helix.....	175
9-1	Two extreme cases of the Φ value analysis.....	190
9-2	Projection of the β -sheet registries of ubiquitin wildtype and mutants	192
9-3	Equilibrium thermal unfolding of three mutants	194
9-4	Transient DVE difference spectra of mutants	196
9-5	Relaxations of mutant <i>l</i>	198
9-6	Experimentally obtained population relaxations of the folded species of mutant <i>l</i>	199
9-7	Relaxations of mutant <i>g</i>	201
9-8	Comparison of the changes of the free energy surface of the wildtype and mutants by T-jump.....	204
10-1	Beam alignment and data processing for transient 2D IR spectroscopy	212
10-2	Equilibrium thermal unfolding of ubiquitin monitored by 2D IR spectroscopy	214
10-3	Transient 2D IR spectra of NMA and ubiquitin.....	216
10-4	Transient 2D IR spectra after initiated by a T-jump from 63°C to 72°C	217
10-5	Transient changes monitored by slices	219
10-6	Comparison of DVE spectra measured and reconstructed from 2D IR spectra	221
10-7	SVD analyses for different frequency blocks of transient 2D IR difference spectra	223
10-8	Changes in anti-diagonal width.....	225
10-9	Comparison of different polarization geometries	228

List of Tables

4-1	Thermal properties of D ₂ O and CaF ₂	78
7-1	Relaxation parameters extracted from the SVD component of the v_{\perp} region	140
7-2	Temperature dependent relaxations in the two-state kinetic regime	142
7-3	Relaxation parameters extracted from the first SVD component of the v_{\parallel} and random coil regions	144
9-1	Thermodynamic parameters of wildtype and mutant ubiquitin.....	195
9-2	Relaxation parameters extracted from the SVD component of the v_{\perp} region of mutant i	197
9-3	Relaxation parameters extracted from the first SVD component of the v_{\parallel} and random coil regions of mutant i	197
9-4	Global stretched exponential fit of the unfolding part.....	200
9-5	Φ_m values for mutant i and l	202

Chapter 1

Introduction

1.1. Protein folding problem: Folding paradigms

Proteins are one of the essential constituents for living organisms. Although the information about the function of proteins is genetically encoded in their amino acid sequence, this function requires a folding of the polypeptide chain into a unique three-dimensional structure. In fact, a long-standing and important question in the biological sciences is how this biopolymer eventually finds its folded structure. By Anfinsen's "thermodynamic hypothesis,"¹ a given amino acid sequence folds spontaneously and finds a conformation having the lowest free energy. From a kinetic perspective, however, this hypothesis is not easily rationalized. As Levinthal pointed out,² a protein has too many conformations to be searched even if we allow only three possible states for each amino acid residue (α , β , and random conformations). Hence, it is impossible for a protein to find a unique structure on a biologically relevant timescale by a random search. Thus some guiding mechanism based on physical principles must be responsible. How fast proteins can fold is a crucial problem because folding and aggregation of proteins are in competition with each other after their synthesis in the cell and misfolding results in a disease.^{3,4}

One important unresolved aspect of the problem is whether there are general mechanistic principles that underlie the folding of all proteins. It is a widely held belief

that common physical interactions within all protein chains and with the surrounding water are an indication that common mechanistic principles will hold broadly for all proteins or at least for certain classes. As a result, several general paradigms for protein folding have developed over the many years that this problem has been studied. Models differ from each other in the description of how the number of conformations is reduced in the initial collapse. At one extreme of the spectrum, the hierarchical *framework model*^{5,6} states that folding starts with the formation of native-like secondary structures. Then, fine tertiary-contacts are formed by following dockings of these pre-formed secondary structures. The other extreme is the *hydrophobic collapse model*,^{7,8} in which a burial of hydrophobic residues induces a non-specific collapse and rearrangements and fine-tunings occur afterwards. However, most of proteins obey an intermediate (or mixed) pathway, which is described by the *nucleation-condensation mechanism*.^{9,10} In this model, the partially formed secondary structures offer an extended folding nucleus. In this nucleus, however, secondary structures are not stable enough without partial tertiary contacts.

From a statistical mechanical perspective, protein folding can be conceptually understood by the free energy landscape theory. In the theory called the ‘New View,’ a folding free energy landscape is smooth and steep towards a native conformation like a funnel,¹¹⁻¹⁶ which is known as the “principle of minimum frustration.” In this theory, the number of conformations to be searched is significantly reduced by rapid collapses, which is followed by minor re-orderings and activated barrier crossings to find a native state.

Then, how can these questions be best addressed by experimentalists? We believe that monitoring the time-evolution of protein structure after preparing it in a non-equilibrium environment would provide a unique description of folding. For this kind of experiment, initiation of folding faster than the dynamics and a structure-sensitive probe are needed. This thesis focuses on these topics as described below.

1.2. Experimental approaches

1.2.1. The two-state folder: Fast folding and unfolding dynamics

Initially it was believed that protein folding involves sequential processes consisting of a series of intermediate states. Therefore, studies had focused on the characterizing those intermediate states. However, this view was modified by the emergence of two-state folders: proteins characterized only by two dominant free energy minima for a folded and unfolded state. This suggested a simple folding principle, which can be described by the two-state kinetics based on the Kramers' theory¹⁷

$$k_f = k_o \exp\left[-\frac{E_a}{RT}\right]. \quad (1.1)$$

Here, k_f is the folding rate constant, k_o is its upper limit, and E_a is a barrier height. The presence of two-state folders suggests that folding mechanism can be understood by characterizing the two-state folding transition state ensemble. Combined with a point mutation using genetic recombination techniques, the transition state can be mapped at the residue level using the Φ value analysis. The Φ value, a ratio of the change in the folding activation free energy to the change in the folding free energy, ranges from 0, when a residue is not involved in the transition state, to 1, when it is fully structured in the transition state.⁹ Discoveries of a diffuse but native-like structure of the transition state were the basis for the nucleation-condensation mechanism.

Two-state folders have also provided the fastest folding timescale that small proteins can reach. Folding times of individual structural units have been measured by a design of simple proteins or peptides having particular secondary structures. Isolated simple secondary structures show ultrafast folding, the timescale of which ranges from tens or hundreds of ns for the loop¹⁸ and the α helix^{19,20} to several μ s for β hairpin.^{21,22} Small α helical proteins of tens of residues fold in several μ s^{23,24} while β sheet proteins of similar size such as the WW domain fold in tens of μ s.²⁵⁻²⁷ (More examples and their timescales can be found in Ref. 28.) Usually, α helical proteins fold about ten times faster than β sheet proteins because of their small contact order²⁹ (average separation in primary sequence between residues in contact). Based on the previous experimental and

theoretical results, the speed limit for a small N -residue polypeptide ($N \sim 100$) is known to be $N/100 \mu\text{s}$.²⁸

1.2.2. Folding and unfolding dynamics

Most protein folding experiments described above, however, have focused on acquiring kinetic information such as the rate of appearance or disappearance of an experimental signature for a particular species, typically on millisecond or longer time scales. These results give information on energetic barrier heights much greater than thermal energy (kT), but say little about how structure changed along the path. As the barrier height in Eq. (1.1) becomes smaller, the folding rate approaches its speed limit (k_0) and folding becomes a downhill process rather than an activated process.³⁰⁻³³ A number of fast folding experiments of proteins and peptides have shown that downhill folding can be initiated with a nanosecond temperature jump (T-jump) as illustrated in Fig. 1-1. Such experiments work in a diffusive regime that allows a freer exploration of available structures, and provide evidence that the relevant molecular time scales for folding is nanoseconds to microseconds.^{20,30,34,35} Of particular interest is obtaining meaningful information on the underlying molecular dynamics of folding by probing transient species directly when downhill folding is initiated and followed with a structure-sensitive probe.

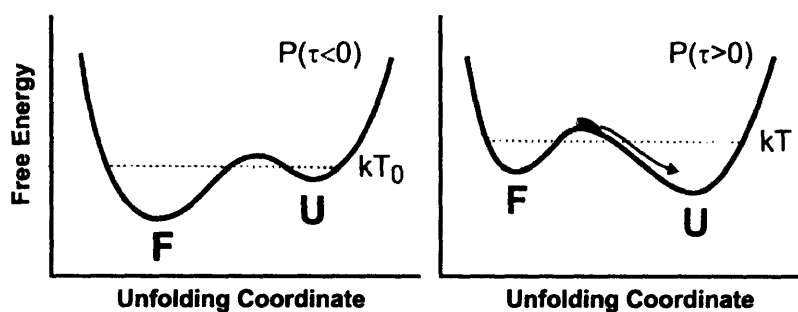


Figure 1-1. Free energy curve and populations of folded and unfolded species before and after a temperature jump from T_0 to T . F and U represent the folded and unfolded states, respectively. Barrier-free unfolding is induced by shifting the position of the barrier towards the folded state.

This thesis reports on such an experiment, a conformationally sensitive probing of downhill unfolding of proteins over nanosecond to millisecond time scales. The dynamics is initiated by a nanosecond T-jump and probed with various structure-sensitive nonlinear infrared spectroscopies. A fast temperature change allows for a small amount of folded species to be populated around the activation energy barrier on the free energy surface. Using time-evolution of these species, the transition state of unfolding is mapped. A small 76-residue protein, ubiquitin, is appropriate for this purpose because it is a fast two-state folder and has all secondary structural motifs.

1.2.3. Biological relevance of *in vitro* folding

One of the concerns of the physical or chemical approaches in this field is their biological relevance. From the biological perspectives, *in vitro* folding might be unrealistic because its folding environment is too simple compared to the crowded intracellular environment. For example, in a live cell, proteins are synthesized sequentially from N- to C-termini and folding may start from N-terminus with attached to the ribosome. However, after synthesized, most proteins are transported into the mitochondria or endoplasmic reticulum (ER) through a membrane, and undergo unfolding and folding cycle again. Then, the folding event in these compartments should be similar to *in vitro* folding. Also, many proteins interact with chaperone proteins that protect protein aggregations.^{36,37} However, chaperone mediated folding is an ATP consuming process, and the ATP hydrolysis takes many seconds. Therefore, involving of chaperone proteins is thought to be in the late phase of folding and the interaction with small and fast folding proteins would be minor.³⁸

1.3. Initiation and probe of fast unfolding dynamics of proteins

1.3.1. Initiation method: Laser temperature jump

The laser temperature jump (T-jump) technique has been one of the fastest methods to initiate protein thermal unfolding or folding by raising the temperature of the solution.^{3,39-43} Since a direct excitation of the overtone stretching vibrations of the O-H ($\lambda = 1.5 \mu\text{m}$) or O-D ($\lambda = 2.0 \mu\text{m}$) of the solvent molecules followed by a rapid vibrational

relaxation induces the temperature change, the timescale is not limited by any mechanical process as are in the rapid mixing experiments, the dead time of which is still tens of microseconds.⁴⁴ The vibrational relaxation within roughly 10 – 100 ps,⁴² induces a T-jump virtually on the same timescale of the laser pulse duration, which is usually less than 10 ns. This rapid T-jump offers a unique opportunity to probe secondary structure dynamics that occur on the ns (α helix) to microsecond (β hairpin) timescale. In addition, the laser T-jump does not require any chemical denaturant or engineering of the sample that are usually needed in most other optical initiation techniques such as the breakage of the cyclized peptide,⁴⁵ pH jump,^{46,47} photodissociation of a ligand,⁴⁸ cis-trans isomerization,^{49,50} or photolysis of the engineered disulfide bridge.⁵¹ As a result, the laser T-jump technique has been widely used for the studies of rapid folders such as small proteins^{23,30,31,34,52,53} and peptides.^{20,21,27,54-57}

1.3.2. Probing method: Infrared spectroscopy

The experimental study of protein folding involves the characterization of a heterogeneous ensemble of structures over timescales ranging from picoseconds to seconds. The vast range of length and time scales involved makes the direct observation of structural coordinates difficult, and ensures that no single experimental technique can capture structural changes at high spatial resolution in solution over all time scales.^{39,58-62} Multi-dimensional nuclear magnetic resonance (NMR) spectroscopy provides an atomic level structural tool in solution,^{63,64} but its use for following the folding of proteins has been limited to time scales from ms to seconds.⁶³ Advances in time-resolved x-ray diffraction promises to reveal the details of time-dependent structural change with atomic resolution in crystals,⁶⁵⁻⁶⁷ and x-ray scattering methods in solution give more limited structural information.⁶⁸ The inherently fast timescales of high frequency spectroscopic techniques in the optical and infrared regimes make them appealing tools, although they lack the atomistic detail of X-ray and NMR. Nonetheless, time-resolved fluorescence,^{3,39,41} infrared,^{49,69,70} Raman,⁷¹ and UV circular dichroism⁷²⁻⁷⁷ have been used with considerable success to reveal folding and denaturing processes in proteins on picosecond to millisecond time scales.

Of these latter methods, infrared spectroscopy is appealing because of its structural sensitivity to nuclear degrees of freedom, its picosecond intrinsic measurement time scale, and no requirement of chromophores or prosthetic groups. IR spectroscopy is sensitive to protein structure through vibrational resonances arising from the polypeptide backbone or side chains. The protein structure, local interactions such as hydrogen bonding, and the electrostatic and mechanical couplings between its various vibrational degrees of freedom determine the resonance frequencies observed in the infrared spectrum. Consequently, the infrared spectrum can be used to reveal structural information about the protein. Even with these potential advantages, traditional protein IR spectroscopy is a largely qualitative method for two primary reasons. First, IR spectra of proteins are generally broad and featureless due to overlapping contributions from many oscillators, making assignment and modeling ambiguous. Also, the quantitative relationships between vibrational resonance frequencies and structure, as encoded in the various couplings between vibrations, are not understood well enough to allow a clear interpretation. However, these topics are presently receiving considerable attention, and there is a rapid pace of development from both the experimental and theoretical perspectives,⁷⁸⁻⁸⁷ including the recent theoretical and experimental efforts to understand how 2D IR spectroscopy can be used to overcome some of the disadvantages of traditional IR spectroscopy of protein amide I transitions.

1.3.3. Nonlinear infrared spectroscopy

1.3.3.1. New structure-sensitive probe

Nonlinear IR spectroscopy has been useful to investigate ultrafast dynamics in the condensed phase on fs to ps timescales.^{83,88-92} In addition, its application can be extended to overcome some of the complications of traditional linear IR spectroscopy. Multiple interactions between the system and a sequence of three incident pulses make it possible to interrogate not only the first excited state but also the second excited state of the vibrational ladder, which gives additional information about the coupling and orientation between transition dipoles, anharmonicities, heterogeneity of the system.⁸⁸ Among various techniques, two-dimensional infrared (2D IR) spectroscopy, which is an infrared

analogue of 2D NMR, provides the most information on vibrational couplings through the formation of cross peaks.^{83,88,89}

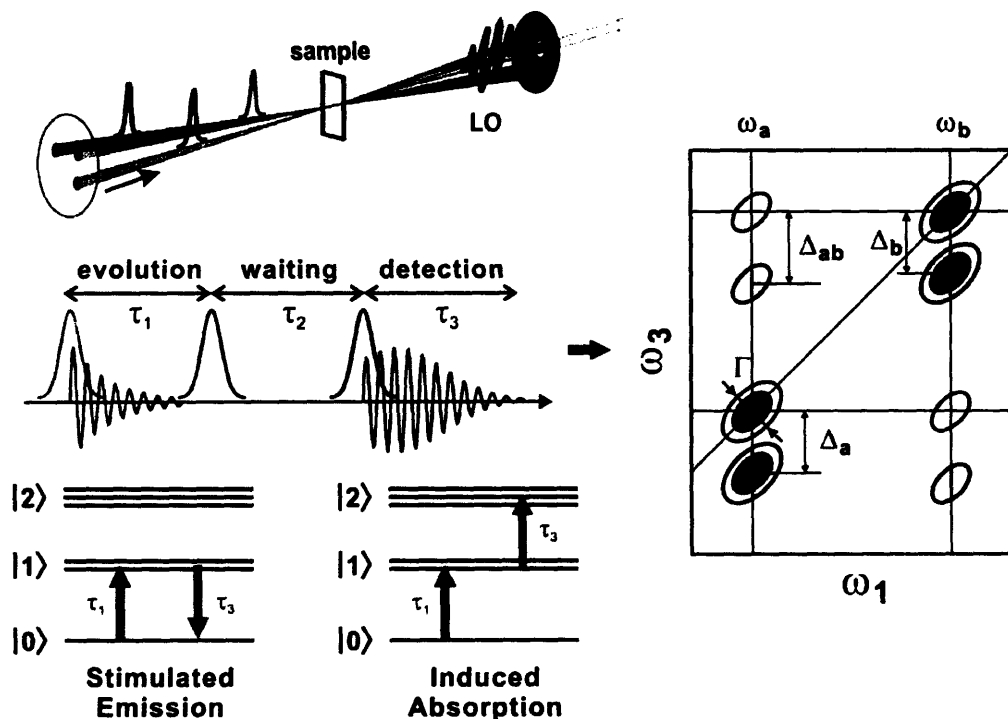


Figure 1-2. 2D IR spectroscopy.

The basic experimental setup, pulse sequence, interactions of pulses with vibrational states, and a typical 2D IR spectrum for a coupled-oscillator system are illustrated in Fig. 1-2. The third-order signal is generated by a sequence of three femtosecond infrared pulses impinging on the sample. A coherent state of the ground and first excited states prepared by the first pulse, evolves during the evolution period, τ_1 until quenched by the second pulse. This process is shown with a black arrow in the vibrational ladder. After a waiting period τ_2 , the third pulse prepares another coherent state. At this time, two different kinds of transition are possible between the ground and the first excited states and between the first and second excited states. The latter is called the ‘induced absorption’ (blue arrow) and the former is called the ‘stimulated emission.’ (red arrow) This coherent state (third-order polarization) evolves again and an electromagnetic wave (the third-order signal) is emitted to the wave-vector matched

direction ($\mathbf{k} = -\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c$), which is determined by the propagation vectors of three incoming pulses as shown in Fig. 1-3a. The phase of the third-order signal oscillates with both τ_1 and τ_3 and its double Fourier transform gives a 2D IR spectrum. The ω_3 axis of the 2D IR spectrum is obtained by a spectral interferometry in τ_3 domain after overlapping the third order signal with an external local oscillator (LO) field, while the transform in τ_1 domain is performed by a numerical Fourier transform of the data collected as a function of τ_1 .

Information that we can obtain from a 2D IR spectrum of a coupled oscillator system is summarized in the right panel of Fig. 1-2. First, the two eigenfrequencies (ω_a , ω_b) are obtained from the locations of the positive peaks resulting from the stimulated emission process (red) in the diagonal axis. Induced absorption peaks are located below the positive peaks, and the anharmonicities (Δ_a , Δ_b) of the vibrational potential can be obtained from this red shift. The sign of the induced absorption peak is negative because of the phase difference of π between the signals by the stimulated emission and induced absorption. Also, from the splitting of the positive and negative cross peaks (Δ_{ab}), the magnitude of the coupling between the two oscillators can be extracted. Assuming the same anharmonicity of the two oscillators ($\Delta_a = \Delta_b = \Delta$), the coupling strength β is related with the splitting as $\Delta_{ab} \approx 4\beta^2\Delta/|\omega_a - \omega_b|^2$.

One unique feature of the 2D IR spectrum can be found through lineshape analysis. In the linear spectrum, the line-broadening mechanism is hardly distinguishable. However, in the 2D IR correlation spectrum, the inhomogeneity can be measured by a comparison of the diagonal ($\omega_1 = \omega_3$) and anti-diagonal ($\omega_1 + \omega_3 = \text{constant}$) widths of a peak. The anti-diagonal is related to the homogeneous linewidth, whereas the diagonal is related to the inhomogeneous distribution (convolved with the homogeneous linewidth).

A 2D IR correlation spectrum is the sum of the rephasing and non-rephasing spectra, which differ by the pulse ordering as shown in Fig. 1-3. Due to the phase conjugation in τ_1 period caused by the different pulse orderings, the spectrum is tilted along the diagonal and anti-diagonal direction for the rephasing and non-rephasing spectrum, respectively. When the system is homogeneous, the amplitude of rephasing and

non-rephasing spectra is same and the correlation spectrum shows a symmetric absorptive shape. On the other hand, in the presence of the inhomogeneity, the rephasing spectrum is more intense than the non-rephasing spectrum because the echo signal (recurrence of vibrational coherences) during τ_3 period makes the signal last longer in the rephasing measurement. This intensity mismatch induces a tilt along the diagonal axis in the correlation spectrum and the inhomogeneity can be measured from the ellipticity of the peak. Also, regardless of the tilt, anti-diagonal width always gives the homogeneous linewidth Γ .

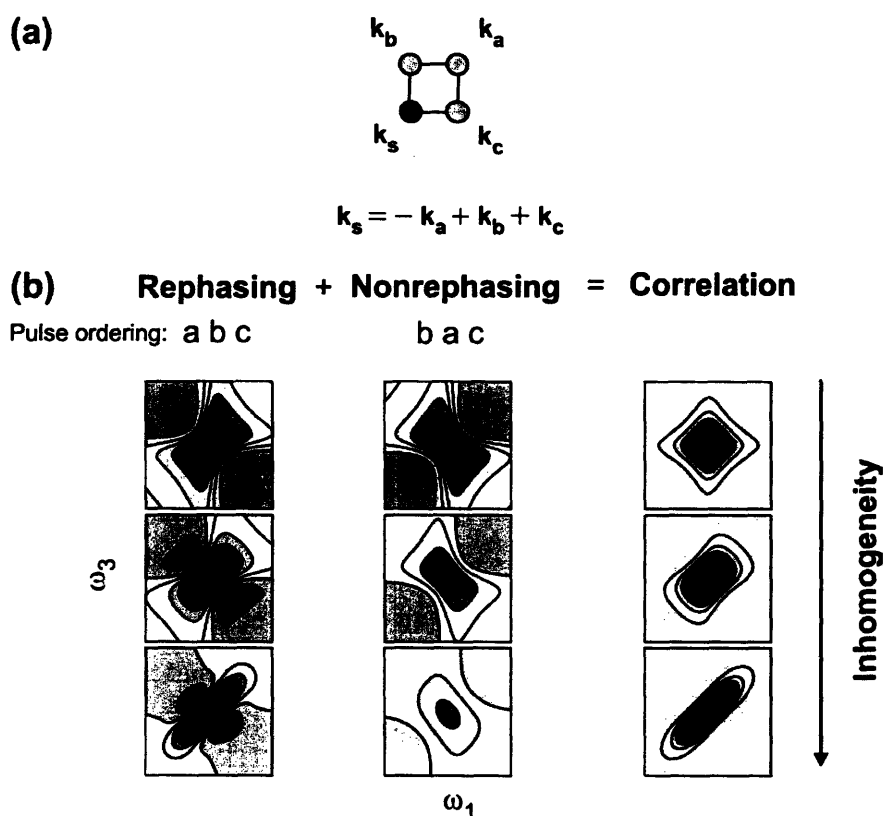


Figure 1-3. Lineshapes in 2D IR spectrum. (a) A boxcar geometry with k -vectors of incoming pulses and the signal. (b) Illustrated are the dependences of the relative intensities of rephasing and non-rephasing spectra and the lineshape of the correlation spectra on the inhomogeneity.

In spite of simpler nonlinear IR experiments, the frequency-dispersed pump probe (DPP) and dispersed vibrational echo (DVE)⁹³⁻⁹⁵ also have considerable power for revealing resonances hidden in traditional IR spectra. Although it is driven by two IR

pulses, DPP signal arises from the same third-order polarization as those of 2D IR spectroscopy. A pump pulse prepares a population state and a subsequent probe pulse detects pump-induced changes after delay τ_2 . DVE is the background free (homodyne) signal obtained from a signal scattered by three time-coincident pulses ($\tau_1 = \tau_2 = 0$). It is the same signal used for 2D IR spectroscopy although without the LO. Both the DPP and DVE signals are referred to as “dispersed” because the signals are observed in the frequency domain after spectral dispersion by the spectrometer to obtain ω_3 components. These simpler techniques are related with the projection of the 2D IR spectrum onto ω_3 axis. These relationships are presented in Chapter 3.

1.3.3.2. Amide I vibrational spectroscopy

Amide I vibrational spectroscopy of proteins has long been used to describe secondary structure content and is increasingly targeted for investigations into the dynamics of protein structures. Of the many possible vibrations, the amide I mode ($1600 - 1700 \text{ cm}^{-1}$) is of particular interest since it is intrinsic to all proteins, and additionally is localized on the peptide backbone and not strongly influenced by sidechains.⁷⁹ The amide I vibration of a single peptide unit is primarily carbonyl stretching, but its vibrational frequency in a protein is sensitive to the type and amount of secondary structures. Strong electrostatic interactions between the many neighboring amide I oscillators of the protein lead to delocalized (excitonic) states, whose absorption frequencies reflect the underlying structural arrangement of oscillators. Well established empirical frequency-structure correlations find that the β sheet contributes both at low (ν_{\perp} mode, $1610 - 1640 \text{ cm}^{-1}$) and high frequencies (ν_{\parallel} mode, above 1680 cm^{-1}), and the α helix and random coil structure are located at $1650 - 1660 \text{ cm}^{-1}$ and $1640 - 1650 \text{ cm}^{-1}$, respectively.^{96,97} However, the broad amide I linewidth arising from structural heterogeneity of the protein hides much of the underlying structure needed for peak assignment. To overcome this problem in infrared spectra, several numerical treatments have been developed such as Fourier self deconvolution,⁹⁷ second derivative technique,⁹⁸ and factor analysis.^{99,100} More recently, two-dimensional infrared spectroscopy of peptides has offered a new way of dissecting congested amide I lineshapes and quantitatively describing the structural origins of the

vibrational spectroscopy.^{80,81,101-103} The 2D IR spectroscopy of proteins is now being used to characterize structure and molecular dynamics of proteins through the delocalized amide I states.^{93,104,105}

1.3.3.3. Application to the anti-parallel β sheet

One particular strength of 2D IR spectroscopy is its sensitivity to the details of anti-parallel β -sheet secondary structure.^{93,106} Couplings between amide I vibrations are revealed in 2D IR spectra of AP β -sheet containing proteins through the formation of a cross peak between the ν_{\perp} and ν_{\parallel} transitions.¹⁰⁶ For large, ordered β -sheets, such as those found in poly-L-lysine at high pH, the 2D IR correlation spectrum shows distinct cross peaks between the two amide I resonances.⁹³ These features are retained in the 2D IR spectra of proteins with AP β -sheets even when other secondary structures are present, although it appears considerably different.⁹³ Since the splitting between the ν_{\parallel} and ν_{\perp} resonances are sensitive to the strength of amide I couplings, the size and geometry of the sheet, and structural or energetic disorder associated with the peptide units within the sheet, these transitions become inhomogeneously broadened in proteins. As a result, interference effects between inhomogeneously broadened diagonal vibrations with cross peaks arising from the ν_{\parallel} and ν_{\perp} modes leads to a characteristic “Z” shaped pattern for the amide I region in the 2D IR spectrum as shown in Fig. 1-4.

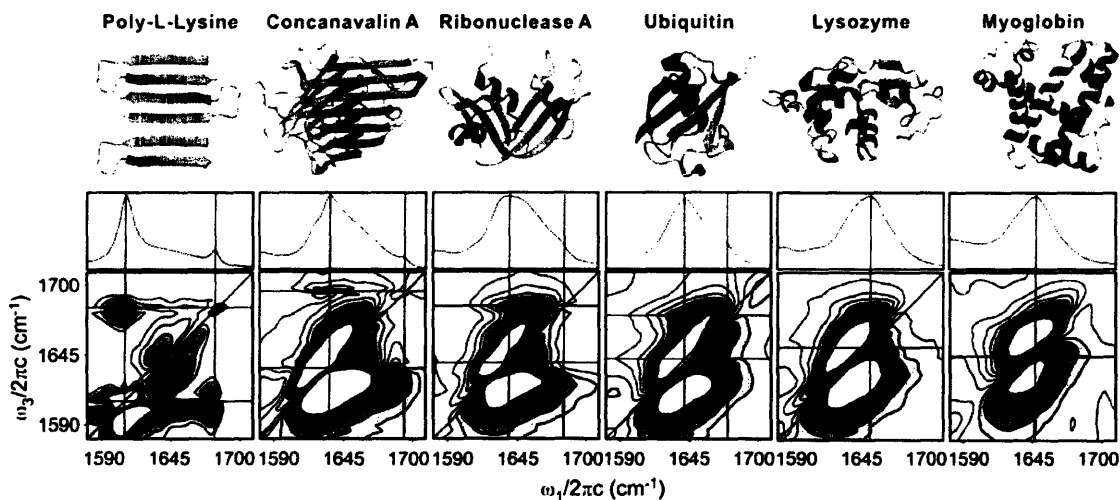


Figure 1-4. 2D IR spectra of proteins with different β -sheet contents.

1.4. Thesis outline

This thesis starts by discussing linear infrared spectroscopy of protein amide I vibrations in Chapter 2. As a basis for the nonlinear IR spectroscopy, modeling, visualization, and characterization of amide I vibrations are provided. Chapter 3 reviews relationships between various third-order nonlinear IR techniques used in this work for the comparison of results in following chapters. Chapter 4 describes all experimental details including the nonlinear infrared setup, T-jump laser setup, control of timing between lasers, and detection schemes. Chapter 5 reports the first application of nonlinear IR spectroscopy to equilibrium thermal unfolding of ribonuclease A (RNase A), in which the unfolding of its β sheet is monitored through the disappearance of cross peaks. From Chapter 6 to 9, transient thermal unfolding of ubiquitin is probed by dispersed vibrational echo spectroscopy. To explain a complicated unfolding dynamics, an unfolding free energy surface is proposed in Chapter 6. This surface is more investigated by temperature-variation and site mutation studies in Chapter 7 and 9, respectively. Also, theoretical calculations of the free energy surface are presented in Chapter 8 to support the interpretations of unfolding behaviors. Finally, Chapter 10 reports the probing of ubiquitin unfolding by transient T-jump 2D IR spectroscopy. This development of transient two-dimensional infrared spectroscopy sheds light on new features buried in congested spectra.

References

- (1) Anfinsen, C. B. *Science* **1973**, *181*, 223.
- (2) Levinthal, C. *J. Chim. Phys.* **1968**, *65*, 44.
- (3) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J. *Acc. Chem. Res.* **1998**, *31*, 745.
- (4) Dobson, C. M. *Nature* **2003**, *426*, 884.
- (5) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1982**, *51*, 459.
- (6) Ptitsyn, O. B.; Rashin, A. A. *Biophys. Chem.* **1975**, *3*, 1.
- (7) Baldwin, R. L. *Trends Biochem. Sci.* **1989**, *14*, 291.
- (8) Dill, K. A. *Biochemistry* **1985**, *24*, 1501.
- (9) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3.
- (10) Daggett, V.; Fersht, A. R. *Trends Biochem. Sci.* **2003**, *28*, 18.
- (11) Baldwin, R. L. *Nature* **1994**, *369*, 183.
- (12) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619.
- (13) Sali, A.; Shakhnovich, E. I.; Karplus, M. *Nature* **1994**, *369*, 248.
- (14) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10.
- (15) Onuchic, J. O.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545.
- (16) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70.
- (17) Kramers, H. A. *Physica* **1940**, *7*, 284.
- (18) Huang, F.; Nau, W. M. *Angew. Chem. Int. Ed.* **2003**, *42*, 2269.
- (19) Thompson, P. A.; Munoz, V.; Jas, G. S.; Henry, E. R.; Eaton, W. A.; Hofrichter, J. *J. Phys. Chem. B.* **2000**, *104*, 378.
- (20) Huang, C.-Y.; Getahun, Z.; Zhu, Y.; Klemke, J. W.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2788.
- (21) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
- (22) Xu, Y.; Oyola, R.; Gai, F. *J. Am. Chem. Soc.* **2003**, *125*, 15388.
- (23) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952.
- (24) Kubelka, J.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2003**, *329*, 625.

- (25) Jager, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373.
- (26) Ferguson, N.; Johnson, C. M.; Macias, M.; Oschkinat, H.; Fersht, A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13002.
- (27) Nguyen, H.; Jager, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3948.
- (28) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76.
- (29) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985.
- (30) Sabelko, J.; Ervin, J.; Gruebele, M. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6031.
- (31) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193.
- (32) Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Munoz, V. *Science* **2002**, *298*, 2191.
- (33) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2006**, *442*, 317.
- (34) Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2527.
- (35) Yang, W. Y.; Gruebele, M. *Biophys. J.* **2004**, *87*, 596.
- (36) Bukau, B.; Horwich, A. L. *Cell* **1998**, *92*, 351.
- (37) Hartl, F. U.; Hayer-Hartl, M. *Science* **2002**, *295*, 1852.
- (38) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573.
- (39) Eaton, W. A.; Muñoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 327.
- (40) Ballew, R. M.; Sabelko, J.; Reiner, C.; Gruebele, M. *Rev. Sci. Instrum.* **1996**, *67*, 3694.
- (41) Gruebele, M.; Sabelko, J.; Ballew, R.; Ervin, J. *Acc. Chem. Res.* **1998**, *31*, 699.
- (42) Callender, R. H.; Dyer, R. B.; Gilmanshin, R.; Woodruff, W. H. *Annu. Rev. Phys. Chem.* **1998**, *49*, 173.
- (43) Wray, W. O.; Aida, T.; Dyer, R. B. *Appl. Phys. B* **2002**, *74*, 57.
- (44) Roder, H.; Maki, K.; Latypov, R. F.; Cheng, H.; Shastry, M. C. R. Early events in protein folding explored by rapid mixing methods. In *Protein folding handbook*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH: Weinheim, 2005.

- (45) Hansen, K. C.; Rock, R. S.; Larsen, R. W.; Chan, S. I. *J. Am. Chem. Soc.* **2000**, *122*, 11567.
- (46) Abbruzzetti, S.; Sottini, S.; Viappiani, C.; Corrie, J. E. T. *Photochem. Photobiol. Sci.* **2006**, *5*, 621.
- (47) Abbruzzetti, S.; Carcelli, M.; Pelagatti, P.; Rogolino, D.; Viappiani, C. *Chem. Phys. Lett.* **2001**, *344*, 387.
- (48) Jones, C. M.; Henry, E. R.; Hu, Y.; Chan, C.-K.; Luck, S. D.; Bhuyan, A.; Roder, H.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11860.
- (49) Bredenbeck, J.; Helbing, J.; Sieg, A.; Schrader, T.; Zinth, W.; Renner, C.; Behrendth, R.; Moroder, L.; Wachtveitl, J.; Hamm, P. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6452.
- (50) Bredenbeck, J.; Helbing, J.; Kumita, J. R.; Woolley, G. A.; Hamm, P. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2379.
- (51) Lu, H. S. M.; Volk, M.; Kholodenko, Y.; Gooding, E.; Hochstrasser, R. M.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 7173.
- (52) Kubelka, J.; Chiu, T. K.; Davies, D. R.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2006**, *359*, 546.
- (53) Brewer, S. H.; Vu, D. M.; Tang, Y.; Li, Y.; Franzen, S.; Raleigh, D. P.; Dyer, R. B. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16662.
- (54) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200.
- (55) Yang, W. Y.; Gruebele, M. *J. Am. Chem. Soc.* **2004**, *126*, 7758.
- (56) Du, D.; Zhu, Y.; Huang, C.-Y.; Gai, F. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15915.
- (57) Dyer, R. B.; Maness, S. J.; Peterson, E. S.; Franzen, S.; Fesinmeyer, R. M.; Andersen, N. H. *Biochemistry* **2004**, *43*, 11560.
- (58) Creighton, T. E. *Biochem. J.* **1990**, *270*, 1.
- (59) King, J. *C&E News* **1989**, *April 10*, 32.
- (60) Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed.* **1998**, *37*, 868.
- (61) Gruebele, M. *Ann. Rev. Phys. Chem.* **1999**, *50*, 485.
- (62) Gruebele, M. *Curr. Opin. Struct. Biol.* **2002**, *12*, 161.

- (63) Balbach, J.; Forge, V.; Lau, W. S.; van Nuland, N. A. J.; Brew, K.; Dobson, C. M. *Science* **1996**, *274*, 1161.
- (64) Dobson, C. M.; Hore, P. J. *Nat. Struct. Biol.* **1998**, *5*, 504.
- (65) Schotte, F.; Lim, M.; Jackson, T. A.; Smirnov, A. V.; Soman, J.; Olson, J. S.; Jr., P. G. N.; Wulff, M.; Anfirud, P. A. *Science* **2003**, *300*, 1944.
- (66) Moffat, K. *Chem. Rev.* **2001**, *101*, 1569.
- (67) Schmidt, M.; Pahl, R.; Srajer, V.; Anderson, S.; Ren, Z.; Ihee, H.; Rajagopal, S.; Moffat, K. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4799.
- (68) Doniach, S. *Chem. Rev.* **2001**, *101*, 1763.
- (69) Dyer, R. B.; Gai, F.; Woodruff, W. H.; Gilmanshin, R.; Callender, R. H. *Acc. Chem. Res.* **1998**, *31*, 709.
- (70) Bredenbeck, J.; Helbing, J.; Behrendt, R.; Renner, C.; Moroder, L.; Wachtveitl, J.; Hamm, P. *J. Phys. Chem. B* **2003**, *107*, 8654.
- (71) Yamamoto, K.; Mizutani, Y.; Kitagawa, T. *Biophys. J* **2000**, *79*, 485.
- (72) Goldbeck, R. A.; Thomas, Y. G.; Chen, E.; Esquerra, R. M.; Kliger, D. S. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2782.
- (73) O'Connor, D. B.; Goldbeck, R. A.; Hazzard, J. H.; Kliger, D. S.; Cusanovich, M. A. *Biophys. J.* **1993**, *65*, 1718.
- (74) Chen, E.; Wittung-Stafshede, P.; Kliger, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 3811.
- (75) Xie, X.; Simon, J. D. *Rev. Sci. Instrum.* **1989**, *60*, 2614.
- (76) Chen, E.; Goldbeck, R. A.; Kliger, D. S. *J. Phys. Chem. A* **2003**, *107*, 8149.
- (77) Lewis, J. W.; Goldbeck, R. A.; Kliger, D. S.; Xie, X.; Dunn, R. C.; Simon, J. D. *J. Phys. Chem.* **1992**, *96*, 5243.
- (78) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379.
- (79) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181.
- (80) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2036.
- (81) Woutersen, S.; Mu, Y.; Stock, G.; Hamm, P. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11254.
- (82) Zanni, M. T.; Gnanakaran, S.; Stenger, J.; Hochstrasser, R. M. *J. Phys. Chem. B* **2001**, *105*, 6520.

- (83) Woutersen, S.; Hamm, P. *J. Phys.: Condens. Mat.* **2002**, *14*, 1035.
- (84) Huang, R.; Kubelka, J.; Barber-Armstrong, W.; Silva, R.; Decatur, S. M.; Keiderling, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 2346.
- (85) Moran, A.; Mukamel, S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 506.
- (86) Cha, S.; Ham, S.; Cho, M. *J. Chem. Phys.* **2002**, *117*, 740.
- (87) Ham, S.; Cho, M. *J. Chem. Phys.* **2003**, *118*, 6915.
- (88) Khalil, M.; Demirdöven, N.; Tokmakoff, A. *J. Phys. Chem. A* **2003**, *107*, 5258.
- (89) Zanni, M. T.; Hochstrasser, R. M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 516.
- (90) Fecko, C. J.; Eaves, J. D.; Loparo, J. J.; Tokmakoff, A.; Geissler, P. L. *Science* **2003**, *301*, 1698.
- (91) Kim, Y. S.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11185.
- (92) Zheng, J. R.; Kwak, K.; Asbury, J.; Chen, X.; Piletic, I. R.; Fayer, M. D. *Science* **2005**, *309*, 1338.
- (93) Demirdöven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. *J. Am. Chem. Soc.* **2004**, *126*, 7981.
- (94) Thompson, D. E.; Merchant, K. A.; Fayer, M. D. *J. Chem. Phys.* **2001**, *115*, 317.
- (95) Merchant, K. A.; Xu, Q.-H.; Thompson, D. E.; Fayer, M. D. *J. Phys. Chem. A* **2002**, *106*, 8839.
- (96) Jackson, M.; Mantsch, H. H. *Crit. Rev. Biochem. Mol.* **1995**, *30*, 95.
- (97) Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469.
- (98) Dong, A.; Huang, P.; Caughey, W. S. *Biochemistry* **1990**, *29*, 3303.
- (99) Lee, D. C.; Haris, P. I.; Chapman, D.; Mitchell, R. C. *Biochemistry* **1990**, *29*, 9185.
- (100) Baumruk, V.; Pancoska, P.; Keiderling, T. A. *J. Mol. Biol.* **1996**, *259*, 774.
- (101) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123.
- (102) Woutersen, S.; Hamm, P. *J. Phys. Chem. B* **2000**, *104*, 11316.
- (103) Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *115*, 7737.
- (104) Chung, H. S.; Khalil, M.; Tokmakoff, A. *J. Phys. Chem. B* **2004**, *108*, 15332.
- (105) Chung, H. S.; Khalil, M.; Smith, A. W.; Ganim, Z.; Tokmakoff, A. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 612.
- (106) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *J. Chem. Phys.* **2004**, *120*, 8201.

Chapter 2

Visualization and characterization of the IR active amide I vibrations of proteins

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

- “Visualization and characterization of the infrared active amide I vibrations of proteins,” H. S. Chung and A. Tokmakoff, *J. Phys. Chem. B*, **110**, 2888 (2006).

2.1. Introduction

The quantitative description of FTIR and 2D IR spectra requires a molecular description of the origin of spectral features. A key advance in the description of amide I spectroscopy came from Torii and Tasumi,¹ who calculated the amide I spectrum of several proteins by diagonalizing a Hamiltonian constructed in the basis of each peptide unit with couplings calculated from transition dipole interactions between sites. More recent improvement involved *ab initio* calculation is used to get a more realistic coupling model between the adjacent residues.²⁻⁴ This relatively simple model, the local amide I Hamiltonian, has had considerable success reproducing the band shape of amide I spectrum.

The eigenstates from such a calculation form the basis for trying to understand the correlations between amide I frequency and protein structure characterized by IR spectra. In particular, they can be used to answer a number of questions. What are the vibrations that we

observe in an experiment at a given frequency? To what degree are these states delocalized? To what degree do they give us information on secondary and tertiary structure? Are the vibrations of secondary structures a good basis set for decomposing protein IR spectra?

Previous studies have addressed some of these questions. The characterization of the phase of various amide I vibrations has been started for the ideal model of the anti-parallel β sheet and the α helix with perfect symmetries.⁵⁻¹¹ The role of energetic disorder in mode symmetry and localization has recently been addressed.¹² Analysis of several proteins indicates that the lowest frequencies observed in the IR spectrum of β proteins are localized to the sheet, and that resonance frequencies for α helical excitations are size dependent.¹³ The visualization of the phases and amplitudes of several vibrational eigenstates having large transition dipole moments for two proteins, myoglobin and flavodoxin, has been reported,¹⁴ indicating that vibrations of different eigenstates are localized either in the α helix or β sheet. Even with the considerable progress, more quantitative analysis methods are needed to interpret the structural origin of vibrational resonance frequencies within the amide I spectrum, particularly in the presence of structural heterogeneity found in proteins. The large number of eigenstates complicates the characterization of spectra, since analysis of a small number of eigenstates with large transition dipole intensity may not be representative of the whole. Also, sequence dependent characterizations of normal modes or eigenstates are not particularly helpful when structural correlations must be characterized in three spatial dimensions.

In this chapter, methods that can be used to address frequency-structure correlations in infrared spectra of proteins are presented. Drawing on the eigenstates of the local amide I Hamiltonian, the structural origin of protein amide I infrared spectroscopy can be characterized using visualization methods that identify the amplitude and phase relationship of amide I oscillators in one, two, and three dimensions. Vibrational amplitudes and phases are obtained from doorway modes^{13,15} rather than eigenstate vibrations to find average nature of the vibrations in a particular frequency region. Doorway modes represent a transformation of a subset of eigenstates into a bright state characteristic of a part of the IR spectrum. The characterization is performed for the six split frequency ranges of the amide I band to find frequency dependent vibrational characters. For the visualization, the vibrational amplitudes and phases of individual peptide oscillators are color-coded directly onto the crystal

structure, and then projected onto a two-dimensional grid for β sheets.^{12,16} These methods can then be quantitatively analyzed with spatial correlation functions for the oscillators within delocalized vibrations, as previously demonstrated in electronic systems.¹⁷ One and two dimensional analyses are performed for the α helix and β sheet, respectively. These correlation analyses effectively present the phase relationship between adjacent vibrations through bond and through space. To quantify the specific vibrational mode character associated with the idealized vibrations of β sheets and α helices, a mode character f is developed based on a phase-associated correlation length. As a reference, the characterizations are first performed for the ideal α helix and β sheet. Then, detailed case studies for real systems are presented for three proteins: concanavalin A, an all β sheet containing protein; ubiquitin, mixed α/β proteins; and myoglobin, an all α protein.

2.2. Methods

2.2.1. Calculation of amide I eigenstates and IR spectrum

To characterize frequency-structure relationships in the amide I spectrum, we calculate the IR spectrum using a sum over amide I eigenstates, and then analyze frequency-dependent eigenstate characteristics using doorway mode analysis for a subset of eigenstates within a given frequency range. The amide I eigenstates and energy eigenvalues are obtained by diagonalizing a local amide I Hamiltonian (LAH) \mathbf{H} ,¹⁶ an approach pioneered by Krimm¹⁸ and Torii¹ and more recently extended by Hamm and Hochstrasser.^{19,20} The LAH is constructed in the site basis of the amide I vibrations for the individual peptide units, drawing on the coordinates from a crystal structure. The energies of the individual sites (diagonal terms) are assigned based on hydrogen bonding criteria, and the vibrational couplings (off-diagonal terms) include through-space and through-bond effects. IR spectra are calculated by a sum over amide I vibrational eigenstates weighted by the transition dipole matrix elements squared and convoluted with a Lorentzian lineshape

$$A(\omega) = \text{Im} \sum_q |\langle q | \mathbf{m} | 0 \rangle|^2 / (\omega - \omega_{q0} - i\Gamma). \quad (2.1)$$

To account for natural small variations in structure or solvation, site energies are averaged over a distribution.

The diagonal elements of the LAH, H_{ii} , consist of site frequencies ω_i for the individual amide I vibrations of the peptide units. The site frequencies are red-shifted $\omega_i = \omega_0 + \delta\omega$ from an unperturbed frequency ($\omega_0 = 1688 \text{ cm}^{-1}$) to reflect hydrogen bonding interactions with other residues or solvent molecules. Three criteria are used to determine the changes of site frequency: (i) hydrogen bonding to the carbonyl oxygen by hydrogen of NH group of another residue, (ii) hydrogen bonding to the carbonyl oxygen by hydrogen of surrounding water molecules, and (iii) hydrogen bonding to the NH group by another residue or solvent molecules. The empirical formula is used to calculate the effect of (i) is $\delta\omega = a(2.6 - r_{\text{OH}})$, where r_{OH} is the O...H hydrogen bond distance in Ångstroms and $a = -30 \text{ cm}^{-1}$.¹⁹ The formation of hydrogen bond is assumed when $r_{\text{OH}} < 2.6 \text{ Å}$ and the angle formed by O...H, and H-N is less than 35° . Because of the insufficient number of water molecules in the crystal structure, the uniform shift of site frequency of -20 cm^{-1} and -10 cm^{-1} are used for (ii) and (iii), respectively.²¹

Off diagonal elements except those of the nearest neighbor pairs are constructed using the empirically parameterized transition dipole coupling model:^{1,18}

$$H_{ij} = \frac{A}{r_{ij}^3} \left(\hat{\mu}_i \cdot \hat{\mu}_j - 3(\hat{\mu}_i \cdot \hat{r}_{ij})(\hat{\mu}_j \cdot \hat{r}_{ij}) \right) \quad (2.2)$$

Following Krimm, we take $A = 580 \text{ cm}^{-1} \text{ Å}^3$. $\hat{\mu}_i$ is the unit vector of the transition dipole moment associated with the amide I vibrational coordinate of the i^{th} peptide site, \hat{r}_{ij} is unit vector from site i to j , and r_{ij} is the separation. The transition dipole of each peptide group is assumed to be located on the C=O bond axis, 0.868 Å from the carbon atom to the oxygen with an orientation of 20° off from the C=O bond axis toward the nitrogen.^{1,7} For interactions between the bonded pairs of adjacent peptide units ($i = j \pm 1$), amide I couplings determined from *ab initio* calculations on the diglycine molecule are used.³

With these elements of the model, the amide I eigenstates are obtained by diagonalizing the LAH

$$\mathbf{H} = \mathbf{TET}^T \quad (2.3)$$

where \mathbf{T} is the transformation matrix formed by the eigenvectors. The x , y , and z components of the transition dipoles connecting eigenstates can be obtained from the $N \times 3$ transition dipole matrix \mathbf{d} representing the x , y , and z components of transition dipoles for the N peptide

sites. $\mathbf{d}^E = \langle q | \mathbf{m} | 0 \rangle$ in Eq. (2.1) is obtained from $\mathbf{T}_q^T \cdot \mathbf{d}$ where \mathbf{T}_q is the q^{th} column of \mathbf{T} . IR spectra are calculated from \mathbf{H} using Eq. (2.1) with a Lorentzian line width of $\Gamma = 5 \text{ cm}^{-1}$. To account for broadening of the spectrum as a result of disorder of the local mode frequency, 10,000 spectra are averaged in which the site energies are chosen from a Gaussian random distribution with width $\sigma = 6 \text{ cm}^{-1}$ centered at the nominal site energy ω_i .

2.2.2. Doorway state calculation

Due to the high density of states, the LAH eigenstates are of limited use for describing and visualizing the vibrations contributing to different frequency ranges within the amide I infrared spectrum. Given a frequency range, one cannot generally select a representative vibrational eigenstate by its strength in the IR spectrum in the presence of several equivalent transition dipole moments. This raises questions about how to average the contribution of several states without overestimating the vibrational character of states that do not strongly contribute to the IR spectrum. We have chosen to apply the doorway state analysis used previously by Torii and Tasumi.^{13,15} We obtain a set of three representative bright states for a specific subset of eigenvectors grouped by frequency through an orthogonal transformation. For analysis of structure-frequency correlations, we divide the amide I spectrum into six frequency ranges. The region between 1640 cm^{-1} and 1680 cm^{-1} is divided into 10 cm^{-1} intervals. Additionally, the modes below 1640 cm^{-1} and above 1680 cm^{-1} form two groups. All the eigenstates within each frequency region constitute the vibrational subset for doorway analysis. The frequency range could be divided in another manner guided by the density of eigenstates and the desired frequency resolution, and retaining a minimum of three eigenstates remains in each region. As the number of eigenstates per interval decreases, the character of the doorway mode approaches that of the eigenstates. If the region is too wide, various vibrational mode characters are overlapped and we lose the information on the frequency-structure relationship. We found the appropriate frequency splitting is $5 - 10 \text{ cm}^{-1}$. In this chapter, we show results of 10 cm^{-1} splitting, but the analysis with 5 cm^{-1} splitting gives similar results.

The relationship between the site basis, eigenstates, and doorway states are depicted in Fig. 2-1. In the transformation for n_i eigenstates within a particular frequency range R_i ,

three bright states (nonzero transition dipole moment) and n_i-3 dark states (zero transition dipole moment) that are orthogonal to the bright states are expressed as a linear combination of eigenstates. (The number of the bright states is determined by the dimensionality of the dipole moment.) The first, second, and third doorway modes are labeled by decreasing transition dipole strength. Although any linear combination of the three bright states is orthogonal to the dark states, choosing one of the orthogonal bright states eliminates interference between them.¹⁵

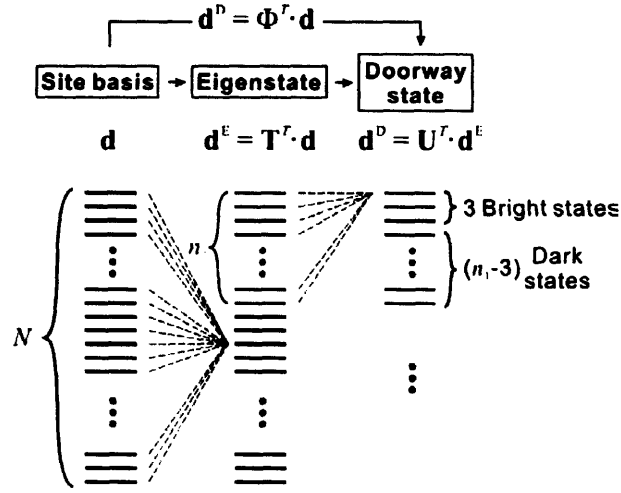


Figure 2-1. Transformation from site basis to doorway states. Diagonalization of the local amide I Hamiltonian (LAH) of N amide I oscillators in the site basis gives N eigenstates. The transform matrix \mathbf{T} consists of the eigenvectors. Each eigenstate is the linear combination of the states in site basis as shown with green dashed lines. Doorway transformation of the subset of n_1 eigenstates (red) in the first frequency region R_1 gives 3 bright states and $n_1 - 3$ dark states. The transformations for the following frequency regions also give three bright states for each frequency region. Note that all the doorway states in the same frequency region are not associated with a definite energy. The orthogonal set of bright states is obtained by SVD and the first doorway mode is used for the visualization and correlation analyses.

We use singular value decomposition (SVD) to find the transformation \mathbf{U}_i that generates the three orthogonal bright states from the eigenstates in the i^{th} frequency region R_i . The $n_i \times 3$ eigenstate dipole matrix, \mathbf{d}_i^E is formed as

$$\mathbf{d}_i^E = \mathbf{T}_i^T \mathbf{d}. \quad (2.4)$$

\mathbf{T}_i is the subset of the transformation matrix \mathbf{T} (Eq. (2.3)) for the eigenstates within R_i . The doorway dipole matrix, \mathbf{d}^D_i can be obtained by the orthonormal transformation \mathbf{U}_i

$$\mathbf{d}^D_i = \mathbf{U}_i^T \mathbf{d}^E_i. \quad (2.5)$$

SVD decomposes \mathbf{d}^E_i into three matrices,

$$\mathbf{d}^E_i = \mathbf{U}'_i \mathbf{S}_i \mathbf{V}_i^T. \quad (2.6)$$

\mathbf{U}'_i and \mathbf{V}_i are $n \times n$ and 3×3 orthonormal matrices, respectively. All elements of \mathbf{S}_i ($n \times 3$) are zero except the three diagonal singular values (s_{11}, s_{22}, s_{33}). Since the transformation to the three orthogonal bright states is unique, Eqs. (2.5) and (2.6) are equivalent. Therefore, $\mathbf{U}'_i = \mathbf{U}_i$ and $\mathbf{d}^D_i = \mathbf{S}_i \mathbf{V}_i^T$, and the first three columns of \mathbf{U}_i are the transformation vectors from eigenstates to the three orthogonal doorway modes. The three column vectors of \mathbf{V}_i consist of unit vectors of the three representative transition dipoles and three nonzero singular values of \mathbf{S}_i are their transition dipole moments.

The transform matrix Φ_i from the site basis to the doorway states in the frequency region R_i is given by rearranging Eqs. (2.4) and (2.5) as

$$\Phi_i = \mathbf{T}_i \mathbf{U}_i. \quad (2.7)$$

Among the three doorway modes, the first mode is chosen to represent the frequency region R because it has the most intense dipole strength although it does not totally dominate the other two modes. In most cases, the dipole moment square of the first doorway mode exceeds 50% of the total. The elements of ϕ_1 , the first column of Φ_i give the amplitudes and phases of local amide I vibrations that constitute the first doorway mode. ϕ_1 is used for the analyses on the frequency dependent vibrational characteristics: the visualization of vibrational phases and calculating the correlation functions, correlation lengths, and secondary structure contents. All calculations are performed averaging over 10,000 disorder realizations ($\sigma = 6 \text{ cm}^{-1}$). Visualization of the vibrational amplitude and phase is presented for the crystal structure without disorder. The assignments of a residue to a particular secondary structure are taken from the data in the PDB file. Since the square of each element of ϕ_1 is the fraction of each local mode vibration, the secondary structure contents are calculated by the summation of the squares of ϕ_1 over the secondary structure indices.

2.2.3. Visualization of amide I vibrations

The amide I doorway modes are visualized using a red to blue color map that encodes the amplitude and phase of each of the amide I sites. Red is positive, blue is negative, and the amplitude scales with the depth of the color. The direct color-coding of the amplitudes on the protein structure is performed by color change of the amino acid residues using the software Pymol (DeLano Scientific LLC). This scheme is effective to show the behavior of vibrational correlations under structural variances in the real proteins. It also provides a unique visualization method for α helix that is not easily projected onto two dimensions like a β sheet. We colored the amino acid residues, for convenience, instead of the actual oscillators which lie in the middle of the two residues forming peptide groups. In other words, the amplitude of the amide oscillator linking residues n and $n+1$ is recorded on the residue n as shown in Fig. 2-2a. Although this results in the shift of the location of the oscillator by a half residue, the color-coding is straightforward for α helix and parallel β sheet because every oscillator is shifted to the same direction in case of unidirectional growth. This shift induces a problem in the antiparallel β sheet. In Fig 1a, the shift of the oscillator 1 and 2 to the left and the shift of the residue 3 and 4 to the right results the mismatch by one residue. To avoid this mismatch, the oscillators of the every other strand in the antiparallel β sheet are shifted, i.e. if the amplitudes of the oscillator 1 and 2 are colored on the residue 1 and 2, the amplitudes of the oscillator 3 and 4 are colored on the residue 4 and 5.

2.3. Results and Discussion

2.3.1. Characterization of the vibrations of secondary structure

As a starting point for the interpretation of protein amide I vibrations, we characterize the doorway modes of idealized secondary structures. Here we consider the antiparallel β sheet and α helix. The empirical frequency-structure relationships indicate that antiparallel β sheets are characterized by a two peak structure in the infrared absorption spectrum at 1630 – 1640 cm^{-1} and $> 1680 \text{ cm}^{-1}$, and that α helix and random coil structure are located at ~ 1650 – 1660 cm^{-1} and 1640 – 1650 cm^{-1} , respectively.^{22,23}

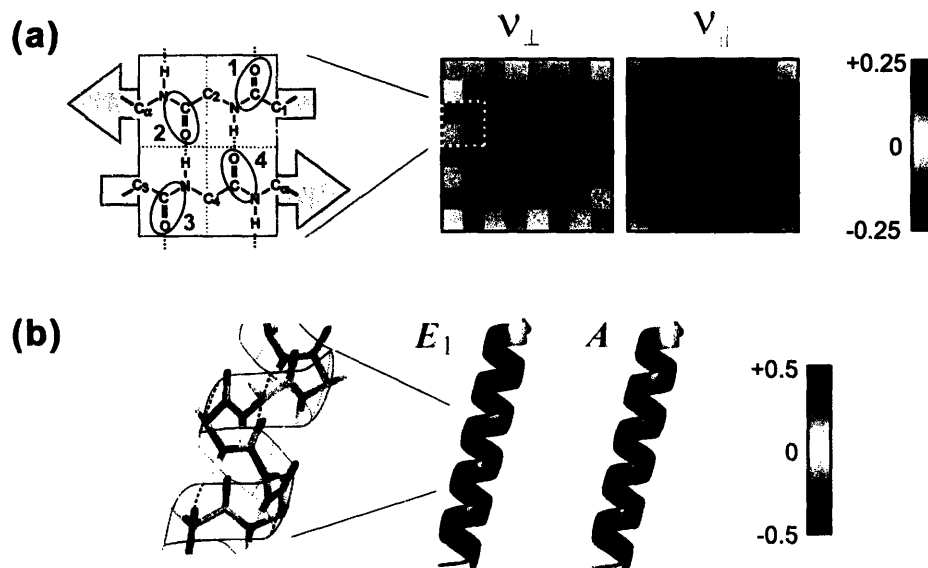


Figure 2-2. (a) Projection of the antiparallel β sheet on a plane. (Left) Four peptide groups in the unit cell are shown in two anti-parallel β strands. Peptide units are numbered with red figures and amino acids are numbered in C_{α} . (Right) Projection of the amplitudes and phases oscillators in the ν_{\perp} and ν_{\parallel} vibrations of the ideal β sheet formed by 8×8 oscillators. (b) Amplitudes and phases of oscillators in the A and E_1 vibrations of the ideal 21-residue α helix.

The split two peak spectrum of the antiparallel β sheet has been explained by the long-range transition dipole coupling and the symmetry of the four oscillator unit cell for an idealized sheet shown in Fig. 2-2a.^{6,7,18} These studies indicate that there are four characteristic vibrations for the antiparallel β sheet that differ in the phase relationships between adjacent oscillators, of which two are observed in the IR spectrum. The lowest frequency transition carries most of the oscillator strength because it involves in phase oscillation of the aligned amide I transition dipoles on adjacent strands, but alternating phase between bonded amide units. Vibrations of this symmetry will be designated ν_{\perp} since the in phase oscillations and the transition moments lie perpendicular to the strands of the β sheet. The other high frequency mode, designated ν_{\parallel} , is weaker because the vibrations of adjacent oscillators are in-phase for bonded units and alternating for oscillators between strands. We use these notations instead of the a^+ and a^- notation used for idealized structures²¹²⁶ because this symmetry is broken for the real β sheet structure, and these assignments also apply to

parallel β sheets and the mixed structure of parallel and antiparallel sheets such as ubiquitin. The vibrational frequency of ν_{\perp} mode is sensitive to the number of the strands because the vertically in-phase relationship of amide I vibration is energetically more favorable in the addition of strands while ν_{\parallel} mode is insensitive if at least one hairpin presents.⁷ Therefore, the splitting between the two modes is larger and the ν_{\perp} transition becomes stronger in the bigger β sheet, and thus the IR spectrum of β sheet containing protein become more asymmetric.

The symmetry of the ν_{\perp} and ν_{\parallel} modes can be visualized by projecting the β -sheet structure onto a two-dimensional grid that represents the spatial placement of nearest neighbors to each amide I oscillator.^{12,16} The horizontal dimension represents the direction of a given strand within the sheet. Each box is color coded to indicate the amplitude and phase of the oscillator in a particular mode. Fig. 2-2a shows a visualization of totally delocalized amide I eigenstates with ν_{\perp} and ν_{\parallel} character for an idealized antiparallel sheet of 8×8 oscillators with degenerate site energies. The vibrational amplitude and phase relationships between oscillators of the two modes are shown with varying depth of red (positive) and blue (negative) colors. These diagrams emphasize that the character of the two modes expected for the sheet differ by either columns or rows retaining constant phase and the other dimension exhibiting alternating phase.

A similar visualization strategy can be adopted for helices. For the amide I vibrations of an ideal α helix, A and E_1 modes are expected,^{5,6,8,11,24,25} The vibrational phases and amplitudes of these two modes for ideal 21-residue α helix are shown on the structure in Fig. 2-2b by projecting the amplitude and phase of the amide I vibrations onto a ribbon diagram of the helix. All of the amide I oscillators are in phase for the A mode. For the E_1 mode, the phase relationship between adjacent oscillators in the primary sequence is such that oscillators every 3.6 residues are in phase, reflecting the fact that there are 18 residues for 5 turns in the unit cell of the helix.

2.3.2. Spatial Correlation Functions

To characterize the frequency-structure relationships for the doorway modes of secondary structures and proteins as distinguished by the phase relationship between

oscillators within the structure, we consider spatial correlation functions that characterize the amplitude and phase relationships between the amide I oscillators and the degree of delocalization. A one-dimensional correlation function for the vibrational amplitude A between the amide I sites as a function of their separation (n) is defined¹⁷

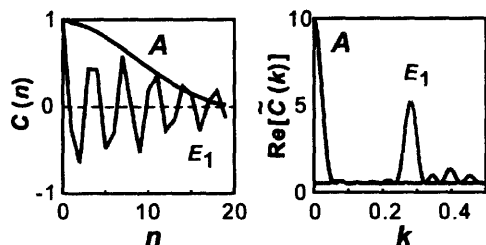
$$C(n) = \left\langle \sum_{j=1}^{N-n} A_j A_{j+n} \right\rangle / \left\langle \sum_{j=1}^N A_j^2 \right\rangle, \quad (2.8)$$

Here N is the number of oscillators, n takes on values from 0 to $N-1$, and A_j is the ϕ_1 element of the j^{th} oscillator. The brackets represent the averaging performed over disorder in the site energies. Additionally, we can characterize the phase relationship with displacement between oscillators through a discrete Fourier transform

$$\tilde{C}(k) = \sum_{n=0}^{N-1} e^{2\pi i k n} C(n) \quad (2.9)$$

where the wavevector $k = 0$ to $1/2$. Since the correlation function is the sum of the cosine functions that peaks at $n = 0$, only the real part of the Fourier transform is meaningful.

(a) α -helix



(b) Antiparallel β -sheet

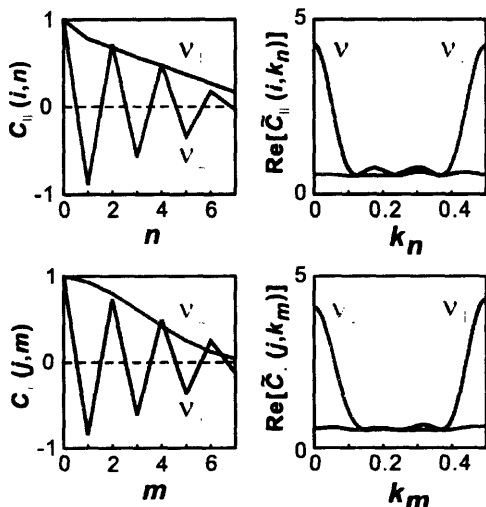


Figure 2-3. Amide I oscillator correlation functions (Left) and their Fourier transforms (Right). (a) The correlation function of the A mode of the ideal α helix of Fig. 2-2b decreases smoothly while that of E_1 mode oscillates with the period of 3.6 residues as shown in the Fourier transform. A and E_1 modes are found in the doorway vibrations of the frequency ranges of 1645 cm^{-1} and 1655 cm^{-1} , respectively. (b) Correlation functions for the ideal 8×8 β sheet (Fig. 2-2a). Row correlations (top) and column correlations (bottom) of the fourth row and column of show the in-phase and out of phase natures of the two vibrations (v_{\perp} and v_{\parallel}) between the adjacent residues and strands. v_{\perp} and v_{\parallel} modes are the doorway vibrations of $< 1640 \text{ cm}^{-1}$ and $> 1680 \text{ cm}^{-1}$, respectively.

In the case of an α -helix, the index j in Eq. (2.8) is chosen to be the oscillator index along the peptide backbone. As pictured in Fig. 2-3a, the amide I A mode will be characterized by constant phase for all oscillators and its correlation function $C_A(n)$ decays monotonically with oscillator separation for a finite helix. The correlation function for the E_1 mode $C_{E_1}(n)$ oscillates at a periodicity given by the pitch of the helix. These modes are clearly distinguished in the Fourier transform $\tilde{C}(k)$, where \tilde{C}_A and \tilde{C}_{E_1} are seen to peak sharply at $k_A = 0$ and $k_{E_1} = 5/18$, respectively.

For the characterization of the delocalization of a particular mode, a correlation length can be defined as

$$L = \frac{2}{N+1} \sum_{n=0}^{N-1} C(n) = \frac{2}{N+1} \tilde{C}(0). \quad (2.10)$$

The factor $2/(N+1)$ normalizes to the number of oscillators and dictates that the value of L varies from 1 for totally delocalized states to 0 for large chains with uncorrelated vibrational excitations. Generally, this expression only gives an accurate estimate for oscillators that are entirely in-phase, as for the A mode. In the case of a periodically varying phase such as the E_1 mode, the correlation length is more properly evaluated from the amplitude of the correlation function at the mode wavevector k_{E_1}

$$L_{E_1} = \frac{2}{N+1} \operatorname{Re} \left[\sum_{n=0}^{N-1} 2e^{-2\pi i k_{E_1} n} C(n) \right] = \frac{4}{N+1} \operatorname{Re} \left[\tilde{C}(k_{E_1}) \right]. \quad (2.11)$$

Since we use a symmetric correlation function $C(n) = C(-n)$, the normalization factor is twice compared to Eq. (2.10) to account for the symmetric peak amplitude at k_{E_1} and $-k_{E_1}$ in $\tilde{C}(k)$.

For the correlation function of the finite idealized α helix of 21 residues (Fig. 2-2b), we find $L = 0.95$ for the A mode and $L(k = 1/3.6) = 0.98$ for the E_1 mode, both of which are close to the maximum correlation length. These correlation lengths also indicate the degree of α helix mode character for the A and E_1 modes can be defined as

$$f_A = \bar{L}(k_A) \quad (2.12)$$

$$f_{E_1} = \bar{L}(k_{E_1})$$

Here, \bar{L} indicates an average over several helices in one protein.

In the case of β sheets, a correlation function in one dimension is inadequate to characterize the symmetry of the vibration, since v_{\perp} and v_{\parallel} character is distinguished by the phase relationship in two dimensions. Drawing on the grid representation of the β sheet doorway modes shown in Fig. 2-2a, we can investigate the spatial correlation function as a function of separation along rows (intra-strand) and columns (inter-strand). If the amplitude of an oscillator in a particular row and column is $A_{i,j}$, then the correlation function for a row i parallel to the β strands is

$$C_{\parallel}(i, n) = \left\langle \sum_{j=1}^{N_i-n} A_{i,j} A_{i,j+n} \right\rangle / \left\langle \sum_{j=1}^{N_i} A_{i,j}^2 \right\rangle. \quad (2.13)$$

N_i is the number of oscillators in the row and n is the separation. A similar correlation function $C_{\perp}(j, m)$ can be defined for a column j perpendicular to the strands. Fig. 2-3b shows the correlation functions for the v_{\perp} and v_{\parallel} modes of a large sheet of 64 amino acids (8×8 oscillators). Their Fourier transforms are peaked at $k_n = 0.5$ and $k_m = 0$ for v_{\perp} and $k_n = 0$ and $k_m = 0.5$ for v_{\parallel} .

In analogy to Eq. (2.10), we also define the correlation length L_i for a given row i , and the correlation length averaged over M rows

$$\bar{L}_{\parallel}(k) = \frac{1}{M} \sum_{i=1}^M L_i(k) \quad (2.14)$$

Here the correlation length $L_i(k) = 2\tilde{C}(k)/(N+1)$ is evaluated at the appropriate mode wavevector. A similar definition for \bar{L}_{\perp} can be obtained from the average over the column correlation lengths of N columns.

As a working definition, the fractional v_{\perp} and v_{\parallel} character can be obtained by multiplying the average row and column correlation lengths with a mode wavevectors of $k =$

0 and 0.5 to account for the delocalization and phase information along both the row and column direction:

$$\begin{aligned}
 f_{\perp} &= \bar{L}_{\parallel}(k=0.5) \cdot \bar{L}_{\perp}(k=0) \\
 f_{\parallel} &= \bar{L}_{\parallel}(k=0) \cdot \bar{L}_{\perp}(k=0.5)
 \end{aligned}
 \tag{2.15}$$

Finally, while our analysis was performed for antiparallel β sheets, the assignment of modes and spatial correlation functions for v_{\perp} and v_{\parallel} character apply equally well to the analysis of IR spectra of parallel β sheets.

2.3.3. IR spectra

Infrared absorption spectra were calculated for four proteins with varying secondary structural content, depicted in Fig. 2-4a. The LAH was constructed drawing on structures from the Protein Data Bank. Concanavalin A (Con A, PDB id 1JBC²⁶) is a β protein with two large extended β sheets. Ribonuclease A (RNase A, 7RSA²⁷) and ubiquitin (1UBQ²⁸) are α/β proteins. Ribonuclease A has two antiparallel β sheets, whereas ubiquitin has one β sheet with mixed parallel/anti-parallel hydrogen-bonding registry. Myoglobin (1WLA²⁹) is an α protein.

Figure 2-4b shows calculated IR spectra for the four proteins. Black bars at the bottom of the each spectrum are dipole moment squares of eigenstates calculated from the crystal structures. Each eigenstate is broadened homogeneously and added to construct the IR spectrum (red curves). The green curves are spectra averaged over 10,000 disorder realizations. Although the calculated spectra do not show quantitative agreement with the experimental spectra (blue curves), the relatively simple models used in this chapter reproduce many of the trends in peak position, linewidth, and lineshape successfully. Perhaps the biggest deviation is the low frequency region of Con A, for which a distinct resonance in the calculation is not observed in the experiment. This results from an overestimation of short range TDC couplings in β sheets. Changes to the LAH model that would improve the comparison are possible, and the parameterization of site frequencies, couplings, and influence of structural or dynamic disorder is presently a topic of considerable research.^{4,30-32}

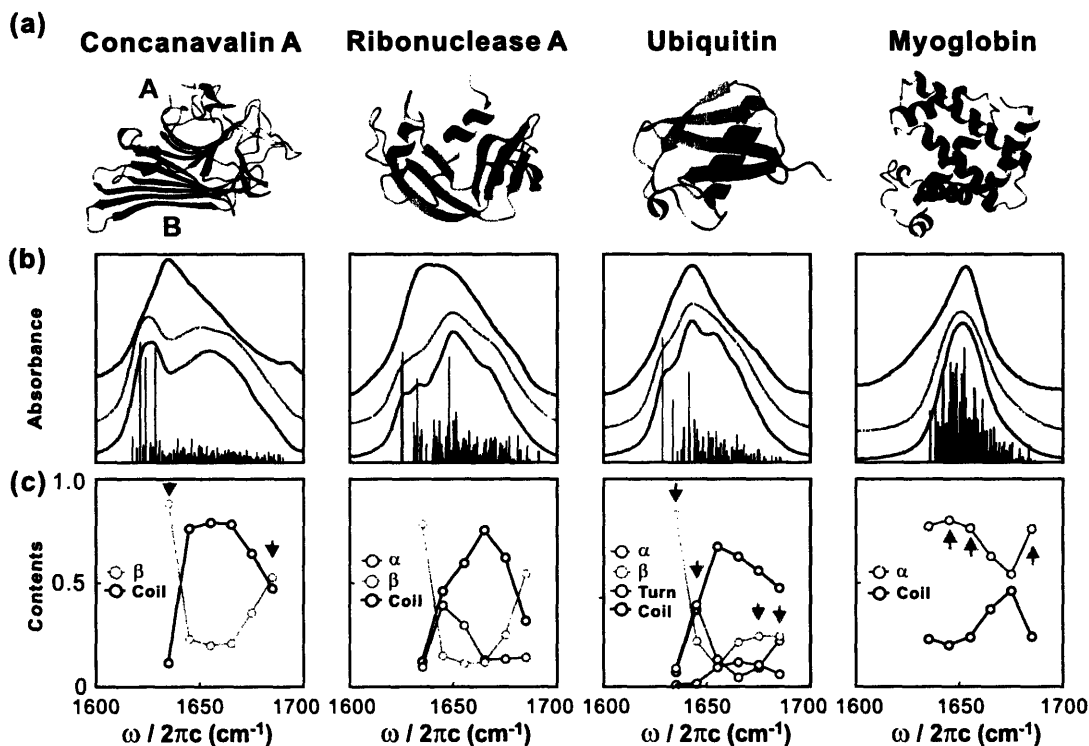


Figure 2-4. (a) Ribbon diagrams of the structures for the proteins investigated. (b) Experimental (blue) and calculated (red, green) infrared spectra of proteins. Averaged spectra with diagonal disorder calculations (green) show smoother features compared to those calculated without disorder (red). Black bars are the square of the dipole strength of amide I eigenstates. (c) Secondary structure content to the doorway modes. β sheet has high contents in the red and blue side of the spectrum. The turn and coil regions lie in the middle and α helix show high probability in the red side between random coil region and β sheet region. Blue and orange arrows indicate the doorway modes where the vibrational phases of β sheets and α helices are analyzed.

For Con A, because of the large content and size of β sheets, the spectrum is asymmetric and the splitting between ν_{\perp} and ν_{\parallel} modes is large as explained in III A. Several eigenstates with very large dipole moments in the red side have the ν_{\perp} vibration character. For the mixed α/β proteins, RNase A and ubiquitin spectra show similar asymmetric shape with less splitting between the two modes, and the more dipole strength in the middle of the amide I band originated from α helix, turn, and random coil structures. The small splittings between ν_{\perp} and ν_{\parallel} modes arise from the small size of β sheet. The myoglobin spectrum shows a more symmetric one-peak structure with a much narrower width.

The secondary structure contents for the doorway modes are shown in Fig. 2-4c. These calculations confirm many of the empirical relations between amide I frequency and

secondary structure content. The β sheet content is particularly high to the red of the absorption ($<1640\text{ cm}^{-1}$, ν_{\perp}). It also peaks to the blue side ($>1680\text{ cm}^{-1}$, ν_{\parallel}), although various structural elements also contribute. The intermediate frequency region is dominated by turn and random coil structures. The α helix region is known as to be overlapped with the random coil region; here we find that it is inclined to the red side ($1640 - 1650\text{ cm}^{-1}$) of the amide I region. The relationships between secondary structure, amide I absorption frequency, and the character of doorway modes is described in detail for three proteins below.

2.3.4. β -Protein: Concanavalin A

The visualization and analysis of amide I doorway modes for the β -protein Con A reveals many similarities to the idealized β sheet. Based on the structure assignments, we expect that doorway modes for the lowest ($< 1640\text{ cm}^{-1}$) and highest ($> 1680\text{ cm}^{-1}$) frequency range, marked with the blue arrows in Fig. 2-4c, correspond to amide I modes with ν_{\perp} and ν_{\parallel} symmetry. Figure 2-5a shows a visualization of these doorway modes on the crystal structure and the two-dimensional projection of the amide I oscillators within the larger β -sheet B. We note the constant phase relationship perpendicular and parallel to the strands of both sheets. We here focus on the analysis of oscillators within sheet B, but the results are similar for sheet A. These visualizations confirm the assignment of the two modes. For the ν_{\perp} mode, the projections of the vibrations show in-phase correlation across the strands (column direction) and alternating phase relation along the strands (row direction) clearly. For the ν_{\parallel} mode, the in-phase and alternating phase relationships hold for the row and column directions, respectively. The ν_{\perp} mode is clearly delocalized over the entirety of both sheets, whereas the amplitude of oscillators in the ν_{\parallel} mode appears patchier and concentrated on the first four strands of sheet B and much of sheet A.

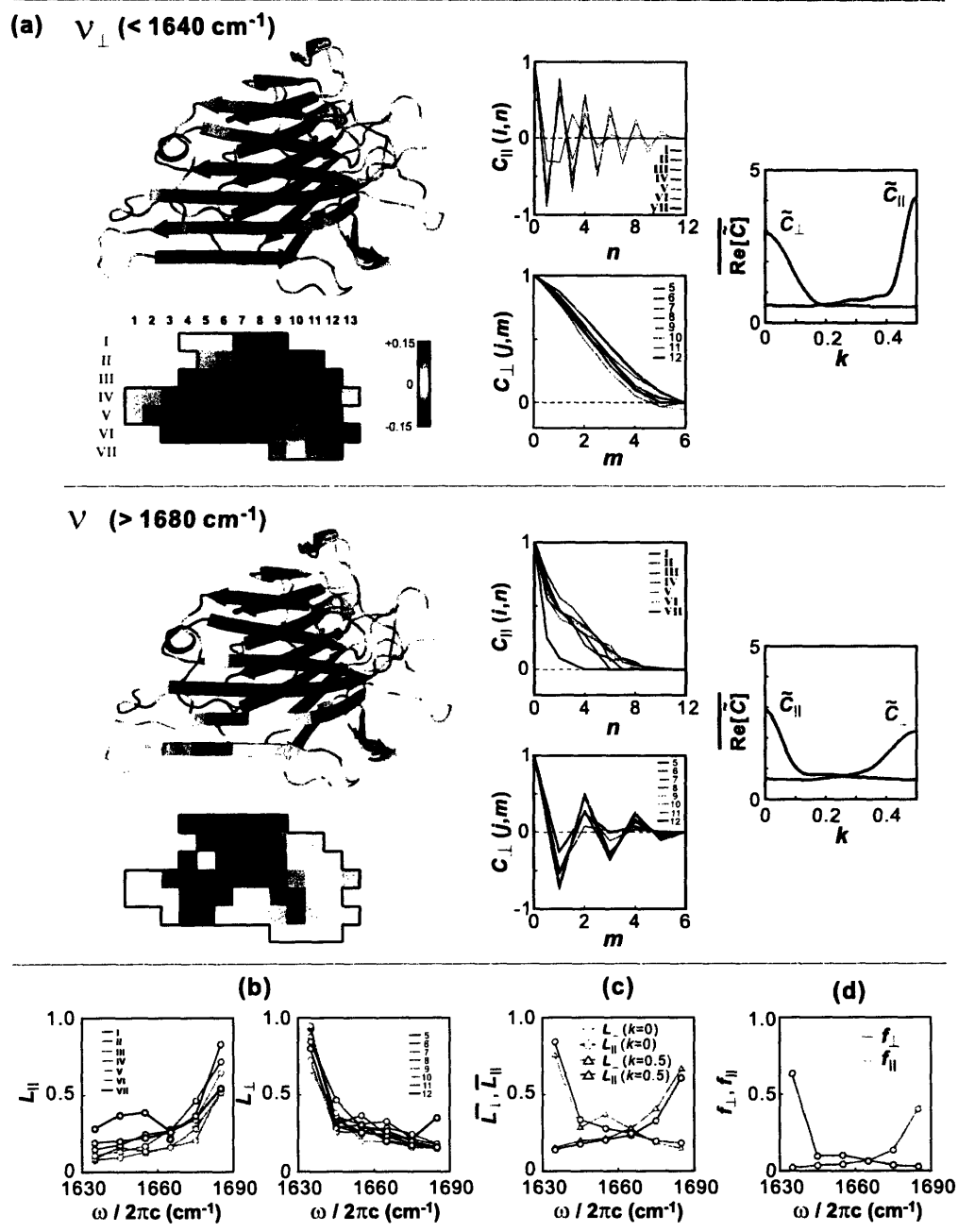


Figure 2-5. Visualization and characterization of the ν_{\perp} and ν_{\parallel} doorway modes for concanavalin A. (a) Vibrational amplitudes and phases of the ν_{\perp} and ν_{\parallel} modes visualized on the crystal structure. The contribution of sheet B is projected onto a two-dimensional grid representing the spatial placement of oscillators within the sheet. Row (m) and column (n) correlation functions, and the Fourier transform of the average are presented. (b) Row and column correlation lengths, (d) average correlation lengths, and (e) fractional ν_{\perp} and ν_{\parallel} mode characters are shown as a function of doorway mode frequency.

Figure 2-5a also shows the oscillator correlation functions of sheet B for the v_{\perp} and v_{\parallel} modes and their Fourier transforms. The v_{\perp} mode row correlation function, C_{\perp} oscillates with every other residue spacing, whereas the column correlation function, C_{\parallel} decreases monotonically. For the v_{\parallel} mode, the row correlation function oscillates and the column correlation function decreases monotonically. Therefore, C_{\perp} and C_{\parallel} of the v_{\parallel} mode resemble C_{\parallel} and C_{\perp} of the v_{\perp} mode, respectively. The overall effect is seen in the averages of the real part of the Fourier transform of rows (\tilde{C}_{\parallel}) and columns (\tilde{C}_{\perp}), for which peaks at $k = 0$ and 0.5 represent in-phase and oscillating features of vibrations, respectively.

From Eq. (2.14), we see that the magnitude of \tilde{C}_{\parallel} and \tilde{C}_{\perp} at $k = 0$ and 0.5 is proportional to the average correlation length along rows or columns, for instance $\bar{L}_{\parallel} = \tilde{C}_{\parallel}(k = 0)$. As can be seen in Fig. 2-5a, the column and row correlation lengths of v_{\perp} mode are longer than those of the v_{\parallel} mode. The results for the ideal β sheet shown in Fig. 2-3b give a similar correlation length both for the v_{\perp} mode and for the v_{\parallel} mode, implying that an irregular phase relationship presents itself more in the v_{\parallel} mode than that of v_{\perp} mode. The origin of the shorter correlation length for v_{\parallel} is the preferential localization of this mode in the upper left part of the phase map of sheet B. The amplitude of v_{\parallel} mode is low in the lower right and the phase relationship is broken while the amplitude and the phase correlation is almost perfect in v_{\perp} mode, which results the fast decays of the envelope of C_{\perp} of column 10, 11, 12 of v_{\parallel} mode. This selective disruption of the vibrational phase results mostly from the mixture of β -sheet oscillators with those in coiled regions in the v_{\parallel} transition. This is not the case for the v_{\perp} mode is located in the far red side of the spectrum and is localized to the β sheet as indicated by the high β content in Fig. 2-4c. In addition, the structure of the lower right region is bent (see Fig. 2-5a) and almost perpendicular to the center of the sheet. The modification of the couplings between the oscillators of this varied structure also disrupts the phase relationship

The row and column correlation lengths provide measures of the v_{\perp} and v_{\parallel} character of the β -sheet vibration. Figure 2-5b shows these as a function of amide I frequency for β sheet B. As expected, the row and column correlation length reach maxima on the blue ($> 1680 \text{ cm}^{-1}$) and the red side ($< 1640 \text{ cm}^{-1}$) of the spectrum which are dominated by the v_{\parallel} and

v_{\perp} mode, respectively. The irregularity of the phases in the lower right region of the sheet leads to short row correlation lengths for strands IV – VI.

Characterization of the mode character and degree of delocalization of β -sheet vibrations requires a phase-sensitive characterization in two dimensions. To account for the constant phase in one dimension and alternating phase in the other dimension for v_{\perp} and v_{\parallel} modes, we draw on the average correlation length \bar{L} parallel and perpendicular to the strands evaluated at the wavevectors $k = 0$ and $k = 0.5$. These are shown as a function of amide I doorway mode frequency in Fig. 2-5c. As expected, $\bar{L}_{\perp}(k = 0)$ and $\bar{L}_{\parallel}(k = 0.5)$ decrease and $\bar{L}_{\parallel}(k = 0)$ and $\bar{L}_{\perp}(k = 0.5)$ increase as the frequency becomes higher. The resemblance of $\bar{L}_{\perp}(k = 0)$ and $\bar{L}_{\parallel}(k = 0)$ with $\bar{L}_{\parallel}(k = 0.5)$ and $\bar{L}_{\perp}(k = 0.5)$ indicates the similar degree of the correlation of the vibrations along both the row and the column direction. Equation (2.15) relates the product of \bar{L}_{\parallel} and \bar{L}_{\perp} , evaluated at the appropriate wavevector, to the β -sheet character of the doorway mode f . f_{\perp} and f_{\parallel} are shown in Fig. 2-5d. It is clear that the two β sheet vibrational characters are separated to the red and the blue side. On the other hand, the middle of the spectrum does not show any of these characters. The larger value of f_{\perp} ($< 1640 \text{ cm}^{-1}$) than f_{\parallel} ($> 1680 \text{ cm}^{-1}$) implies that v_{\perp} mode is more preserved than v_{\parallel} mode as indicated by other phase properties.

2.3.5. α -Protein: Myoglobin

The vibrations of an infinite idealized α helix will have A and E_1 symmetry.^{5,6,24,33} Since all local vibrations are in-phase for the A mode, the transition dipole moment is parallel to the helical axis. The degenerate E_1 mode consists of linear combinations of vibrations with periodic phases of $2\pi n/m$, where m is the number of oscillators per turn (3.6 for α helix) and n is the index of the oscillator. For a helix of finite length, though the degeneracy of E_1 mode is broken and higher order of E type vibrations become IR active, the shape of the spectrum is determined by the most intense A mode.²⁴ Previously, a doorway mode analysis of all amide I eigenstates was performed for the α helices of myoglobin¹³ and the result shows that the first and second doorway modes had A and E_1 mode character, respectively.

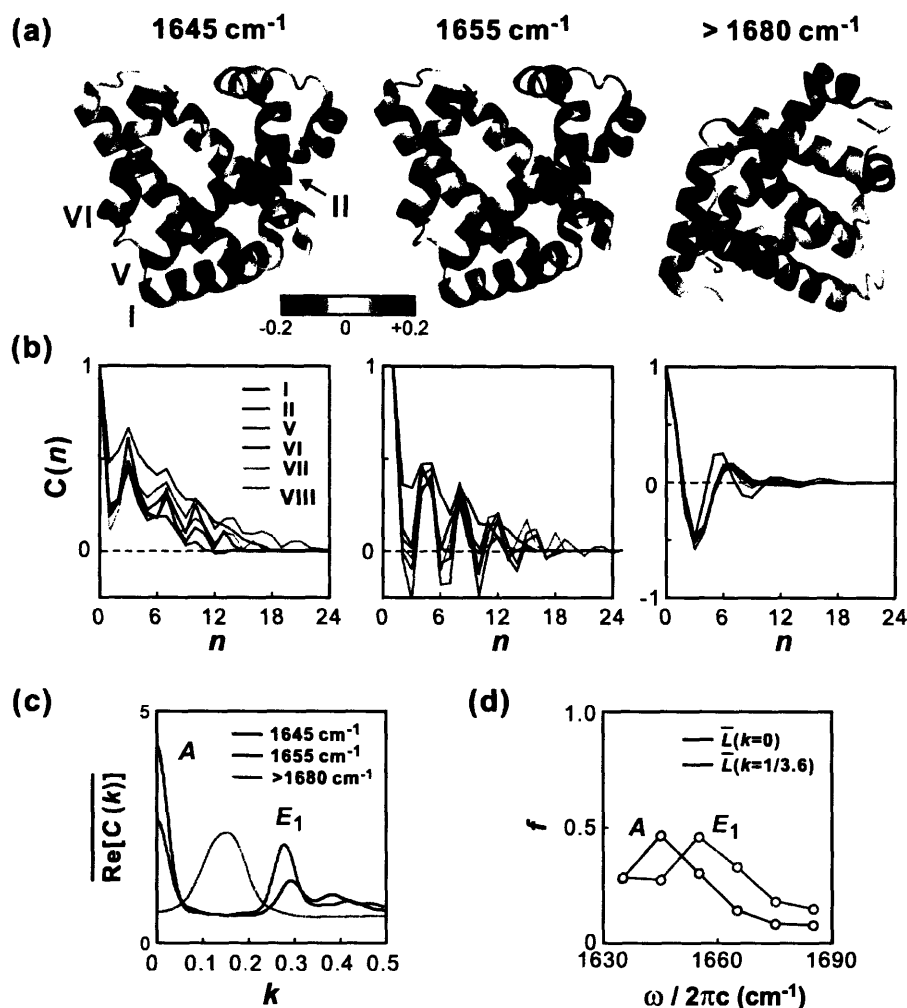


Figure 2-6. (a) Visualization and characterization of three doorway modes for myoglobin (1645 cm^{-1} , 1655 cm^{-1} , and $>1680\text{ cm}^{-1}$, marked with orange arrows in Fig. 2-4). Amplitudes and phases of oscillators are encoded on the structure, with helices numbered. (b) Helix correlation functions for amide I oscillators along the various helices. (c) The average of the Fourier transformed correlation functions for six helices are shown. (d) The fractional A and E_1 mode character as a function of amide I frequency given by the average correlation length at $k=0$ and $k=1/3.6$, respectively.

For the α -helical protein myoglobin, we analyzed the six helices (I, II, V-VIII) with more than 12 residues to investigate the frequency-structure correlation for amide I doorway modes. The results are shown in Fig. 2-6 for vibrations in the three frequency regions as marked with orange arrows in Fig. 2-4c. For the 1645 cm^{-1} doorway mode visualized on the structure of Fig. 2-6a, oscillators are in phase, although one can observe a slight periodic amplitude variation in the helix in almost every three residues in helix I, V, and VI. Similar periodicity can be found in other helices. This periodic feature characteristic of the E_1 mode

appears in the correlation function (Fig. 2-6b) atop an overall monotonic decrease characteristic of the A mode. The mixed mode character is highlighted in the Fourier transforms shown in Fig. 2-6c, which shows a high $k = 0$ peak and weaker broad peak around $k \sim 0.33$. As a result, the dipole moment of this frequency region is relatively parallel to the helical axis. The opposite sign of vibrations of the helix V arise from the opposite growing direction to those of helix I and VI. Therefore, the transition dipole moments of the three helices are relatively parallel.

The average correlation lengths representing the A and E_1 mode character are shown in Fig. 2-6d. The vibrations of 1645 cm^{-1} and 1655 cm^{-1} show the highest A and E_1 mode characters, respectively. The 10 cm^{-1} difference between these peaks matches well with the predicted and observed value of 10 cm^{-1} or less.^{18,25,34} However, compared to the smooth and oscillating correlation functions of the ideal A and E_1 mode in Fig. 2-3a, the correlation functions of 1645 cm^{-1} and 1655 cm^{-1} regions look like the mixtures of the two ideal correlation functions. We thus see a relatively low peak mode characters of ~ 0.5 compared to those of the ideal helix that exceed 0.9. This is partially a result of the variation of partial A and E_1 mode character between different helices for a given frequency range. For example, the correlation function of helix II in 1655 cm^{-1} show more A like character than most helices at 1645 cm^{-1} . The origin of this can be understood in terms of the near-perpendicular alignment of dipoles within helix II with helices I and V. Therefore, E_1 and, to a lesser extent, A character is spread over all frequencies even though the A and E_1 mode character peaks sharply at 1645 cm^{-1} and 1655 cm^{-1} respectively. Torii and Cho have found that E mode character peaks at lower amide I frequency than A character.^{13,14} This discrepancy is a result of the LAH model; both studies employed degenerate site energies and transition dipole coupling.

2.3.6. α/β -Protein: Ubiquitin

In evaluating the structure-frequency correlation for the mixed α/β protein ubiquitin, we analyzed the helix and sheet character of the amide I doorway modes. For sheets, Fig. 2-7a shows the vibrational amplitude and phase correlations for v_{\perp} and v_{\parallel} doorway modes found at the red ($<1640 \text{ cm}^{-1}$) and blue (1675 cm^{-1}) sides of the amide I spectrum. We note that in ubiquitin the β sheet is not purely anti-parallel, with strands I and V lying parallel to

each other, but the phase relationships characteristic of anti-parallel β sheets still holds. Vibrations are in-phase for oscillators aligned across strands and alternating for those along backbone for the v_{\perp} symmetry mode, and vice versa for the v_{\parallel} mode. Also, the degree of phase correlation of v_{\parallel} mode is weaker than that of v_{\perp} mode as a result of mixing with other structural components. Figure 2-7a indicates that the $<1640\text{ cm}^{-1}$ doorway mode is mostly v_{\perp} localized on the sheet, with weak mixing with an helix vibrations, whereas the 1675 cm^{-1} mode contains not only the v_{\parallel} sheet vibration, but the influence of the helix and coils. In Fig. 2-7b, average correlation lengths are shown for the β sheet, showing similar features to Con A for the $k = 0$ and $k = 0.5$ correlation lengths. Both $\bar{L}_{\perp}(k = 0)$ and $\bar{L}_{\parallel}(k = 0.5)$ for the red side v_{\perp} vibration show similar correlation lengths with that of Con A. The β -sheet correlations are distinctly weaker on the blue side of the line. This again is due to mixing and interference of β -sheet oscillators with other structures and the relatively small β sheet content of ubiquitin.

Even though only one helix is present, the α helical character of ubiquitin doorway mode vibrations shows striking similarities to myoglobin. A high fraction of A and E_1 mode character (f_A and f_{E_1}) is observed at 1645 cm^{-1} and 1655 cm^{-1} , respectively. The doorway modes for these frequency ranges are predominately localized on the helix. This is illustrated for the 1645 cm^{-1} doorway mode in Fig. 2-7c. The excitation is largely localized to the helix, although some mixing with the v_{\perp} vibration of the β sheet is apparent. The Fourier transform of the correlation function for amide I oscillators along the peptide sequence is strongly peaked at $k=0$ indicating the predominant A symmetry to the mode. Analyses of the remaining doorway modes also show similar trends and correlation functions to those seen in myoglobin. The doorway mode of the blue side ($>1680\text{ cm}^{-1}$) shown in Fig. 2-7c contains a $k = \frac{1}{2}k_{E_1}$ α -helix excitation similar to that observed in myoglobin.

From the analysis for ubiquitin, we find that β sheet and α helical characters that are found in β -sheet protein (conA) and α -helical protein (myoglobin) are preserved even though ubiquitin is α/β mixed protein. For the lower frequency transitions, the calculated doorway modes appear to be mainly localized to secondary structures, and separated by frequency. This independence may result from the small interference because of the long distance

between β sheet and α helix that is separated by the hydrophobic core. Even so, the highest frequencies of the amide I band appear to far more mixed between the sheet, helix, and random coil regions.

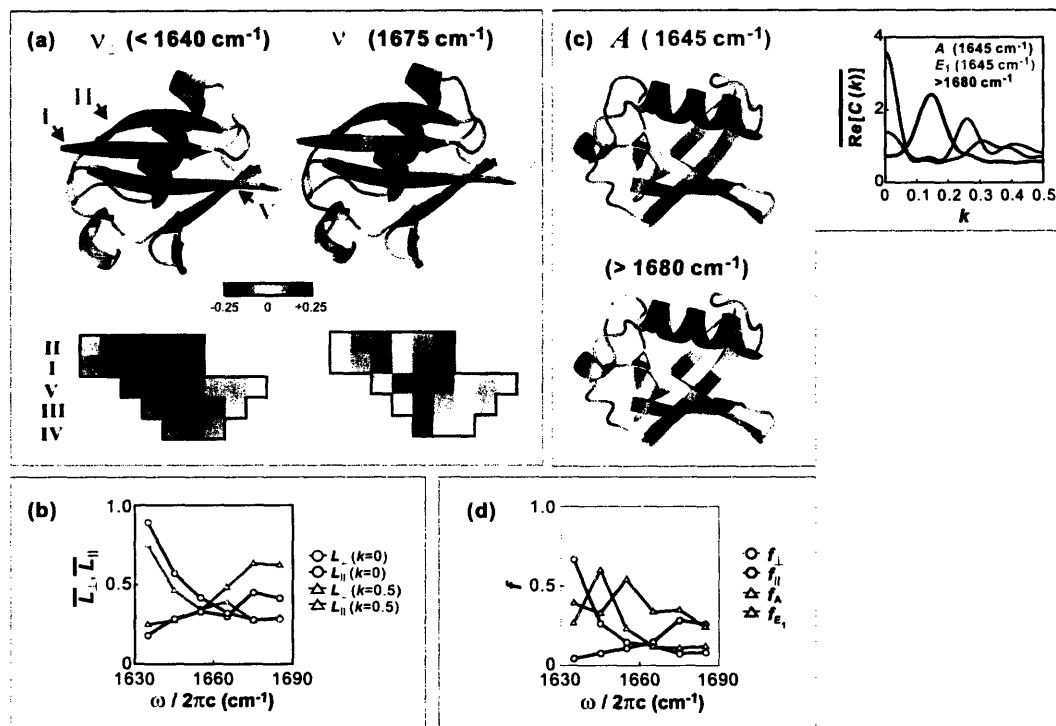


Figure 2-7. Phase correlation of vibrations of ubiquitin. (a) Amplitudes and phases of vibrations are encoded on the crystal structure and projected on a two-dimensional plane for v_{\perp} and v_{\parallel} modes of β sheet of ubiquitin. (b) Averages of the normalized correlation lengths for the β sheet are shown. (c) Vibrational amplitudes of A mode and the mode of $>1680 \text{ cm}^{-1}$ of α helix are encoded in the crystal structure. The Fourier transform shows a sharp peak at $k=0$ and close to $1/(2 \times 3.6)$, respectively. (d) v_{\perp} and v_{\parallel} character for the β sheet and A and E_1 mode character for the α helix.

2.4. Conclusions

The relationship between amide I absorption frequency and the vibrations of secondary structural elements were investigated through the spatial amplitude and phase correlations between amide I oscillators in proteins. The phases and amplitudes are extracted from a bright doorway mode, rather than the eigenstate vibrations, to more generally describe the vibrations contributing to the IR spectrum in six frequency regions. The mode character f defined with respect to a correlation length is useful to summarize the relation of specific

vibrational characters with frequency regions. Generally, the vibrational characters are not strongly localized at a particular frequency, though different modes peak at characteristic frequencies. The β sheet absorption is split into the extreme high- and low-frequency sides of the amide I band. The ν_{\perp} mode on the red side ($<1640\text{ cm}^{-1}$) is strongly localized within the β sheet while the blue side ν_{\parallel} mode is mixed with other structural components. The A and E_1 mode of the α helix are located in the 1645 and 1655 cm^{-1} regions, respectively, with a fair amount of mixing with themselves and with other components. The result for the α/β protein ubiquitin showed the preservation of the same vibrational characteristic for all the four modes. The red side is dominated by the ν_{\perp} mode of the β sheet structure and other regions result from the superposition of various secondary structure vibrational modes. As shown in the case of the helix II of the myoglobin the interference between adjacent secondary structures can shift the location of the mode characters.

Generally speaking, doorway mode vibrations and thus the IR spectrum involve vibrations that are largely localized on secondary structures at lower amide I frequencies, but involve mixtures of various secondary structure modes on the high frequency side of the line. It appears that within the constraints of our current coupling model, the four characteristic vibrational modes (ν_{\perp} , ν_{\parallel} , A , E_1) plus a random coil contribution may be an effective basis set for describing the amide I vibrations of proteins. We find that the simple superposition of the vibrational characters works well for the protein having well separated secondary structures like ubiquitin but the interference effect may not be negligible in the presence the strong tertiary contacts.

The calculation model used in this chapter is relatively simple, and several avenues for improvement exist. The unperturbed amide I frequency was chosen for the best match to the experimental spectrum and the solvent interaction is considered uniformly in the red shift of the site frequency. The ongoing work of many groups is targeted at parameterizing a quantitatively accurate LAH. Various improved models are being developed to get more accurate site frequency by considering the local electric field³⁰ or electrostatic potential^{31,32} drawing on *ab initio* calculations. For the coupling model, transition charge coupling⁴ by oscillating charges located at the atoms of peptide groups can be used to account for the short-range interactions. In addition, in the case of the real protein, dynamics are certain to be important in the spectroscopy, and the static average implied by Eq. (2.1) may not adequately

characterize the spectrum. Femtosecond to picosecond fluctuations in the solvent, side chains, and periphery of the protein could lead to a number of effects, including narrowing of the statically calculated spectrum, motional narrowing of subsets of the eigenstates we use here, and exciton relaxation within the amide I band. IR spectra calculated directly from the Fourier transform of a dipole correlation function for the LAH model, although computationally taxing, would allow nonadiabatic effects to be characterized.

Despite the needs for the improvement of the model, the correlation characteristics are still applicable for calculation based in the local vibrational mode basis. The correlation between spectral changes observed in the experiments and the changes of calculated parameters with a proper model for structural dynamics will provide a quantitative way of interpretation of the structural changes on the molecular level. The visualization and spectral characterization techniques will find considerable applicability in the interpretation of infrared and 2D IR data of proteins, and well as in the comparison of experimental and simulation studies of folding, unfolding or binding of proteins.

Acknowledgments

We thank Arend Dijkstra and Jasper Knoester for stimulating discussions and suggestions on several aspects of characterizing delocalized states.

References

- (1) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379.
- (2) Torii, H.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 81.
- (3) Ham, S.; Cho, M. *J. Chem. Phys.* **2003**, *118*, 6915.
- (4) Hamm, P.; Woutersen, S. *Bull. Chem. Soc. Jpn.* **2002**, *75*, 985.
- (5) Higgs, P. W. *Proc. R. Soc. London, Ser. A* **1953**, *133*, 472.
- (6) Miyazawa, T. *J. Chem. Phys.* **1960**, *32*, 1647.
- (7) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *J. Chem. Phys.* **2004**, *120*, 8201.
- (8) Ham, S.; Hahn, S.; Lee, C.; Kim, T.-K.; Kwak, K.; Cho, M. *J. Phys. Chem. B* **2004**, *108*, 9333.
- (9) Lee, C.; Cho, M. *J. Phys. Chem. B* **2004**, *108*, 20397.
- (10) Abramavicius, D.; Zhuang, W.; Mukamel, S. *J. Phys. Chem. B* **2004**, *108*, 18034.
- (11) Wang, J.; Hochstrasser, R. M. *Chem. Phys.* **2004**, *297*, 195.
- (12) Dijkstra, A. G.; Knoester, J. *J. Phys. Chem. B* **2005**, *109*, 9787.
- (13) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *97*, 92.
- (14) Choi, J. H.; Ham, S.; Cho, M. *J. Chem. Phys.* **2002**, *117*, 6821.
- (15) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *97*, 86.
- (16) Chung, H. S.; Khalil, M.; Smith, A. W.; Ganim, Z.; Tokmakoff, A. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 612.
- (17) Didraga, C.; Knoester, J. *J. Chem. Phys.* **2004**, *121*, 10687.
- (18) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181.
- (19) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123.
- (20) Woutersen, S.; Hamm, P. *J. Phys. : Condens. Matter* **2002**, *14*, R1035.
- (21) Torii, H.; Tatsumi, T.; Tasumi, M. *Mikrochim. Acta [Suppl.]* **1997**, *14*, 531.
- (22) Jackson, M.; Mantsch, H. H. *Crit. Rev. Biochem. Mol.* **1995**, *30*, 95.
- (23) Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469.
- (24) Nevskaya, N. A.; Chirgadze, Y. N. *Biopolymers* **1976**, *15*, 637.
- (25) Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *115*, 7737.
- (26) Parkin, S.; Rupp, B.; Hope, H. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **1996**, *52*, 1161.

- (27) Wlodawer, A.; Svensson, L. A.; Sjolín, L.; Gilliland, G. L. *Biochemistry* **1988**, *27*, 2705.
- (28) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
- (29) Maurus, R.; Overall, C. M.; Bogumil, R.; Luo, Y.; G., M. A.; Smith, M.; Brayer, G. D. *Biochim. Biophys. Acta* **1997**, *1341*, 1.
- (30) Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. *J. Chem. Phys.* **2004**, *121*, 8887.
- (31) Kwac, K.; Cho, M. *J. Chem. Phys.* **2003**, *119*, 2247.
- (32) Bour, P.; Keideling, T. A. *J. Chem. Phys.* **2003**, *119*, 11253.
- (33) Barth, A.; Zscherp, C. *Q. Rev. Biophys.* **2002**, *35*, 369.
- (34) Lee, S.-H.; Krimm, S. *Biopolymers* **1998**, *46*, 283.

Chapter 3

Projection relationships in nonlinear infrared spectroscopy

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

- “Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding,” **H. S. Chung**, M. Khalil, and A. Tokmakoff, *J. Phys. Chem. B*, **108**, 15332 (2004).

In this thesis, protein unfolding is investigated with various third-order nonlinear IR spectroscopies. Perhaps, two-dimensional infrared experiments are the most complete characterization of the nonlinear response function that carries with it the information on the molecular structure and dynamics of the system. They are therefore the most sensitive third-order probe for complex or congested spectra, such as the amide bands of proteins. However, 2D IR spectroscopy is somewhat limited to the equilibrium experiments. Although unique features of transient 2D IR spectroscopy are presented in Chapter 10, analyses remain in a qualitative level compared to other simple techniques especially for describing the protein dynamics such as timescales and relaxation shapes of population

changes. This is mainly due to the technical complicity and longer data collection time of 2D IR spectroscopy. As nonlinear probes, dispersed vibrational echo (DVE) and dispersed pump-probe (DPP) experiments should also be sensitive to much of the underlying structure and dynamics, without many of the technical challenges associated with the interferometric detection required for time-domain 2D IR experiments. DVE and DPP experiments reflect signals that arise from interactions with multiple vibrations, and projected onto one frequency axis.

In this chapter, a brief review of the relationships between 2D IR and projected spectroscopies are presented.

3.1. Third-order response function of 2D IR spectroscopy

Traditional linear infrared spectroscopy characterizes the absorption of radiation through the resonant driving of charge displacements during vibrational motion by a single infrared field. The process of absorption can be thought of as a signal field radiated by an induced polarization (the vibrational dipoles) destructively interfering with the transmitted excitation field. Quantum mechanically, this polarization is proportional to a linear response function

$$\mathbf{R}^{(1)}(\tau) = (-i/\hbar) \langle [\mu(\tau), \mu(0)] \rangle, \quad (3.1)$$

where μ is the dipole operator. The absorption spectrum is related to the Fourier transform of the response function.

The nonlinear experiments used here can be expressed in terms of the third-order nonlinear response function that describes the vibrational dynamics of a system subject to transition dipole interactions with three sequential electric fields:^{1,2}

$$\mathbf{R}^{(3)}(\tau_3, \tau_2, \tau_1) = (-i/\hbar)^3 \left\langle \left[\left[[\mu(\tau_3 + \tau_2 + \tau_1), \mu(\tau_2 + \tau_1)], \mu(\tau_1) \right], \mu(0) \right] \right\rangle \quad (3.2)$$

The response function has information on molecular structure in the form of vibrational coordinates, couplings, and dipole orientations, and the dynamics of the system that arise from interactions with its environment. For each experiment, a sequence of three femtosecond infrared fields, separated by sequential time delays τ_1 and τ_2 , leads to a nonlinear polarization that radiates a signal into a wavevector matched direction during the detection period τ_3 . For dispersed vibrational echo (DVE) experiments, the

background-free signal is frequency dispersed, and the power in each Fourier component of the radiated signal field is detected. In pump-probe experiments, two interactions are derived from a single pump pulse ($\tau_1=0$), and a third interaction with a time-delayed probe pulse leads to the radiation of a signal collinear with the transmitted probe. The transmitted probe is frequency dispersed, and the pump-induced change in the probe intensity is characterized as a function of the detection frequency.

For 2D IR experiments, the radiated signal field during τ_3 is characterized through heterodyne detection for repeated measurements as a function of τ_1 . A two-dimensional Fourier transform leads to a 2D IR spectrum in the conjugate variables (ω_1, ω_3) which respectively represent those vibrational transitions that were initially excited and then ultimately detected. The 2D IR correlation spectrum $\tilde{\mathbf{S}}_{\mathbf{C}}$ is obtained from the sum of two complementary spectra, the rephasing and non-rephasing spectra, $\tilde{\mathbf{S}}_{\mathbf{C}} = \tilde{\mathbf{S}}_{\mathbf{R}} + \tilde{\mathbf{S}}_{\mathbf{NR}}$.^{2,3} For rephasing signals, the phase acquired by coherences during the evolution period, $e^{i\Omega\tau_1}$, is the conjugate of that for the detection period, $e^{-i\Omega\tau_3}$, whereas coherences in non-rephasing measurements evolve with the same phase during both τ_1 ($e^{-i\Omega\tau_1}$) and τ_3 ($e^{-i\Omega\tau_3}$). The 2D IR spectra of rephasing and nonrephasing signals have phase-twisted lineshapes, elongated along the diagonal and anti-diagonal directions, respectively, but the 2D IR correlation spectrum has proper absorptive 2D line shapes.^{2,3} Nonetheless, the rephasing and nonrephasing spectra interpreted separately can reveal resonances hidden in the correlation spectrum, or suppress unwanted spectral information.

Each of the experiments is a measure of the third-order nonlinear signal field ($\mathbf{E}_{\text{sig}}^{(3)}$) generated by a third-order nonlinear polarization ($\mathbf{P}^{(3)}$), which is radiated into the phase-matched direction \mathbf{k}_s . In the limit of delta-function input pulses and perfect phase-matching conditions $\mathbf{P}^{(3)} \propto \mathbf{R}^{(3)}$, and the signal field can be expressed as a function of the three experimental time periods (τ_1 , τ_2 , and τ_3) following three consecutive interactions with the infrared fields:

$$\mathbf{E}_{\text{sig}}^{(3)}(\mathbf{k}_s, \tau_1, \tau_2, \tau_3) \propto i\mathbf{P}^{(3)}(\mathbf{k}_s, \tau_1, \tau_2, \tau_3) \quad (3.3)$$

The rephasing and non-rephasing signals $\mathbf{E}_R^{(3)}$ and $\mathbf{E}_{NR}^{(3)}$ differ by their wave-vectors \mathbf{k}_R and \mathbf{k}_{NR} , respectively. Note that each signal field contains both dichroic (in-phase) and birefringent (in-quadrature) contributions. For a resonant experiment, we would expect the dichroic experiment to be dominant, and we make this assumption in the following discussion.

The experimental characterization of the femtosecond nonlinear IR signal field is achieved by using a heterodyne-detection scheme, using a well-characterized local oscillator (LO) field \mathbf{E}_{LO} . For the 2D IR experiments presented here, the signal and LO fields are spectrally dispersed in a monochromator, which effectively Fourier transforms the fields in the detection time period (τ_3) to its Fourier transform pair, ω_3 . Ignoring any timing errors, the experimentally obtained 2D IR correlation spectrum can be written as:

$$\tilde{\mathbf{S}}'_C(\omega_1, \tau_2, \omega_3) \propto \text{Re } \mathcal{F} \left[\mathbf{E}_{LO}(\omega_3) \mathbf{E}_R^{(3)}(\tau_1, \tau_2, \omega_3) + \mathbf{E}_{LO}(\omega_3) \mathbf{E}_{NR}^{(3)}(\tau_1, \tau_2, \omega_3) \right] \quad (3.4)$$

where the Fourier transform is defined as $\mathcal{F}(Z) = \int_{-\infty}^{+\infty} Z e^{i\omega_1 \tau_1} d\tau_1$. In the above expression, the correlation spectrum is expressed in terms of the rephasing and non-rephasing signals, and the LO field is taken to be timed (in-phase) with the third pulse. The correlation spectrum represents only the real part of the complex 2D IR spectrum

$$\tilde{\mathbf{S}}_C(\omega_1, \tau_2, \omega_3) = \tilde{\mathbf{S}}'_C(\omega_1, \tau_2, \omega_3) + i\tilde{\mathbf{S}}''_C(\omega_1, \tau_2, \omega_3) \quad (3.5)$$

where,

$$\tilde{\mathbf{S}}''_C(\omega_1, \tau_2, \omega_3) \propto \text{Im } \mathcal{F} \left[\mathbf{E}_{LO}(\omega_3) \mathbf{E}_R^{(3)}(\tau_1, \tau_2, \omega_3) + \mathbf{E}_{LO}(\omega_3) \mathbf{E}_{NR}^{(3)}(\tau_1, \tau_2, \omega_3) \right]. \quad (3.6)$$

3.2. Projection relationships

The two-pulse DPP signal represents a heterodyne detected nonlinear signal, in which the transmitted probe field (\mathbf{E}_{pr}) acts as an intrinsic in-phase LO:

$$\tilde{\mathbf{S}}_{DPP}(\tau_1 = 0, \tau_2, \omega_3) \propto \text{Re} \left[\mathbf{E}_{pr}(\omega_3) \mathbf{E}_{PP}^{(3)}(\mathbf{k}_{pr}, \tau_1 = 0, \tau_2, \omega_3) \right] \quad (3.7)$$

Here the pump-probe signal field has both rephasing and non-rephasing contributions

$$\mathbf{E}_{PP}^{(3)}(\tau_1 = 0, \tau_2, \omega_3) = \mathbf{E}_R^{(3)}(\tau_1 = 0, \tau_2, \omega_3) + \mathbf{E}_{NR}^{(3)}(\tau_1 = 0, \tau_2, \omega_3) \quad (3.8)$$

For a two-pulse DPP experiment, the first two interactions are with the same pump pulse and $\tau_1=0$. Using the projection-slice theorem, we can relate the DPP and DVE signals to projections of the 2D IR correlation spectrum along ω_1 . It has been established that in the limit of weak sample absorbance, the dispersed pump-probe measurements can be obtained from the 2D IR correlation spectrum by projecting it onto ω_3 :⁴

$$\tilde{\mathbf{S}}_{\text{DPP}}(\tau_2, \omega_3) = \frac{\mathbf{E}_{\text{pr}}(\omega_3)}{\mathbf{E}_{\text{LO}}(\omega_3)} \int_{-\infty}^{\infty} \tilde{\mathbf{S}}'_{\text{C}}(\omega_1, \tau_2, \omega_3) d\omega_1 \quad (3.9)$$

Here, $\tilde{\mathbf{S}}_{\text{C}}$ can be replaced either $\tilde{\mathbf{S}}_{\text{R}}$ or $\tilde{\mathbf{S}}_{\text{NR}}$ because $\mathbf{E}_{\text{R}}^{(3)}(\tau_1=0, \tau_2, \omega_3)$ and $\mathbf{E}_{\text{NR}}^{(3)}(\tau_1=0, \tau_2, \omega_3)$ in Eq. (3.8) are identical at $\tau_1 = 0$. The ratio of the \mathbf{E}_{pr} and \mathbf{E}_{LO} spectra in Eq. (3.9) would cancel only in the limit of weak absorption and matched dispersion in the sample and LO arms of the interferometer. Eq. (3.9) forms the basis for “phasing” methods used to correct for timing uncertainties in our heterodyne-detected experiments.⁴

Similar relationships exist for the DVE measurement, a homodyne signal given as the square of the third-order nonlinear signal field. For the case presented here where $\tau_1 = 0$, the signal is a degenerate four-wave mixing signal in which all excitation pulses are timed to one another, so that both rephasing and non-rephasing signals contribute

$$\tilde{\mathbf{S}}_{\text{DVE}}(\tau_1=0, \tau_2, \omega_3) \propto \left| \mathbf{E}_{\text{R}}^{(3)}(\tau_1=0, \tau_2, \omega_3) + \mathbf{E}_{\text{NR}}^{(3)}(\tau_1=0, \tau_2, \omega_3) \right|^2 \quad (3.10)$$

Since the DVE is a homodyne measurement, the DVE spectrum can be related to the absolute value square of the projection of the complex 2D IR spectrum along ω_1

$$\tilde{\mathbf{S}}_{\text{DVE}}(\tau_1=0, \tau_2, \omega_3) \propto \frac{1}{I_{\text{LO}}(\omega_3)} \left| \int_{-\infty}^{\infty} \tilde{\mathbf{S}}'(\omega_1, \tau_2, \omega_3) + i\tilde{\mathbf{S}}''(\omega_1, \tau_2, \omega_3) d\omega_1 \right|^2 \quad (3.11)$$

In this equation, the normalization by the intensity spectrum of the local oscillator corrects for the additional convolution with the local oscillator field used in measuring $\tilde{\mathbf{S}}$. As the case of DPP in Eq. (3.9), $\tilde{\mathbf{S}}$ can be either $\tilde{\mathbf{S}}_{\text{R}}$ or $\tilde{\mathbf{S}}_{\text{NR}}$. However, for the 2D IR correlation data, $\tilde{\mathbf{S}}'_{\text{C}} = \tilde{\mathbf{S}}'_{\text{R}} + \tilde{\mathbf{S}}'_{\text{NR}}$ and $i\tilde{\mathbf{S}}''_{\text{C}} = i\tilde{\mathbf{S}}''_{\text{R}} - i\tilde{\mathbf{S}}''_{\text{NR}}$ due to the phase accumulation of

the opposite sign during τ_1 period ($e^{\pm i\Omega\tau_1}$) for the rephrasing and non-rephrasing measurements.

As a test of these projection relationships, DPP and DVE spectra are reconstructed from 2D IR spectra of RNase A and ubiquitin and compared with measured spectra in Fig. 3-1.

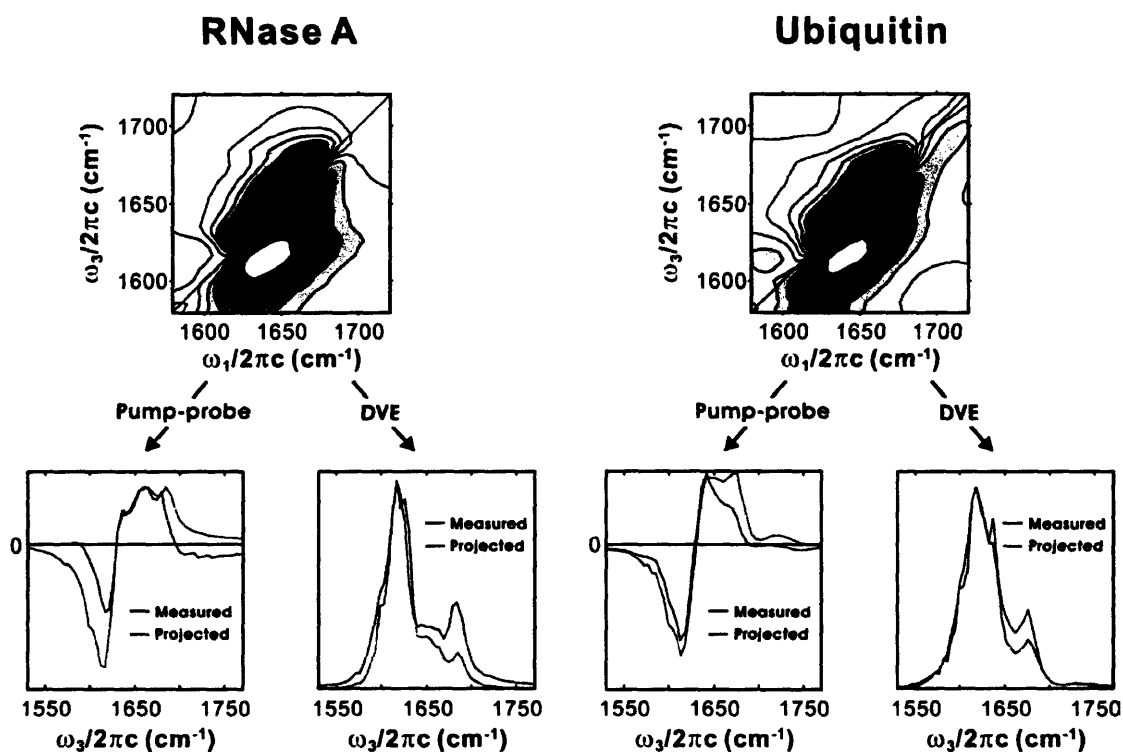


Figure 3-1. Projection relationships between the 2D IR spectrum, pump-probe, and DVE spectra. Measured and reconstructed pump-probe and DVE spectra are compared for RNase A and ubiquitin. All spectra were taken at 25°C.

References

- (1) Sung, J.; Silbey, R. J. *J. Chem. Phys.* **2001**, *115*, 9266.
- (2) Khalil, M.; Demirdöven, N.; Tokmakoff, A. *J. Phys. Chem. A* **2003**, *107*, 5258.
- (3) Khalil, M.; Demirdöven, N.; Tokmakoff, A. *Phys. Rev. Lett.* **2003**, *90*, 47401.
- (4) Gallagher Faeder, S. M.; Jonas, D. M. *J. Phys. Chem. A* **1999**, *103*, 10489.

Chapter 4

Experimental setups and data analysis

This chapter describes the experimental instrumentation and data analysis used in the following chapters. The emphasis is on the design of the transient temperature jump experiment. The laser alignment and experimental conditions are presented in each chapter since they vary somewhat. Also, the singular value decomposition (SVD) method used in data analysis of equilibrium and transient experiments is described.

4.1. Generation of infrared pulses

The layout of the laser system generating femtosecond infrared pulses is shown in the Fig. 4-1. An 82 MHz femtosecond pulse train is obtained from the mode-locked Ti:Sapphire oscillator (Tsunami, Spectra-Physics), which is pumped by the continuous-wave frequency-doubled Nd:YVO₄ laser (Millenia V, Spectra-Physics) with a power of 3.74 W at 532 nm. The oscillator pulse is centered at 790 nm with a FWHM of 16 nm. One third of the output of 400 – 450 mW (82 MHz) is used as a seed for the Regenerative Ti:Sapphire amplifier (Spitfire, Spectra-Physics). The seed pulse is picked off into the cavity by a Pockel cell (pc1) with a repetition rate of 1 kHz, which is synchronized with the Q-switched frequency-doubled Nd:YLF laser (Evolution X, Spectra-Physics) that pumps the amplifier with the power of 6.5 – 7 W at 527 nm. The amplified pulse (FWHM of 95 – 100 fs) is kicked out from the cavity by another Pockel cell (pc2) after

16 – 17 round trips (160 – 170 ns) in the cavity. An output of about 650 – 700 mW is used for the generation of the infrared pulses in the optical parametric amplifier (OPA).

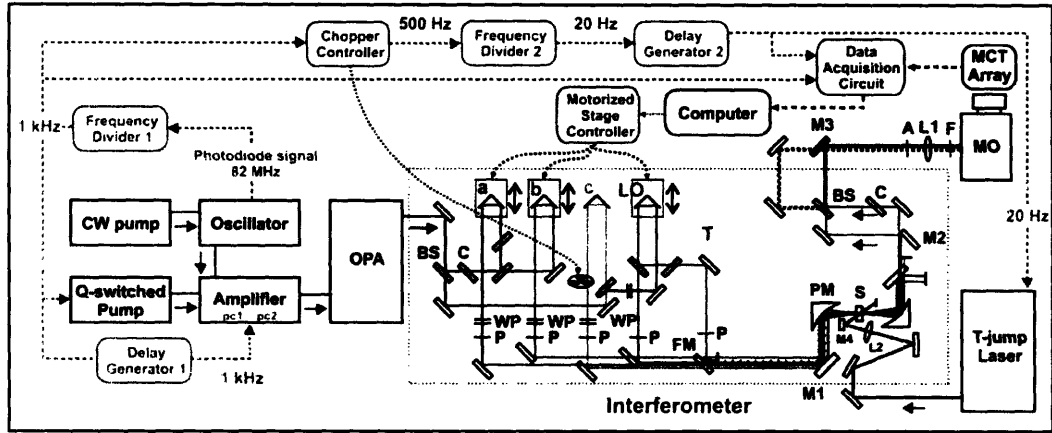


Figure 4-1. Experimental layout. Layout of the laser components, detector, controller, and flowchart of the frequency division are shown. Rectangular boxes with sharp corners represent lasers and those with rounded corners represent electronic devices including a computer and detector. The flow of the electronic triggers is represented by the dashed arrows connecting the electronic components; BS (orange rectangle bar), beamsplitter; C (green rectangle bar), compensator; LO, local oscillator; T, tracer; P (–), polarizer; WP (=), half-wave plate; FM, flipper mirror; PM, parabolic mirror; S, sample; A, analyzer; F, low-pass filter; L1, CaF₂ lens ($f = 11$ cm); L2, BK7 lens ($f = 10$ cm); MO, monochromator.

The infrared pulses are obtained by the difference frequency mixing of the signal and idler from the double stage OPA system based on the type II β -barium borate (BBO) crystal ($\theta = 27^\circ$, 3 mm-thick).¹ The seed pulse is formed by the continuum generation from the Sapphire crystal pumped with about 1% of the amplifier output. 15% and 85% of the remaining amplifier output are used to pump the BBO crystal in the first and the second stage of the amplification, respectively. After the first stage amplification, the signal is separated out and the idler is used for the second stage amplification, which gives a signal ($\lambda_s = 1.40 \mu\text{m}$, horizontal polarization) and idler ($\lambda_i = 1.82 \mu\text{m}$, vertical polarization) of 80 – 100 mW and 40 – 50 mW, respectively. The signal and idler are separated and recombined for the control of the relative timing and focused (fused silica, $f = 25$ cm) onto the difference frequency mixing crystal (AgGaS₂, type II, $\theta = 41.5^\circ$, $\phi = 0^\circ$, 8×8×1 mm, Eksma) located 5 – 10 cm away from the focal point toward the focusing lens. Finally, 90 fs (FWHM) IR pulses of 3 μJ centered at 6 μm with a FWHM of 160

cm^{-1} are generated, (Fig. 4-2) which are collimated by a KBr lens ($f = 25$ cm). Two 3 mm-thick Ge windows are used at the Brewster angle for the second order chirp compensation. These windows also remove the remaining signal and idler fields and combine the IR pulses with a He-Ne laser for the alignment. A 1:2 telescope collimates and expands the beam to 11 cm in diam. at 90% transmission.

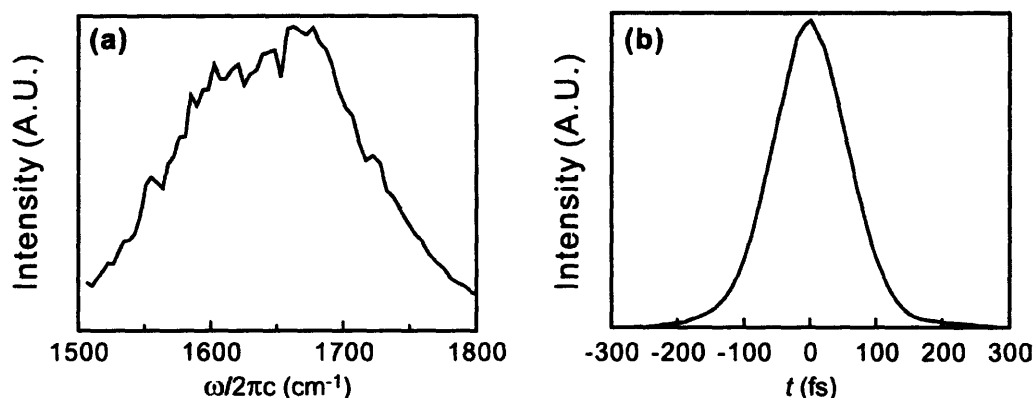


Figure 4-2. Typical characteristics of a femtosecond IR pulse. (a) Pulse spectrum and (b) cross-correlation (second harmonic generation) between beam c and tracer.

4.2. Time delay control and generation of the third order signal

A single femtosecond pulse from the OPA is divided into four identical pulses by 4 mm-thick 50:50 ZnSe beamsplitters (BS1516Z50050S, Rocky Mountain Instrument). ZnSe compensators (WI1012Z550BA2+, Rocky Mountain Instrument) are used to maintain the same chirp character between the transmitted and reflected pulses. Each femtosecond pulse travels to and is reflected back by a retroreflector (3" in diam., PLX). The positions of the retroreflectors for beam a,b, and LO (tracer) are controlled by the motorized linear stages (ANT-50L, Aerotech) to adjust the relative delay between pulses with a resolution of 10 nm (0.067 fs), an accuracy of 300 nm (2 fs), and repeatability of 50 nm (0.33 fs). Beam c is chopped at 500 Hz. The fourth beam is further split into the local oscillator (LO) and the tracer (T). The tracer beam is only sent to the sample by the flipper mirror (1" in diam., FM) when the pump-probe signal is collected. The polarization of each beam is controlled by a wire-grid polarizer (IGP229-25HER-0921, Moletron) and MgF_2 half-wave plate (MWPMFA2-22-6M, Karl Lambrecht Corp.) pair. The energy of the identical pulses a, b, and c after the polarizer and half-wave plate pair

is 0.15 μJ . All beams are focused onto the sample by a gold-coated off-axis (90°) parabolic mirror (3" in diam., $f = 10$ cm, A8037-308, Janos Technology) within the spot size of 100 – 110 μm in diam. at 90% transmission.

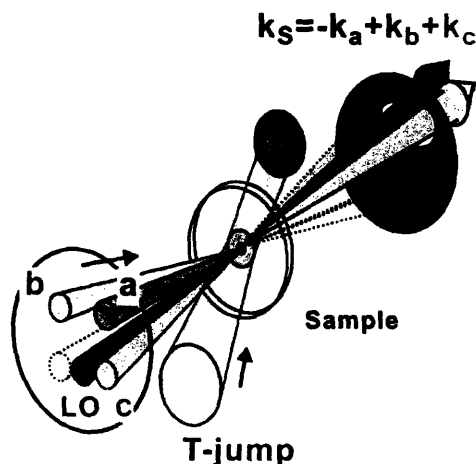


Figure 4-3. Enlarged illustration of the sample cell region. The spot size of the T-jump laser is 500 μm in diam. (90% transmission), which is larger than the spot size of the probe pulses (100 μm in diam.). The third order signal is generated by the interaction of the three femtosecond pulses a, b, and c, and emitted to the wave vector matched direction $\mathbf{k}_s = -\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c$. Three incoming pulses are blocked by the mask to reduce the scattering. The local oscillator pulse passes through the sample at the same spot as the other three pulses.

As shown in the enlarged figure of the sample cell region (Fig. 4-3), the three beams (a, b, and c) and the tracer propagate parallel to each other centered at the four corners of the 1 inch square box before being focused. The nonlinear signal field is generated in the wavevector-matched direction $\mathbf{k}_s = -\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c$, which is ideally the same direction as the tracer. The LO is sent between the tracer and beam c. Since the possible nonlinear signal propagating to the LO direction is formed by an even order process, its intensity should be minor. Also, focusing the LO onto the same spot as the other three beams (a, b, and c) may induce a pump-probe signal to the LO direction. However, we found that this signal is negligible because of the time delay of 35 ps between the LO and three beams at the focal point on the sample when the LO is overlapped temporally with the third order signal at the array detector. This delay results from the small difference in the path length between the third order signal and LO from

the sample cell to the detector. (For the equilibrium 2D IR measurements in Chapter 5 and 6, LO was not sent through the sample.)

After passing through the sample cell, the LO and the third order signal are allowed to propagate while beam a, b, and c are masked. The LO is combined with the signal by a 50:50 beamsplitter after being picked off by M2. (Fig. 4-1) The reflected third order signal and transmitted LO pair (solid lines) are focused onto the upper stripe of the dual stripe (2×64) array MCT detector (IR-0144, Infrared Systems Development) after being dispersed by a 190-mm monochromator (Triax 190, Jobin Yvon) with a 40 lines/mm grating. The other pair (dashed lines) propagate over the mirror (M3) and are focused onto the lower stripe. The analyzer (A) determines the polarization of the third order signal and the low-pass filter (F) removes the scattered T-jump pulse. A CaF_2 lens (L1, $f = 11$ cm) focuses the third order signal and LO pairs onto the slit (0.2 mm) of the monochromator.

4.3. 2D IR and DVE spectroscopy

2D IR data was taken either in the parallel (ZZZZ) or in the crossed (ZZYY) polarization geometry. Rephasing ($\mathbf{k}_s = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 \equiv \mathbf{k}_R$) and non-rephasing ($\mathbf{k}_s = \mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_3 \equiv \mathbf{k}_{NR}$) experiments are performed for $\tau_2=0$ or 100 fs by altering the time sequence of the \mathbf{k}_a and \mathbf{k}_b pulses. Arrays of dispersed heterodyned signals with a spectral resolution of $\sim 4 \text{ cm}^{-1}$ in the ω_3 dimension are collected as a function of τ_1 either in 4 fs or in 14 fs steps (undersampling) for both the rephasing and non-rephasing configurations. The data is collected for τ_1 delays ranging from 0 to 2.1 ps and from 0 to 1.2 ps for the rephasing and non-rephasing experiments, respectively. A Fourier transform along the τ_1 axis yields the individual 2D rephasing and non-rephasing spectra, and the sum of these gives the 2D IR correlation spectrum. The actual frequency ω_1 for the data collected by undersampling is obtained by reflecting the transformed frequency (ω_{1u}) to the Niquist frequency (ω_N) as

$$\omega_1 = 2\omega_N - \omega_{1u} . \quad (4.1)$$

The resolution in the ω_1 dimension after Fourier transformation is 0.5 cm^{-1} .

The treatment of the raw data to yield properly “phased” 2D IR correlation spectra that correct for timing or phase errors has been described before,² but we use a slightly different method for the data presented here. The phasing procedure involves using the dispersed pump-probe trace at $\tau_2=0$ or 100 fs (according to τ_2 of the 2D IR spectrum) to find the error in the τ_1 timing ($\Delta\tau_1$) for the rephasing and non-rephasing signals collected as a function of ω_3 and τ_1 . We make use of the fact that the slices corresponding to $\tau_1=0$ of the rephasing and non-rephasing data matrices are each equal to the dispersed pump-probe array and identical to each other.³ This allows us to correct for the finite values of $\Delta\tau_1$ for the individual rephasing and non-rephasing signals.

However, for the data obtained by undersampling, this procedure is not applicable because the τ_1 step (14 fs) is too large to find a small $\Delta\tau_1$ by an interpolation. Instead, we used the projection of the Fourier transformed 2D IR spectrum along ω_1 . The real part of the projection is the same as the slice at $\tau_1=0$ when it is phased correctly. (see Chapter 3) To find the error, $\Delta\tau_1$, we multiply the phase factor, $\exp(i\omega_1\Delta\tau_1)$ to the complex 2D IR spectrum and compare the real part of the projection to the pump-probe data. The timing errors of the rephasing and non-rephasing spectrum are found separately.

Dispersed pump-probe spectra are obtained using a tracer beam as the probe and E_c as the pump at $\tau_2=0$ or 100 fs.

The DVE measurements reflect background-free four wave mixing signals generated with time-coincident pulses ($\tau_1=\tau_2=0$). The spectra were obtained in the all-parallel (ZZZZ) polarization geometry by dispersing the homodyne signal (without LO) in the monochromator and measuring the integrated power at a given frequency component with the array detector.

4.4. T-jump laser

The T-jump prior to the transient probing is created by a nanosecond excitation of the OD stretch overtone of the D₂O molecule by the idler ($\lambda = 1.98 \mu\text{m}$) from a BBO-based optical parametric oscillator (OPO) pumped by the output of a 20 Hz, frequency-doubled Q-switched Nd:YAG laser (YG981c, OPOTEK). The vibrational relaxation on

the picosecond timescale raises the temperature of the solution on the same time scale of the T-jump pulse.⁴ The idler beam is expanded and collimated using a telescope consisting of a pair of plano-convex MgF₂ coated BK7 lenses ($f = 10$ and 15 cm). The $5 - 6$ mJ, 8 ns T-jump pulse is obtained and is focused (L2, $f = 10$ cm, BK7) to a $500 \mu\text{m}$ diameter (90% transmission) at the sample. This leads to a uniform T-jump of $\sim 10 - 12^\circ\text{C}$ in the $100 \mu\text{m}$ diameter spot probed by the mid-IR pulses. Since only $5 - 10\%$ of the radiation is absorbed by D₂O, the T-jump is also uniform along the propagation direction. The fluctuation of the energy of the T-jump pulses is 3% in standard deviation. The increase of transmission of the $6 \mu\text{m}$ probe pulse is useful both for the spatial and for the temporal overlap with the T-jump pulse. The spatial overlap is achieved simply by maximizing the transmission of the probe pulse. The rough timing between the T-jump and probe pulses is found by the fast Si photodiode (EOT) and the fine overlap is found using the rising curve of the transmission. (Fig. 4-6 (c)) The delay time 0 is set at the point where the transmission increase is half of the maximum.

4.5. Control of the delay between T-jump and probe pulses

A train of the photodiode signal of the Ti:Sapphire oscillator pulses with a 82 MHz repetition rate is used as the clock signal for the synchronization. (The first row in Fig. 4-4) The second row in the figure shows the triggering scheme of the Ti:Sapphire amplifier. The 82 MHz clock signal is divided down to 1 kHz by the frequency divider 1 (Medox, E-O Inc.), which triggers the Q-switched pump laser. After a $3.81 \mu\text{s}$ delay, the two Pockel cells in the amplifier (pc1 and pc2) are triggered successively with a relative delay of 170 ns to insert the seed oscillator pulse into the regenerative cavity and release the amplified pulse used to generate the $6 \mu\text{m}$ probe pulse in the OPA (Gaussian-shaped purple lines in the second row). It takes about 20 ns for the probe pulse to hit the sample after pc2 is fired. The delays between the triggers for the Q-switched pump laser, pc1 and pc2 are controlled by the delay generator 1 (DG535, Stanford Research Systems). In the third row, the 500 Hz electronic signal from the chopper controller (3501, New Focus), which is synchronized to the 1 kHz signal, is further divided down to 20 Hz by the frequency divider 2 (PRL-260NT, Pulse Research Lab). The flash lamp and the Q-switch

of the T-jump laser are triggered with a fixed delay of 340 μs , which causes the T-jump pulse to be always fired later than the probe pulse 1. Therefore, the next probe pulse (probe pulse 2) arriving 1 ms later is used as the first probe pulse after a T-jump. The delay, τ , between the T-jump laser pulse (Gaussian-shaped red line in the third row) and the probe pulse 2 is controlled from -100 ns to 660 μs by adjusting the delay between the trigger signals for the Q-switched pump (in the second row) and for the flash lamp (in the third row) as $660 \mu\text{s} - \tau$ with delay generator 2 (DG535, Stanford Research Systems). The overall timing jitter is ± 2 ns. In the fourth row, the probe pulse trains and the T-jump pulses are shown on the longer time scale. The details of the triggering procedure explained above are the magnification of the small region in the fourth row marked using dashed lines. Note that there are 50 probe pulses between the two T-jump pulses.

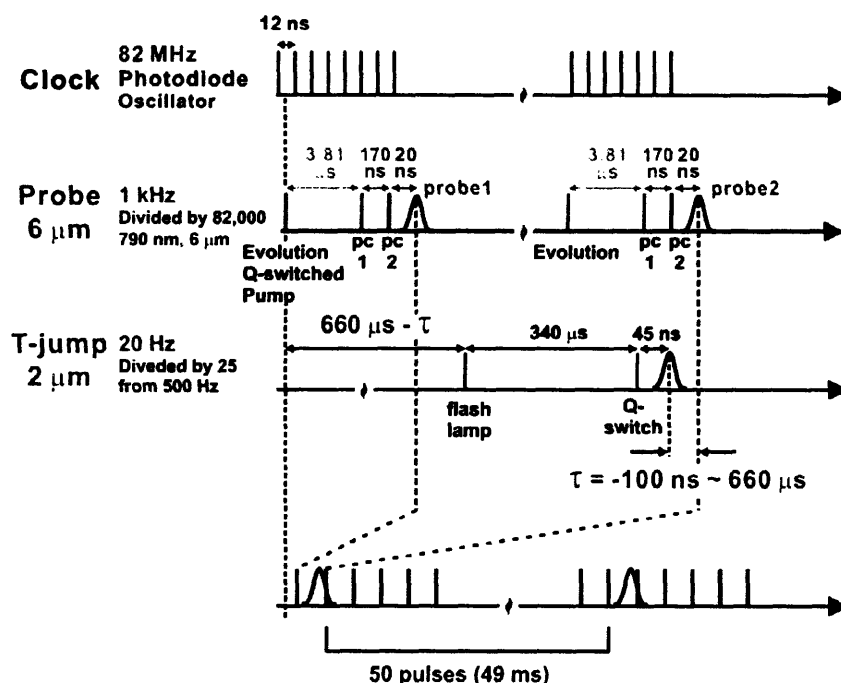


Figure 4-4. Control of the time delay. The first row is the 82 MHz photodiode signal from the Ti:Sapphire oscillator, which serves as a reference clock. The clock signal is divided down to 1 kHz to trigger the Q-switched pump laser and two pockel cells (pc1 and pc2) of the Ti:Sapphire amplifier. (second row) In the third row, the flash lamp and Q-switch of the T-jump laser are triggered at 20 Hz. The delay, τ , between T-jump pulse and the first probe pulse (probe pulse 2) is controlled from -100 ns to 660 μs . The 50 probe pulses between two T-jump pulses are shown in the fourth row.

4.6. Collection of transient data

We performed the balanced heterodyne detection using the dual stripe array (2×64). Because of the relative phase difference of 180° between the reflected and transmitted pairs of the third order signal and the LO (solid and dashed lines after the beamsplitter in Fig. 4-1), the heterodyned parts of the signal of the two stripes have opposite signs as

$$\begin{aligned} S^U &= |E_{LO} + E_s|^2 = |E_{LO}|^2 + 2 \operatorname{Re}[E_{LO}E_s^*] + |E_s|^2 \\ S^L &= |E_{LO} - E_s|^2 = |E_{LO}|^2 - 2 \operatorname{Re}[E_{LO}E_s^*] + |E_s|^2 \end{aligned} \quad (4.2)$$

where S^U and S^L are the signal intensity measured by the upper and lower stripe, and E_{LO} and E_s are the electric field of the LO and the third order signal, respectively. In principle, the LO intensity and the homodyne signal (DVE) should be canceled out by subtracting S^L from S^U in Eq. (4.2). In the real measurement, however, there is a mismatch of the LO spectrum in the two stripes which gives an offset. The fluctuation of this offset induces a fair amount of noise even after the Fourier transform. Therefore, chopping at 500 Hz is still needed to subtract LO intensity from each stripe. Then, subtracting the signals of the two stripes is effective at removing the baseline fluctuation of the heterodyned part to improve the signal to noise ratio, as shown in Section 4.9. For the DVE measurement, signals of the two stripes are identical and are simply added.

$$\begin{aligned} S^{2D} &= S^U - S^L \\ S^{DVE} &= S^U + S^L \end{aligned} \quad (4.3)$$

To obtain transient difference spectra, the 50 data sets by the $6 \mu\text{m}$ pulses after the T-jump are processed independently as shown in Fig. 4-5. The spectra by the 49th and 50th pulses are used for the reference when the temperature is completely relaxed. The first probe pulse ($n = 1$) is delayed by τ from the T-jump pulse, and thus the delay time of the following pulse n is $\tau + (n - 1)$ ms. S_0 and S_π are either the homodyne signal (S^{DVE}) or the heterodyne signal (S^{2D}) when the chopper phase is 0° and 180° , respectively. The closed or open circles indicate if the beam c is blocked or not by the chopper. When beam c is blocked, the LO is detected for the 2D measurement while background electrical noise is collected in the DVE measurement. (When the Q-switch of the T-jump laser is

fired, we found that a significant amount of electrical noise is collected through the detector cables. Shielding the cables with grounded aluminum foil removes most of the noise but a subtraction is still needed for a DVE data collection.) The difference of the open and closed pair ($S_0(\tau, n) - S_\pi(\tau, n)$) gives the transient spectrum at each delay. Finally, the transient difference spectrum is obtained by subtracting the reference spectrum as

$$\begin{aligned} \Delta S(\tau, n) &= [S_0(\tau, n) - S_\pi(\tau, n)] - [S_0(\tau, 49) - S_\pi(\tau, 49)], \quad \text{for odd } n \\ &= [S_\pi(\tau, n) - S_0(\tau, n)] - [S_\pi(\tau, 50) - S_0(\tau, 50)], \quad \text{for even } n \end{aligned} \quad (4.4)$$

The signals of open and closed circles should be subtracted by the references of open and closed circles, respectively, in the same chopper phase of the same cycle. The subtraction of the probed signal in S_0 by the reference in S_π (or vice versa) gives a lot of noise because of the drift of the laser intensity during the period of cycling the chopper phase. (several seconds) Therefore, two reference signals are needed for the even and odd n .

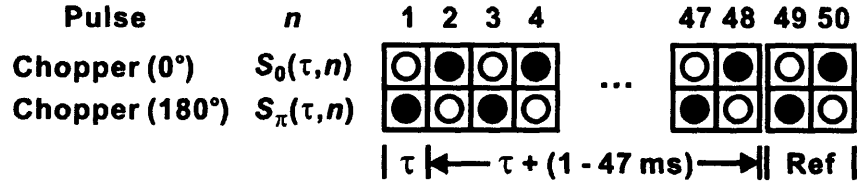


Figure 4-5. Detection scheme of transient data. A sequence of the nonlinear signals is labeled as n after T-jump. The first signal ($n = 1$) is electronically delayed by τ and the following pulses are delayed by $\tau + (n - 1)$ ms. S_0 is either the homodyne signal (DVE) or the heterodyned signal (2D) when the chopper phase is 0° . S_π is the signal when the chopper phase is 180° . Open and closed circles indicate the presence and absence of the third order signal, respectively. The difference of the open and closed pair ($S_0(\tau, n) - S_\pi(\tau, n)$) gives the transient spectrum at each delay. The 49th and 50th signals are used as reference signals for the signals of odd and even n , respectively.

For the transient non-rephasing spectrum, after collecting S_0 for 5 seconds (5000 laser shots), the chopper phase is changed to 180° and S_π is collected for 5 seconds at each τ_1 delay. The chopper phase is changed every 2500 laser shots for the rephasing spectrum. This cycle is repeated 3 times before moving the τ_1 delay. For the DVE measurement, the signals of 5000 and 2000 laser shots are collected for S_0 and S_π ,

respectively. 20 cycles are averaged for each delay τ , and the whole delay set is repeated 5 – 10 times.

4.7. Temperature-controlled sample cell

Samples were placed in a temperature-controlled brass cell (Appendix 4-A) consisting of two 1-mm-thick and 1-inch-diam. CaF₂ windows separated by a 50 μ m-thick Teflon spacer. The temperature of the cell was regulated to $\pm 0.1^\circ\text{C}$ by a circulating water bath. The quoted temperatures are those measured on the CaF₂ window by a thermocouple.

4.8. Calculation of temperature relaxation

Millisecond re-equilibration following the nanosecond temperature jump is characterized by monitoring the transmission change of a 6 μ m probe field. The result is compared with the calculation in Fig. 4-6. The coordinate system of the sample cell consisting of two CaF₂ windows is also shown. The dissipation of the heat is calculated by solving the diffusion equation numerically, (Appendix 4-B)

$$\frac{\partial}{\partial t} \Delta T(r, z, t) = \alpha (\nabla_r^2 + \nabla_z^2) \Delta T(r, z, t), \quad (4.5)$$

where $\Delta T(r, z, t)$ is the temperature difference from the equilibrium value (temperature of the brass sample cell) and α is the heat diffusivity of the material. The thermal properties of D₂O^{5,6} and the CaF₂^{6,7} window are summarized in Table 4-1.

	D ₂ O	CaF ₂
Density, d (g/cm ³)	1.09 (60°C)	3.18 (25°C)
Specific Heat, C_v (J/g·K)	4.00 (60°C)	0.854 (0°C)
Thermal Conductivity, k_c (J/cm·s·K)	6.26×10^{-3} (60°C)	9.2×10^{-2} (50°C)
Thermal diffusivity, $\alpha = k_c/d \cdot C_v$ (cm ² /s)	1.44×10^{-3}	3.4×10^{-2}

Table 4-1. Thermal properties of D₂O and CaF₂.

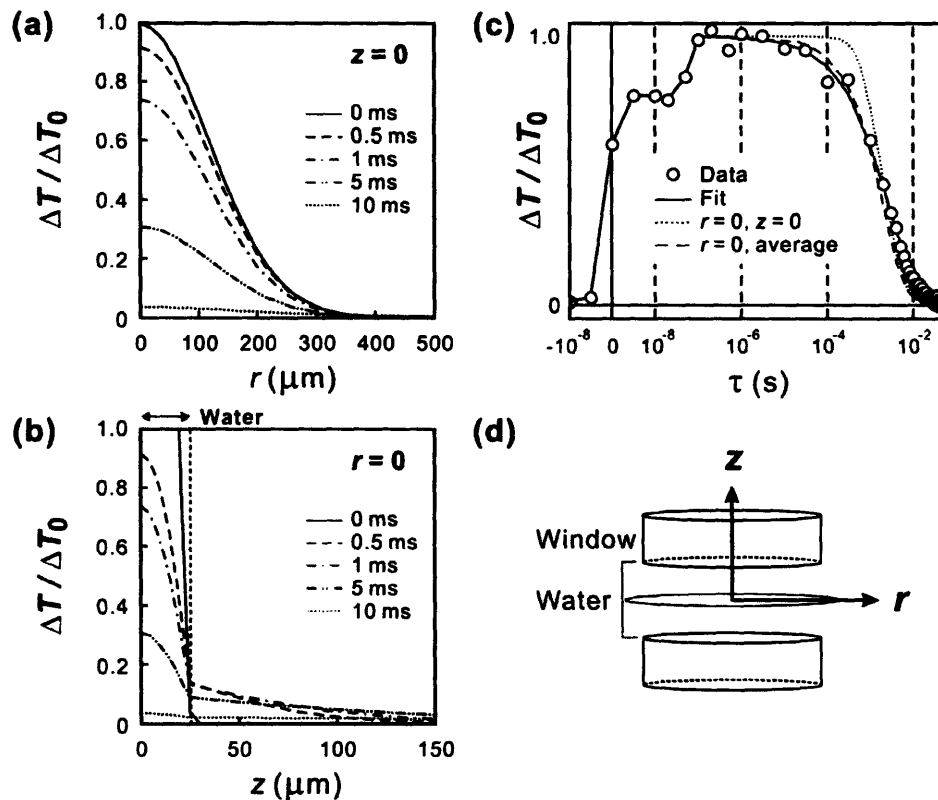


Figure 4-6. Calculation and measurement of the temperature relaxation in the sample cell. The initial radial distribution of the T-jump is given as the Gaussian profile of the T-jump laser. For the z direction, the initial temperature change is constant for water and 0 for the window. Normalized temperature changes (a) along the radial direction ($z = 0$) and (b) the z direction ($r = 0$) are shown at several time delays. (c) Measured temperature relaxation (open circles) and the stretched exponential fit (solid line, red, $\tau = 2.8$ ms, $\beta = 0.67$) are compared with the calculated relaxation profiles at the center of the sample cell ($r = 0, z = 0$, dotted line, green) and averaged over the z direction ($r = 0$, dashed line, blue). The profile is measured by the transmission change at 1622 cm^{-1} . Note that the timescale is linear from -10 ns to 10 ns and log scale afterwards. (d) Spatial coordinates of the sample cell.

The normalized distribution of the temperature difference along the radial direction (r) and the z direction are shown in Fig. 4-6 (a) and (b), respectively. The initial distribution of ΔT along r is given by the intensity profile of the T-jump laser, which is a Gaussian shape with a radius of $250\text{ }\mu\text{m}$ at 90% transmission. The initial distribution along z is assumed to be constant since the absorption of the T-jump laser by solvent is small (5 – 10%). Heat transfer from the window to the air is ignored and the temperature of the edge of the window is kept constant. Since the thermal conductivity of the D_2O is

much smaller than that of the CaF₂ window, heat is transferred from D₂O to the window along the z direction and spreads out through the window. Therefore, the shape of the distribution along the radial direction does not change but the amplitude decreases at the center of the cell ($z = 0$) in Fig. 4-6 (a).

The calculation and the experimental result of the temperature relaxation is compared in Fig. 4-6 (c). As expected from the rapid vibrational relaxation of D₂O, the temperature jumps instantaneously within 10 ns. (Note the linear scale of the time axis from -10^{-8} to 10^{-8} s and log scale afterwards.) However, there is another increase of the transmission between 10 ns to 100 ns. We believe that this is the result of a propagation of the density wave rather than the real temperature change. Since the rapid increase of the temperature changes the density of the solvent, a shock wave is formed.⁸ This wave may act as a lens to modulate the probe beam intensity with a minor spectral modulation. Since the wave will propagate at the speed of sound in D₂O, about 1500 m/s,⁹ it will take 30 – 40 ns to escape the region of the probe beam, which is a comparable timescale to the transmission change between 10 ns to 100 ns. Therefore, the data is fitted from 100 ns after the wave has propagated out of the probe region completely. A stretched exponential fit ($\exp\left[-(\tau/\tau_r)^\beta\right]$) of the experimental result gives $\tau_r = 2.8$ ms and $\beta = 0.67$. However, this result is not matched well with the calculated temperature relaxation at the center of the sample cell. ($r = 0$ and $z = 0$, dotted green line) The slower calculated relaxation induces a deviation on the hundred of μ s time scale. This mismatch can be explained by the temperature gradient along the z direction as shown in Fig. 4-6 (b). Since the probe beam also propagates along the z direction, the relaxation curve should be integrated along z direction. Then, the broad distribution of the relaxation induces the stretch of the time scale. Finally, the calculated temperature change averaged over z (dashed blue line) shows a similar stretched exponential shape.

4.9. Balanced detection

To show how effectively the balanced detection reduces the baseline noise caused by the fluctuation of the local oscillator intensity, the oscillating heterodyned signals of a

single channel ($\omega_3 = 1634 \text{ cm}^{-1}$) of the upper stripe (red solid line) and the lower stripe (green dashed line) are independently shown as a function of τ_1 in Fig. 4-7. The non-rephasing signal (ZZYY) of ubiquitin is collected at 25°C with a regular sampling step of 4 fs. Oscillations of the signal of the two individual stripes show a phase difference of 180°. The difference of the two stripes is also shown with a bold line (blue).

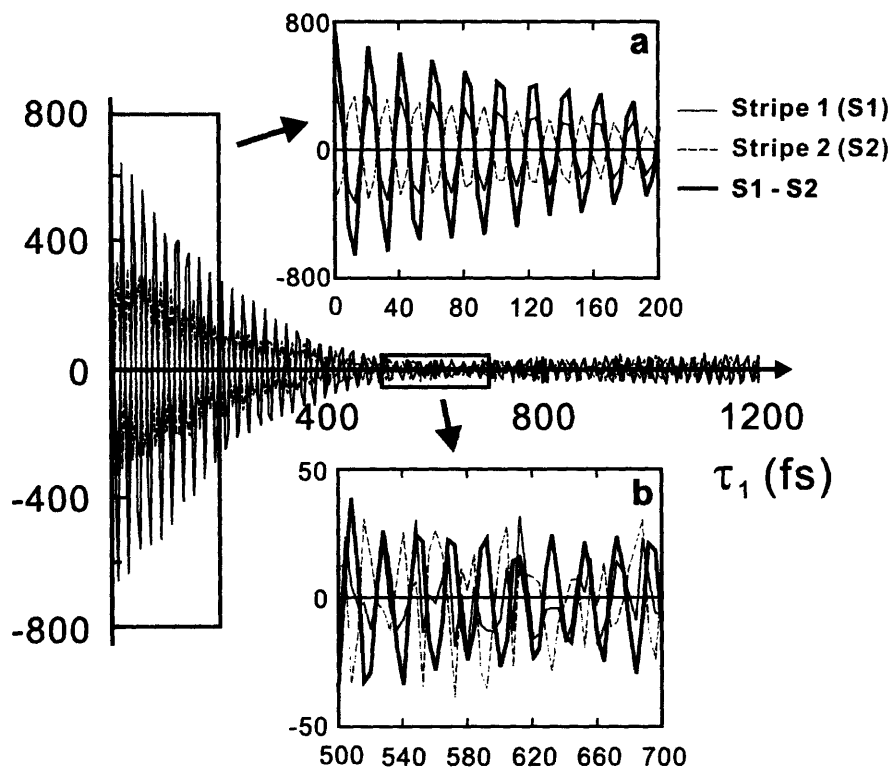


Figure 4-7. Balanced heterodyne detection. Heterodyned nonrephasing signals of ubiquitin at $\omega_3 = 1634 \text{ cm}^{-1}$ are shown along the delay between the first and second femtosecond pulses, τ_1 . Signals of the two stripes are shown with solid (red) and dashed (green) lines and the difference is shown with bold solid line (blue). Inset (a) and (b) are the magnification of the regions that are indicated with red rectangles from 0 to 200 fs and from 500 to 700 fs, respectively.

When the signal is big enough, the oscillations in both of the channels are regular and anti-correlated against each other as shown in the inset (a). As a result, the amplitude of the difference is roughly twice as much as those of the individual stripes. However, as signal intensity becomes small and similar to the magnitude of the noise level, the two oscillations don't appear to be correlated (inset (b)). The amplitude of the difference

signal is comparable to those of the individual stripes. However, the difference signal shows much more regular oscillations with a stable baseline close to 0, which means that the subtraction of the signals of the two stripes removes the baseline fluctuations. As a result, the signal to noise ratio is improved by a factor of 13. This signal to noise ratio is calculated from the comparison of the relative standard deviations of 500 shots of the individual stripes and their difference.

4-10. Future directions

Although the improvement in the signal to noise ratio by suppressing the baseline fluctuation using a dual stripe array makes it possible to resolve spectral changes of 0.5% or less in amplitude, the better time resolution and better signal to noise ratio can be achieved with further development.

The most important part of the synchronization between the 20 Hz T-jump and the 1 kHz probe is the stability of the 82 MHz reference clock signal. This implies that the stability of the Ti:Sapphire oscillator is crucial because the repetition rate is determined by the cavity length of the laser. Usually, the laser is stable when the temperature and humidity of the room are controlled well and the timing jitter is small. The 2ns jitter comes mostly from the poor electronic components of the chopper controller providing 500 Hz square pulses. Using a more stable electronic circuit, the timing error can be reduced to less than 1 ns.

Stage calibration is another way to improve the data quality. The inaccuracy of about 2 fs in the stepping of the stage will cause noise even though the period of the 6 μm pulse (20 fs) is relatively long compared to the inaccuracy. If the stepping inaccuracy is random, it will simply increase the noise in the spectrum. However, if there is drift, it distorts the lineshape of the 2D IR spectrum.¹⁰ The stage calibration using the interference fringe of a He-Ne laser will provide a better accuracy of 0.15 fs.¹⁰

As mentioned in Section 4.8, one problem in the current experiment is the temperature gradient formed in the sample cell along the z direction. This gradient produces heterogeneity of the sample species and complicates the analysis. The temperature gradient is induced by the higher thermal conductivity of the CaF₂ window compared to that of the solvent. One of the possible solutions for this problem is to use a

different window material, the thermal conductivity of which is similar to that of D₂O. However, in this case, the temperature relaxation will be slow and will not be completed within 50 ms. To prevent re-heating the sample before the temperature relaxation, the circulation of the sample is necessary though the information on the ms time scale is lost. Currently, numerous simple circulating methods for small sample volumes are available.¹¹

Finally, the usage of this technique is not limited to the temperature jump experiment. Based on the technique presented here, any combination of a nanosecond initiation using Nd:YAG laser system and nonlinear spectroscopic detection at a 1 kHz probing rate will be possible.

4.11. Singular value decomposition

Singular value decomposition (SVD) is commonly used in this thesis to determine changing spectral components in a temperature-dependent or time-dependent series of spectra. This is performed by diagonalization of rectangular spectral matrices.¹²⁻¹⁴ In this thesis, SVD is used for constructing melting curves for equilibrium thermal unfolding and obtaining relaxation profiles from transient T-jump data.

The temperature-dependent (time-dependent) spectral data matrix is built by arranging temperature (time) components in columns (n) and spectral components in rows (m)

$$\mathbf{S} = \begin{pmatrix} s(\omega_1, \zeta_1) & s(\omega_1, \zeta_2) & \cdots & s(\omega_1, \zeta_n) \\ s(\omega_2, \zeta_1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ s(\omega_m, \zeta_1) & \cdots & \cdots & s(\omega_m, \zeta_n) \end{pmatrix}, \quad (4.6)$$

where ζ is temperature T or delay time τ . Mathematically, an $m \times n$ spectral matrix \mathbf{S} can be decomposed as

$$\mathbf{S} = \mathbf{U}\mathbf{P}\mathbf{C}^T, \quad (4.7)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n]$, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_n]$, and $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_n)$, ($p_1 \geq p_2 \geq \dots \geq p_n \geq 0$). $\{\mathbf{u}_j\}$ and $\{\mathbf{c}_j\}$ are orthonormal column vector sets, the length of

which length are m and n , respectively. Then, a spectrum \mathbf{S} at temperature T or time τ can be expressed with the sum of component spectra as

$$\mathbf{S}(\zeta) = \sum_{i=1}^n c_i(\zeta) \boldsymbol{\sigma}_i, \quad (4.8)$$

where c_i is a column components of \mathbf{c}_i , and $\boldsymbol{\sigma}_i (= p_i \mathbf{u}_i)$ is the i th component spectrum. Then, a ζ -dependent profile of the i th component can be obtained by plotting $c_i(\zeta)$ versus ζ . For example, $c_i(T)$ is the temperature dependent profile and $c_i(\tau)$ is the temporal profile of the i th component.

4.11.1. Equilibrium experiments

Since the singular value p_i is sorted in decreasing order, in most cases, only the first two or three components are meaningful. In the two-state unfolding transition, particularly, the spectral changes can be expressed by a linear combination of the first and second components as

$$\mathbf{S}(T) = c_1(T) \boldsymbol{\sigma}_1 + c_2(T) \boldsymbol{\sigma}_2. \quad (4.9)$$

Here, the first component spectrum ($\boldsymbol{\sigma}_1$) represents the average spectral shape while the second component spectrum ($\boldsymbol{\sigma}_2$) represents the difference between the spectra of the folded and unfolded states. Therefore, in ideal two-state transition, $c_1(T)$ is constant and $c_2(T)$ is a sigmoid shape. The thermal melting curve in following chapters are obtained from $c_2(T)$ in the two-state transition limit.

4.11.2. Transient experiments

Transient data of T-jump experiments in following chapters are analyzed after subtracting the reference spectrum. In the SVD analysis of these transient difference spectra, the spectral change is incorporated in the first component rather than the second component because the averaged feature is already subtracted out. The higher order terms mostly contain experimental noise in the two-state transition limit as

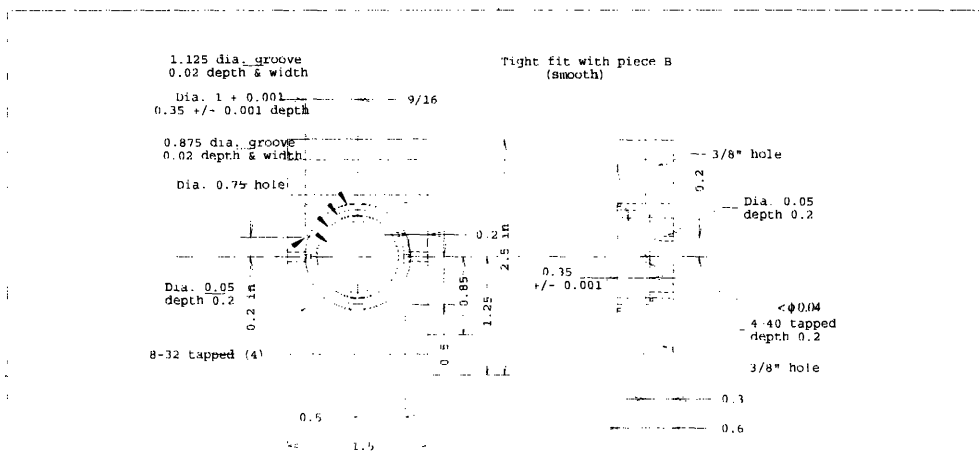
$$\Delta \mathbf{S}(\tau) = c_1(\tau) \boldsymbol{\sigma}_1 + \text{Residual}. \quad (4.10)$$

Therefore, a temporal profile can be obtained from $c_1(\tau)$.

References

- (1) Demirdöven, N.; Khalil, M.; Golonzka, O.; Tokmakoff, A. *Opt. Lett.* **2002**, *27*, 433.
- (2) Khalil, M.; Demirdöven, N.; Tokmakoff, A. *J. Phys. Chem. A* **2003**, *107*, 5258.
- (3) Gallagher Faeder, S. M.; Jonas, D. M. *J. Phys. Chem. A* **1999**, *103*, 10489.
- (4) Callender, R. H.; Dyer, R. B.; Gilmanshin, R.; Woodruff, W. H. *Annu. Rev. Phys. Chem.* **1998**, *49*, 173.
- (5) kazavchinskii, Y. Z.; Kessel'man, P. M.; Kirillin, V. A.; Rivkin, S. L.; Sheindlin, A. E.; Shpil'rain, E. E.; Sychev, V. V.; Timrot, D. L. *Heavy water Thermophysical properties*; Israel program for scientific translations: Jerusalem, 1971.
- (6) *CRC Handbook of chemistry and physics*; 87 ed.; Lide, D. R., Ed.; Taylor & Francis: Boca Raton, 2006.
- (7) Musikant, S. *Optical materials*; Marcel Dekker: New York, 1985; Vol. 6.
- (8) Wray, W. O.; Aida, T.; Dyer, R. B. *Appl. Phys. B* **2002**, *74*, 57.
- (9) Wilson, W. D. *J. Acoust. Soc. Am.* **1961**, *33*, 314.
- (10) Loparo, J. J.; Roberts, S. T.; Tokmakoff, A. *J. Chem. Phys.* **2006**, *125*, 194521.
- (11) Bredenbeck, J.; Hamm, P. *Rev. Sci. Instrum.* **2003**, *74*, 3188.
- (12) Shrager, R. I.; Hender, R. W. *Anal. Chem.* **1982**, *54*, 1147.
- (13) Hofrichter, J.; Henry, E. R.; Sommer, J. H.; Deutsch, R.; Ikeda-Saito, M.; Yonetani, T.; Eaton, W. A. *Biochemistry* **1985**, *24*, 2667.
- (14) Chen, W.-G.; Braiman, M. S. *Photochem. Photobiol.* **1991**, *54*, 905.

Appendix 4-A. Drawing of brass sample cell



Piece A Material : Brass

Tolerance : +/- 0.005

Unit : inch

TITLE

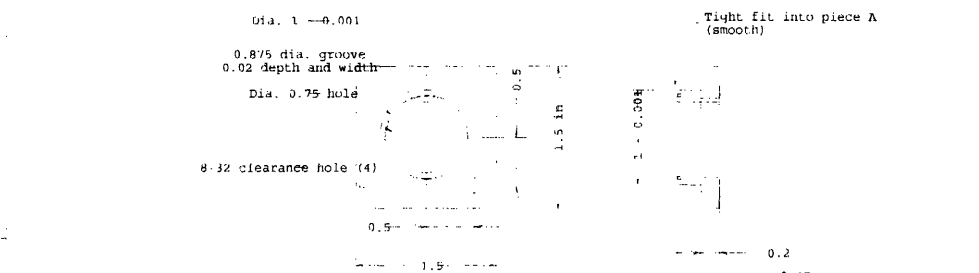
SIZE CASE CODE DWG NO

REV

A

SCALE

SHEET



Piece B Material : Brass

Tolerance : +/- 0.005

Unit : inch

TITLE

SIZE CASE CODE DWG NO

REV

A

SCALE

SHEET

Appendix 4-B. Solving heat diffusion equation

Crank-Nicholson method

The heat diffusion equation in cylindrical coordinates is

$$\frac{\partial}{\partial t}U(r, z, t) = D\nabla^2U(r, z, t) = \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} \right) U(r, z, t) + \frac{\partial^2}{\partial z^2} U(r, z, t), \quad (4.B1)$$

where $U(r, z, t)$ is the spatial and temporal temperature profile and D is the thermal diffusion constant. To numerically evaluate this differential equation, three steps are required to keep the accuracy to second order

$$U_{j,m}^{n+\frac{1}{2}} = U_{j,m}^n + \frac{1}{2} \alpha_r \delta_r^2 U_{j,m}^{n+\frac{1}{2}} + \frac{1}{2} \alpha_z \delta_z^2 U_{j,m}^n \quad (4.B2)$$

$$U_{j,m}^{n+1} = U_{j,m}^{n+\frac{1}{2}} + \frac{1}{2} \alpha_r \delta_r^2 U_{j,m}^{n+\frac{1}{2}} + \frac{1}{2} \alpha_z \delta_z^2 U_{j,m}^{n+1} \quad (4.B3)$$

$$U_{j,m}^{n+1} = U_{j,m}^{n+\frac{1}{2}} + \frac{1}{2r} \beta_r \delta_r U_{j,m}^n \quad (4.B4)$$

where $U_{j,m}^n = U(r_0 + j \cdot \Delta r, z_0 + m \cdot \Delta z, n \cdot \Delta t)$,

$$\alpha_r = D\Delta t / (\Delta r)^2, \quad \alpha_z = D\Delta t / (\Delta z)^2, \quad \beta_r = D\Delta t / \Delta r,$$

$$\delta_r^2 U_{j,m}^n = U_{j+1,m}^n - 2U_{j,m}^n + U_{j-1,m}^n$$

$$\delta_z^2 U_{j,m}^n = U_{j,m+1}^n - 2U_{j,m}^n + U_{j,m-1}^n.$$

$$\delta_r U_{j,m}^n = U_{j+1,m}^n - U_{j-1,m}^n$$

When $1 < j < j_{\text{end}}$, rearranging Eq. (4.B2) gives

$$-\frac{1}{2} \alpha_r U_{j-1,m}^{n+\frac{1}{2}} + (1 + \alpha_r) U_{j,m}^{n+\frac{1}{2}} - \frac{1}{2} \alpha_r U_{j+1,m}^{n+\frac{1}{2}} = U_{j,m}^n + \frac{1}{2} \alpha_z \delta_z^2 U_{j,m}^n.$$

When $j=1$, $\delta_r^2 U_{j,m}^n = 2(U_{j+1,m}^n - U_{j,m}^n)$ and Eq. (4.B2) is

$$(1 + \alpha_r) U_{j,m}^{n+\frac{1}{2}} - \alpha_r U_{j+1,m}^{n+\frac{1}{2}} = U_{j,m}^n + \frac{1}{2} \alpha_z \delta_z^2 U_{j,m}^n.$$

Then, the matrix representation of Eq. (4.B2) is

$$\begin{pmatrix} \alpha_r + 1 & -\alpha_r & 0 & & \\ -\frac{1}{2}\alpha_r & \alpha_r + 1 & -\frac{1}{2}\alpha_r & & \\ 0 & -\frac{1}{2}\alpha_r & \alpha_r + 1 & & \\ & & & \ddots & \\ & & & & \ddots \end{pmatrix} \begin{pmatrix} U_{1,m}^{n+\frac{1}{2}} \\ U_{2,m}^{n+\frac{1}{2}} \\ U_{3,m}^{n+\frac{1}{2}} \\ \vdots \end{pmatrix} = \begin{pmatrix} U_{1,m}^n + \frac{1}{2}\alpha_z\delta_z^2 U_{1,m}^n \\ U_{2,m}^n + \frac{1}{2}\alpha_z\delta_z^2 U_{2,m}^n \\ U_{3,m}^n + \frac{1}{2}\alpha_z\delta_z^2 U_{3,m}^n \\ \vdots \end{pmatrix},$$

where the RHS is given by the previous time step. Then U in the step $n+1/2$ can be obtained. Eq. (4.B3) can be solved with similar way. When $1 < m < m_{\text{end}}$,

$$-\frac{1}{2}\alpha_z U_{j,m-1}^{n+1} + (1 + \alpha_z)U_{j,m}^{n+1} - \frac{1}{2}\alpha_z U_{j,m+1}^{n+1} = U_{j,m}^{n+\frac{1}{2}} + \frac{1}{2}\alpha_r \delta_r^2 U_{j,m}^{n+\frac{1}{2}}.$$

When $m=1$, $\delta_z^2 U_{j,m}^n = 2(U_{j,m+1}^n - U_{j,m}^n)$ and Eq. (4-B3) is

$$(1 + \alpha_z)U_{j,m}^{n+1} - \alpha_r U_{j,m+1}^{n+1} = U_{j,m}^{n+\frac{1}{2}} + \frac{1}{2}\alpha_r \delta_r^2 U_{j,m}^{n+\frac{1}{2}}.$$

Eq. (4.B4) can be solved simply by itself with a condition when $j=1$ as given below

$$U_{1,m}^{n+1} = U_{1,m}^{n+1} + 2\alpha_r (U_{2,m}^n - U_{1,m}^n).$$

Finally, the boundary condition gives

$$U_{j,m}^n = 0, \quad j = j_{\text{end}},$$

$$U_{j,m}^n = U_{j,m-1}^n, \quad m = m_{\text{end}}.$$

Numerical code

```
% relax_r1.m -----
Calculating heat diffusion of the sample cell with 50 micron spacer between two 1" diam. and 1
mm thick CaF2 windows
% -----
clear all

tmax=0.05; % last time points: 50 ms

A=10; % Absorption coefficient (doesn't affect the result)
E0=16.8/2; % Intensity of the T-jump laser (doesn't affect the result)
w=250e-4/1.073; % 1/e^2 radius (86% transmission)
energy=2*E0*pi/2*w^2; % Energy of the T-jump laser (doesn't affect the result)
% density=1; % Density of water
density=1.0906; % Density of D2O at 60 oC
% c=4.2; % Specific heat of water
c=4.0026; % Specific heat of D2O at 60 oC
d=50e-4; % Space between windows
T0=A*E0/density/c;
```



```

Tmax=T0*(1+10^(-A*d));           % T-jump magnitude (doesn't affect the result)
kw=6.25e-3;                       % Thermal conductivity of D2O at 60 oC
kwin=9.2e-2;                       % Thermal conductivity of CaF2 window at 50 oC
denwin=3.179;                      % density of the CaF2 window at 25 oC
cwin=0.854;                        % Specific heat of CaF2 window at 0 oC
aw=kw/density/c;                   % Diffusion coefficient of water
awin=kwin/denwin/cwin;             % Diffusion coefficient of window

dr=20e-4;                          % steps in radial direction (cm)
dz=5e-4;                          % steps in z (perpendicular to the window) (cm)
dt=20e-6;                          % time step (sec)

saveid=cat(2,1:5,10:5:50,75:25:250,300:50:2500); % time points to be saved

rmax=2.54/2;                       % end point (r = 1/2 inch)
z1max=50e-4/2;                    % thickness of water /2
z2max=0.1;                         % thickness of window

r=0:dr:rmax;
z=0:dz:z1max+z2max;
lz1=z1max/dz+1;
lz=length(z);
lr=length(r);

u=zeros(lr,lz);                   % temperature matrix
for m=1:lr;                       % initial distribution (gaussian along r and const. along z, window is 0
    u(m,1:lz1)=Tmax*exp(-2*r(m)^2/w^2);
end
u(:,lz1)=u(:,lz1-1).*(aw/(aw+awin)); % temperature of the boundary of water and window

figure;
subplot(1,2,1);plot(r(1:60),u(1:60,1));axis square
subplot(1,2,2);plot(z,u(1,:));axis square

% Crank-Nicholson method -----
alphanwin=awin*dt/dr^2;
alphazwin=awin*dt/dz^2;
betarwin=awin*dt/dr;
alphanw=aw*dt/dr^2;
alphazw=aw*dt/dz^2;
betarw=aw*dt/dr;

mat1win=zeros(lr);
mat1w=zeros(lr);
for jj=2:lr-1;
    mat1win(jj,jj-1:jj+1)=[-alphanwin/2,alphanwin+1,-alphanwin/2];
    mat1w(jj,jj-1:jj+1)=[-alphanw/2,alphanw+1,-alphanw/2];
end
mat1win(1,1:2)=[alphanwin+1,-alphanwin];
mat1w(1,1:2)=[alphanw+1,-alphanw];
invmat1win=inv(mat1win(1:end-1,1:end-1));
invmat1w=inv(mat1w(1:end-1,1:end-1));

mat2=zeros(lz);
for m=2:lz1-1;
    mat2(m,m-1:m+1)=[-alphazw/2,alphazw+1,-alphazw/2];

```

```

end
mat2(1,1:2)=[alphazw+1,-alphazw];
mat2(lz1,lz1-1:lz1+1)=[-alphazw/2,1+alphazw/2+alphazwin/2,-alphazwin/2];
for m=lz1+1:lz-1;
    mat2(m,m-1:m+1)=[-alphazwin/2,alphazwin+1,-alphazwin/2];
end
mat2(end-1,end-2:end-1)=[-alphazwin/2,alphazwin/2+1];
tinvmat2=(inv(mat2(1:end-1,1:end-1)))';

kk=1;
for k=1:tmax/dt;
    u0=u;
    m=1;vec1=u0(1:lr-1,m)+(alphazw).*(u0(1:lr-1,m+1)-u0(1:lr-1,m));
    u1(1:lr-1,m)=invmat1w*vec1;
    for m=2:lz1-1;
        vec1=u0(1:lr-1,m)+(alphazw/2).*(u0(1:lr-1,m+1)-2.*u0(1:lr-1,m)+u0(1:lr-1,m-1));
        u1(1:lr-1,m)=invmat1w*vec1;
    end
    for m=lz1+1:lz-1;
        vec1=u0(1:lr-1,m)+(alphazwin/2).*(u0(1:lr-1,m+1)-2.*u0(1:lr-1,m)+u0(1:lr-1,m-1));
        u1(1:lr-1,m)=invmat1win*vec1;
    end
    u1(:,lz1)=u1(:,lz1-1).*(aw/(aw+awin))+u1(:,lz1+1).*(awin/(aw+awin));
    u1(lr,:)=0;
    jj=1;vec21=u1(jj,1:lz1-1)+(alpharw).*(u1(jj+1,1:lz1-1)-u1(jj,1:lz1-1));
    vec23=u1(jj,lz1+1:lz-1)+(alpharwin).*(u1(jj+1,lz1+1:lz-1)-u1(jj,lz1+1:lz-1));
    vec22=vec21(end)*(aw/(aw+awin))+vec23(1)*(awin/(aw+awin));
    u2(jj,:)=cat(2,vec21,vec22,vec23)*tinvmat2;
    for jj=2:lr-1;
        vec21=u1(jj,1:lz1-1)+(alpharw/2).*(u1(jj+1,1:lz1-1)-2.*u1(jj,1:lz1-1)+u1(jj-1,1:lz1-1));
        vec23=u1(jj,lz1+1:lz-1)+(alpharwin/2).*(u1(jj+1,lz1+1:lz-1)-2.*u1(jj,lz1+1:lz-1)+u1(jj-1,lz1+1:lz-1));
        vec22=vec21(end)*(aw/(aw+awin))+vec23(1)*(awin/(aw+awin));
        u2(jj,:)=cat(2,vec21,vec22,vec23)*tinvmat2;
    end
    for jj=2:lr-1;
        u(jj,1:lz1-1)=u2(jj,1:lz1-1)+(betarw/2/r(jj)).*(u0(jj+1,1:lz1-1)-u0(jj-1,1:lz1-1));
        u(jj,lz1+1:lz-1)=u2(jj,lz1+1:lz-1)+(betarwin/2/r(jj)).*(u0(jj+1,lz1+1:lz-1)-u0(jj-1,lz1+1:lz-1));
    end
    jj=1;u(jj,1:lz1-1)=u2(jj,1:lz1-1)+(betarw/dr*2).*(u0(jj+1,1:lz1-1)-u0(jj,1:lz1-1));
    u(jj,lz1+1:lz-1)=u2(jj,lz1+1:lz-1)+(betarwin/dr*2).*(u0(jj+1,lz1+1:lz-1)-u0(jj,lz1+1:lz-1));
    u(:,lz1)=u(:,lz1-1).*(aw/(aw+awin))+u(:,lz1+1).*(awin/(aw+awin));
    u(lr,:)=0;u(:,lz)=u(:,lz-1);

    if k==saveid(kk);
        timewwin(kk)=k*dt;
        k*dt
        uwwin(:,k)=u./Tmax;
        kk=kk+1;
    end
end
end

save timewwin_real.dat timewwin -ascii
save rwin_real.dat r -ascii
save zwin_real.dat z -ascii
save uwwin_real uwwin

```

Chapter 5

Nonlinear IR spectroscopy of conformational change of RNase A during thermal unfolding

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

- “Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding,” **H. S. Chung**, M. Khalil, and A. Tokmakoff, *J. Phys. Chem. B*, **108**, 15332 (2004).

5.1. Introduction

Nonlinear IR spectroscopies offer the possibility of overcoming some of the complications with traditional linear IR spectroscopy in the protein folding research. Perhaps the most general and powerful method is two-dimensional infrared (2D IR) spectroscopy, which in a manner analogous to 2D NMR spectra spreads vibrational resonances over two frequency dimensions, revealing vibrational couplings through the formation of cross peaks.¹⁻³ But simpler nonlinear IR experiments that derive from the same nonlinear response function, such the frequency-dispersed pump probe (DPP) and dispersed vibrational echo (DVE),⁴⁻⁶ should also have considerable power for revealing resonances hidden in traditional IR spectra. The investigations described in this chapter

aim to demonstrate this power of 2D IR, DPP and DVE experiments as structurally sensitive methods for the study of protein folding and denaturing. Conformational changes that accompany the thermal denaturing and folding of ribonuclease A (RNase A) are investigated with particular emphasis on the changes associated with β -sheets. 2D IR experiments are used to follow the denaturing of the protein, showing the loss of cross-peaks and other spectral features that form a “Z”-shaped contour profile characteristic of AP β -sheets. This indicates a disruption of inter-strand amide I couplings that is consistent with the melting of the β -sheet. The loss of native β -sheet structure is also revealed in DPP and DVE spectra, which are related to projections of the complex 2D IR spectrum as described in Chapter 3.

RNase A is a 124-residue enzyme which catalyzes the degradation of RNA. An abundance of detailed structural information and studies of folding exist for RNase A, making it an excellent model system to test new experimental methods for probing the structural dynamics of protein folding and unfolding. The main secondary structural elements of RNase A include a V-shaped antiparallel β -sheet with three strands on one side and four on the other, and three helices.^{7,8} Previous IR, Raman, and UV CD studies have shown that the thermal denaturation of RNase A is a reversible process with a melting temperature of approximately 63°C, and that this melting involves loss of β -sheet content.⁹⁻¹³ Previous studies have shown that the thermally denatured states of RNase A retain a sizeable amount of non-random structure.^{11,14-16} This is not altogether surprising, as the disulfide bonds, which hold the protein together, remain intact at high temperatures.¹⁷ This view is supported by the measurement of the radius of gyration of the thermally denatured states, which increases from 20 Å to 30 Å under reducing conditions.¹⁸ However, the comparison of the FTIR spectra of thermally and chemically denatured protein has shown that this residual structure is not same as the native structure.¹⁹ Recent FTIR and CD measurements have revealed that the folding/unfolding of RNase A is not a simple two-state process, but also shows a pre-transition around 47 °C.⁹ It has been suggested that the pre-transition involves the unwinding of the helix at the N terminus along with the destabilization of the β -strand involving residues 43-49.

5.2. Experimental

General schemes of data collection for 2D IR, DVE, and DPP spectroscopy are described in Chapter 4, but specific details are presented in this section. For the 2D IR measurement, dispersed heterodyned signals with a spectral resolution of $\sim 4 \text{ cm}^{-1}$ in the ω_3 dimension are collected as a function of τ_1 in 4 fs steps for both rephasing and non-rephasing configurations. The data is collected for τ_1 delays ranging from -0.06 to 2 ps and from -0.06 to 1 ps for the rephasing and non-rephasing experiments respectively. The resolution in the ω_1 dimension after Fourier transformation is 0.5 cm^{-1} . The procedures to find properly “phased” 2D IR correlation spectra are described in section 4.3.

Dispersed pump-probe spectra were obtained in the crossed polarization geometry using the tracer beam as the probe and E_3 as the pump, and setting $\tau_2=0$. By chopping the pump-beam, the differential probe transmission through the sample is observed on the 64-channel single stripe MCT array detector (IR-0144, Infrared Systems Development). The DVE measurements reflect background-free four wave mixing signals generated with time-coincident pulses ($\tau_1 = \tau_2 = 0$). The spectra were obtained in the all-parallel (ZZZZ) polarization geometry by dispersing the signal in the monochromator and measuring the integrated power at a given frequency component with the array detector. FTIR spectra were collected on a Mattson Infinity Gold FTIR spectrometer with a frequency resolution of 2 cm^{-1} .

The RNase A sample was purchased from Sigma Aldrich (R5500). It was dissolved in a 20 mM D_2O phosphate buffer solution of $\text{pH}^* = 7$ (uncorrected pH) for a final concentration of 16.5 mg/mL. This solution allowed for reversible thermal denaturing and folding of the protein when tested by FTIR. The sample had a peak OD of 0.2 for the 2D IR and dispersed pump-probe experiments, and 0.5 for the dispersed echo experiments. All the samples were incubated at 60°C for 1 hr prior to performing the experiments to allow for H/D exchange.

5.3. Results and Discussion

5.3.1. 2D IR spectroscopy of proteins with antiparallel β -sheets

Empirical relationships have existed for some time linking AP β -sheet structure in proteins with a pair of resonances in the amide I spectral region: a strong one which lies between 1610 and 1640 cm^{-1} and a weak one between 1680 and 1700 cm^{-1} .²⁰⁻²² While these modes can be clearly distinguished in β -sheet aggregates or proteins with high AP β -sheet content, they are more difficult to see in proteins where different secondary structural elements are present or where there is considerable conformational disorder within the β -sheet regions of the proteins.

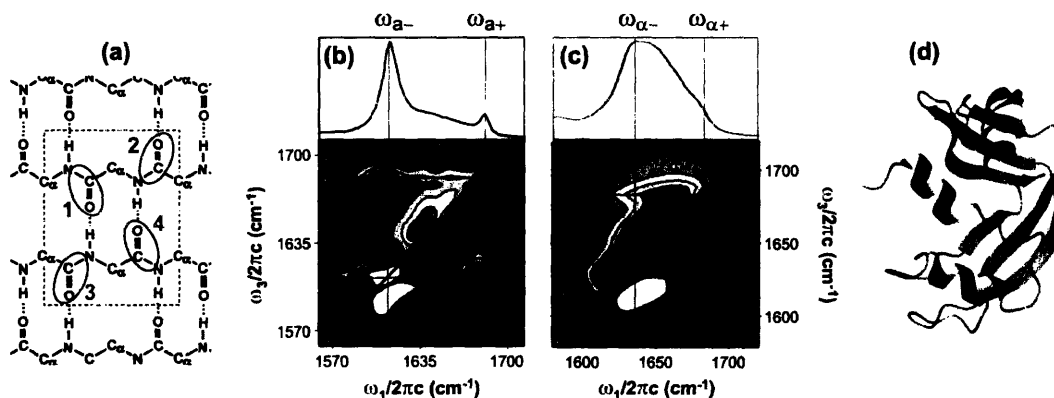


Figure 5-1. (a) A schematic showing the unit cell for an idealized antiparallel β -sheet, identifying four amide I oscillators (primarily CO stretch with CN stretching and other contributions). For the idealized structure, there are four vibrational states which involve the symmetric (*s*) and asymmetric (*a*) stretching of the 1-4 and 2-3 dimer pairs. The IR oscillator strength is carried by the *a*+ and *a*- modes, involving in-phase (+) and out-of-phase (-) vibration between the 1 and 2 oscillators.²³ The *s* states are only weakly IR allowed. (b) The linear FTIR and 2D IR correlation spectra of poly-L-lysine at $\text{pH}^*=11.4$ in the amide I region. Nineteen equally spaced contours are plotted between -20% and 20% of the peak maximum. The resonance frequencies of the two dominant IR active transitions of an extended pleated antiparallel β -sheet are labeled. (c) The linear and 2D IR amide I correlation spectra of RNase A at 25°C in the crossed polarization geometry. Twenty-one equally spaced contours are drawn from -60% to 60% of the maximum amplitude in the 2D spectrum. (d) Ribbon diagram of the crystal structure of RNase A (PDB entry 7RSA).⁷

We have found that the two amide I features are revealed in 2D IR spectra of antiparallel β -sheet containing proteins through the formation of a cross peak between the low and high frequency transitions.^{23,24} The clearest example is the 2D IR correlation spectrum of poly-L-lysine at high pH, a model system for studying AP β -sheet structure, which shows cross peaks between the two amide I resonances at 1611 cm^{-1} and 1680 cm^{-1} (Fig. 5-1b).²⁴ The 2D IR spectrum shows two diagonal and two cross peaks, each composed of doublets of opposite sign. This is the spectrum expected for two vibrational eigenstates that arise from coupled anharmonic vibrations.¹ As originally described by Miyazawa,²⁰ these two observed amide I transitions can be understood in terms of the structure and vibrational couplings for an idealized AP β -sheet.²³ The dominant transitions can be predicted from the symmetry of the four oscillator unit cell pictured in Fig. 5-1. The IR oscillator strength is carried by two vibrational modes ($a+$ and $a-$) involving antisymmetric stretching (a) of the 1-4 and 2-3 dimer pairs, with in-phase (+) and out-of-phase ($-$) vibration between the 1 and 2 oscillators. The more intense, lower frequency transition corresponds to the $a-$ vibration, whereas the weaker high frequency transition arises from the $a+$ vibration.

The 2D IR spectra of proteins with AP β -sheets also show this structure, although it appears different.²⁴ The position and shape of the spectral features are sensitive measures of the vibrational couplings within the sheet, the size and geometry of the sheet, and structural or energetic disorder associated with the peptide units within the sheet.^{23,24} The vibrations of a finite or disordered sheets may lead to overlapping transitions of slightly varying symmetry from the $a-$ and $a+$ designation used for the idealized sheet, and for that reason we label the resonances $\alpha-$ and $\alpha+$ in proteins. The frequency of the low frequency $\alpha-$ transition is sensitive to the size and geometry of the β -sheet leading to considerable inhomogeneous broadening in proteins with conformational disorder. This is observed as the stretching of the $\alpha-$ diagonal and cross peaks. On the other hand, the $\alpha+$ transition is insensitive to disorder.

These effects are apparent in the 25°C 2D IR correlation spectrum of RNase A shown in Fig. 5-1c. The clear eight-peak structure seen in polylysine is not observed. One of the “cross-peaks” between the two vibrations of the AP β -sheets is seen as a ridge

along $\omega_3 = 1684 \text{ cm}^{-1}$. This arises from a distribution of α^- frequencies and correspondingly narrow distribution of α^+ frequencies, expected for a protein with β -sheet structures of varying length, strand number, and conformation (Fig. 5-1d). The presence of the other ($\omega_1 > \omega_3$) cross peak is indicated by the decrease in intensity of the negative band elongated along the diagonal axis, and constructive interferences that stretch the negative feature along $\omega_3 \approx 1620 \text{ cm}^{-1}$. The spectral resonances along the diagonal, associated with the α -helices, turns and random coils, are diagonally elongated indicative of strong inhomogeneous broadening. The net effect, now observed in several β - and α/β -proteins,²⁴ is that interference effects between the negative and positive lobes of the diagonal and cross peaks arising from a narrow α^+ transition and a broadened α^- transition, combined with diagonally elongated resonances from additional amide transitions, leads to a characteristic “Z”-shaped contour profile for the amide I region in the 2D IR spectrum.

Despite suffering from spectral congestion, it is clear to see that the 2D IR correlation spectrum provides considerable additional information compared to the FTIR spectrum by accessing a second frequency dimension. 2D IR spectra reveal the underlying two-peak structure associated with the antiparallel β -sheets, and separate these from other components such as diagonal resonances near $1645\text{-}1650 \text{ cm}^{-1}$ that presumably arise from the α -helices and/or disordered regions.

5.3.2. FTIR spectra

Figure 5-2 shows the temperature dependent FTIR spectra of RNase A at pH* 7, in the range of 49°C - 72°C . The spectrum at 49°C is asymmetric with a broad maximum in the region of 1636 cm^{-1} and a weak shoulder at 1680 cm^{-1} . Both these regions correspond to the main amide I transitions of the AP β -sheet structure. As the temperature increases, the overall shape of the spectrum becomes more symmetric and the spectrum shifts to higher frequencies, so that the main peak is centered at 1649 cm^{-1} at 72°C . The net absorbance change between high and low temperatures, seen in Fig. 2b, shows that there is a loss of intensity at two transition frequencies that correspond closely to those expected for the α^+ and α^- modes. The overall changes are consistent with the

loss of two transitions, a 26 cm^{-1} wide band at $\omega_{\alpha-} = 1632\text{ cm}^{-1}$ and a 10 cm^{-1} band centered at $\omega_{\alpha+} = 1682\text{ cm}^{-1}$, together with an increase of a 74 cm^{-1} -wide Gaussian spectral component centered at 1649 cm^{-1} . If we correlate these observations to changes in the secondary structural elements of RNase A using empirical relations, this change reflects a depletion of the β -sheet structure and an increase in the α -helical, turns and/or random coil structures. The $\alpha+$ transition is insensitive to the disorder of the beta sheet, and comparison of its 10 cm^{-1} line width with the 26 cm^{-1} width of the structure-sensitive $\alpha-$ transition forms a measure of the structural inhomogeneity of the native β -sheet. The maximum change observed in the FTIR difference spectra ($\Delta T = 25^\circ\text{C}$), corresponds to $\sim 30\%$ of the maximum absorbance in the amide I resonance at 50°C . Further analysis of the spectra using the singular value decomposition method is given in Section 5.3.5.

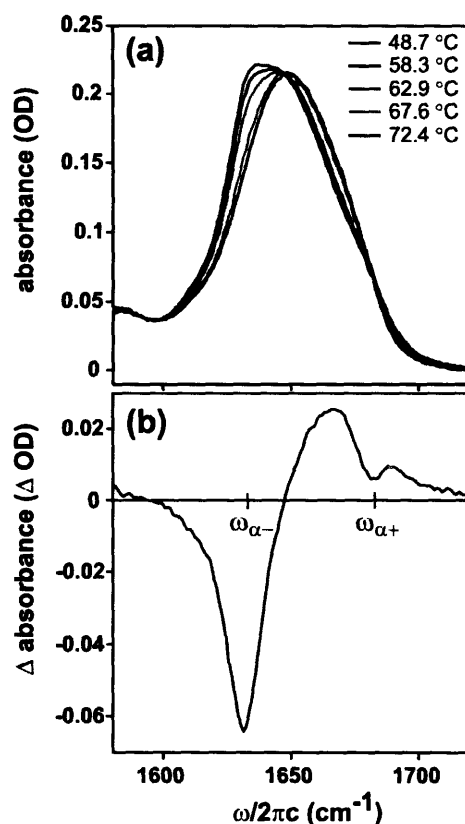


Figure 5-2 (top) Temperature-dependent amide I FTIR spectra of the thermal denaturing of RNase A with temperatures color coded in the figure. (bottom) Difference of 72.4°C and 48.7°C FTIR spectra with the frequencies of the characteristic β -sheet transitions labeled.

5.3.3. 2D IR spectra

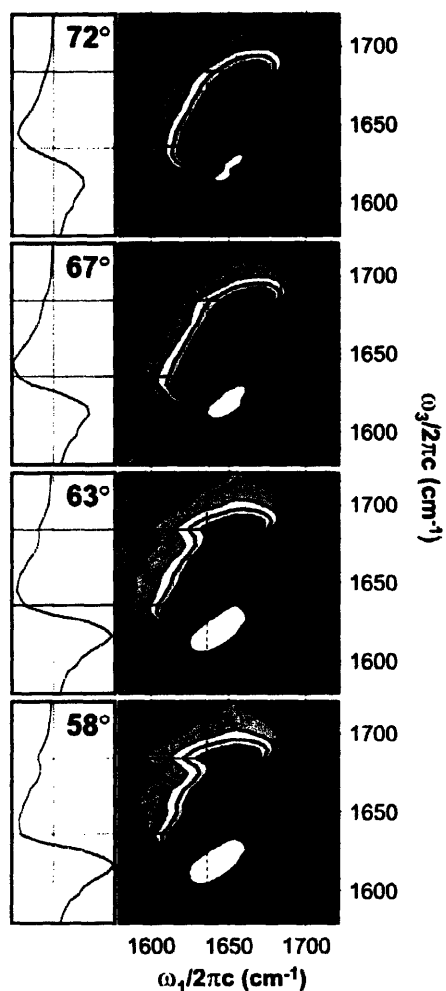


Figure 5-3. Temperature-dependent 2D IR correlation spectra \tilde{S}'_C of RNase A obtained in the crossed-polarization geometry. Temperatures from top to bottom are 72.2, 67.4, 62.7, and 57.9°C. The left panel of each 2D IR spectrum displays the slice along ω_2 corresponding to $\omega_1=1636 \text{ cm}^{-1}$ (dashed line). All 2D IR spectra have been normalized to the maximum positive amplitude of the 25 °C spectrum. Twenty-one equally spaced contours are drawn from -60% to 60% of the maximum amplitude for all the 2D IR correlation spectra. The largest changes in the 2D IR spectra are observed between 63 and 72°C.

Figure 5-3 shows the temperature-dependent 2D IR spectra of RNase A between 58°C and 72°C, the temperature range over which the largest changes are observed in the FTIR spectra. Only small changes are observed between the native state at 25 °C (Fig. 5-

1) and the 58°C spectrum. Spectra in this range show the signatures of AP β -sheet structure through the stretching of cross peaks and constructive interference along ω_1 , leading to the “Z”-shaped contour profile discussed earlier. As the temperature increases above 58°C, the dominant peak on the diagonal axis ($\omega_1=\omega_3$) shifts its maximum to higher frequency, its anti-diagonal width increases, the Z-shaped contour profile disappears, and the overall spectral shape becomes more symmetric. The disappearance of the “Z” indicates a loss of the stretched cross peak between the $\alpha+$ and $\alpha-$ transitions of the AP β sheet structures as the protein denatures. This is consistent with disrupting the amide I vibrational couplings due to the melting of β -sheets. The disappearance of the cross peak is also seen in the slices of the 2D IR spectra taken along $\omega_1 = 1636 \text{ cm}^{-1}$ and plotted adjacent to each spectrum. At 72°C, there is no distinct cross-peak seen along ω_3 for $\omega_1 = 1636 \text{ cm}^{-1}$. Instead, the 72°C spectrum appears as a symmetric doublet of diagonally elongated resonances characteristic of a classic inhomogeneously broadened transition that is characteristic of a highly disordered system. The small residual elongation observed for $\omega_3 = \omega_{\alpha+}$ and $\omega_{\alpha-}$ indicates that the denatured state is not completely random, and that residual AP hydrogen bonding configurations between amide oscillators on adjacent strands are still present in this denatured state. This should be expected since four disulfide bridges stabilize the β -sheets in RNase A.

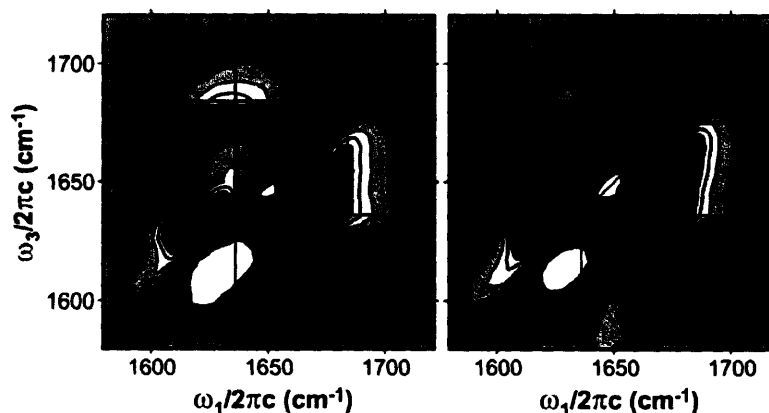


Figure 5-4. (Left) 2D IR difference spectrum, $\Delta\tilde{S}'_C = \tilde{S}'_C(25^\circ\text{C}) - \tilde{S}'_C(72^\circ\text{C})$ (Right) Second component $\sigma^{(2)}$ of the singular value decomposition of temperature dependent 2D IR correlation spectra. Each spectrum has been normalized to its maximum intensity and 21 equally spaced contours are plotted from -40% to 40%

Perhaps the clearest evidence for the melting of β -sheet structure is observed in the difference of 2D IR correlation spectra at 25°C and 72°C is shown in Figure 5-4. This spectrum highlights the disappearance of the eight-peak structure expected for the $\alpha+$ and $\alpha-$ transitions of AP β -sheets, in addition to the shift of spectral intensity to the blue on thermal denaturing. The difference spectrum also gives an indication of the shape of the elongated $\alpha+/\alpha-$ cross peaks in the native state. The spectrum indicates that 2D IR spectroscopy is sensitive to AP β -sheet structures in proteins and that the $\alpha+/\alpha-$ features of β -sheets are still clearly present in α/β globular proteins such as RNase A.

Further evidence for the presence of a cross peak in the native state and its disappearance on thermal denaturing is seen in the absolute value non-rephasing (NR) spectra. 2D IR NR spectra at 25°C and 72°C, which show similar effects to the correlation spectra, are shown in Figure 5-5. The additive phase relationship in τ_1 and τ_3 allow nonrephasing spectra to suppress inhomogeneous lineshapes along the diagonal axis.¹ The NR spectrum at 25°C reveals the stretched shape of the $\alpha+/\alpha-$ cross peak expected for β -sheets with intrinsic disorder in the native state due to variation in the number of strands in the sheet, the variation in hydrogen bonding geometry in the sheet, and deviation of the sheet from planarity.²⁴ Due to its insensitivity to inhomogeneous broadening, there is considerably less amplitude along the diagonal in the 2D IR NR spectrum between 1640-1670 cm^{-1} . Transitions in this region are commonly associated with random coils, turns and α -helices. Random coils would likely be suppressed in the NR spectrum, but the distinct peak is observed at (1660, 1650) cm^{-1} could very likely arise from the α -helices in RNase A. The *A*-mode for the amide I vibrations of the helices would be expected to carry most of the oscillator strength. As the protein is denatured by raising the temperature to 72°C, we note a loss of the stretched cross peaks (particularly for $\omega_1 < \omega_3$), loss of the diagonal feature at 1650 cm^{-1} , an overall blue shift in the peak signal, and a more symmetric lineshape. Additionally, the integrated amplitude of the spectrum decreases greatly, also reflecting a much more inhomogeneous system at high temperature. These observations are all consistent with the disruption of the vibrational couplings due to melting of β -sheets and a more random thermally denatured state.

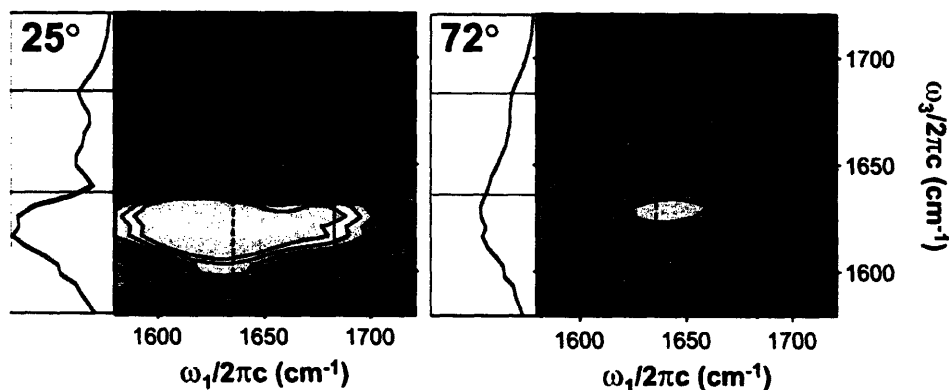


Figure 5-5. Absolute value non-rephasing spectra $|\tilde{S}_{NR}|$ of RNase A for the native and thermally denatured states obtained in the crossed-polarization geometry. The left panel of each spectrum displays the slice along $\omega_3 = 1636 \text{ cm}^{-1}$. The 72°C spectrum has been normalized to the maximum positive amplitude of the spectrum at 25°C . Eleven equally spaced contours are drawn from 0-60% of the maximum amplitude of 25°C spectrum.

5.3.4. Dispersed nonlinear signals: Pump-probe and vibrational echoes

While 2D IR spectra are perhaps the most revealing infrared probes of protein vibrations, they are technically demanding and time-consuming to acquire in comparison to other nonlinear experiments. Thus, as alternate probes, dispersed pump-probe and vibrational echo signals are appealing, particularly for application to transient studies of protein folding. In addition to being easier to acquire, they can be interpreted in terms of projections of 2D data onto the ω_3 axis. As discussed above, for the DPP measurement this is a projection of the real part of the complex correlation spectrum, and in the case of the DVE, the absolute square value of the projection of the complex correlation spectrum. In either case, the combination of multiple negative and positive features in the 2D IR spectrum make it clear that these projections cannot be interpreted simply and will usually involve the interference between various signal contributions. Nonetheless, for either experiment, the 2D IR correlation spectra of RNase A in Fig. 5-3 suggest that the two peak α^+/α^- structure indicative of β -sheets should be revealed.

Figure 5-6 shows the temperature-dependent DPP and DVE spectra obtained at $\tau_2=0$ and $\tau_1 = \tau_2=0$ respectively. At low temperatures, the DPP spectra show signatures of three features, which can be assigned in terms of a projection of the 2D IR correlation spectrum. Two sharper features below ω_{α^+} and ω_{α^-} arise from projections of the

oppositely signed ridges of the “Z” pattern, and an additional resonance from the diagonal feature is observed at 1660 cm^{-1} . As the temperature increases, the three-peak structure in the positive lobe starts to disappear, the overall amplitude decreases, and the positive and negative lineshapes look symmetric. This directly corresponds to the spectral changes in the 2D IR correlation spectra where the Z-shaped contour profile disappear at high temperatures leaving two almost symmetric and oppositely signed peaks along the diagonal.

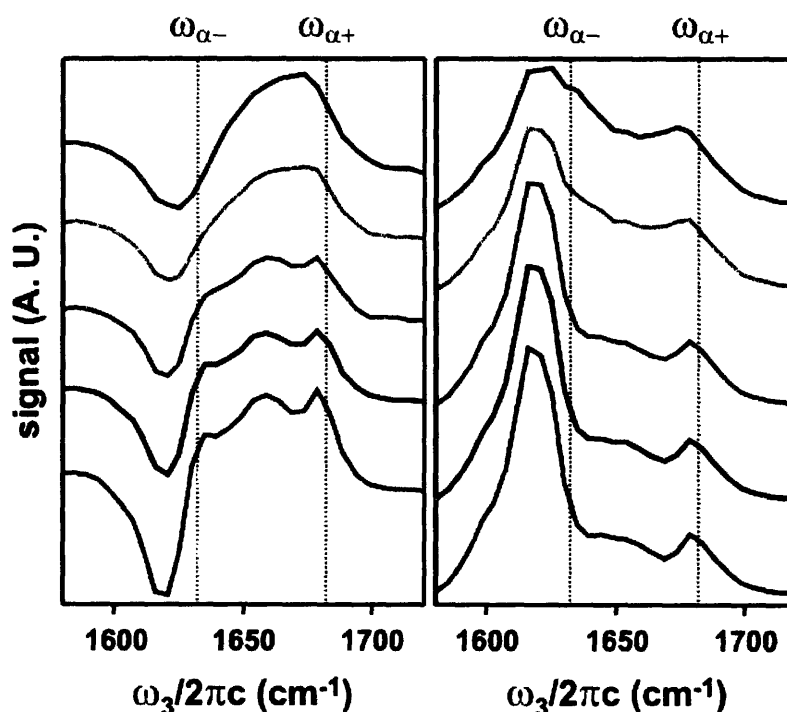


Figure 5-6. Temperature-dependent dispersed pump-probe (left) and dispersed vibrational echo (right) spectra obtained at $\tau_2=0$ and $\tau_1=\tau_2=0$ respectively. The temperatures of the traces increase from bottom to top corresponding to: 48.2 °C, 57.9 °C, 62.7 °C, 67.4 °C and 72.2 °C.

The DVE spectra show similar dramatic changes. At the lowest temperature, we see a two-peak structure with peaks at 1618 cm^{-1} and 1684 cm^{-1} , which looks remarkably similar to the FTIR spectrum of poly-L-lysine at high pH. Due to interference effects, it is not straightforward to relate a projection of the 2D IR spectrum to the DVE signal. Nonetheless, one can see that the peak at 1684 cm^{-1} corresponds to the high frequency amide I mode of the AP β -sheet, while the low frequency peak results from the

interference of the positive and negative features in the 2D IR spectra. This peak can be thought of arising from the projection of the base of the “Z” onto the ω_3 axis. The nonlinear scaling of the transition dipole moment in the DVE spectra results in far greater temperature-dependent changes as opposed to the linear FTIR spectra. As the temperature increases, we see that the relative ratio of the low and high frequency features in the DVE spectra drops, the low frequency peak moves to $\sim 1630 \text{ cm}^{-1}$, high frequency peak to $\sim 1679 \text{ cm}^{-1}$, and the spectral amplitude in the central region around 1650 cm^{-1} increases.

Although, the spectra displayed in Fig. 5-6 involve only one frequency axis, their origin in a nonlinear experiment leads to features that are more sensitive to the changes in the amide I region accompanying the thermal denaturation of RNase A than the traditional linear FTIR spectra shown in Figure 5-2. Any third-order nonlinear IR experiment on multiple coupled vibrations will have a signal that arises from transitions to multiple singly and doubly excited vibrational states, and will thereby be sensitive to vibrational couplings. Even with the projections involved, the DPP and the DVE spectra can be related to the more intuitive 2D IR spectra that encode the vibrational couplings between peptide vibrations making them a more sensitive probe of underlying protein structure than traditional FTIR.

5.3.5. SVD analysis of spectral changes

The amide I band in FTIR spectroscopy has been used to follow the denaturation of proteins due to temperature, pressure and chemical denaturants. Often times, the changes in the FTIR spectrum are too minute and difficult to correlate to structural changes. This has led to various spectral decomposition techniques to try and glean structural information, in particular the amount of secondary structure in the denatured state versus the native state.²⁵ In a similar manner, we have applied a singular value decomposition analysis to interpret the temperature-dependent spectral changes observed in the linear and nonlinear infrared probes of RNase A denaturing.

Singular value decomposition (SVD) can be used to analyze the common spectral components to a temperature-dependent or other series of linear and nonlinear spectra using techniques for diagonalization of rectangular matrices.²⁶⁻²⁸ The temperature-

dependent spectral data matrix is built by arranging temperature components in columns and spectral components in rows. 2D IR data matrices are converted to a one-dimensional array by appending rows one after another. Then, the spectrum at temperature T can be expressed as a linear combination of a series of SVD component spectra

$$\mathbf{S}(T) = \sum_{i=1}^n c_i(T) \boldsymbol{\sigma}_i . \quad (5.1)$$

The coefficients c_i form the temperature-dependent amplitudes of a series of common spectral components $\boldsymbol{\sigma}_i$ for the series of temperature-dependent spectra. The first component c_1 reflects the mean spectral amplitudes for all temperatures, whereas the second spectral component c_2 reflects the difference in spectral amplitude between temperatures. Higher components reflect increasingly subtle variations in the spectral features between temperatures. Since the second component coefficient c_2 is the most sensitive to the spectral change, it is an effective parameter for constructing melting curves for the thermal unfolding of proteins. For the case of two-state folding, the temperature variation of c_2 can be used to analyze the folding thermodynamics.

The second spectral components $\boldsymbol{\sigma}_2$ obtained from the SVD of the temperature-dependent FTIR, DVE and DPP spectra are shown in Fig. 5-7 together with the difference spectra between the high and low temperature spectra. In each case, the $\boldsymbol{\sigma}_2$ are similar to the difference spectra and show features that clearly reflect the concerted loss of α^+ and α^- transitions on increasing the temperature. In the case of the FTIR and the DVE there is also an increase of amplitude in the 1650-1660 cm^{-1} range. These observations are consistent with loss of β -sheet structure and increase of the random component. We find that the amplitude of the third SVD component for FTIR and DVE are very small, 0.7% and 1.5% relative to $\boldsymbol{\sigma}_1$, whereas for the DPP the amplitudes of $\boldsymbol{\sigma}_2$ and $\boldsymbol{\sigma}_3$ are comparable, 15% and 12%, respectively.

The second SVD component of the temperature-dependent 2D IR spectra is shown in Fig. 5-4 together with the difference of 2D IR spectrum at 25°C and 72°C. The two spectra are very similar, both showing the disappearance of the eight-peak structure expected for diagonal and cross peaks of α^+ and α^- transitions of antiparallel β -sheets.

The SVD analysis indicates that in large part the dominant spectral changes observed through 2D IR over the entire temperature range are consistent with the disappearance of β -sheet structure. More so than the difference spectrum, the SVD analysis also emphasizes the growth of an extra diagonal feature in the $\sim 1650\text{ cm}^{-1}$ region. Both spectra have similarity to the poly-L-lysine spectrum except the absence of α^+ mode in the diagonal slice ($\sim 1684\text{ cm}^{-1}$) because of its small dipole strength. The existence of the strong broad diagonal peaks between α^+ and α^- reflects the increase of random structure portion at higher temperature. Note that cross peaks are clearly shown in the off diagonal regions. Those in the lower right region ($\omega_1 > \omega_3$) are stronger than those in the upper left region ($\omega_3 > \omega_1$) because of the constructive and destructive interference with the wings of the strong α^- mode peaks.

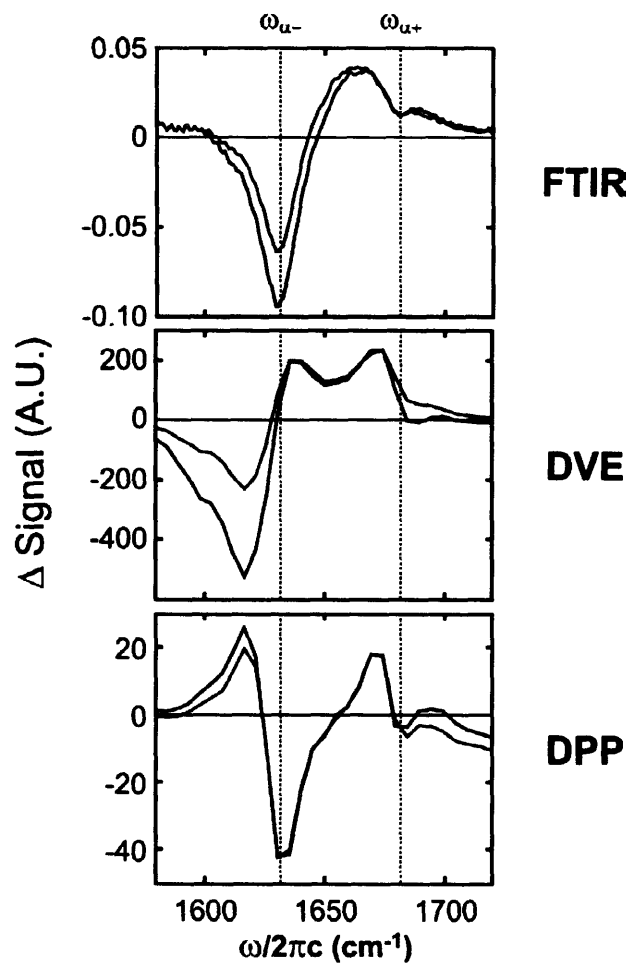


Figure 5-7. Difference spectra between the highest and lowest temperature $S(T_n) - S(T_l)$ (blue) and the second SVD component spectra, σ_2 (red) of FTIR, DVE, and DPP.

The melting curves derived from the temperature-dependent amplitude of the second SVD component c_2 are shown in Fig. 5-8. While the shapes of the melting curves vary somewhat between experiments, all show that the protein denaturation occurs at approximately 65°C. The melting temperatures (mid point) obtained from FTIR and DVE are 66°C and 65°C, respectively. The SVD analysis of the four experiments indicates that the 65°C transition, common to all experiments, is associated with the melting of β -sheet structure. Although the DPP and 2D IR spectra are not sampled at as many temperatures as the FTIR and DVE, it is clear that there are differences in the melting curves for temperatures <60°C. Note that the melting curve of DVE deviates from single sigmoid shape and shows evidence of a pretransition at 43°C. A similar curve is obtained from FTIR. The existence of pretransition has been observed by FTIR, CD and Raman spectroscopy, and is attributed to the partial unfolding of helix II with modification of the β -strand 43-49.^{9,11}

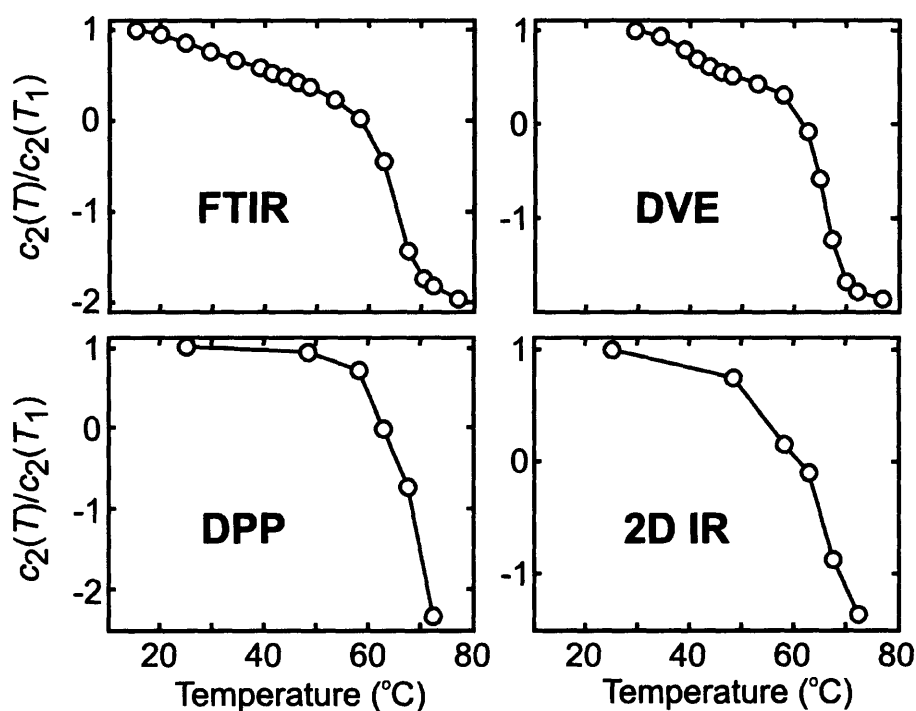


Figure 5-8. Thermal melting curves obtained from the scaled second components $c_2(T)/c_2(T_1)$ of the FTIR, DVE, DPP, and 2D-IR spectra. $c_2(T_1)$ is the coefficient of the spectrum at the lowest temperature T_1 .

In the case of two-state folding kinetics, the temperature-dependent series should be well-described by only the first and second SVD component, reflecting continuous amplitude changes between the spectra of reactants and products. For this limit, when $c_1(T)$ remains constant and c_i components for $i > 2$ are negligible, the normalized second SVD component follows the fractional change of one of the two equilibrium components. Then, c_2 can be used in a thermodynamic analysis, as discussed in the appendix. This is only the case when the signal amplitude scales linearly with the sample concentration, which is the case for FTIR, 2D IR, and DPP experiments. For background free DVE experiments, this analysis is not generally applicable but possible when the changes are small compared to the average spectrum and can be linearly expanded (see appendix). In the case of RNase A, the melting curves in Fig. 5-8 indicate that the folding is not two-state when characterized over the entire temperature range. If we limit our analysis to temperatures above 50°C, the FTIR results are described by first and second components, and the enthalpy and entropy of denaturation obtained is consistent with previous results.^{9,29,30}

5.4. Conclusions

Thermal denaturing of AP β -sheet secondary structure in RNase A has been investigated using a number of nonlinear infrared probes of the amide I transitions. Each of these experiments is sensitive to changes in protein secondary structure on thermal denaturing. 2D IR spectroscopy offers the most information about protein conformation, and is particularly sensitive to AP β -sheet structure through the cross peaks between the α^- and α^+ transitions. These cross peaks persist even for distorted β -sheets in α/β -proteins such as RNase A. The distinct eight-peak structure in the 2D IR difference spectrum of native and denatured states indicates that the vibrational couplings between amide I oscillators in AP β -sheet interstrand registry disappear. This is consistent with melting of the β -sheet in the thermal denaturing process, although the highest temperature spectrum indicates some residual AP β -sheet configurations.

The DVE and DPP spectra also show clear features associated with the β -sheets. These experiments are simpler to perform than 2D IR measurements, and reveal more

underlying features in the amide I region than conventional FTIR. As a result, the thermal denaturing of β -sheets is observed more clearly in these two methods compared to FTIR, even though both are presented as a spectrum in one frequency axis. The DVE and DPP experiments are related to the complex 2D IR correlation spectrum through projection relationships. Nonetheless, the interpretation of these signals is not trivial as a result of interference effects between negative and positive features in projections of the 2D IR spectrum, and nonlinear scaling of the signal in population. Melting curves for RNase A denaturing were obtained from SVD analysis of the nonlinear spectra. These reveal a pre-transition at 43°C followed by the main denaturing of β -structure of RNase A at 65°C.

Acknowledgements

We thank Vanessa De Souza for her assistance acquiring FTIR spectra at an early stage in this work, Nurettin Demirdöven for experimental assistance, and Martin Gruebele for suggesting SVD analysis of the 2D IR spectra.

References

- (1) Khalil, M.; Demirdöven, N.; Tokmakoff, A. *J. Phys. Chem. A* **2003**, *107*, 5258.
- (2) Woutersen, S.; Hamm, P. *J. Phys. : Condens. Matter* **2002**, *14*, R1035.
- (3) Zanni, M. T.; Hochstrasser, R. M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 516.
- (4) Demirdöven, N.; Khalil, M.; Golonzka, O.; Tokmakoff, A. *J. Phys. Chem. A* **2001**, *105*, 8025.
- (5) Thompson, D. E.; Merchant, K. A.; Fayer, M. D. *J. Chem. Phys.* **2001**, *115*, 317.
- (6) Merchant, K. A.; Xu, Q.-H.; Thompson, D. E.; Fayer, M. D. *J. Phys. Chem. A* **2002**, *106*, 8839.
- (7) Wlodawer, A.; Svensson, L. A.; Sjolín, L.; Gilliland, G. L. *Biochemistry* **1988**, *27*, 2705.
- (8) Santoro, J.; Gonzalez, C.; Marta, B.; Neira, J. L.; Nieto, J. L.; Herranz, J.; Rico, M. J. *Mol. Biol.* **1993**, *229*, 722.
- (9) Stelea, S. D.; Pancoska, P.; Benight, A. S.; Keiderling, T. A. *Protein Sci.* **2001**, *10*, 970.
- (10) Reinstädler, D.; Fabian, H.; Backmann, J.; Naumann, D. *Biochemistry* **1996**, *35*, 15822.
- (11) Chen, M. C.; Lord, R. C. *Biochemistry* **1976**, *15*, 1889.
- (12) Takeda, N.; Kato, M.; Taniguchi, Y. *Biochemistry* **1995**, *34*, 5980.
- (13) Hermans, J.; Scheraga, H. A. *J. Am. Chem. Soc.* **1961**, *83*, 3283.
- (14) Matheson, R. R.; Scheraga, H. A. *Biochemistry* **1979**, *18*, 2437.
- (15) Matthews, C. R.; Froebe, C. L. *Macromolecules* **1981**, *14*, 452.
- (16) Seshadri, S.; Oberg, K. A.; Fink, A. L. *Biochemistry* **1994**, *33*, 1351.
- (17) Klink, T. A.; Woycechowsky, K. J.; Taylor, K. M.; Raines, R. T. *Eur. J. Biochem.* **2000**, *267*, 567.
- (18) Sosnick, T. R.; Trehwella, J. *Biochemistry* **1992**, *31*, 8329.
- (19) Fabian, H.; Mantsch, H. H. *Biochemistry* **1995**, *34*, 13651.
- (20) Miyazawa, T.; Blout, E. R. *J. Am. Chem. Soc.* **1961**, *83*, 712.
- (21) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181.
- (22) Barth, A.; Zscherp, C. *Q. Rev. Biophys* **2002**, *35*, 369.

- (23) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *J. Chem. Phys.* **2004**, *120*, 8201.
- (24) Demirdöven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. *J. Am. Chem. Soc.* **2004**, *126*, 7981.
- (25) Jackson, M.; Mantsch, H. H. *Crit. Rev. Biochem. Mol.* **1995**, *30*, 95.
- (26) Shrager, R. I.; Hendler, R. W. *Anal. Chem.* **1982**, *54*, 1147.
- (27) Hofrichter, J.; Henry, E. R.; Sommer, J. H.; Deutsch, R.; Ikeda-Saito, M.; Yonetani, T.; Eaton, W. A. *Biochemistry* **1985**, *24*, 2667.
- (28) Chen, W.-G.; Braiman, M. S. *Photochem. Photobiol.* **1991**, *54*, 905.
- (29) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
- (30) Pace, C. N.; Grimsley, G. R.; Thomas, S. T.; Makhatadze, G. I. *Protein Sci.* **1999**, *8*, 1500.

Appendix 5A: Thermodynamics of two-state folding from SVD analysis

Consider the case that the spectrum $\mathbf{S}(T_i)$ at all temperatures T_i can be expressed with the first and second SVD components σ_1 and σ_2 ,

$$\mathbf{S}(T_i) = c_1(T_i) \sigma_1 + c_2(T_i) \sigma_2. \quad (5.A1)$$

If the spectral amplitude depends linearly on concentration, then in the case of two-state folding, the spectrum can also be expressed by the linear combination of the spectra of lowest temperature (T_1) and highest temperature (T_n):

$$\mathbf{S}(T_i) = N_F(T_i) \mathbf{S}(T_1) + N_U(T_i) \mathbf{S}(T_n). \quad (5.A2)$$

Here $N_F(T_i)$ and $N_U(T_i)$ are the fractional population of the folded and unfolded species respectively. By substituting Eq. (5.A1) into Eq. (5.A2), we obtain

$$N_F(T_i) = \frac{c_2(T_n) - c_2(T_i)}{c_2(T_n) - c_2(T_1)}, \quad (5.A3)$$

which is the normalized second SVD coefficient. Therefore, approximate values of the enthalpy and entropy of denaturation can be obtained by fitting the equilibrium constant from the second SVD component ($K = N_U/N_F$) to the thermodynamic relation $RT \log K = -\Delta H + T\Delta S$.

When the measured signal is not linearly proportional to concentration, as is the case for background-free (homodyne) nonlinear experiments like the DVE, above relation does not generally hold, but is valid when spectral changes with temperature are small. From Eq. (3.11) we define

$$\mathbf{S}_{\text{DVE}} = |\mathbf{S}_{\text{E}}|^2 \propto \left| \int_{-\infty}^{\infty} \mathbf{S}_{\text{C}}(\omega_1, \tau_2, \omega_3) d\omega_1 \right|^2, \quad (5.A4)$$

where \mathbf{S}_{E} is the integration of the 2D correlation spectrum over ω_1 . Since \mathbf{S}_{E} is linear in the concentration, and $N_F(T_i) + N_U(T_i) = 1$,

$$\mathbf{S}_{\text{E}}(T_i) = N_F(T_i) \mathbf{S}_{\text{E}}(T_1) + N_U(T_i) \mathbf{S}_{\text{E}}(T_n) = \mathbf{S}_{\text{E}}(T_1) + N_U(T_i) \Delta \mathbf{S}_{\text{E}}, \quad (5.A5)$$

where $\Delta \mathbf{S}_{\text{E}}$ is the difference spectra between lowest and highest temperature. In the limit that $|\Delta \mathbf{S}_{\text{E}}| \ll |\mathbf{S}_{\text{E}}|$, we obtain the same linear equation for the temperature-dependent changes to the DVE spectrum using population conservation ($N_F(T_i) + N_U(T_i) = 1$) and Eq. (5.A5)

$$\begin{aligned}
\mathbf{S}_{\text{DVE}}(T_i) &= \mathbf{S}_{\text{DVE}}(T_1) + 2N_U(T_i) \text{Re} \left[\mathbf{S}_E(T_1) \cdot \Delta \mathbf{S}_E^* \right] + N_U(T_i)^2 |\Delta \mathbf{S}_E|^2 \\
&= N_F(T_i) \mathbf{S}_{\text{DVE}}(T_1) + N_U(T_i) \mathbf{S}_{\text{DVE}}(T_n) + \left(N_U(T_i)^2 - N_U(T_i) \right) |\Delta \mathbf{S}_E|^2. \quad (5.A6) \\
&\approx N_F(T_i) \mathbf{S}_{\text{DVE}}(T_1) + N_U(T_i) \mathbf{S}_{\text{DVE}}(T_n)
\end{aligned}$$

Here, the small third term that is ignored in the second line is incorporated in the third component of SVD analysis.

Chapter 6

Conformational changes during the nanosecond to millisecond unfolding of ubiquitin

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

- “Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin,”
H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim, and A. Tokmakoff, *Proc. Natl. Acad. Sci. USA*, **102**, 612 (2005).

6.1. Introduction

Most protein folding experiments measure kinetics: the rate of appearance or disappearance of an experimental signature for a particular species, typically on millisecond or longer time scales. These results give information on the height of energetic barriers much higher than thermal energy, but say little about how structure changed along the path. A number of fast folding experiments of proteins and peptides have shown that downhill folding, in which evolution is governed by energy barriers $\lesssim kT$, can be initiated with a nanosecond T-jump. Such experiments work in a diffusive regime that allows a freer exploration of available structures, and provide evidence that the relevant molecular time scales for folding is nanoseconds to microseconds.¹⁻⁴ If downhill folding can be initiated and followed with a structure-sensitive probe, then

meaningful information can be obtained on the underlying molecular dynamics of folding. We report here on such an experiment, a conformationally sensitive probing of downhill unfolding of ubiquitin over nanosecond to millisecond time scales following a T-jump.

In this chapter, conformational changes during the thermal unfolding of ubiquitin are studied both at equilibrium, with temperature-dependent 2D IR and DVE spectroscopy as in Chapter 5, and transiently, with DVE spectroscopy following a nanosecond T-jump. Ubiquitin, a small 76-residue protein, has been one of the best model systems for the protein folding research since it has all common structural motifs such as α -helix, β -sheet, turns, and hydrophobic core, but does not have any extra disulfide bond, metal cofactor, or a prosthetic group. Equilibrium experiments suggest that the system obeys a simple two-state folding kinetics, but transient experiments reveal a more complex relaxation behavior spanning six decades in time. From 100 ns to 0.5 ms, we observe a non-exponential, sequential loss of two β -sheet transitions that originate in delocalized β -sheet vibrations. On millisecond time scales we cross into a kinetic regime, characterized by a concerted 3 ms exponential unfolding. An energy landscape represented in terms of native contacts between different strands of the β -sheet is proposed to explain the experiments.

6.2. Methods

Nonlinear IR experiments for the equilibrium and transient measurements are described in Chapter 4 and 5. Data is collected with a single strip MCT array detector as Chapter 5. T-jumps of $\sim 12^\circ\text{C}$ from the initial temperature $T_0 = 58^\circ\text{C}$ is obtained with a T-jump pulse of 6 mJ. Infrared spectra are calculated as described in Chapter 2.

Ubiquitin was purchased from Sigma Aldrich (U6253). It was dissolved in D_2O solution, adjusted with DCl to pH* 1, and incubated at 60°C for 1 hr prior to the experiments to allow for H/D exchange. The protein concentrations were 33 mg/mL for DVE measurements and 17 mg/mL for 2D IR and FTIR measurements, which result in peak optical densities of 0.5 and 0.25, respectively. Temperature-dependent spectral changes were found to be reversible up to 85°C by FTIR.

6.3. Results

6.3.1. Equilibrium Thermal Unfolding

The conformational changes on thermal unfolding of ubiquitin as monitored by temperature-dependent FTIR, 2D IR, and DVE spectra are shown in Fig. 6-1. FTIR absorption spectra, shown in Fig. 6-1a, are asymmetric, with a peak at 1642 cm^{-1} and a shoulder at 1676 cm^{-1} . Although broad and quite featureless, the asymmetry is suggestive of the two-peak structure that arises from amide I vibrational couplings in anti-parallel β -sheets.^{5,6} The strong band ($1630\text{-}1650\text{ cm}^{-1}$) has a transition moment roughly perpendicular to the β -strands and we label it ν_{\perp} , whereas the transition moment of the weak band ($\sim 1680\text{ cm}^{-1}$) is near parallel to the strands and designated ν_{\parallel} . As the temperature is raised from 58°C to 77°C , a very small decrease in the peak intensity and increase in the $1650\text{-}1660\text{ cm}^{-1}$ range is observed.

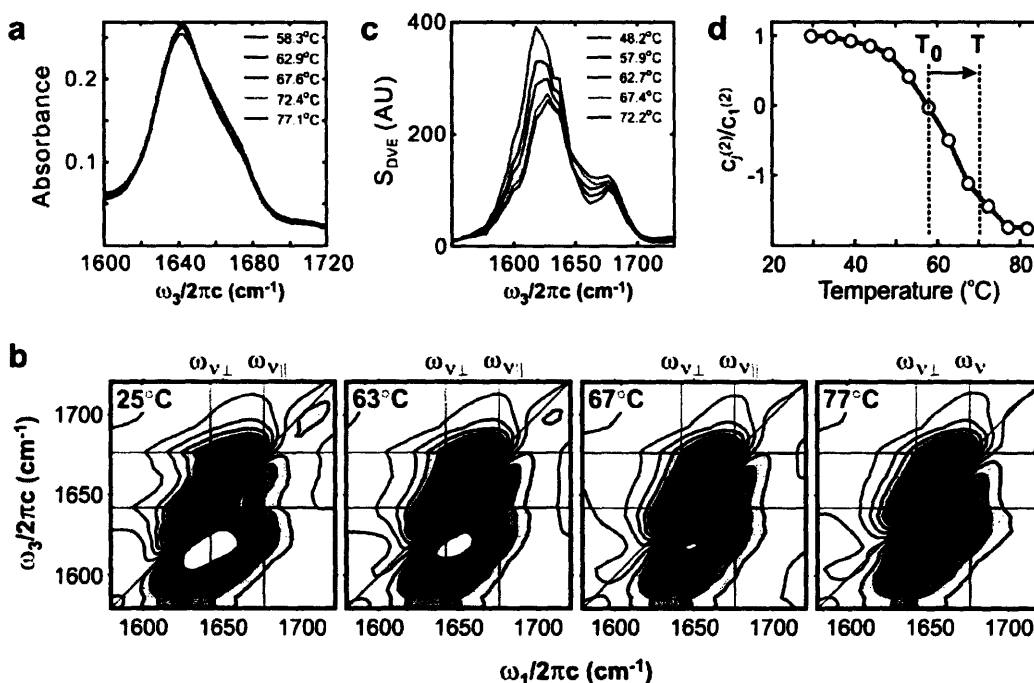


Fig. 6-1. Temperature-dependent IR spectroscopy of ubiquitin unfolding. (a) FTIR spectra from 58 to 77°C . (b) 2D IR spectra from 25°C to 77°C . 21 equally spaced contours are drawn from -60% to 60% of the maximum amplitude of 25°C spectrum. (c) DVE spectra from 48 to 72°C . (d) Thermal melting curve constructed from normalized SVD second component coefficients $c_2(T)/c_2(T_1)$ of temperature-dependent DVE spectra. Vertical dashed lines indicate the T-jump window.

The vibrational transitions within the congested amide I lineshape are more apparent in 2D IR spectra of ubiquitin, shown from 25°C to 77°C in Fig. 6-1b. In these spectra, the ω_1 dimension marks the frequency of the initially excited vibration, and ω_3 is the detection frequency for vibrational transitions induced by the initial excitation. Red contours originate from fundamental transitions ($v = 1 \rightarrow 0$) and blue contours are anharmonically shifted out-of-phase induced absorptions to two quantum states ($v = 1 \rightarrow 2$). The characteristic ν_{\perp} and ν_{\parallel} transitions underlying the congested FTIR spectrum are apparent in the 2D IR spectrum from the stretched ridges at $\omega_3 = \omega_{\nu_{\parallel}}$ and $\omega_{\nu_{\perp}}$. The ν_{\perp} diagonal peak is diagonally elongated due to the inhomogeneous broadening that results from structural disorder within the β -sheet and overlap with resonances from other secondary structure. The ν_{\parallel} diagonal peak does not appear because of its weak transition dipole moment, but it is apparent through a cross peak between the ν_{\perp} transition and ν_{\parallel} transition that is elongated into a ridge along $\omega_3 = \omega_{\nu_{\parallel}}$. The cross peak is elongated parallel to the ω_1 axis because the ν_{\perp} transition is sensitive to the size and disorder of the β -sheet, whereas the ν_{\parallel} transition is not.^{7,8} Another cross peak at $(\omega_1, \omega_3) \approx (\omega_{\nu_{\parallel}}, \omega_{\nu_{\perp}})$ is present as an elongation of the negative peak along $\omega_3 \approx \omega_{\nu_{\parallel}}$ due to interference with a negative peak anharmonically shifted below the diagonal axis. The features of diagonal elongation and ω_1 -stretching of cross peaks leads to a characteristic Z-shape in the 2D IR spectra of β -sheet proteins.⁸

As the temperature is raised, we observe a concerted blue shift of the ν_{\perp} diagonal peak and shrinking of the cross peak ridges. The overall shape as the temperature is raised tends toward a diagonally elongated doublet. These observations are similar to those seen for the unfolding of ribonuclease A,⁹ and are consistent with melting of the β -sheet structure in ubiquitin. At 77°C, the cross peak still persists as a weak protrusion from the diagonal peak indicating the existence of the residual β -sheet secondary structure.

Temperature dependent DVE spectra of ubiquitin are shown in Fig. 6-1c. The 48°C spectrum shows two distinct peaks at 1621 cm^{-1} and 1677 cm^{-1} . These resonances can be interpreted in terms of the features in the 2D IR spectra, since the DVE spectrum is the absolute value squared of the projection of the complex 2D IR spectrum onto the

ω_3 axis.⁹ As a result of the projection along the elongated $\omega_3 = \omega_{v\perp}$ negative ridge and the $\omega_3 = \omega_{v\parallel}$ cross peak ridge, two distinct peaks are observed in the DVE spectrum, which are distinct signatures of the ubiquitin β -sheet. More detailed features are observed in the DVE spectra and the spectral changes as the temperature is raised are much more dramatic than those of FTIR spectra. The peak maximum (v_{\perp}) shifts from 1621 cm^{-1} to 1630 cm^{-1} and the peak intensity decreases.

A melting curve for ubiquitin can be obtained by singular value decomposition (SVD) of the equilibrium temperature-dependent DVE spectra.⁹ More than 99% of the spectra can be reconstructed by linear combinations of the first and second component SVD spectra and the scaled second component coefficient at each temperature is used for the melting curve. This melting curve (Fig. 6-1d) shows a sigmoid shape with a melting temperature of $\sim 61^\circ\text{C}$, suggesting simple two-state unfolding kinetics.

6.3.2. Transient spectral change

DVE spectra were used to select an initial temperature of $T_0=58^\circ\text{C}$, leading to the 12°C T-jump window shown in Fig. 6-1d. The resulting transient DVE difference spectra are shown in Fig. 6-2a for nanosecond to millisecond delays. Also shown is the equilibrium difference spectrum ΔS_{∞} , which should be recovered for long times following a step T-jump. Figure 6-2b shows a semi-log plot of the response of three frequency channels: two at the peaks associated with features originating in the v_{\perp} and v_{\parallel} β -sheet transitions, and one in the intermediate spectral region commonly associated with random coils and α -helices. The time-dependent spectral changes consist of five roughly sequential stages: (1) the initial rise of the whole spectrum, (2) a μs drop of the signal in $\omega_3 = \omega_{v\perp}$ region, (3) the drop of the high frequency signal including the $\omega_3 = \omega_{v\parallel}$ region, (4) a ms unfolding in the β -sheet region resembling the equilibrium difference spectrum, and (5) refolding. The unfolding can also be distinguished by non-exponential relaxation for times prior to $\sim 0.5\text{ ms}$, followed by exponential behavior for longer times.

The initial rise of the difference spectrum for $\tau < 100\text{ ns}$ results mostly from the small temperature-dependent increase of transmittance in the D_2O background at $6\text{ }\mu\text{m}$. Also, these early delays could be influenced by acoustic or shock waves that result from

the rapid T-jump. However, since the 100 ns DVE difference spectrum for $\omega_3 < 1650 \text{ cm}^{-1}$ is distinctly different from the equilibrium DVE spectrum at T_0 , there must also be an additional fast response from the protein. Local solvation effects with a subset of solvent exposed amide vibrations would be expected on the ps time scale.

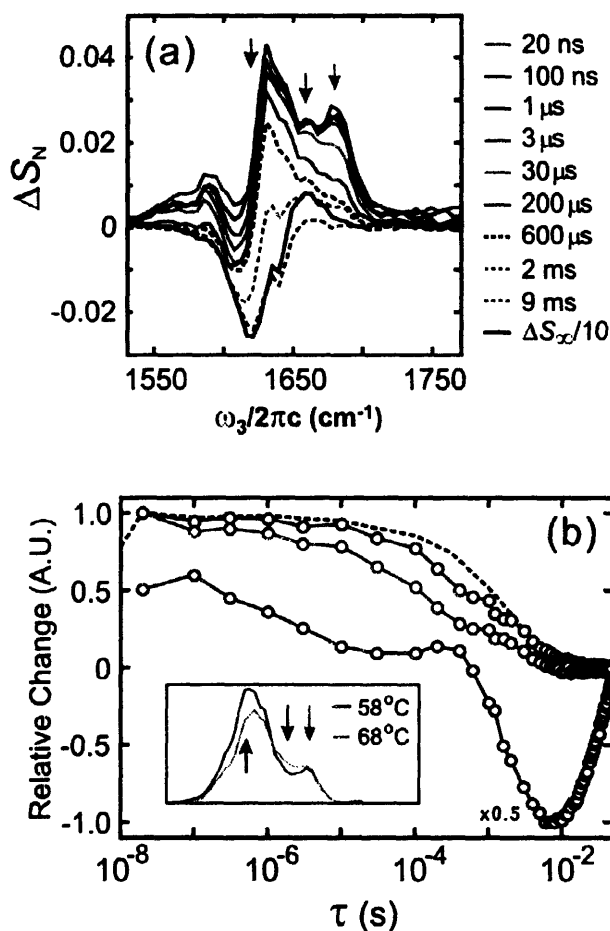


Fig. 6-2. (a) DVE difference spectra normalized to the peak signal $\Delta S_N = \Delta S / S_{\text{DVE}}(T_0, \omega_3 = 1621 \text{ cm}^{-1})$ at several time delays from the T-jump. The black solid curve shows the equilibrium difference spectrum $\Delta S_\infty = [S_{\text{DVE}}(T) - S_{\text{DVE}}(T_0)] / S_{\text{DVE}}(T_0, \omega_3 = 1621 \text{ cm}^{-1})$ between 68 and 58°C scaled by 0.1. (b) The normalized time-dependent DVE response $\Delta S_\tau = \Delta S(T, \tau) / \Delta S(T, \tau = 20 \text{ ns})$ plotted for three frequency channels. These channels, labeled on the equilibrium DVE spectra at T and T_0 in the inset, correspond to ν_\perp ($\omega_3 = 1621 \text{ cm}^{-1}$, blue open circle), ν_\parallel ($\omega_3 = 1677 \text{ cm}^{-1}$, green open circle) and $\omega_3 = 1658 \text{ cm}^{-1}$, red open circle). Temperature relaxation profile determined from the transmittance change through D₂O at 6 μm is shown in black.

The non-exponential decrease of the ν_{\perp} region ($\omega_3 \approx 1621 \text{ cm}^{-1}$) is observed on the 3 μs time scale while other spectral regions remain constant. The ν_{\perp} frequency is sensitive to interstrand vibrational couplings and the number of strands within the β -sheet over which the vibration is delocalized.^{7,8} The presence of the major changes in this region indicates that the change results from a breakdown of interstrand interactions due to denaturing of the sheet. A stretched exponential fit of the normalized data from 20 ns to 100 μs ($\Delta S \propto \exp[-(\tau/\tau_0)^{\beta}]$) gives $\tau_0 = 2.6 \mu\text{s}$ and $\beta = 0.72$ implying unfolding through multiple routes.

Subsequently, the intensity of other regions decreases non-exponentially on 10 μs to $\sim 0.5 \text{ ms}$ time scales. In this time window, the ν_{\perp} region remains unchanged. For the $\omega_3 = 1677$ and 1658 cm^{-1} relaxation traces in Fig. 6-2b, a fit to a sum of stretched exponential for fast components and single-exponential fit for slow components gives $\tau_1 = 80 \mu\text{s}$, $\beta = 0.55$ and $\tau_2 = 2.8 \text{ ms}$, and $\tau_1 = 210 \mu\text{s}$, $\beta = 0.74$ and $\tau_2 = 4.0 \text{ ms}$, respectively. Non-exponential relaxation during the μs time window has previously been observed in the refolding of a cold denatured ubiquitin mutant.¹

After approximately 0.5 ms, the amplitude in the ν_{\perp} region decreases exponentially with a $\sim 3 \text{ ms}$ time constant reaching the maximum change at 6 ms. This process is also observed at the other probing frequencies, and represents a major concerted melting of the β -sheet. This drop in ν_{\perp} amplitude would continue, however the T-jump re-equilibrates with a 2 ms time constant. The maximum change appears after the temperature re-equilibrates. At this point the transient difference spectrum has a similar shape to the equilibrium difference spectrum ΔS_{∞} , although only 10% of its amplitude. For times longer than 10 ms, the temperature of the solution has re-equilibrated at T_0 . On this time scale, the T-jump appears as an impulse, and the protein refolds with a 20 ms time constant.

6.4. Discussion

2D IR spectra and DVE spectra show distinct signatures of a two-peak structure that is characteristic of polypeptide strands in anti-parallel hydrogen-bonding registry.¹⁰ Idealized anti-parallel β -sheets show two infrared active bands ($a+$ and $a-$) whose

symmetry leads to transition dipole moments roughly parallel and perpendicular to the strands.^{7,8} Strong inter-strand couplings and vibrational delocalization across strands lead to a significant red shift of the lower frequency α^- resonance as strands are added to a growing sheet.⁷ However, the frequency and amplitude of the α^+ transition is insensitive to the size or geometry of the sheet, remaining approximately the same for structures with two or more strands.

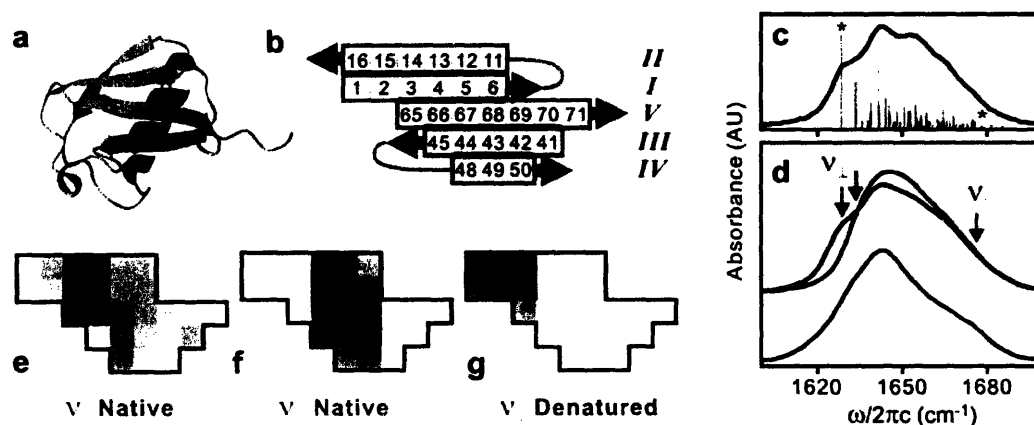


Fig. 6-3. (a) Ribbon diagram of the crystal structure of ubiquitin (PDB entry 1UBQ)¹¹. (b) Two-dimensional projection of the β -sheet with boxes numbering the peptide groups in the primary structure, and indicating the numbering of strands. Amide I couplings reduced on unfolding are indicated in red and are related to the native hydrogen bonding contacts $n(III,IV,V)$. The contacts $n(I,II)$ are shown in purple. (c) Calculated IR spectrum and stick figure of the dipole moment square for amide I eigenstates in the native structure obtained from the local amide I Hamiltonian. (d) The calculated IR absorption spectra of native structure (blue) and partially denatured structure (red) including disorder in the site energies compared with the experimental FTIR spectrum (black). (e-g) Color maps of the amplitude and phase of representative amide I vibrational eigenstates for the folded and denatured states. Red and blue colors represent opposite sign of the vibrational phase, and the depth of the color reflects the amplitude. (e,f) States from native structure with ν_{\parallel} and ν_{\perp} character, labeled with asterisks in Fig. 6-3c and green arrows in (d). (g) State with ν_{\perp} character in unfolded calculation showing localization on strand $I-II$ hairpin. Corresponds to transition labeled with red arrow in (d).

It is not immediately apparent that these features should be present in the 2D IR spectra, since the β -sheet of ubiquitin is not entirely anti-parallel. We investigated the assignment using calculations of the IR absorption spectrum of ubiquitin drawing on the crystal structure¹¹ and using a local amide I Hamiltonian.^{5,12} A planar projection of the β -

sheet of ubiquitin (Fig. 6-3) shows it is composed of an N-terminal hairpin (strands *I* and *II*) and three additional anti-parallel strands (*III*, *IV*, *V*), with parallel registry between strand *I* and the C-terminal strand *V*. The calculated IR spectrum using the native structure has more peaks than the experimental spectrum (Fig. 6-3c); however, we find that by averaging over a modest variation of the amide I energies ($\sigma = 6 \text{ cm}^{-1}$), the breadth and asymmetry of the experimental lineshape are reproduced (Fig. 6-3d). Visualization of the amide I eigenstates obtained using the crystal structure confirms that vibrations with $a+$ and $a-$ character are expected on the high and low frequency sides of the line, respectively. Figures 6-3e and 3f show color maps of the vibrational amplitude and phase of the local amide I vibrations in the β -sheet contributing to amide I eigenstates on the high and low frequency sides of the absorption line. For the low frequency (1628 cm^{-1}) eigenstate, the oscillators aligned between strands are in-phase but out-of-phase along the strand, the signature of $a-$ character.⁸ For the high frequency mode (1679 cm^{-1}), oscillators along the strands are in-phase, as expected for $a+$ vibrations. Although these are only two eigenstates, they reflect the broader trends in the spectrum.

Fragmentation studies, multidimensional NMR experiments, and molecular dynamics simulation on ubiquitin have indicated that strands *III*, *IV*, and *V* are less conformationally stable than strands *I* and *II* and the α -helix.¹³⁻¹⁷ We therefore postulate that the 3 μs blue shift of the ν_{\perp} resonances with no concerted change in ν_{\parallel} , can be explained as preferential unfolding of strands *III-V* and a localization of the ν_{\perp} transition in the strand *I-II* hairpin. This observation can be tested by modeling changes in the amide I IR absorption spectrum upon unfolding of strands *III-V*. For the calculation of the spectrum of the partially denatured state, the interstrand amide I couplings between strand *I* and *V*, *V* and *III*, and *III* and *IV* are reduced by 75%. These interactions are labeled in red in Fig. 6-3b. (For a direct displacement of strands, this reduction in coupling corresponds to a 60% increase in interstrand distances). Figure 6-3d shows a comparison of the modeled FTIR spectra for native and denatured states with the experimental FTIR spectrum at 55°C. As the couplings are reduced, the ν_{\perp} region shows a significant blue shift while the ν_{\parallel} region ($1670\text{-}1680 \text{ cm}^{-1}$) remains unchanged. The

localization of the v_{\perp} excitation in the strand *I-II* hairpin in the denatured state is confirmed by visualizing the lowest energy eigenstate in Fig. 6-3g.

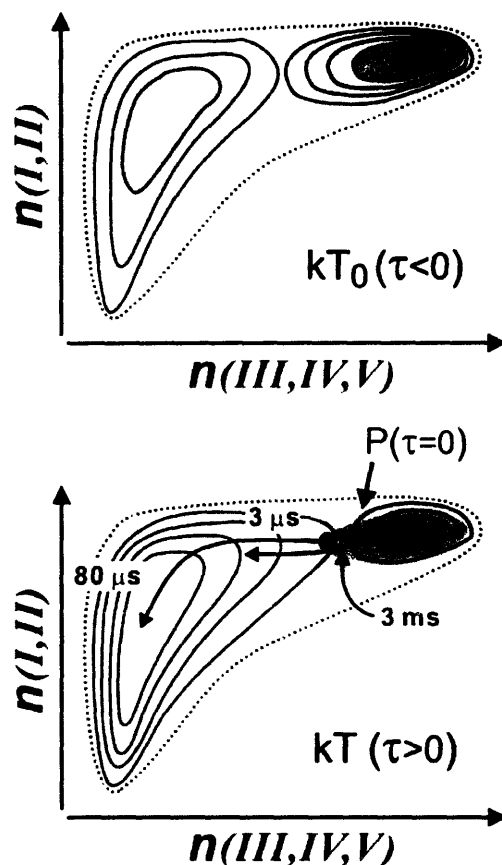


Fig. 6-4. (a) An illustration of a reduced free energy surface at kT_0 that would be obtained by averaging over all degrees of freedom except the native hydrogen bond contacts to strands *III-V* and to strands *I-II*. (b) Same free energy surface after T-jump at kT . The equilibrium is shifted to the unfolded side, and the barrier is shifted partially toward the native state. Shaded region shows initial population at $\tau=0$ with representative trajectories. A small portion of proteins with partial unfolding of the *III-V* region (red dot) unfold along a barrierless trajectory, whereas most re-equilibrate in a native-like state prior to activated barrier crossing (green dot).

On the basis of the modeling we propose a microscopic picture for ubiquitin unfolding in the T-jump experiments. The molecular dynamics of unfolding can be described in terms of trajectories on a microscopic free energy landscape that is a function of numerous configurational coordinates.¹⁸⁻²⁰ Due to their sensitivity to the interstrand couplings that change dramatically over the Ångstrom scale, the amplitude

and resonance frequency of the ν_{\parallel} and ν_{\perp} transitions can be directly related to native inter-strand contacts. We imagine that averaging the energy landscape over much of its roughness and concentrating on the native hydrogen-bonding contacts to strands *III-V* and the contacts in the strand *I-II* hairpin would lead to an energy surface at kT_0 with the characteristics illustrated in Fig. 6-4a. The surface has a native state energy minimum that retains most of the contacts found in the crystal structure, and an extended minimum for the unfolded state that retains some residual structure. The contours indicate the preferential thermal stability of strands *I* and *II* over *III*, *IV*, and *V*. The initial temperature kT_0 is already quite elevated so that a considerable fraction of the molecules are unfolded. Although this representation is just a proposal, in the future this free energy surface can in principle be obtained through modeling of our 2D IR spectra.

The T-jump prepares the system with a distribution of structures that is mostly centered near the native state (Fig. 6-4b). The shift of the barrier to the native side reflects the tendency toward a less stable native state and more stable denatured ensemble after the T-jump. For a relatively small partially denatured sub-ensemble (red dot in Fig. 6-4b), there is no significant barrier to unfolding. The T-jump places this sub-ensemble at or near the transition state for unfolding, and we observe 1-500 μs nonexponential unfolding that can be described as diffusion in a multidimensional configurational space.²⁰⁻²²

On the early 3 μs time scale, the rise in thermal energy in the protein and solvent allows increased configurational flexibility in the more weakly interacting and partially denatured strands *III-V*. Breaking of native contacts between strands reduces inter-strand vibrational couplings, localizing the ν_{\perp} vibrational excitation on strands *I-II* and blue shifting the ν_{\perp} transition. The fastest time scales correspond to those considered the speed limit for protein folding.²³⁻²⁶ In the second stage, the remaining hairpin involving the more stable strands *I-II* unfolds. The ν_{\parallel} transition decays only when the final inter-strand interactions in hairpins are disrupted, resulting in the relaxation of ν_{\parallel} region on a 80 μs time scale. The somewhat longer time scale of relaxation in the intermediate frequency regime commonly associated with random coils and α -helices (1658 cm^{-1}) is more difficult to interpret, but probably reflects both unfolding of anti-parallel contacts and the helix, and the growth of denatured amide I species. With our present methods, this central region could be influenced by interference effects arising from the projection of

overlapping resonances in the nonlinear DVE response. A more detailed picture of this region can be obtained in the future using the 2D IR experiment as a transient structural probe.²⁷

The majority of the population re-equilibrates within the native-like minimum (green dot in Fig. 6-4b), and then unfolds as a 3 ms kinetic process that appears as a concerted unfolding. The time scale separation between short time non-exponential and long time exponential relaxation processes does not necessarily reflect unfolding along different (parallel) coordinates. The trajectory is similar to the fast unfolding, but because of the higher barrier, this appears as the two-state kinetics seen in the equilibrium temperature-dependent experiments. This observation lends further support to the two-state model for ubiquitin folding. Previous rapid-mixing kinetic studies of ubiquitin refolding on millisecond to second time scales with NMR, fluorescence quenching, and circular dichroism have debated whether folding proceeds through an intermediate state or follows two-state kinetics.²⁸⁻³³ We emphasize that these long time kinetics probe crossing of barriers $\gg kT$, whereas our experiments differ by revealing fast downhill evolution over a landscape with energy corrugation $\lesssim kT$.

The ability of 2D IR and DVE spectroscopy to make measurements of downhill diffusion with conformational sensitivity indicates that these experiments characterize an unfolding dynamics: A time-resolved measure of how nuclear configuration for the protein is changing. We differentiate this from a kinetics measurement that characterizes the rate of appearance or disappearance of folded or denatured state. The structural information is not atomistic. Rather, these experiments characterize vibrations that are delocalized across extended regions of the protein, giving insight into the conformational changes to this extended region. These coordinates of mesoscopic scale are particularly appropriate for the study of large scale conformational changes in processes such as folding or binding, where reaction coordinates are similarly collective. We also note that the conformational information is statistical in the sense that the experiment still presents an average over the influence of many degrees of freedom not directly observed in the experiment. These attributes suggest that the methods presented here open a number of new avenues for making comparisons between fast folding experiments and molecular dynamics simulations.

Acknowledgement

We thank Arend Dijkstra and Jasper Knoester for discussions on characterizing amide I vibrations of β -sheets, and for sharing a preprint. We thank Minhaeng Cho for suggestions about assigning amide I site energies in β -sheets, and sharing *ab initio* calculations of amide I couplings found in Ref. 34.

References

- (1) Sabelko, J.; Ervin, J.; Gruebele, M. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6031.
- (2) Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2527.
- (3) Yang, W. Y.; Gruebele, M. *Biophys. J.* **2004**, *87*, 596.
- (4) Huang, C.-Y.; Getahun, Z.; Zhu, Y.; Klemke, J. W.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2788.
- (5) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181.
- (6) Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469.
- (7) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *J. Chem. Phys.* **2004**, *120*, 8201.
- (8) Demirdöven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. *J. Am. Chem. Soc.* **2004**, *126*, 7981.
- (9) Chung, H. S.; Khalil, M.; Tokmakoff, A. *J. Phys. Chem. B* **2004**, *108*, 15332.
- (10) Miyazawa, T. *J. Chem. Phys.* **1960**, *32*, 1647.
- (11) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
- (12) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123.
- (13) Jourdan, M.; Searle, M. S. *Biochemistry* **2000**, *39*, 12355.
- (14) Harding, M. M.; Williams, D. H.; Woolfson, D. N. *Biochemistry* **1991**, *30*, 3120.
- (15) Cox, J. P. L.; Evans, P. A.; Packman, L. C.; Williams, D. H.; Woolfson, D. N. *J. Mol. Biol.* **1993**, *234*, 483.
- (16) Stockman, B. J.; Euvrard, A.; Scahill, T. A. *J. Biomol. NMR* **1993**, *3*, 285.
- (17) Alonso, D. O. V.; Daggett, V. *J. Mol. Biol.* **1995**, *247*, 501.
- (18) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545.
- (19) Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed.* **1998**, *37*, 868.
- (20) Dill, K. A. *Protein Science* **1999**, *8*, 1166.
- (21) Succi, N. D.; Onuchic, J. N.; Wolynes, P. G. *Proteins* **1998**, *32*, 136.
- (22) Gruebele, M. *Annu. Rev. Phys. Chem.* **1999**, *50*, 485.
- (23) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11615.

- (24) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76.
- (25) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193.
- (26) Wang, T.; Zhu, Y.; Gai, F. *J. Phys. Chem. B* **2004**, *108*, 3694.
- (27) Bredenbeck, J.; Helbing, J.; Behrendt, R.; Renner, C.; Moroder, L.; Wachtveitl, J.; Hamm, P. *J. Phys. Chem. B* **2003**, *107*, 8654.
- (28) Briggs, M. S.; Roder, H. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2017.
- (29) Khorasanizadeh, S.; Peters, I. D.; Roder, H. *Nat. Struct. Biol.* **1996**, *3*, 193
- (30) Krantz, B. A.; Sosnick, t. r. *Biochemistry* **2000**, *39*, 11696.
- (31) Gladwin, S. T.; Evans, P. A. *Folding & Design* **1996**, *1*, 407.
- (32) Qin, Z.; Ervin, J.; Larios, E.; Gruebele, M.; Kihara, H. *J. Phys. Chem. B* **2002**, *106*, 13040.
- (33) Larios, E.; Li, J. S.; Schulten, K.; Kihara, H.; Gruebele, M. *J. Mol. Biol.* **2004**, *340*, 115.
- (34) Ham, S.; Cho, M. *J. Chem. Phys.* **2003**, *118*, 6915.

Chapter 7

Temperature dependent unfolding dynamics of wildtype ubiquitin

7.1. Introduction

In chapter 6, we proposed a free energy surface of ubiquitin unfolding, which is mapped by the two reaction coordinates that are related to the two vibrational modes of the β sheet. The unfolding dynamics on this surface are described based on the sequential unfolding scenario in which the less stable β strands (III-V) unfold earlier than the remaining stable β hairpin (I, II). This scenario has been supported by several studies investigating the stability of the α helix and β strands I and II, which constitute the N-terminus of the protein. Fragment studies have shown that this region (residue 1-35) is stable.^{1,2} The transition state has also been directly mapped by the Φ value analysis with various mutants, in which the transition state is localized in the same stable region, the α helix and the β strand I and II.^{3,4} In addition, nuclear magnetic resonance studies support the stability of this regions by revealing that the native contacts are preserved in the 60% methanol and water mixture.⁵⁻⁷ This state is called A-state and regarded as a possible intermediate state. Ubiquitin is thought to be a two-state folder but the presence of the

intermediate state is still controversial because the appearance of the burst phase in some of the refolding experiments depends on the probing method,^{8,9} the time resolution of the instrument,¹⁰⁻¹² and the folding environment such as the salt concentration and temperature.¹³

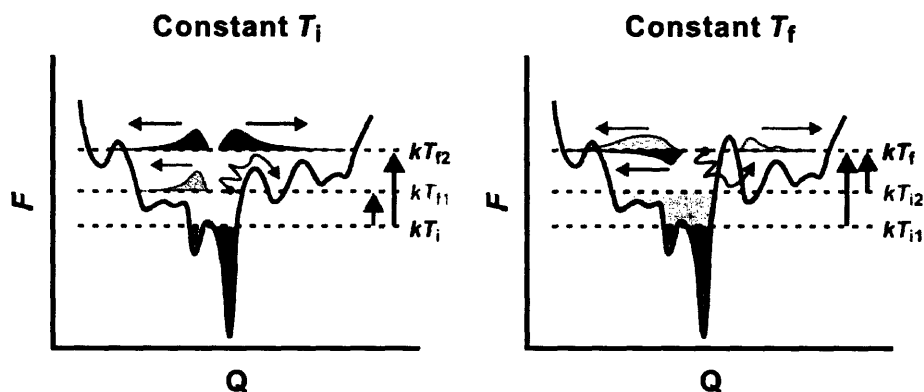


Figure 7-1. Temperature dependent T-jump experiment. F and Q are the free energy and a one-dimensional configurational coordinate, respectively. Dashed horizontal lines indicate the accessible thermal energy level at each temperature. The unfolding relaxations of the barrierless μ s process and the barrier-crossing ms process are represented by a straight arrow with an asymmetric Gaussian-shaped profile and a wiggled arrow across a barrier on the thermal energy level, respectively. The two free energy surfaces were chosen to illustrate the possible cases described below. (Left) T-jump from the same initial temperature. The ms component in the right side at T_{f1} is converted to the μ s component at T_{f2} because of the disappearance of a barrier, which results in the increase of the amplitude of the μ s component. (Right) T-jump to the same final temperature. The small population in the right side at T_{f2} shows a non-exponential relaxation after the T-jump. Also, a broader population in the left side gives a bigger amplitude μ s time scale change.

The free energy surface including the transition state can be investigated by observing the changes following given perturbations either on the environment or on the protein. The perturbing methods can be either a temperature change of the solution or site-specific mutation on the protein. In this chapter, the unfolding dynamics are investigated by the former method, changing the initial and final temperature of the T-jump: the fixed T-jump width, the fixed initial temperature (T_i), and the fixed final temperature (T_f). The expected temperature-dependent unfolding relaxations are shown with a simple one-dimensional free energy curve in Fig. 7-1. In the fixed T_i experiment,

the unfolding relaxation of the T-jump to the lower final temperature (T_f) will have two component: μ s unfolding to the barrier-free region in the left side and ms unfolding through a barrier in the right side of the free energy curve. This situation is similar to what is described in Chapter 6. With a larger T-jump, more configurational space is accessible without crossing a barrier and the amplitude of the μ s component will increase. In the fixed T_f experiment, the change in the absolute population will decrease with T_i because of the decrease of the T-jump width. However, the relative amplitude of the μ s relaxation will increase because of the wider distribution of the population at higher T_i (blue region) that gives more chances to access the barrier free space in the free energy curve. Therefore, we expect the relative amplitude of the fast non-exponential relaxation increases with T_i and T_f .

We have performed data analysis of these experiments based on two-state kinetics, but modified by adding a fast component of small amplitude on the μ s timescale in the unfolding phase. The temperature dependence of the amplitude and timescale of this fast unfolding phase reveals the barrier-free unfolding rate and the relative amount of population trapped in the transition state region by the T-jump. Changes in the barrier height can be obtained from the temperature dependent kinetic analysis of the slow refolding phase. A deviation from the quasi-two-state kinetics is found at high temperature and this strange relaxation is interpreted as the increase in the heterogeneity of the transition state caused by shifting from sequential unfolding to global unfolding through multiple routes.

A supporting calculation of the free energy surface, comparison with molecular dynamics simulation, and site-specific mutant studies will be presented in the following chapters.

7.2. Methods

The experiment performed in this chapter is a T-jump followed by a dispersed vibrational echo (DVE) probe, which is similar to that of Chapter 6. The generation and detection of the third order signal, the delay control between the T-jump pulse, and the DVE probe were explained in Chapter 4. The DVE measurement is related to the squared

absolute value of the projection of the complex 2D IR spectrum onto the ω_3 axis, as explained in Chapter 3. In spite of the one-dimensional projection of a 2D IR spectrum, it is still sensitive to the vibrational coupling between the amide I oscillators. To observe the temperature-dependent dynamics, the temperature is jumped from various initial temperatures (T_i) with similar T-jump widths (ΔT). T-jumps of various widths from the same T_i or to the same final temperature (T_f) were also performed, which makes it possible to separate the effect of the initial and the final temperature.

For a variable T-jump width, the pulse energy is changed by varying the delay between the flash lamp and the Q-switch trigger in the T-jump laser. The pulse energy decreases to 83% and 64% of the maximum at the delays of 350 μ s and 360 μ s, respectively. In these cases, the actual pulse output is delayed additionally from the Q-switch trigger by 4 and 8 ns, and these delays are compensated by moving the whole delay set of the T-jump laser for synchronization with the 6 μ m probe pulse.

The same wildtype ubiquitin sample is used as in Chapter 6. The concentration of the protein is 30 mg/mL. The solution is slightly different in the concentration of DCl from that of Chapter 6. We keep the concentration of the DCl at 0.35% (1/100 dilution) rather than make a pH 1 solution because control of the concentration is easier than pH. As a result, the melting temperature is slightly higher than that of Chapter 6.

7.3. Results

7.3.1. Thermodynamic analysis

Since the solution composition is slightly different from that of Chapter 6, we obtained the new thermodynamic parameters from equilibrium DVE measurements. Fig. 7-2a and 2b show the temperature dependent DVE spectra and their differences in about 10°C interval. A melting curve (Fig. 7-2c) is constructed by fitting to the second SVD component with a two-state model,¹⁴

$$K = \frac{P_U}{P_F} = \exp(-\Delta G(T)/RT) \quad (7.1)$$

$$\begin{aligned}\Delta G(T) &= \Delta H(T) - T\Delta S(T) \\ &\approx -\Delta S_m(T - T_m) - \frac{\Delta C_p}{2T_m}(T - T_m)^2,\end{aligned}\quad (7.2)$$

where the temperature dependent enthalpy and entropy of unfolding are given as $\Delta H = \Delta H_m + \Delta C_p(T - T_m)$ and $\Delta S = \Delta S_m + \Delta C_p \ln(T/T_m)$, respectively. ΔH_m and ΔS_m are the enthalpy and entropy of unfolding at the melting temperature $T_m (= \Delta H_m/\Delta S_m)$. The heat capacity change, ΔC_p is assumed to be constant. The higher-order terms of the temperature in the Taylor expansion of ΔS are ignored. P_U and P_F represent the fractional population of the unfolded and folded state. ($P_U + P_F = 1$) The spectra at 34°C and 82°C are assumed to be the representative of the folded and unfolded species. The fitting gives $T_m = 64.0^\circ\text{C}$, $\Delta H_m = 40.9$ kcal/mol, $\Delta S_m = 121$ cal/mol·K, and $\Delta C_p = 1264$ cal/mol·K.

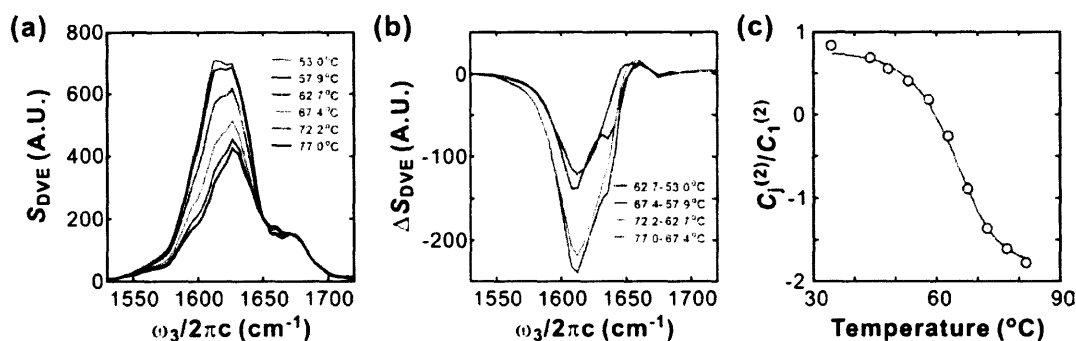


Figure 7-2. Equilibrium thermal unfolding of ubiquitin. (a) Temperature dependent DVE spectra. (b) DVE difference spectra. (c) Melting curve (red line) obtained by a fitting to the second SVD component (open circles).

7.3.2. Ultrafast responses of the solvated region

In Chapter 6, it is mentioned that a fast response includes the transmission increase across the spectrum and an additional change in the red side attributed to the local changes in the hydrogen-bonding environment of the solvated residues. To confirm this conjecture, we performed the T-jump experiment on N-methylacetamide (NMA). Since NMA is the simplest peptide system that has only one amide I oscillator, the peak position and the lineshape are only determined by the hydrogen-bonding environment.

The dependence of the amide I frequency on the strength of the hydrogen bonds to water molecules have been investigated.^{15,16} The weakened hydrogen bonds make the vibrational force constant larger, which induces a blue shift of the spectrum and depletion of intensity in the red side of the spectrum. Ubiquitin is also investigated by T-jump at low temperature (25°C) to observe similar responses with minimal structural change.

Fig. 7-3a shows the transient DVE difference spectra of NMA at six different delays after a T-jump from 58°C to 69°C. The difference spectra ($\Delta S_{\text{DVE}}(\omega, \tau)$) between the transient DVE signal at delay τ ($S_{\text{DVE}}(\omega, \tau)$) and the reference spectrum ($S_{\text{DVE}}^{\text{Ref}}(\omega)$) are normalized to the maximum of the reference spectrum as

$$\Delta S_{\text{N}}(\omega, \tau) = \frac{S_{\text{DVE}}(\omega, \tau) - S_{\text{DVE}}^{\text{Ref}}(\omega)}{\text{Max}[S_{\text{DVE}}^{\text{Ref}}(\omega)]} = \frac{\Delta S_{\text{DVE}}(\omega, \tau)}{\text{Max}[S_{\text{DVE}}^{\text{Ref}}(\omega)]}. \quad (7.3)$$

The reference spectrum at 58°C in the upper panel is obtained by summing the spectra of the 49th and 50th reference pulses following the T-jump pulse. (Fig. 4-5) The two peaks of the reference spectrum at 58°C do not originate from two different resonances but from the fundamental ($0 \rightarrow 1$) and overtone ($1 \rightarrow 2$) transitions. As expected, in addition to the increase of the signal over the whole frequency range by the transmission increase, there is a depletion of the intensity in the red side of the spectrum around 1600 cm^{-1} at $\tau = 200 \text{ ns}$. The red dashed line with an arrow indicates the maximum of the reference spectrum and shows its red shift. Therefore, we conclude that this change results from the rapid disruption of the hydrogen bond between the amide I oscillator and solvent molecules.

The response of ubiquitin after a T-jump from 25°C to 37°C is similar to that of NMA as shown in Fig. 7-3b. The blue shift of the red side of the spectrum indicates the depletion of the intensity. The similarity of the transient DVE difference spectrum at $\tau = 100 \text{ ns}$ with the equilibrium difference spectrum between 25°C and 35°C ($\Delta S_{\text{Eq}} = (S_{\text{DVE}}(35^\circ\text{C}) - S_{\text{DVE}}(25^\circ\text{C})) / \text{max}[S_{\text{DVE}}(25^\circ\text{C})]$) both in the shape and in the amplitude and no further changes in the spectral shape with delay time suggest that most of the changes at this temperature are finished within 100 ns. The following transient changes are purely due to the buffer transmission change. Since 100 ns is too short period

for significant structural changes of proteins, the spectral change of ubiquitin for $0 < \tau < 100$ ns most likely arises from the same phenomena as that of NMA. We therefore assign these changes to decrease in the coordination and the strength of the hydrogen bond between the solvent exposed residues of the protein and water molecules.

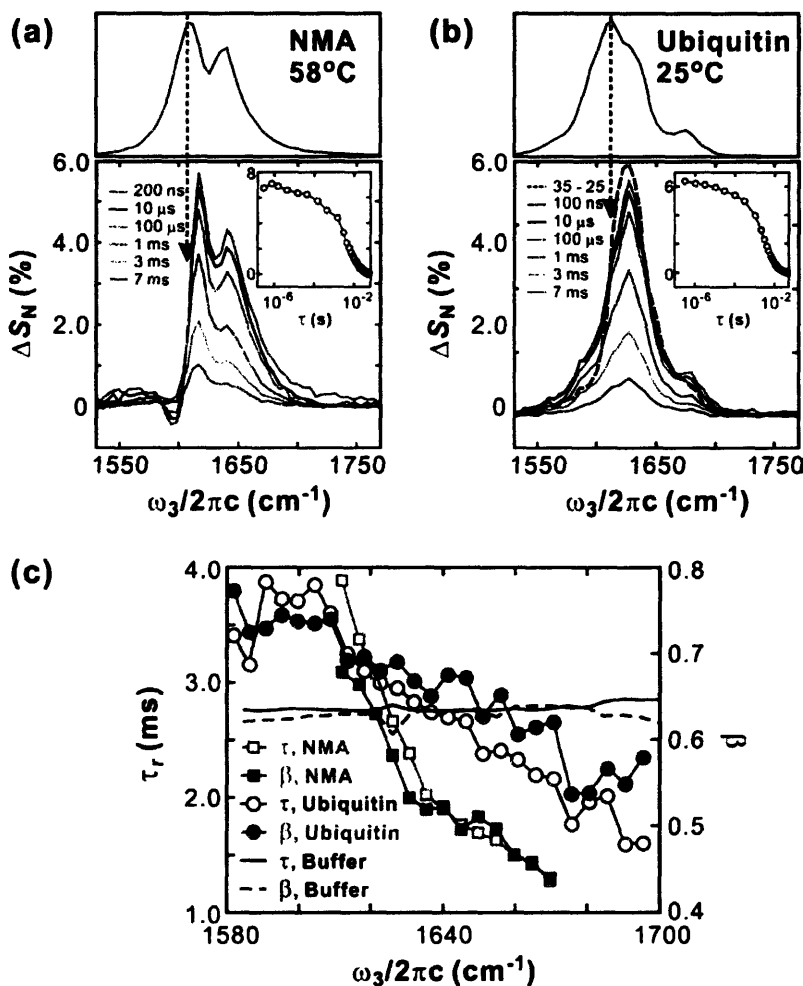


Figure 7-3. Comparison of transient changes of NMA and ubiquitin. Transient spectra are shown after the T-jump from (a) $T_i = 58^\circ\text{C}$ for NMA and (b) $T_i = 25^\circ\text{C}$ for ubiquitin. Equilibrium DVE spectra at the initial temperatures are plotted in upper panels. Also shown in the insets are the normalized relaxation profiles at 1618 cm^{-1} . For ubiquitin, the equilibrium difference spectrum ($35^\circ\text{C} - 25^\circ\text{C}$) is shown with a black dashed line with the transient spectra for comparison. (c) Frequency dependent time constant (τ) and β for the stretched exponential fit of the relaxations of NMA, ubiquitin, and the transmission of the DCI buffer solution are compared.

To characterize the relaxation behavior and the temperature equilibration of the system, a stretched exponential fit is performed for each channel using Eq. (7.4),

$$U(\tau) = \exp[-(\tau/\tau_r)^\beta]. \quad (7.4)$$

The fitting parameters are summarized in Fig. 7-3c. The frequency dependence is very small for the transmission change of the buffer solution while both the time constant and β decrease at high frequency region both for NMA and ubiquitin. It is not clear what affects the faster decay of the blue side but the low level of signal in the blue side and the nonlinearity of the DVE signal may affect those parameters.

7.3.3. Transient spectral changes of ubiquitin unfolding

Fig. 7-4 shows normalized transient DVE difference spectra, $(\Delta S_N(\omega, \tau))$, following T-jumps from three different initial temperatures with a width of 11 – 12°C. Also, the temporal profile of the three frequency regions, ν_\perp , ν_\parallel , and random coil regions, as indicated with arrows, are plotted in Fig. 7-4. In this plot, the relative change is obtained by dividing the DVE signal at each frequency with the reference signal of the same frequency. $(\Delta S_{DVE}(\omega, \tau)/S_{DVE}^{Ref}(\omega))$ In this section, the frequency dependent responses are described qualitatively. Quantitative analyses using the SVD of the three frequency blocks shown with color bars above the spectra are presented in the following sections.

As explained in Chapter 6, in general, the spectral changes of the three regions occur over several decades of time. Changes appear in the ν_\perp region from several μ s to ms timescales, while the relaxations of the ν_\parallel and random coil regions occur from 100 μ s to ms. After 5 – 6 ms, refolding occurs as a result of the temperature re-equilibration of the sample.

Temperature-dependence relaxation can be seen in the ν_\perp region for both the μ s and ms components. The relative amplitude of the μ s component becomes bigger as the temperature is raised, which reflects an increase of trapped population at the unfolding barrier or equivalently a broader accessible configurational space for the unfolded state at higher temperature, as shown in Fig. 7-1. The amplitude of the ms component also grows and becomes faster with temperature because of the increase of the unfolding rate with

temperature. The changes of the ν_{\parallel} and random coil regions are the same as those described in Chapter 6 except those of the T-jump from 67°C, for which relaxation is slow or even increases for the random coil region. The deviation of the relaxations of the T-jump at $T_i = 67^\circ\text{C}$ will be discussed more in Section 7.3.5.

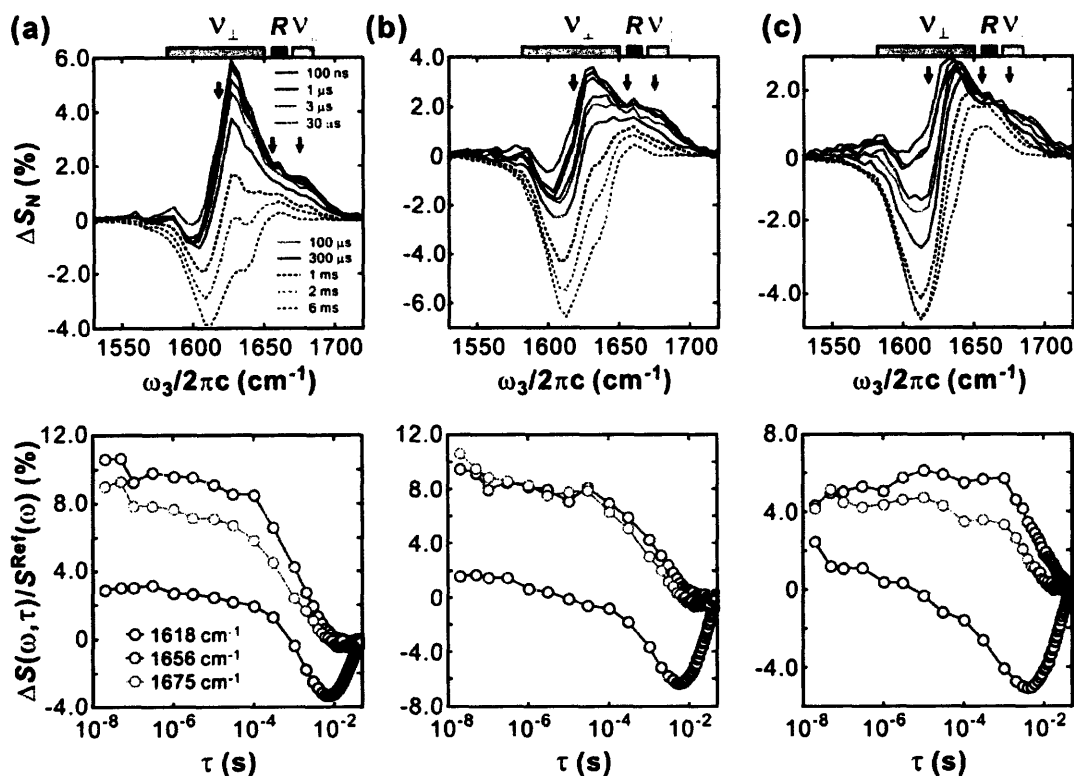


Figure 7-4. Temperature dependent transient thermal unfolding of ubiquitin. (Upper) DVE difference spectra normalized to the maximum of the reference DVE spectrum are shown for several delay times at different initial temperatures, (a) 58°C, (b) 63°C, and (c) 67°C. Arrows indicate the frequency channel plotted with the same colors in the lower panels. Color bars above each of the spectrum mark the frequency range where the SVD analyses are performed. (Lower) Spectral relaxations of three frequency components are shown. Each time trace is normalized by the reference DVE intensity at its frequency.

7.3.4. Analysis of unfolding in ν_{\perp} region

We use the relaxation profile of the ν_{\perp} region to extract the thermodynamic parameters of the activation energy barrier. In this spectral range, the signal is intense enough to distinguish the small amplitude of the μs relaxation, the ms unfolding part, and

the refolding part. This barrier information along with thermodynamic parameters of the protein obtained in Section 7.3.1 is also used in the calculation of the population change during thermal unfolding and refolding in Section 7.4.

The analysis of protein unfolding from the spectral responses is, however, complicated by the temperature relaxation because the temperature-dependent transmission change of the probe beam alters the spectral responses. The unfolding from several hundred μs to several ms is mostly affected by the temperature relaxation, the timescale of which is 2 – 3 ms. As shown in Fig. 7-4a, the spectral changes in the ν_{\perp} region are within $\pm 6\%$ of the reference spectrum. Therefore, we assume that both the spectral changes from the protein structure and temperature are in the linear regime even though the DVE measurement is a nonlinear probe. Although the complete deconvolution of the responses by structural changes and by the temperature change is not possible at the current signal-to-noise level of the experiment, we minimize the temperature effect by redefining the DVE difference using the temperature corrected difference spectrum at 100 ns delay as

$$\Delta S_{\text{DVE}}'(\omega, \tau) = \Delta S_{\text{DVE}}(\omega, \tau) - \Delta S_{\text{DVE}}(\omega, \tau = 100 \text{ ns}) \cdot U(\tau). \quad (7.5)$$

Here, $U(\tau)$ ($= \exp[-(\tau/\tau_r)^\beta]$) is the temperature relaxation profile obtained from the SVD analysis of the transmission change of the tracer beam for the whole frequency range. ($\beta = 0.68$, $\tau_r = 2.7$ ms) In other words, $\Delta S_{\text{DVE}}'(\omega, \tau)$ is a double difference spectrum with respect to the temperature corrected difference at $\tau = 100$ ns. We chose the 100 ns spectrum as a baseline because of the shock wave effect that might be present in the probed region at earlier delays as mentioned in Chapter 4.

The relaxation of the ν_{\perp} region is constructed from the first SVD component ($C_j^{(1)}$) of $\Delta S_{\text{DVE}}'(\omega, \tau)$ over the frequency range from 1582 cm^{-1} to 1651 cm^{-1} . The SVD analysis has advantages over the analysis on a single frequency component in the representation of the averaged spectral responses resulting in a better signal-to-noise ratio. Also, the similarity in the spectral shapes of the first SVD component spectra and the difference spectra of the equilibrium measurement as shown in Fig. 7-5 implies that

the remaining temperature relaxation effect in Eq. (7.5) is contained in the higher-order SVD components.

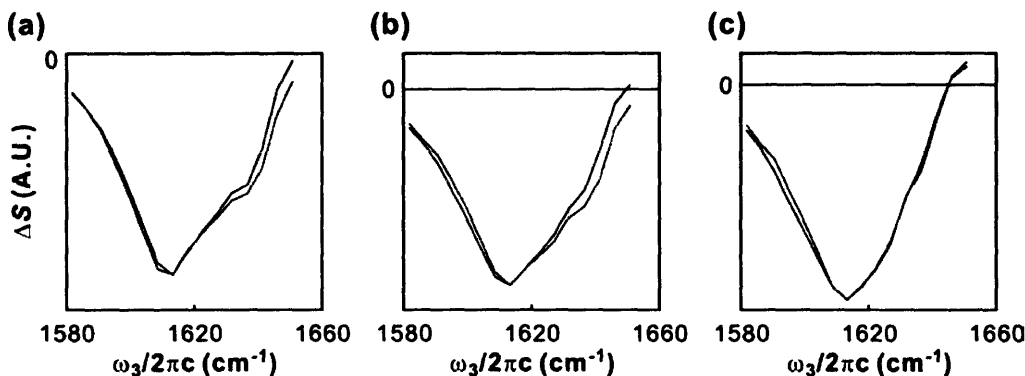


Figure 7-5. Comparison of the equilibrium DVE difference spectra (blue) and the first SVD component spectra (red) of the transient data (1582 – 1651 cm^{-1}) at three initial temperatures, (a) 58°C, (b) 63°C, and (c) 67°C. Each spectrum is scaled to its minimum.

Using this modified relaxation profile, the unfolding and folding parameters are extracted as shown in Fig. 7-6a. The fitting is performed separately for the unfolding and refolding time scales. The unfolding is described by a stretched exponential and an exponential decay for the μs part and ms part, respectively,

$$C_j^{(1)} = A_{\mu\text{s}} \exp\left[-\left(\tau_j/\tau_{\mu\text{s}}\right)^\beta\right] + A_{\text{ms}} \exp\left[-\tau_j/\tau_{\text{ms}}\right], \quad (7.6)$$

where τ_j is the j th delay time ranging from 100 ns to the time when $C_j^{(1)}$ reaches its minimum (6 ms for the $T_i = 63^\circ\text{C}$ data). 100 ns is chosen again as the earliest delay that is free from the possible shock wave effect. The μs relaxation provides the timescale ($\tau_{\mu\text{s}}$), the heterogeneity (β), and the amplitude ($A_{\mu\text{s}}$) of the fast unfolding component that is transiently trapped around the transition state of the free energy surface by a rapid temperature change. The slow ms component provides the timescale (τ_{ms}) of the relaxation through the activation barrier as explained in Chapter 6. However, the temperature decay during the ms relaxation induces the continuous changes of the unfolding and folding rate. Therefore, in the fitting procedure for the unfolding part, the exponential function is used just to guide the slow relaxation part for the best fitting of the fast relaxation part rather than to extract the timescale of the ms component.

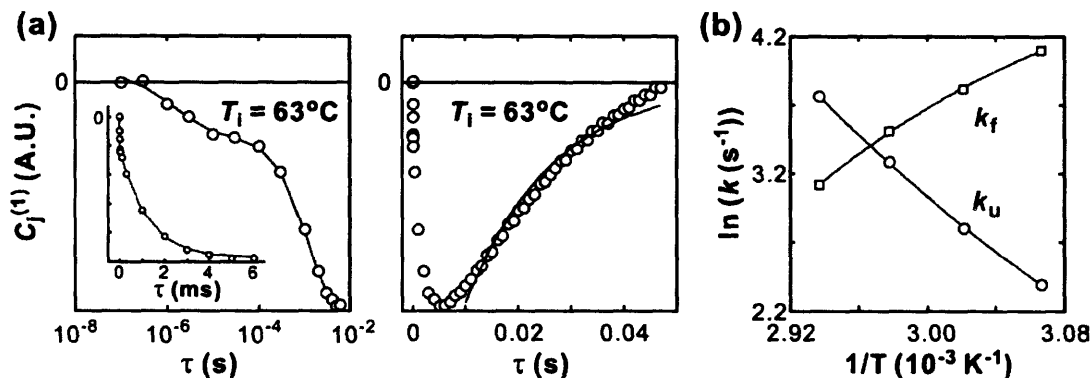


Figure 7-6. Extracting relaxation parameters. (a) (Left) Fast components are obtained by fitting the data with a stretched exponential (μs) and an exponential (ms) functions from 100 ns to 6 ms. The inset shows the decay with a linear timescale. The fast relaxation component will appear as the burst phase in many experiments, where the time resolution is longer than the dynamics. (Right) Refolding part is fitted with a single exponential from 10 ms to 47 ms. ($T_i = 63^\circ\text{C}$) (b) Arrhenius plot of the folding and unfolding rates obtained from the refolding part of the relaxation.

The fast unfolding parameters are summarized in Table 7-1. In the fixed T-jump width experiment (upper block), as the initial temperature becomes higher, the relaxation (τ_a) is getting slower while the heterogeneity is not changed. Although the number of data points from ns to μs is too small to distinguish a timescale difference of a single μs , the big increase of the timescale of the T-jump from $T_i = 67^\circ\text{C}$ is an obvious deviation from the other data. This big increase implies the abrupt changes around the transition state region, which are discussed in Section 7.4 and Chapter 8. The relative amplitude of change, $A_{\mu\text{s,rel}}$ ($A_{\text{ms,rel}}$) is obtained from the multiplication of $A_{\mu\text{s}}$ (A_{ms}) in Eq. (7.6) and the ratio of the amplitude of the first SVD component spectrum to that of the equilibrium difference spectrum,

$$A_{\text{rel}} = A \cdot \frac{\text{Min} \left[S_{\text{SVD}}^{(1)} \right]}{\text{Min} \left[\Delta S_{\text{Eq}} \right]}, \quad (7.7)$$

Here, $S_{\text{SVD}}^{(1)}$ is the first SVD component spectrum and ΔS_{Eq} is the equilibrium DVE difference spectrum. ($\Delta S_{\text{Eq}} = S_{\text{DVE}}(T_f) - S_{\text{DVE}}(T_i)$) The relative change is obtained from the ratio of the minima of $S_{\text{SVD}}^{(1)}$ and ΔS_{Eq} because their spectral shape is similar as shown in Fig. 7-5. Both of the relative amplitudes of the fast component and the total change

increase with the initial temperature. The bigger amplitude of the ms component at higher temperature obviously results from the increase of the unfolding rate. The relative amplitude of the fast component can be converted to the relative trapped population among the total folded population, $P_{\mu s}/P_F$, using the ratio of the SVD component spectra, the equilibrium difference spectra (Fig. 7-5), and the population changes calculated from the equilibrium thermodynamic quantities. (Note that we assume that the transiently trapped species are part of the folded species.) The complete procedure can be found in Section 7.4. The relative trapped population increases with temperature but it is not clear if the effect is by the initial temperature, the final temperature, or both.

$T_i - T_f$ (°C)	β	$\tau_{\mu s}$ (μs)	$A_{\mu s,rel}$ (%)	$A_{\mu s,rel} + A_{ms,rel}$ (%)	$P_{\mu s}/P_F$ (%)
Fixed ΔT					
53.0 – 65.0	0.63	1.1	2.2	7.1	1.0
57.9 – 69.9	0.44	0.7	2.3	10.6	1.6
62.7 – 74.3*	0.62	1.8	4.8	17.0	3.8
67.4 – 77.9	0.61	8.0	6.9	20.1	5.9
Fixed T_f					
57.9 – 69.9	0.44	0.7	2.3	10.6	1.6
59.8 – 69.4	0.45	0.9	3.4	13.6	2.1
61.7 – 69.1*	0.48	2.1	4.4	16.4	2.4

Table 7-1. Relaxation parameters extracted from the SVD component of the ν_{\perp} region (1582 – 1651 cm^{-1}). β and τ_a are the stretched exponential ($\exp[-(\tau/\tau_{\mu s})^{\beta}]$) parameters for the μs relaxation part. $A_{\mu s,rel}$ and $A_{ms,rel}$ are the relative amplitudes of the μs and ms relaxation compared to the expected change when the temperature is kept constant at the final temperature of the T-jump. $P_{\mu s}/P_F$ is the relative population change during the μs relaxation to the folded population at the initial temperature. The upper block of the table shows the T-jump with a fixed width at various initial and final temperatures. The lower block is the result of T-jumps to the same final temperature with variable widths. The two data marked with an asterisk (*) represent the T-jump from the similar initial temperatures.

The experiment with variable T-jump widths separates the effect of T_i and T_f . The dependence on T_i can be found in the experiment with the same T_f . As summarized in the lower block of the Table 7-1, the transiently trapped population increases with T_i . Also, the comparison the two data marked with an asterisk show the effect of T_f from a similar T_i . As we predicted in Fig. 7-1, the relative amplitude of the fast component increases both with T_i and with T_f though a direct comparison of the 1D free energy curve and the 2D free energy surface is not straightforward. The increase of the relative population for the fast phase in the constant T_f experiment can be explained in a similar manner, that more population visits the transition state region at higher T_i . The same trend in the constant T_i experiment implies a bigger shift of the barrier by a wider T-jump.

Since the ms kinetics in the unfolding part are complicated by the temperature changes during unfolding, we use the refolding part for the kinetic analysis. The relaxation rate can be assumed constant at the initial temperature after 10 ms because only 9% of the heat from the T-jump remains at 10 ms, which corresponds to 1°C. Then, the relaxation should appear as a single exponential decay in the two-state kinetics. The right panel in Fig. 7-6a shows an exponential fitting of the refolding part from 10 ms to 47 ms, and the temperature dependent kinetic parameters are summarized in Table 7-2. We can determine the individual folding and unfolding rate, k_f and k_u using the relationships with the relaxation rate ($1/\tau_r = k_f + k_u$) and the equilibrium constant ($K = k_u / k_f$). The Arrhenius plot is shown in Fig. 7-6b with fits to the Kramers model¹⁷ using equations below,^{18,19}

$$k_u = \frac{k_o}{\eta(T)} \exp \left[-\frac{E_0 + E_1(T - T_m) + E_2(T - T_m)^2}{RT} \right]. \quad (7.8)$$

$$k_f = k_u \exp(\Delta G/RT)$$

Here, the activation energy is expanded to the square term of the temperature as ΔG in Eq. (7.2). $\eta(T)$ is the relative viscosity of D₂O to the viscosity of water at 20°C.²⁰ We used the fitted values using the formula $\eta(T) = 0.292 + 2.00 \cdot \exp(-T(^{\circ}\text{C}) / 27.96)$. The barrier free rate k_o is considered to be the relaxation by diffusion of the polypeptide

chain. This rate should be similar to our fast unfolding relaxation rate, τ_a , which is the barrierless transition. Therefore, we use $k_o = 0.5 \mu\text{s}^{-1}$ ($k_o/\eta(T) \sim 1 \mu\text{s}^{-1}$), and this is also similar to the speed limit of the folding rate ($= (1 \mu\text{s}^{-1}) \cdot (N/100)$), where N is the number of residues of the protein.²¹ As a result, the fitting gives $E_0 = 7.0$ kcal/mol which is the barrier height at the melting temperature T_m . $E_1 = -38$ cal/mol·K and $E_2 = -0.89$ cal/mol·K² give the temperature-dependent change of the unfolding barrier, which corresponds to 300 - 500 cal/mol over the 10°C change.

T (°C)	τ_r (ms)	k_f (s ⁻¹)	k_u (s ⁻¹)
53.0	14.0	60.2	11.0
57.9	16.1	45.5	16.5
62.7	16.7	33.4	26.6
67.4	15.3	22.5	43.0

Table 7-2. Temperature dependent relaxations in the two-state kinetic regime.

7.3.5. Unfolding monitored in other frequency regions

As shown in Fig. 7-2 and 7-4, the equilibrium DVE intensity in the blue side of the spectrum is about a quarter of that of the ν_{\perp} region. The spectral changes following T-jump are also small and dominated by the temperature-dependent transmission changes. Therefore, the subtraction of the transmission change as done in the previous section leaves signals too small to perform a meaningful analysis. Instead, we apply the SVD without subtracting the temperature relaxation effect and simply obtain the approximate time scale of the structural relaxation by stretched exponential fitting.

Fig. 7-7 shows the spectral relaxations of the ν_{\parallel} and random coil regions. Stretched exponential fittings are shown with solid lines and the temperature relaxation profiles are shown with dashed lines of the same colors for comparison. The temperature profile is normalized to the maximum of each of the fitted lines. The relaxations of $T_i =$

58°C and 63°C for both frequency regions are faster than the temperature relaxation of 2.7 ms. The beginning of the spectral responses due to structural changes appear as the deviation starting at about tens of microseconds, which is a similar time scale presented in Chapter 6. The fitting parameters are summarized in Table 7-3. The relaxation is faster in ν_{\parallel} region than in the random coil region over all the temperature ranges. The difference in the timescale of the two frequency regions may arise from the sign of the change. The equilibrium difference shown in Fig. 7-2b, indicates that the net change is positive in the random coil region while the change is negative in the ν_{\parallel} region. These changes appear on top of the temperature relaxation, then, the signal of the ν_{\parallel} region decays faster than that of the random coil region. The relaxation rate of the random coil region for the T-jump from $T_i = 58^\circ\text{C}$ and 63°C are similar to that of 25°C in Fig. 7-3, which indicate the changes in the random coil region are mostly due to the temperature relaxation. On the other hand, the relaxation of 1.1 ms in the ν_{\parallel} region is much faster than that of 25°C , which indicates the negative changes in the ν_{\parallel} region. The difference is clearer in the plot with a linear time scale in the inset.

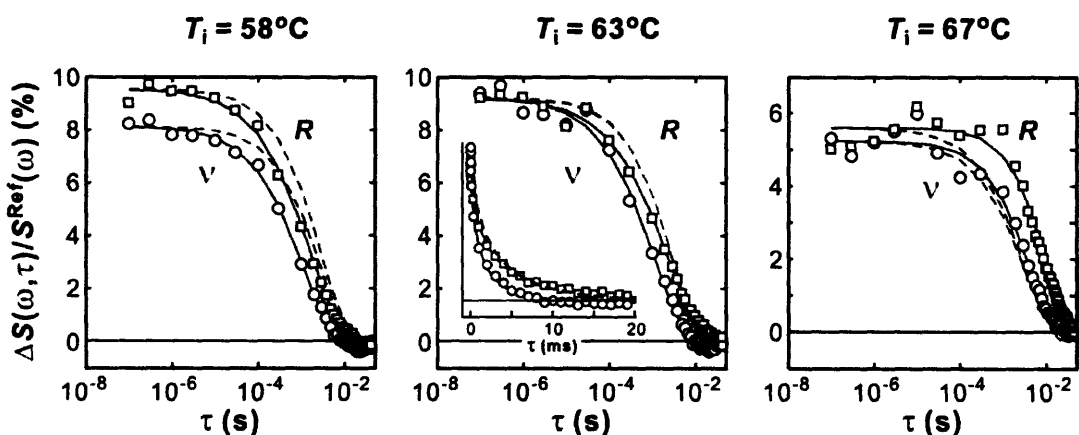


Figure 7-7. Spectral relaxations of the ν_{\parallel} and random coil region at three different initial temperatures. Open circles and squares represent the relative amplitude of the first SVD component spectra of ν_{\parallel} mode ($1670 - 1690 \text{ cm}^{-1}$) and random coil ($1656 - 1666 \text{ cm}^{-1}$) regions, respectively. Stretched exponential fits are shown with green and red lines. Also, the temperature relaxation profiles are shown with dashed lines. (Inset) Relaxations are plotted with linear time scale.

The relaxation of these two frequency regions for the highest T-jump at $T_i = 67^\circ\text{C}$ are different from those of the low temperature data. As shown in the relaxation profile both in Fig. 7-4c and in Fig. 7-7, the spectral response of the random coil region is increasing rather than decreasing until 1 ms, which indicates a positive change in this frequency region. This positive change on top of the temperature relaxation gives a longer relaxation time than that of the temperature relaxation (2.7 ms). This positive change also affects the relaxation of the ν_{\parallel} region because of their proximity in the spectrum, and the relaxation time of the ν_{\parallel} region is also longer than the temperature relaxation. It is not clear yet if this abrupt change is related to the strange behavior of the ν_{\perp} region at $T_i = 67^\circ$ discussed in Section 7.4.1.

$T_i - T_f$ ($^\circ\text{C}$)	β_{\parallel}	τ_{\parallel} (ms)	β_r	τ_r (ms)
53.0 – 65.0	0.56	1.1	0.57	1.8
57.9 – 69.9	0.62	1.1	0.59	1.6
62.7 – 74.3	0.58	1.1	0.55	2.2
67.4 – 77.9	0.75	3.9	0.86	9.7

Table 7-3. Relaxation parameters extracted from the first SVD component of the ν_{\parallel} ($1670 - 1690 \text{ cm}^{-1}$) and random coil ($1656 - 1666 \text{ cm}^{-1}$) regions.

7.4. Discussion

7.4.1. Population change during temperature relaxation

In our T-jump and DVE probe experiment, the analyses of the fast unfolding part from ns to several μs and the slow refolding part after 10 ms are relatively easy because the temperature is constant at T_i and T_f , respectively. However, the spectral changes on the timescales between those two regions are hard to model because of their convolution with the temperature change. Nevertheless, if we can convert the spectral changes to shifts in population, the relaxation can be calculated by solving rate equations. In this

section, the folded population changes obtained from the spectral relaxation of the v_{\perp} region are compared with the calculated ones.

From the thermodynamic analysis on the equilibrium experiment, the fractional folded population ($P_F(T)$) at T_i can be calculated using Eq. (7.1) and (7.2). The population change from this initial value can be obtained in a similar way to obtain Eq. (7.7), comparing the SVD component spectrum and its relaxation with the equilibrium DVE difference spectrum as

$$\Delta P_{F,j} = C_j^{(1)} \frac{\text{Min} \left[S_{\text{SVD}}^{(1)} \right]}{\text{Min} \left[\Delta S_{\text{Eq}} \right]} \Delta P_{F,\text{Eq}} \quad (7.9)$$

Here, $C_j^{(1)}$ is the coefficient and the first SVD component and $\Delta P_{F,\text{Eq}}$ is the equilibrium population difference. ($\Delta P_{F,\text{Eq}} = P_F(T_f) - P_F(T_i)$) Subscript j indicates the j th delaytime.

The population relaxation is also calculated using the temperature-dependent folding and unfolding rate constants that are obtained in Section 7.3.4. In the two-state kinetics, the transient population change can be calculated simply by solving the rate equation, $dP_F / dt = -k_U P_F + k_F P_U$. However, the relaxation shown in this chapter does not follow simple two-state kinetics because of the presence of the fast unfolding phase. Nevertheless, in our unfolding scenario, the fast nonexponential relaxation is also unfolding from the part of the folded state. Also, after the fast unfolding phase is completed, the relaxation obeys the two-state kinetics. Therefore, the rate equation is modified slightly by adding a nonexponential feeding term,

$$\begin{aligned} \frac{dP_F}{dt} &= \frac{d}{dt} \left(P_{\mu\text{s}}(t) \right) - k_U (P_F - P_{\mu\text{s}}(t)) + k_F (1 - P_F) \\ P_{\mu\text{s}}(t) &= P_{\mu\text{s}}(0) \exp \left[- \left(t / \tau_{\mu\text{s}} \right)^{\beta} \right] \end{aligned} \quad (7.10)$$

$P_{\mu\text{s}}(0)$ is the population initially positioned in the barrier region and is described by the parameters in Table 7-1. Eq. (7.10) is solved numerically with the Runge-Kutta method. The population change starts from $P_F(T_i)$ calculated from Eq. (7.1) and (7.2). The temperature dependent folding and unfolding rate constants are obtained from Eq. (7.8).

Fig. 7-8 shows the experimental and calculated population changes. The initial population is indicated with solid horizontal line (green). Because of the slow relaxation

in the refolding phase, the initial folded population is not fully recovered before the next T-jump 50 ms later. Therefore, the population reaches a steady state value marked with dashed horizontal line. The offset is determined by iterating the calculation starting from the final population of the previous run. The calculation converges within 3 iterations, which corresponds to 150 ms in the experiment. The amplitude of the μ s relaxation is bigger than this offset, which confirms that the fast relaxation is not an artifact. Also, since most of the excess population right before the next T-jump remains in the unfolded well, the spectral change by the redistribution within the well induced by the next T-jump would be small.

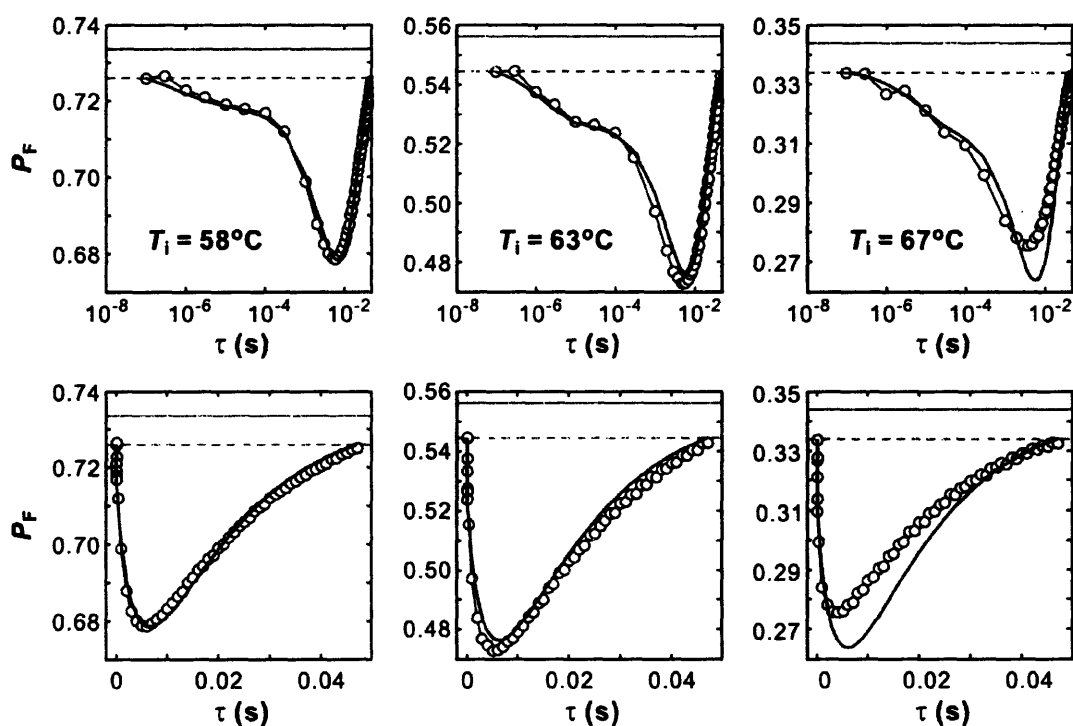


Figure 7-8. Folded population changes after a T-jump. Blue circles are experimental values and red lines are the calculated changes. Green solid and dashed lines represent the folded populations before temperature jump and at the steady state after temperature jump, respectively. Upper and lower panels are the same except the time axes that are log and linear scales, respectively.

The experimental and calculated population changes are well matched in the T-jumps from $T_i = 58^\circ\text{C}$ and 63°C , which indicate that our two-state model with fast component addition is valid in this temperature range. However, at $T_i = 67^\circ\text{C}$, the two

relaxations deviate from each other from 100 μs to several ms. The calculated relaxation indicates that the unfolding should proceed longer and further than the experimental value as shown more clearly in the plot with a linear time scale. The mismatch of this time scale and amplitude of unfolding can be attributed to the inaccuracy of the extrapolated thermodynamic parameters. The parameters for the barrier, E_0 , E_1 , and E_2 are obtained using the refolding part of the relaxation, the highest temperature of which is 67°C . However, the temperature in the unfolding part is much higher than 67°C . ($T = 71^\circ\text{C}$ at 3 ms) Therefore, if there is a bigger change of the free energy surface than what is expected from the extrapolation of Eq. (7.8), then the calculated population relaxation will not be valid any more. Another big change in the relaxation at the highest temperature is the relaxation shape of unfolding. The time scales of the fast and slow unfolding relaxations of the two low temperatures are well distinguished from each other as about 1 μs and several ms. On the other hand, the relaxation time of the fast part for the data of $T_i = 67^\circ\text{C}$ becomes slower as shown in Table 7-1 while that of the slow part becomes faster. As a result, the two phases are merged into one global stretched exponential decay, which is an indication of the changes in the unfolding mechanism. This is discussed in the next section.

7.4.2. Deviation from the two-state kinetics

The time scale of the fast component is believed to be a molecular timescale,^{22,23} in other words, the speed limit of the protein folding.²¹ The molecular timescale of 2 μs has been reported for the five-helix bundle protein λ_{6-85} ,²² which is similar to our observation of the fast phase of 1 μs . The relative amplitude of this fast phase of λ_{6-85} compared to that of the main unfolding phase exceeds 20% at low temperature where the folding rate is fast. This large amplitude of the fast phase results from the small barrier height, calculated using the ratio of the rate of the fast phase and the main phase to be $\sim 1.5 kT$ (~ 1 kcal/mol). However, the barrier height for ubiquitin is 7 kcal/mol. In this case, the population around the barrier region is much smaller than the 2 – 5% that we observed. Therefore, the presence of an intermediate state along the unfolding (folding) pathway with a small barrier cannot be ruled out.

The deviation from the two-state kinetics is also found at the time when the unfolding and refolding phase meet. In the relaxation of the v_{\perp} region, as the temperature of the solution cools down, the relaxation turns from the unfolding phase to the refolding phase after reaching a certain minimum at which the folding and unfolding rates are balanced. ($k_u P_F = k_f P_U$) In the linear timescale plots in Fig. 7-8, the time of this balanced position for the calculated curve differs by 2 – 3 ms from that of the experimental curve especially for the data at $T_i = 67^{\circ}\text{C}$. In the limit of detailed balance, the ratio of the folded and unfolded population should be the same as the ratio of k_u and k_f at the balanced position. Therefore, if the balanced position is formed earlier where k_u is bigger, the folded population should be smaller. However, in the comparison at $T_i = 67^{\circ}\text{C}$, the folded population of the experiment is bigger than the calculated value at the balanced position in spite of its earlier formation. Therefore, one can imagine the presence of an intermediate state that breaks the detailed balance and forms a steady state.

Nevertheless, the merged feature of the μs and ms relaxations at $T_i = 67^{\circ}\text{C}$ implies that the unfolding mechanism is changing with temperature and that the possible intermediate state is a flexible state having a very small barrier rather than a well-defined state. Also, the structure of this state should be very similar to the folded state since at least more than three strands of the β sheet need to be intact to show an unfolding both in the v_{\perp} region and in the v_{\parallel} regions. No indication of the intermediate state in the unfolding curve in Fig. 7-2c also supports the similarity of the possible intermediate state and the folded state. Therefore, we believe that our quasi-two-state analysis assuming that the folded state includes the intermediate state is still valid.

Using the temperature dependent free energy change, a temperature Φ value (Φ_T) analysis has been performed for a small three-stranded β -sheet domain (WW domain).^{18,19} As an analogy of the mutational Φ value, Φ_T has been defined with the ratio of the temperature gradient of the folding activation energy (ΔG_F^{\ddagger}) to the folding free energy (ΔG_F), which is given as below using Eqs. (7.2) and (7.8),

$$\Phi_T(T) = \frac{\partial \Delta G_F^\ddagger(T)/\partial T}{\partial \Delta G_F(T)/\partial T} = \frac{E_1 + \Delta S_m + (2E_2 + \Delta C_p/T_m) \cdot (T - T_m)}{\Delta S_m + (\Delta C_p/T_m) \cdot (T - T_m)} \quad (7.11)$$

$$= \frac{83 + 1.97(T(^{\circ}\text{C}) - T_m(^{\circ}\text{C}))}{121 + 3.75(T(^{\circ}\text{C}) - T_m(^{\circ}\text{C}))}$$

Also, Φ_T ranges from 0 to 1, and the higher the Φ_T value, the closer the topology of the transition state to that of the folded state.

As discussed in Ref. 18, Φ_T should increase with temperature because the transition state is energetically closer to the folded state in the reaction coordinate at higher temperature, which results in a higher Φ value. However, Eq. (7.11) shows a decreasing trend ranging from 0.77 at 53°C to 0.67 at 67°C. This seems contradictory to our unfolding picture, based on the shift of the barrier to the folded side after T-jump. In Eq. (7.11), the temperature dependence of Φ_T is determined mostly by the sign of E_2 , which is the sign of the curvature of the unfolding rate constant k_u . As shown in Fig. 7-6b, the number of data points is small and the opposite trend might be attributed to experimental error, but it can also be a result of the increase of the heterogeneity of the transition state. In the normal case of increasing Φ_T , the relative slope of the folding activation energy should be larger than that of the folding free energy in Eq. (7.11), which is opposite in our data. However, the relationship between the Φ value and the location of the barrier or the similarity of the structures between the transition state and the folded state is valid on the projection of the free energy surface in one dimension. If there is a change of the transition state from a well-defined pathway to multiple routes on the two-dimensional free energy surface at high temperature, the apparent barrier height can decrease as a result of the projection even though the barrier height of each route increases. In other words, although the folding rate of individual routes decreases, the overall rate, which is the weighted sum of individual rates, can increase with temperature. The global non-exponential unfolding feature with a longer unfolding time but a bigger amplitude of the fast phase at $T_i = 67^{\circ}\text{C}$ supports the heterogeneity of the transition state. This increase in the heterogeneity of the transition state can be derived from the

definition of the Φ_T value. Eq. (7.11) can be express with the ratios of the entropy changes as²⁴

$$\Phi_T(T) = \frac{\partial \Delta G_F^\ddagger(T)/\partial T}{\partial \Delta G_F(T)/\partial T} = \frac{\Delta S_F^\ddagger(T)}{\Delta S_F(T)}. \quad (7.12)$$

Here, ΔS_F^\ddagger and ΔS_F indicate the entropy change of the transition state and folded state, respectively. These values are negative because entropy decreases with folding. Therefore, decrease in their ratio, the Φ_T value, indicates that ΔS_F^\ddagger is less negative, which infers that the entropy (heterogeneity) of the transition state is bigger at higher temperature. The multiple transition states of ubiquitin have also been reported using ψ value analysis.²⁵ Therefore, instead of introducing an intermediate state, the deviation in the high temperature data can be explained by the changes in the free energy surface. This feature will be discussed more by following calculations in Chapter 8.

7.5. Conclusion

In this chapter, the free energy surface of ubiquitin, which is proposed in Chapter 6, is investigated using temperature dependent T-jump experiment. By adjusting the T-jump width, the effect of the initial and final temperature on the fast unfolding phase, which is the speed limit of the protein folding, can be observed separately. From the amplitude of this fast unfolding phase, the relative folded population around the barrier region is obtained as a function of T_i and T_f . This population increases both with T_i and with T_f , and both trends are the expected behaviors on our proposed free energy surface as shown in Fig. 7-1. Two-state kinetic analysis on the slow refolding part gives thermodynamic information about the barrier height. By a combination of the fast and slow unfolding and folding parameters, a quasi-two-state kinetic analysis is performed to calculate the population changes of the folded state for the region where the temperature is being changed. This calculation coincides with the experimentally obtained population changes at low temperature but deviations are found in the T-jump from 67°C to 78°C. From the merged feature of the fast and slow unfolding phase at this high temperature and the decreasing trend of the Φ_T value with temperature, we conclude that the changes

of the transition state from a well-defined pathway to multiple routes can result in the deviation. More interpretations for this new aspect of the free energy surface at high temperature will be followed by a free energy calculation and comparison with the molecular dynamics simulation in Chapter 8. Also, the transition state will be further investigated by point mutation studies in Chapter 9.

References

- (1) Cox, J. P. L.; Evans, P. A.; Packman, L. C.; Williams, D. H.; Woolfson, D. N. *J. Mol. Biol.* **1993**, *234*, 483.
- (2) Jourdan, M.; Searle, M. S. *Biochemistry* **2000**, *39*, 12355.
- (3) Sosnick, T. R.; Dothager, R. S.; Krantz, B. A. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17377.
- (4) Went, H. M.; Jackson, S. E. *Protein Eng. Des. Sel.* **2005**, *18*, 229.
- (5) Harding, M. M.; Williams, D. H.; Woolfson, D. N. *Biochemistry* **1991**, *30*, 3120.
- (6) Stockman, B. J.; Euvrard, A.; Scahill, T. A. *J. Biomol. NMR* **1993**, *3*, 285.
- (7) Cordier, F.; Grzesiek, S. *Biochemistry* **2004**, *43*, 11295.
- (8) Qin, Z.; Ervin, J.; Larios, E.; Gruebele, M.; Kihara, H. *J. Phys. Chem. B* **2002**, *106*, 13040.
- (9) Larios, E.; Li, J. S.; Schulten, K.; Kihara, H.; Gruebele, M. *J. Mol. Biol.* **2004**, *340*, 115.
- (10) Khorasanizadeh, S.; Peters, I. D.; Roder, H. *Nat. Struct. Biol.* **1996**, *3*, 193.
- (11) Krantz, B. A.; Sosnick, T. R. *Biochemistry* **2000**, *39*, 11696.
- (12) Krantz, B. A.; Mayne, L.; Rumbley, J.; Englander, S. W.; Sosnick, T. R. *J. Mol. Biol.* **2002**, *324*, 359.
- (13) Went, H. M.; Benitez-Cardoza, C. G.; Jackson, S. E. *FEBS Letters* **2004**, *567*, 333.
- (14) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
- (15) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123.
- (16) Torii, H.; Tatsumi, T.; Tasumi, M. *J. Raman Spectros.* **1998**, *29*, 537.
- (17) Kramers, H. A. *Physica* **1940**, *7*, 284.
- (18) Crane, J. C.; Koepf, E. K.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2000**, *298*, 283.
- (19) Jager, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373.
- (20) Hardy, R. C.; Cottingham, R. L. *J. Chem. Phys.* **1949**, *17*, 509.
- (21) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76.

- (22) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193.
- (23) Yang, W. Y.; Gruebele, M. *Biophys. J.* **2004**, *87*, 596.
- (24) Ervin, J.; Gruebele, M. *J. Biol. Phys.* **2002**, *28*, 115.
- (25) Krantz, B. A.; Dothager, R. S.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *337*, 463.

Chapter 8

Calculation of free energy surface of ubiquitin unfolding

8.1. Introduction

In the description of the protein folding and unfolding dynamics on the free energy surface, one of the most difficult and crucial parts is to choose the best reaction coordinates onto which the huge number of degrees of freedom are projected. Good coordinates are easily determined when the folding process follows one of the folding paradigms: the nucleation-growth,¹ hydrophobic collapse,^{2,3} or framework model.^{4,5} For example, the order of secondary structural formations is the best coordinate for the framework model in which the formation of the local secondary structures precedes that of tertiary structure. In the hydrophobic collapse model, the native contacts between the hydrophobic side chains can be a good coordinate. However, most proteins undergo a mechanism between these two extremes, which can be described by a unifying mechanism, the nucleation condensation model.⁶ In this model, folding occurs via a transition state having both partial secondary- and tertiary-structural interactions. This implies the diversity of the reaction coordinates depends on the relative importance of these interactions and the stability of the motifs. Therefore, in some cases, the combination of several structural motifs will be the best coordinate. Also, the interactions

in one motif can provide multiple coordinates, as what we proposed in Chapter 6, for the β strands in ubiquitin

From the experimental point of view, the reaction coordinates are determined by what the probe is most sensitive to. For example, in Chapter 6, the free energy surface is constructed with the two regions of the β sheet. However, if folding occurs along orthogonal coordinates such as the α helix or the hydrophobic core, which the experiment is not sensitive to, the probed dynamics will be complicated. Therefore, more supporting information is needed either from other kinds of experiments or simulations.

Ideally, if we know the folding process in atomic detail, the best free energy surface can be obtained by projecting it onto a minimum number of essential coordinates. In this sense, the all-atom molecular dynamics (MD) simulation⁷ is one of the most informative theoretical tools. However, the free energy surface is the representation of an ensemble of states and cannot be formed by a single folding or unfolding trajectory. In spite of the amazing growth of computational power, it is still impossible to obtain a sufficient number of trajectories to map the free energy surface with MD simulations. A distributed-computing technique provides a fair number of trajectories⁸⁻¹⁰ but it is still limited to small proteins or peptides. To overcome this limit, simplified models have been used to simulate folding. These models include the lattice model,^{11,12} a minimalist model considering C_α and solvent molecules,¹³ and the Monte-Carlo simulation method with Go potential¹⁴ including interactions between all heavy atoms and backbone torsions.

Instead of simulating folding trajectories, over the last decade, simple methods have been developed to calculate the free energy of a protein based on the native three-dimensional structure.¹⁵⁻¹⁸ These models allow only two states for the amino acid residues, native and non-native similar to the Ising model for the spin system. The possible permutations of these folded and unfolded residues constitute the ensemble of microstates, and the stability of each configuration is determined by the interaction energy given by the native structure and the entropic cost to fix folded residues in the native conformation. The partition function that is obtained in this procedure provides all the thermodynamic quantities and possible kinetics. The simplicity of the model allows

us to fill the gap between the ensemble experiment and the atomically-detailed MD simulation without running a time-consuming simulation.

In this chapter, we generate the unfolding free energy surface of ubiquitin with this statistical mechanical model and describe the relaxation phenomena presented in Chapter 7 on this free energy surface.

8.2. Calculation methods

8.2.1. The Munõz-Eaton model

Our calculation of the free energy surface is based on a model originally proposed by Munõz and Eaton for small peptides and proteins.^{16,19} The Munõz-Eaton (ME) model is analogous to the Ising model for the spin system in statistical mechanics. As a $\frac{1}{2}$ spin in a directional magnetic field has only two states (up and down), the ME model allows only two states for each amino acid residue: folded (F) and unfolded (U). The permutations of F and U states for the residues generate the ensemble of the configurations (microstates) of the protein. The energy of a particular microstate can be calculated using the favorable native interactions between the folded residues and the unfavorable entropic cost that is needed to fix the conformation of a residue. Then, the partition function can be obtained by summing the weighting factors over the possible microstates. The free energy surface can be obtained by the projection of this partition function onto a proper coordinate such as number of native residues or number of native contacts.

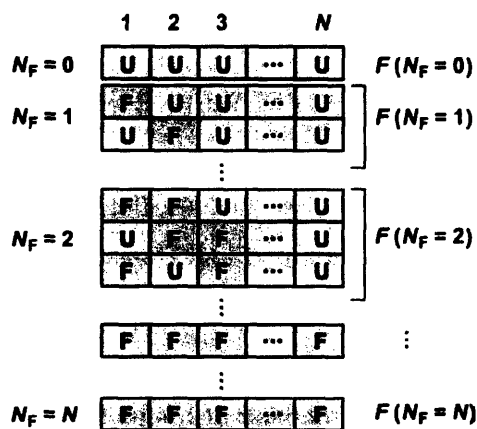


Figure 8-1. Possible states in the ME model. Each residue is allowed only to have a folded (F) or unfolded (U) state.

The calculation of the ME free energy is illustrated in Fig. 8-1. Illustrated is a case for a protein with N residues that can take on several possible combinations of F and U states as a function of the number of folded residues, N_F . There are N microstates for $N_F = 1$, $N(N-1)/2$ microstates for $N_F = 2$, and so on. Calculating a partition function for all possible states by a simple enumeration is not possible unless the number of residues is small. For the case of ubiquitin, the number of possible states is $2^{76} \approx 7.6 \times 10^{22}$. Therefore, simplifying approximations are needed to make the problem tractable. In their original work, Munõz and Eaton simplified the model by allowing the native interactions only between the residues that are located in the same block of consecutive folded residues (folded stretch) in the primary sequence of proteins and limiting the number of folded stretches. Depending on the number of folded stretches allowed, the approximation is called the single, double, and triple native stretch approximation (SSA, DSA, and TSA) as shown in Fig. 8-2. The free energy and the weighting factor for a microstate having a folded stretch from residue i to $i+j-1$ is given as¹⁵

$$F_{ji} = \sum_{\text{contacts}} \varepsilon - T \sum_{k=i}^{i+j-1} \Delta s_k \quad (8.1)$$

$$w_{ji} = \exp(-F_{ji}/RT)$$

Here, j is the index for the number of folded residues in a folded stretch. The first term of the free energy is the summation of the interaction energy for all possible native contacts occurring in the folded stretch of the microstate. The interaction energy ε is added for each of the native contacts between two residues in the folded stretch. Following one of the criteria in the Ref. 15, two residues are assumed to be in a native contact when the distance between C_α atoms of those residues is less than 7 Å. The distance is calculated from the atomic coordinates of the crystal structure. (PDB id: 1UBQ)²⁰ Also, the contacts between the residue k and $k+1$ or k and $k+2$ are excluded because of their natural proximity. The second term, Δs_k in Eq. (8.1) is the entropic cost by fixing the conformation of residue k in its folded state. For the simplicity, the same value of ε is used for all native contacts and the same value of Δs is used for all residues.

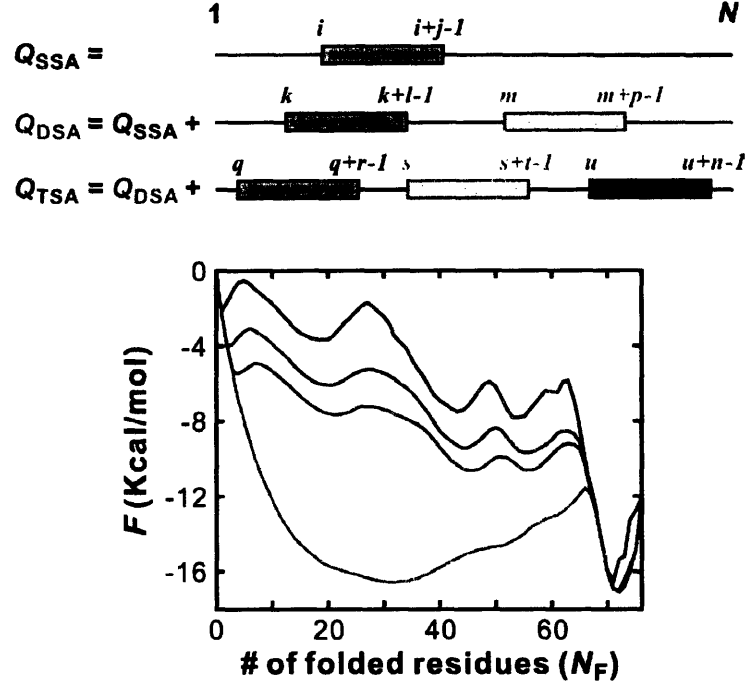


Figure 8-2. 1D free energy curve of the ME model. Free energy curves of ubiquitin at 58°C is calculated as a function of the number of folded residues by SSA (blue), DSA (green), and TSA (red) and the exact model (purple) with parameters of $\epsilon = -0.477$ Kcal/mol and $\Delta s = -1.78$ cal/mol·K.

In our experimental conditions, 13 positive charges are present in ubiquitin because all amine groups are charged in acidic condition. It has been reported that the stability of the protein depends on the surface charges and anions.^{21,22} Therefore, we incorporate the charge repulsion term in addition to Eq. (8.1),

$$F_{ji}' = \sum_{\text{contacts}} \epsilon + \sum_{\text{charged}} \frac{e^2}{4\pi\epsilon_r\epsilon_0} \left(\frac{1}{r} - \frac{1}{l_c} \right) - T \sum_{k=i}^{i+j-1} \Delta s_k \quad (8.2)$$

$$w_{ji}' = \exp\left(-F_{ji}'/RT\right)$$

In the second term of Eq. (8.2), r is the distance between the two charged atoms and l_c is the contour length connecting the two charged residues n_1 and n_2 , given by $l_c = (3.6 \text{ \AA}) \times (n_2 - n_1)$. l_c is the upper limit of the distance between two charged residues, which gives an offset of the repulsion energy. When $r > l_c$, the charged term is not included, which is possible because r is determined by the location of the charged atom while l_c is determined by the location of the C_α . The spatial coordinates of the charged atoms are

also obtained from the crystal structure. e and ϵ_0 is the electric charge and the vacuum permittivity, respectively. The relative permittivity ϵ_r is fixed to 80. The empirical parameters ϵ and Δs are determined by comparing the calculated temperature-dependent equilibrium constants with the experimental values as shown in Section 8.3.1.

Using the weighting factor w_{ji} obtained above, partition functions can be expressed as

$$\begin{aligned} Q_{\text{SSA}} &= 1 + \sum_{j=1}^N \sum_{i=1}^{N-j+1} w_{ji} \\ Q_{\text{DSA}} &= Q_{\text{SSA}} + \sum_{j=1}^N \sum_{l=1}^{j-1} \sum_{k=1}^{N-l-1} w_{lk} \sum_{m=l+k+1}^{N-p+1} w_{pm} \\ Q_{\text{TSA}} &= Q_{\text{DSA}} + \sum_{j=1}^N \sum_{r=1}^{j-2} \sum_{q=1}^{N-r-3} w_{rq} \sum_{t=1}^{j-r-1} \sum_{s=1}^{N-t-1} w_{ts} \sum_{u=r+q+s+1}^{N-n+1} w_{nu} \end{aligned} \quad , \quad (8.3)$$

where $p \equiv j - l$, $n \equiv j - r - t$, and N is the number of residues in the protein. The indices for native stretches are as those shown in Fig. 8-2. DSA includes SSA, and TSA includes both of SSA and DSA.

The one-dimensional free energy curve can be easily obtained from the partition function in Eq. (8.3) by replacing j with the number of folded residues N_F as a coordinate,

$$\begin{aligned} F_{N_F} &= -RT \ln \left(\sum_{i=1}^{N-N_F+1} w_{N_F i} + \sum_{l=1}^{N_F-1} \sum_{k=1}^{N-l-1} w_{lk} \sum_{m=l+k+1}^{N-p+1} w_{pm} \right. \\ &\quad \left. + \sum_{r=1}^{N_F-2} \sum_{q=1}^{N-r-3} w_{rq} \sum_{t=1}^{N_F-r-1} \sum_{s=1}^{N-t-1} w_{ts} \sum_{u=r+q+s+1}^{N-n+1} w_{nu} \right) \quad , \quad (8.4) \end{aligned}$$

The first summation represents SSA and the addition of the second and third summation accounts for DSA and TSA.

Alternatively, the exact solution of the ME model has been obtained using the transfer matrix formalism by Bruscolini and Pelizzola.¹⁷ The method generates the partition function by multiplying a series of rectangular transfer matrices, the size of which increases from 1×2 to $(N+1) \times (N+2)$. Simply saying, $(j+1) \times (j+2)$ matrix transfers a row vector ($1 \times (j+1)$) consisting of a sum of weighting factors for all possible folded and unfolded combinations of residues ranging from 1 to j to the next row vector

of $1 \times (j+2)$ by associating the status of residue $j+1$. In this exact solution, there is no limit on the number of folded stretches though it still limits the native interactions to those within the same native stretch. To track the number of folded residues, a dummy variable (λ) is multiplied to the weighting factor. Then, the partition function is given with a polynomial of λ , the power of which represents the number of folded residues. Using the coefficient of the polynomial, the one-dimensional free energy curve can be constructed as

$$Q_{\text{Exact}} = 1 + \sum_{N_F=1}^N w_{N_F} \lambda^{N_F}, \quad (8.5)$$

$$F_{N_F} = -RT \ln(w_{N_F}).$$

As an example, a procedure for a 4-residue protein is shown in Appendix 8A. The complete theory can be found in Ref. 17.

The projected one-dimensional free energy curves are compared in Fig. 8-2. Firstly, differences in the folded well are very small. The results of DSA and TSA overlap the exact calculation and even the SSA is very similar to the others. On the other hand, substantial differences between the models are observed on the unfolded side. These features can be rationalized by the different stabilizing factors for the two sides. As the number of folded residues approaches its maximum N , the energetic term becomes more important because the number of native contacts increases. In the models above, however, the native interactions are allowed only for those in the same folded stretch and most of the stabilized microstates originate from SSA. Therefore, all the free energy curves of the folded side are similar in Fig. 8-2. On the other hand, in the unfolded side the number of native contacts is small. Therefore, this region is dominated by the diversity (entropy) of the ensemble. Obviously, the number of microstates becomes larger with the number of native stretches allowed and the exact calculation shows the most stable unfolded state.

We also construct a two-dimensional free energy surface to compare it with our experimentally proposed free energy surface in Chapter 6. The partition function is projected onto the two coordinates, the number of folded residues in the strand I and II ($N_F(\text{I-II})$) and in the strand III to V ($N_F(\text{III-V})$), which are determined by the residues 1 –

17 (I-II) and the residues 42 – 49 and 66 – 71 (III-V), respectively. In the approximated calculation, the projection is performed by adding weighting factors that have the same number of folded residues in these two coordinates during the enumeration in Eq (8.3). The weighting factor and the free energy for $(N_F(\text{I-II}), N_F(\text{III-V})) = (i, j)$ are expressed as

$$\begin{aligned} w(i, j) &= \sum_{\{k\} \in \{(i, j)\}} w_{\{k\}} \\ F(i, j) &= -RT \ln(w(i, j)) \end{aligned} \quad (8.6)$$

Here, $w_{\{k\}}$ is the weighting factor for the microstates of the folded index set $\{k\}$, in which the number of folded residues in strand I-II and III-V are i and j , respectively.

In the exact calculation, two dummy variables are introduced for the residues of the two coordinates.

$$\begin{aligned} Q_{\text{Exact}}(i, j) &= 1 + \sum_{i, j} w_{\text{Exact}}(i, j) \lambda^i \nu^j, \\ F_{\text{Exact}}(i, j) &= -RT \ln(w_{\text{Exact}}(i, j)). \end{aligned} \quad (8.7)$$

The reaction coordinates for the proposed free energy surface are not the number of residues but the number of native contacts between the β strands. Nevertheless, the free energy surface should be similar because native contacts occur in the model only when two residues are in the folded state. Fig. 8-3 shows the 2D free energy surfaces that are constructed by the ME model with and without approximations. As in the case of 1D curves, the native state is similar and located in the upper right corner where the number of folded residues is high in both coordinates. However, the differences in the unfolded regions of the approximate and exact calculations are clearer than those of the 1D free energy curves. The unfolded states in the three approximated calculations are located in the lower left corner where the number of folded residues is close to 0, which is unphysical because even in the denatured state there will be some contacts between the residues. In addition, the presence of a big barrier at $N_F(\text{III-V}) = 7$ for all $N_F(\text{I-II})$ indicates the insufficient sampling of the unfolded state in this region. In the exact calculation, the unfolded well is located in this region instead of a barrier.

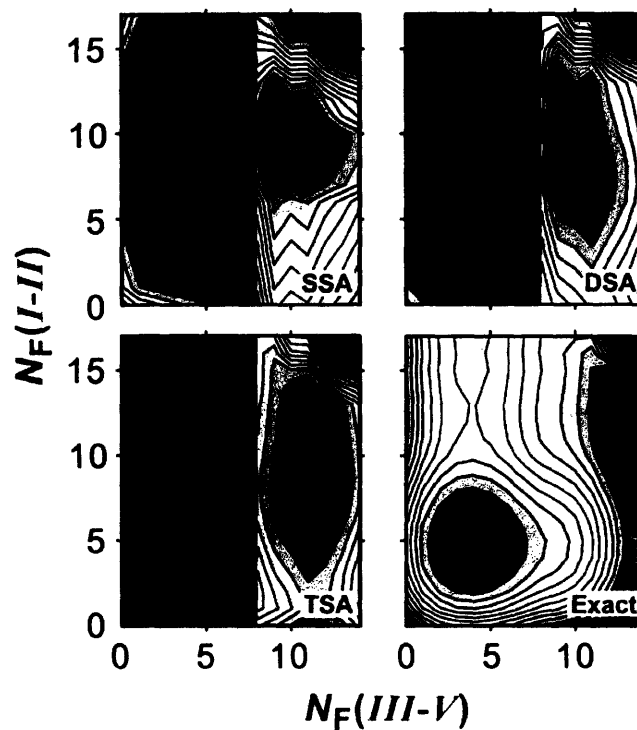


Figure 8-3. Comparison of the approximated and exact calculation by 2-D free energy surfaces at 58°C.

8.2.2. Beyond the ME model: Consideration of partially folded states

As described in the previous section, the unfolded side is well described only by the exact calculation. However, for the folded side, there is a limitation in the model itself. The ME model does not allow native interactions between the residues of different native stretches. Since it is based on the linear amino acid sequence, for the native interactions between the residues far from each other in the primary sequence, all intervening residues must be folded. As a result, the probability for a pair of residues to be folded decreases as a function of the separation in the sequence. This limitation is not a big problem for the small peptides such as a single α helix or β hairpin, the folding of which can be described by a zipper model. However, in the presence of the tertiary contacts that exist in most globular proteins, the three-dimensional proximity becomes important as well and should be given equal consideration as the primary sequence proximity. For example, in Fig. 8-4, the strands I and V of ubiquitin are located at the two ends of the linear sequence but they are one of the closest strands in the 3D structure of

the protein. In this case, the stability of strand I and V are underestimated in the calculation of the previous section especially for the folded state.

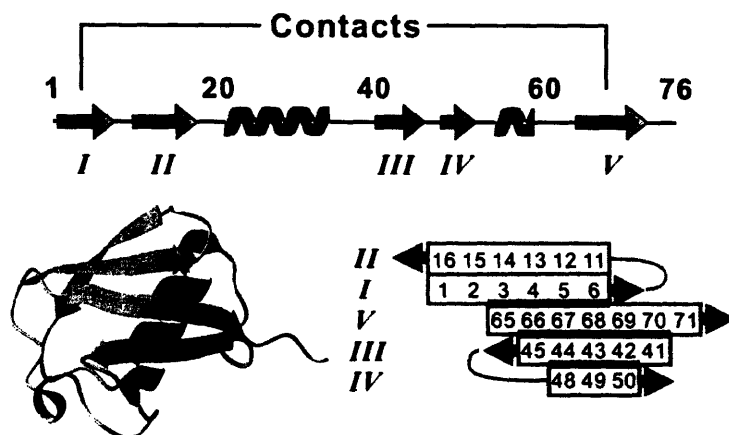


Figure 8-4. Linear sequence, structure and the β -sheet registry of ubiquitin.

To overcome the limitation of this primary-sequence-based model, a modified DSA calculation has been used.¹⁶ As shown in Fig. 8-5a, two possible states are considered for a sequence having two folded stretches. As in normal DSA, the state shown on the left side has no native contacts between different folded stretches as normal DSA. The state on the right side is allowed to have native interactions between these two native stretches separated by a disordered loop. The latter state can be regarded as the partially folded state in which only small portions of the protein are unfolded and the structure is mostly intact. In this case, in addition to the interaction energy between residues in different folded stretches, an additional entropic cost is needed to fix the two ends of the disordered loop connecting the folded stretches. The partition function is the sum of these two cases and is given by¹⁶

$$Q_{\text{DSA+L}} = Q_{\text{DSA}} + Q_{\text{L}}$$

$$Q_{\text{L}} = \sum_{j=1}^N \sum_{l=1}^{j-1} \sum_{k=1}^{N-l-1} \sum_{m=l+k+1}^{N-p+1} w_{lk} w_{pm} v_{lk;pm}, \quad (8.8)$$

where $p \equiv j - l$. $v_{lk;pm}$ is the weighting factor determined by the native interactions between the two different folded stretches and the loop formation entropy as

$$\begin{aligned}
v_{lk;pm} &= \exp\left(-\Delta G_{lk;pm}/RT\right) \\
\Delta G_{lk;pm} &= \sum_{\substack{\text{contacts} \\ lk \leftrightarrow pm}} \varepsilon - T\Delta s_{\text{loop}}.
\end{aligned} \tag{8.9}$$

Indices are the same as those shown in Fig. 8-2. The loop entropy for the Gaussian chain is given as¹⁶

$$\Delta s_{\text{loop}} = R \ln \left(\frac{2\pi a^3/3}{\left(2\pi \langle r^2 \rangle / 3\right)^{3/2}} \right), \tag{8.10}$$

where a is the distance between the location of C_α of the two amino acid residues at the end of the folded stretches that are connected by the disordered loop. $\langle r^2 \rangle$ is proportional to the contour length of the loop as $\langle r^2 \rangle = C_n n l^2$, where $n+1$ is the number of residues in the loop and $l = 3.6 \text{ \AA}$. The Flory characteristic ratio C_n is fixed to 6.0 in our calculation while it is one of the empirical parameters that is determined by fitting in Ref. 16. The free energy surface constructed in this manner is shown in Fig. 8-5c, and it is not so different from that of Fig. 8-3 except for the extended folded region, which is attributed to the fact that the DSA+L is still not sufficient to map the unfolded state.

As the conceptual opposite to the primary-sequence-based model, we constructed a free energy surface based on tertiary contact by introducing a three-dimensional (3D) growth model that is similar to the nucleation growth folding paradigms, the result of which is shown in Fig. 8-5b. In this case, more partially folded configurations can be counted by allowing native interactions between native stretches of more than two. The partition function Q_{3D} of this 3D growth model is constructed in the following way. The folded residues grow from a folding nucleus as shown in Fig. 8-5b. Practically, a series of folding spheres are generated centered at a particular residue acting as a nucleus, which is always folded. For a given sphere of number s , the microstate are generated by permuting the status of each of the residues between the sphere $s-1$ and s . The residues inside the sphere $s-1$ are fixed to the folded state, and the residues outside the sphere s are fixed to the unfolded state. The free energy of each microstate is calculated by including all native

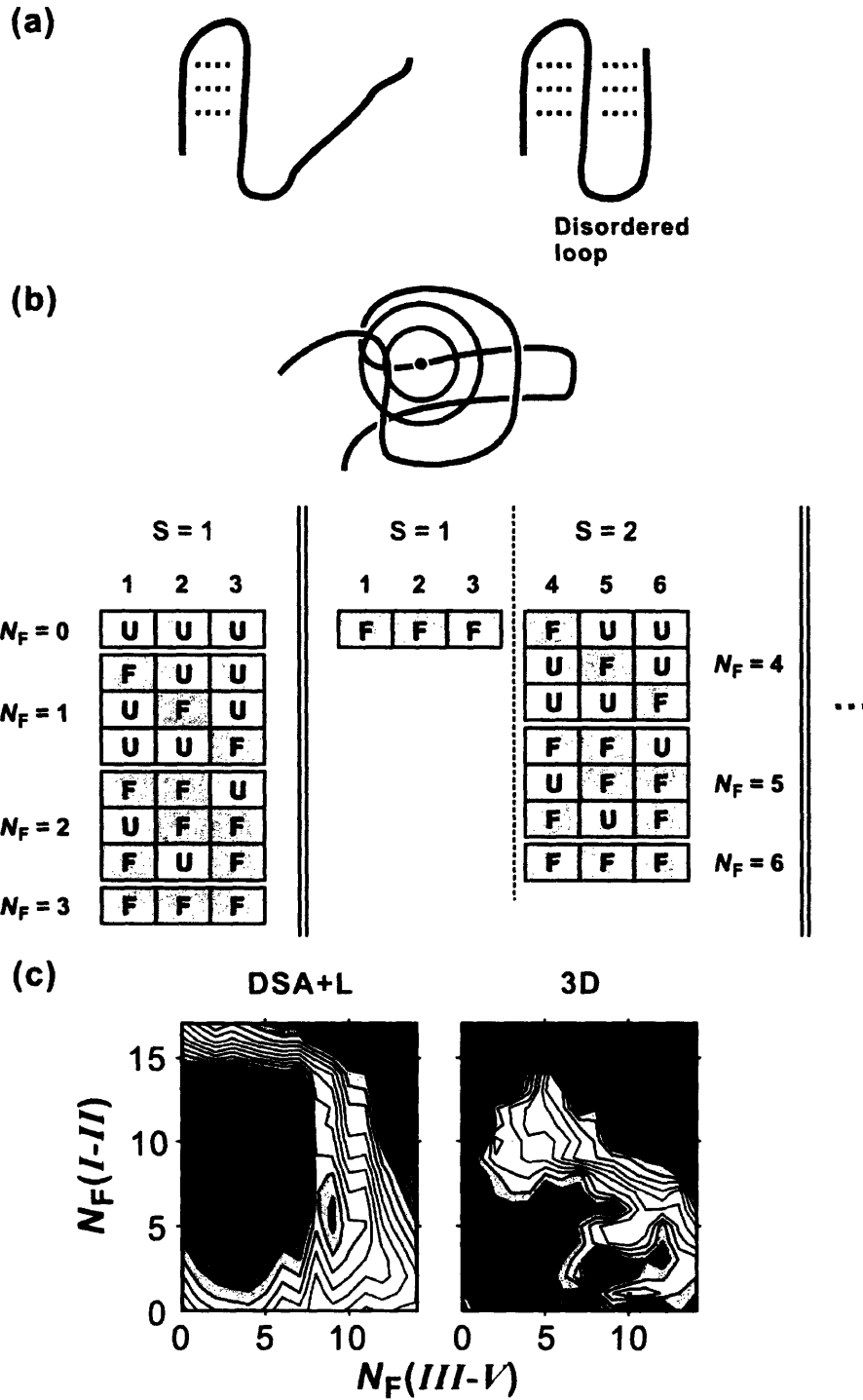


Figure 8-5. Addition of native interactions between different native stretches. (a) Double stretches connected by a disordered loop. Red and gray lines are the folded and unfolded stretch and green dashed lines represent the native contacts between the folded residues. (b) 3D growth model. (c) 2D free energy surface constructed by DSA+L and 3D growth model.

interactions between folded residues and the loop entropy given as Eq. (8.10) for all disordered loops that connect native stretches. The first sphere is given by the initial radius of r_0 . After calculating all the microstates in the first sphere, next spheres are generated one after another by increasing the radius by dr until the sphere reaches the outmost residue. Finally, the whole procedure is repeated with a new initial folding nucleus from residue 1 to N . We used the parameters of $r_0 = 6 \text{ \AA}$ and $dr = 0.75 \text{ \AA}$. The shape in the folded well depends slightly on these parameters. However, the free energy surface in Fig. 8-5c shows the same problems as DSA+L because too few configurations are counted in the unfolded side. Also similarly, the folded side is extended by the addition of partially folded species. The best counting of tertiary contacts can be achieved by finding an exact solution to calculate the interactions between all possible combinations of native stretches connected by disordered loops without a limit on the number of stretches. However, such a solution does not exist yet.

Instead, we can take advantage of the physically appealing features from each model. The unfolded side is described well by the exact ME calculation. The folded side is also relatively well described by the DSA+L and the 3D growth model. Therefore, we introduce a hybrid partition function

$$\begin{aligned} Q_{\text{EX+L}} &= Q_{\text{Exact}} + Q_{\text{L}} \\ Q_{\text{EX+L+3D}} &= Q_{\text{Exact}} + Q_{\text{L}} + Q_{\text{3D}}' \end{aligned} \quad (8.11)$$

Here, the modified partition function of the 3D growth model Q_{3D}' contains the microstates having more than one disordered loop to avoid double counting of the same microstate in Q_{Exact} or Q_{L} .

The results of the hybrid model are summarized in Fig. 8-6. As we expected, the addition of interaction energy between different native stretches connected by a disordered loop does not change the unfolded well located in the lower left side because the unfolded states are governed by entropy rather than interaction energy. In other words, the unfolded state is the bigger ensemble of the energetically unfavorable conformations while folded states are smaller ensemble of the energetically favorable conformations. In a similar manner, our addition of microstates having more disordered loops based on the 3D growth model, which is also biased to the folded side, is valid. In

all three models, the folded well is located on the right side but it becomes wider and the barrier position moves towards the unfolded side as a result of adding the partially folded states. The same phenomena are observed in the 2D free energy surfaces. The addition of the partially folded states widens the folded well.

The Matlab code for these calculations can be found in Appendix 8B.

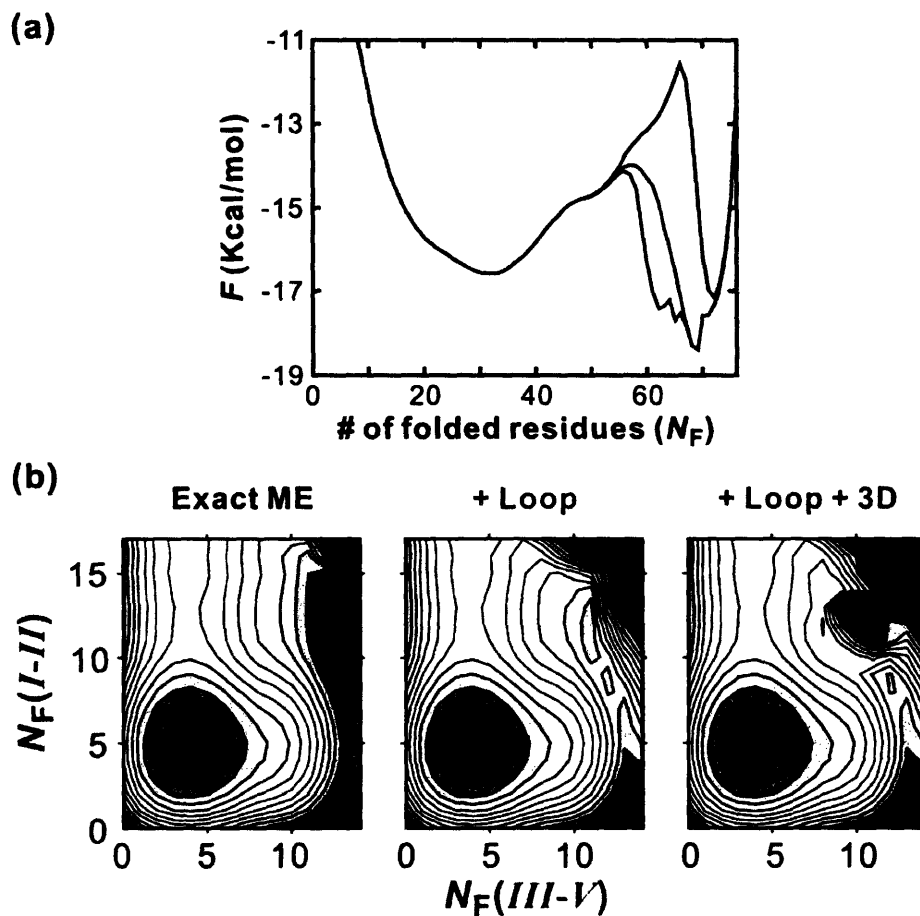


Figure 8-6. Comparison of the hybrid model. (a) 1D free energy curve for the exact ME model (blue), the addition of the DSA with disordered loop (green), and the further addition of the 3D model (red). (b) 2D free energy surfaces calculated by each of the models.

8.3. Results and discussion

8.3.1. Determination of the empirical parameters

In the above calculations, there are several empirical parameters that are determined by comparison to the experimental data such as the temperature-dependent

equilibrium constant (K_{eq}), the folding rate (k_f), and the unfolding rate (k_u). Using K_{eq} , we can obtain the interaction energy ε and the entropy to fix the conformation of a single residue Δs . The barrier height and position is sensitive to the loop entropy. However, since the barrier height is coupled to the pre-exponential factor of the Arrhenius equation, we fix the Flory characteristic ratio C_n to 6.0 for simplicity and describe the kinetic behavior qualitatively.

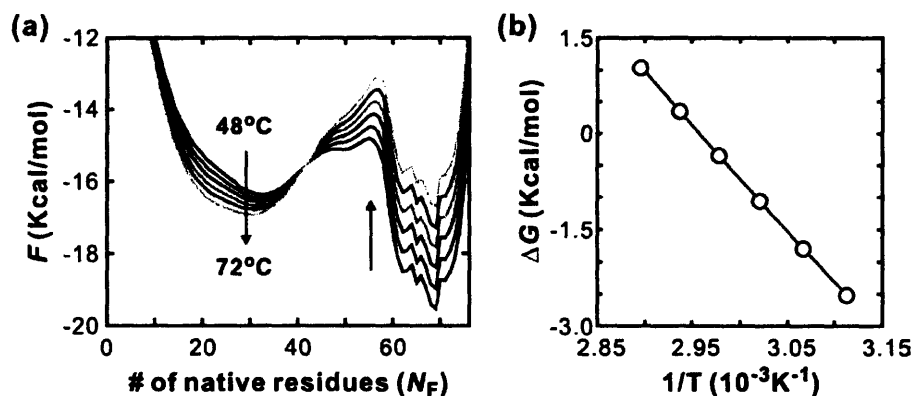


Figure 8-7. Determination of empirical parameters by comparing with experimental data. (a) Temperature dependent 1D free energy curves. (b) Comparison of the free energy of unfolding from the calculation (open circle) and experiment (line), over the temperature range 48°C – 82°C.

The determination of ε and Δs are shown in Fig. 8-7. In the 1D free energy curve, the left and right sides relative to the barrier at $N_F = 55 - 57$ are assumed to be the unfolded and folded states, respectively. From the ratio of the integrated populations, the free energy of unfolding is calculated at each temperature as

$$\Delta G = -RT \ln \left(\frac{\sum_{N_F < N_{Bar}} w_{N_F}}{\sum_{N_F > N_{Bar}} w_{N_F}} \right), \quad (8.12)$$

$$w_{N_F} = \exp(-F_{N_F}/RT).$$

where N_{Bar} is the position of the barrier. The calculated free energy change is linear with the inverse temperature as shown in Fig. 8-7b. However, the experimental free energy change that is obtained by Eqs. (7.1) and (7.2) is not linear over the wide range of accessible temperatures. Therefore, we take the linear portion (48°C – 82°C) of the

experimental result and fit it to the equation $\Delta G(T) = \Delta H - T\Delta S$ without the heat capacity term ΔC_p , which results in $\Delta H = 50.1$ kcal/mol and $\Delta S = 148.2$ kcal/mol·K. The resulting fitted line is shown in Fig. 8-7b. Comparison of the calculated values from Eq. (8.12) with these experimentally obtained free energy changes gives the empirical parameters, $\varepsilon = -0.477$ kcal/mol and $\Delta s = -1.78$ cal/mol·K. These values are used for all calculations in this chapter for the comparison between models.

8.3.2. Unfolding pathways

Calculation of the free energy can also be used to provide evidence for the unfolding pathways observed in our experiments. This is done by plotting the folding probability of each residue as shown in Fig. 8-8. For a given location $(N_F(\text{I-II}), N_F(\text{I-II})) = (i, j)$ on the free energy surface, the folding probability of a specific residue m is obtained by summing the weighting factors of all microstates in which the residue is folded,

$$p_m^F(i, j) = \frac{\sum_{\{k\} \in \{(i, j)\}} w_{\{k\}} p_{\{k\}, m}}{Q_{\text{EX+L+3D}}(i, j)} \quad (8.13)$$

$$Q_{\text{EX+L+3D}}(i, j) = \sum_{\{k\} \in \{(i, j)\}} w_{\{k\}}$$

Here, the summation is performed over the states $(\{k\})$, in which number of folded residues in strand I-II and III-V are i and j , respectively. $p_{\{k\}, m}$ takes 0 (unfolded) or 1 (folded) according to the status of the residue m in the state $\{k\}$. These probabilities are mapped by color-coding on the projection of the β sheet shown in Fig. 8-4. The probabilities are mapped for eight points on the free energy grid: four points in the folded well (F1 – F4), two points in the transition state (T1 and T2), and two points in the unfolded region (U1 and U2).

In the folded well, F1 is the fully folded state as the folding probabilities of all residues in the β sheet are unity. The other three states are partially folded states. F2 is the state where the turn region of the β hairpin of strands III and IV is unfolded while other regions are almost intact. F3 shows that the turn region of the β hairpin of strands I and II is unfolded. Another partially folded state F4 shows that both turn regions are unfolded as expected by the location of the well. This is consistent with the fact that the

turn region is usually more flexible than the middle part of the β sheet. Also, the turn region of ubiquitin is exposed to the solvent.

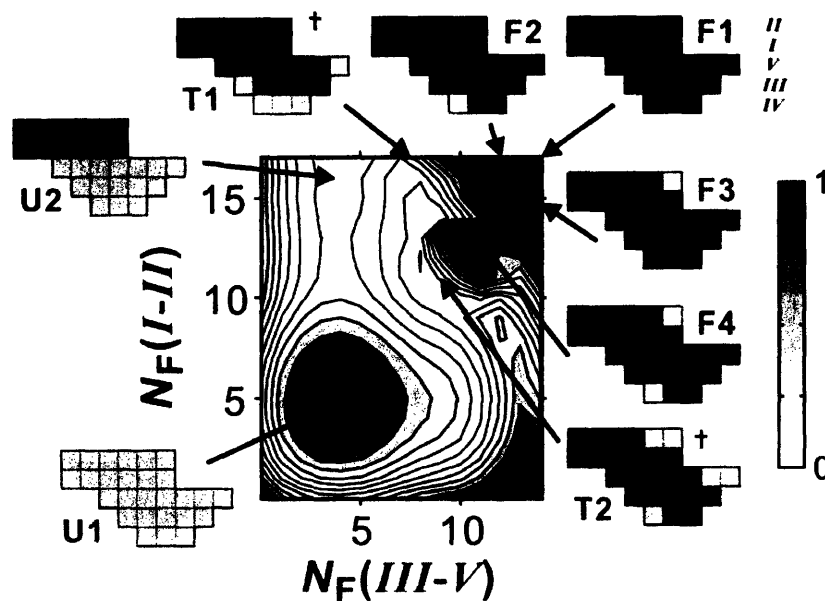


Figure 8-8. Folding probability of the residues in the β sheet at several locations on the free energy surface. (58°C)

T1 is thought to be one of the transition states because it sits atop the lowest barrier height on the surface. In this state, strands I and II are highly ordered while strand IV and the turn region of strands III and IV are mostly unfolded. Strands IV and V are also partially unfolded as indicated by the lower folding probabilities. The unfolding pathway through T1 is followed by the state U2 where the strands III-V are unfolded and strands I and II are still intact. Finally, it reaches the fully unfolded state U1. The folding probability is close to 0.5 for all residues in U1, which implies that the unfolded state is an ensemble of random combinations of the folded and unfolded residues. Another possible unfolding pathway is through the state T2, which has a lower probability due to a higher barrier compared to T1 at 58°C. From the proximity of the location, one can expect that unfolding along this pathway starts from the partially folded state F4 where both of the turn regions are unfolded. As expected, the transition state T2 resembles F4, which shows that both of the turn regions are less stable than other regions. In addition, the folding probability of the whole β sheet is lowered and the end of strand V, which is connected to the extremely flexible C-terminus of the protein, is almost unfolded. At a

first look, the unfolding through T2 seems like a concerted process in which the stability of each residue decreases together. However, the folded probability of each residue is the averaged value for the whole ensemble of microstates, each of which has well defined folded and unfolded regions at different locations of the β sheet. Therefore, unfolding through T2 is a heterogeneous process. On the other hand, the process through T1 is sequential, as is observed in the previous experiments^{23,24} and simulation.²⁵ It will be shown that unfolding through T2 becomes the more probable pathway at high temperature.

8.3.3. Free energy changes by T-jump

The changes of the 2D free energy surface before and after the T-jump are compared in Fig. 8-9a. The shape of the free energy surface in the folded and unfolded region is similar, but the position and the height of the barrier are changed as a function of the final temperature, which may result in the temperature-dependent unfolding dynamics observed in Chapter 7. If we focus on the T-jump from $T_i = 58^\circ\text{C}$ to $T_f = 70^\circ\text{C}$, the lowest temperature pair, the barrier is located at $(N_F(\text{I-II}), N_F(\text{III-V})) = (17, 7)$ at $T_i = 58^\circ\text{C}$, which is the T1 state in Fig. 8-8 where the stable hairpin region (strands I and II) is intact and the other strands are partially unfolded. After the T-jump, the free energy surface at $T_f = 70^\circ\text{C}$ shows that the barrier is shifted to the folded side, which is consistent with our prediction in Chapter 6 and 7. The population changes during the T-jump ($\Delta P = P(T_i) - P(T_f)$) are also plotted with a log scale in Fig. 8-9b. The excess population in the right upper region relaxes into the lower left region. The trapped population in our interpretation may be located in the region around the barrier although the population is smaller than what we observed in the experiment in chapter 7. The changes in free energy and population by a T-jump from 63°C to 74°C are similar to those by a T-jump from 58°C to 70°C and are also presented in Fig. 8-9.

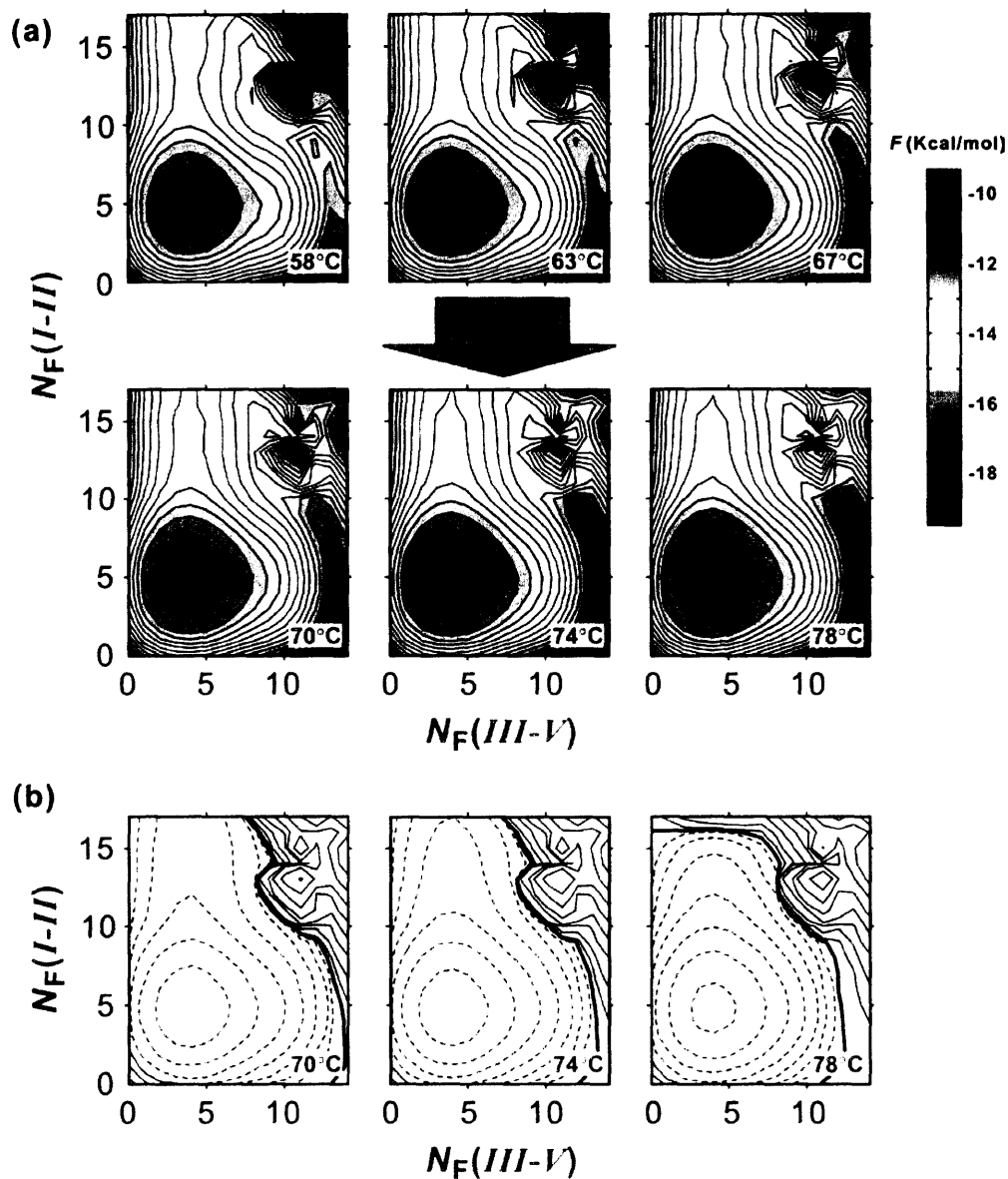


Figure 8-9. Changes in the free energy surface and populations. (a) The changes of the free energy surfaces before and after the T-jump from three different initial temperatures. (b) The population differences ($\Delta P = P(T_i) - P(T_f)$) before and after the T-jump are shown. Solid and dashed contours represent the positive and negative changes, respectively. Contours are plotted in log scale with an increment of $10^{-0.5(11-n)}$ for the n th contour from $\Delta P = 0$ indicated with a bold solid contour.

However, in the T-jump from 67°C to 78°C, the highest temperature pair, the changes are qualitatively different. Before the T-jump, the barrier is still located where the strands I and II are intact. (T1 state) On the other hand, after the T-jump, the barrier heights of the T1 and T2 states become similar. In other words, unfolding occurs via

multiple transition states. Considering this change, the non-exponential relaxation observed in the T-jump for this temperature interval can be interpreted. The μ s and ms relaxations are well separated in the T-jump to the two low final temperatures 70°C and 74°C while the two unfolding phases are merged into a single non-exponential relaxation at $T_f = 78^\circ\text{C}$. As shown in the free energy surface at 78°C, the shift of the barrier from T1 to T2 can be interpreted as a shift in the unfolding dynamics from the sequential process to a balance with the more heterogeneous process. Therefore, this increased unfolding heterogeneity results in a global non-exponential relaxation. Also, the timescale of the μ s part should be longer than that in the sequential unfolding because more residues are involved in unfolding through T2. The population changes in this T-jump range are also different from those of other T-jumps. The region of the excess population is more extended to fill both the T1 and T2 regions, which results in an increased amplitude of the μ s component and non-exponential relaxation in the experiment.

The unfolding through these new routes seems contradictory to the previous experimental results that support the sequential unfolding. However, these results are obtained from the experiments at room temperature while what we observed is a temperature dependent change of the unfolding route. In our interpretation, the unfolding also occurs through a sequential path at low temperature. For the comparison of unfolding at high temperature, we utilize the result of the unfolding molecular dynamics (MD) simulation performed at high temperature by the Daggett group.²⁶ The MD simulation was performed at 498 K to generate unfolded conformations for the refolding simulation. We traced the two unfolding MD trajectories onto our free energy surface as shown in Fig. 8-10. Trajectories 1 and 2 correspond to the unfolding trajectories D1 and D2 in Ref. 26. The state of each residue is determined by the ϕ and ψ angles. A residue is assumed folded when the deviations of both of the angles from those of the crystal structure are less than 50°. The two starting structures correspond to the red dot on each of the free energy surfaces. The initial structure of trajectory 1 is almost identical to the crystal structure (1UBQ) while that of trajectory 2 is slightly unfolded. As expected, both of the trajectories show that unfolding occurs through the transition state T2 though the protein in trajectory 2 stays in the region, where strand I and II are unfolded, for a fair amount of time. Although the temperature of MD simulation cannot be directly

connected to the real temperature, 498 K is high enough to support the alternative unfolding route at high temperature.

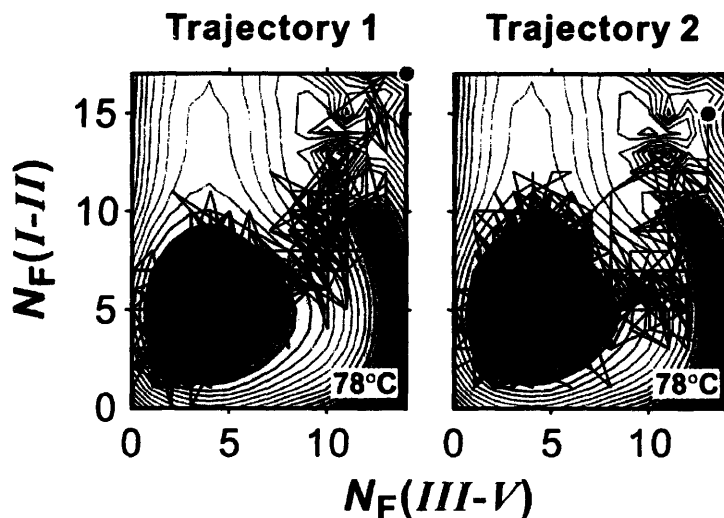


Figure 8-10. Projection of the unfolding trajectories of MD simulations onto the free energy surface.

8.3.4. Orthogonal coordinates

Our experimental probe is not yet sensitive to the unfolding of the α helix and the disruption of the hydrophobic core. Therefore, our conclusion may miss some of the important folding coordinates, and the projection of the populations onto non-ideal coordinates may result in the μ s relaxation and the non-exponential relaxation at high temperature. For example, a missed coordinate can be the number of folded residues in the α helix. The investigation on this orthogonal coordinate should be followed with an experiment sensitive to the changes in the α helix. Recently, unfolding of the α helix has been successfully monitored by the changes in amide I-II cross peaks due to the hydrogen-deuterium exchanges.²⁷ In this section, we will check the possibility of unfolding through the α helix with calculations.

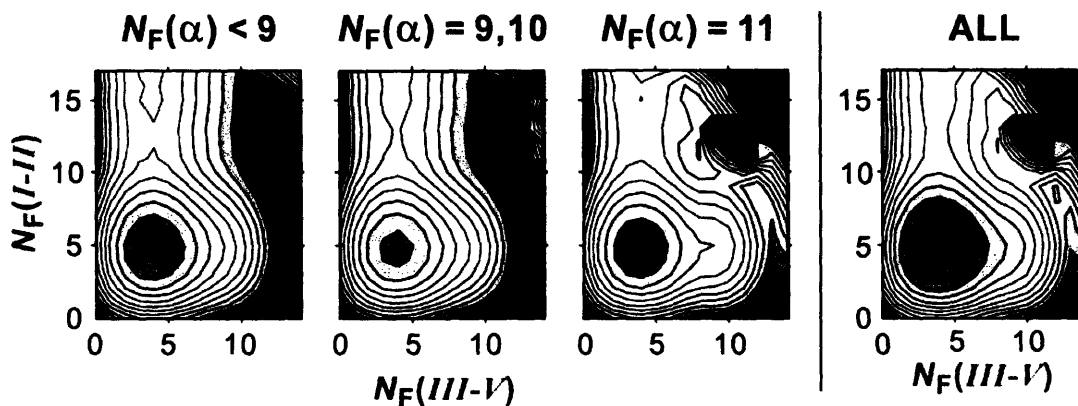


Figure 8-11. Free energy surfaces projected onto the number of folded residues in the α helix. (58°C)

As an extension of a 2D free energy surface, a 3D free energy surface can be constructed, the third coordinate of which is the number of folded residues in the α helix.

$$w(N_F(\text{I-II}), N_F(\text{III-V}), N_F(\alpha)) = \sum_{\{k \in \{N_F(\text{I-II}), N_F(\text{III-V}), N_F(\alpha)\}\}} w_{\{k\}} \quad (8.14)$$

$$F(N_F(\text{I-II}), N_F(\text{III-V}), N_F(\alpha)) = -RT \ln(w(N_F(\text{I-II}), N_F(\text{III-V}), N_F(\alpha)))$$

There are 11 residues in the α helix of ubiquitin. (residue 23 – 33) Instead of showing all twelve 2D surfaces as a function of the number of folded residues $N_F(\alpha)$, the free energy surfaces are projected for the three regions of $N_F(\alpha)$: $N_F(\alpha) < 9$, $N_F(\alpha) = 9, 10$, and $N_F(\alpha) = 11$ and shown in Fig. 8-11. Surprisingly, the folded state including the partially folded states, F2, F3, and F4 are not well-formed until all the residues in the α helix are folded. ($N_F(\alpha) = 11$) This implies that 3D contacts between the α helix and β sheet are crucial to the stability of the β sheet. This is consistent with the previous experimental evidence that the α helix and strands I and II form the folding core.²⁸⁻³² The partially folded state in the surface where the α helix is mostly folded ($N_F(\alpha) = 9, 10$) is the F3 state where the turn region between strands I and II is unfolded. The 3D structure of the protein in Fig. 8-4 shows that the C-terminus of the α helix is close to the β turn between strands I and II. Therefore, the free energy surface shows the slight unfolding of this region. Nevertheless, in conclusion, the α helix is extremely stable, and its unfolding is not at all involved in the unfolding transition state.

8.4. Limitations of the ME model and its improvement

As mentioned earlier, one of the limitations of the ME model is its assumption that native interactions occur only between residues in the same folded stretches. This is partially solved by introducing DSA with a disordered loop and the 3D growth model. However, an exact model with these physical features is still needed, if analytically tractable. In addition, there is another serious limitation in the model: non-native interactions are totally excluded. During folding or unfolding, one may find non-native states, which are stabilized by hydrogen bonding or hydrophobic interactions. The presence of these states makes the free energy surface rugged by forming additional wells on- or off-pathway of folding and unfolding. The wandering of the second MD trajectory in the high-energy region shown in Fig. 8-10 could result from these non-native interactions. This frustration of the free energy surface would increase with the protein size, and the model would not be valid for large proteins. Ignoring non-native interactions also misses some important features of the unfolded state. In the calculations shown in this chapter, the unfolded state is an ensemble of random combinations of the folded and unfolded residues. However, non-native interactions will stabilize some residual structures in the unfolded state. The presence of residual structures is observed in thermal unfolding of RNase A and ubiquitin by the existence of the cross peaks between ν_{\perp} and ν_{\parallel} transitions at high temperature as shown in Chapter 5 and 6. The characterization of the unfolded state is important because these residual structures may act as a folding nucleus in a real systems.

In spite of these limitations and its simplicity, the qualitative description given by the ME model for the non-exponential relaxation that is observed in the experiment at high temperature and the favorable comparison with the MD simulation demonstrate its utility. Although the quantitative prediction is still lacking several quantities such as the folding and unfolding rates, the amplitude of the μ s relaxation, and the barrier height, there are many possible ways to improve the model such as separating the hydrophobic interactions and the hydrogen bonding effects, and assigning different entropic costs according to the secondary structures. Improvement of these empirical parameters will provide better descriptions for the kinetic parameters.

References

- (1) Wetlaufer, D. B. *Proc. Natl. Acad. Sci, USA* **1973**, *70*, 697.
- (2) Dill, K. A. *Biochemistry* **1985**, *24*, 1501.
- (3) Baldwin, R. L. *Trends Biochem. Sci.* **1989**, *14*, 291.
- (4) Ptitsyn, O. B.; Rashin, A. A. *Biophys. Chem.* **1975**, *3*, 1.
- (5) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1982**, *51*, 459.
- (6) Daggett, V.; Fersht, A. R. *Trends Biochem. Sci.* **2003**, *28*, 18.
- (7) Daggett, V. *Curr. Opin. Struct. Biol.* **2000**, *10*, 160.
- (8) Shirts, M.; Pande, V. S. *Science* **2000**, *290*, 1903.
- (9) Snow, C. D.; Qiu, L.; Du, D.; Gai, F.; Hagen, S. J.; Pande, V. S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4077.
- (10) Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6456.
- (11) Karplus, M.; Sali, A. *Curr. Opin. Struct. Biol.* **1995**, *5*, 58.
- (12) Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7*, 29.
- (13) Brooks, C., L., III *Acc. Chem. Res.* **2002**, *35*, 447.
- (14) Mirny, L.; Shakhnovich, E. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 361.
- (15) Munoz, V.; Eaton, W. A. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11311.
- (16) Henry, E. R.; Eaton, W. A. *Chem. Phys.* **2004**, *307*, 163.
- (17) Bruscolini, P.; Pelizzola, A. *Phys. Rev. Lett.* **2002**, *88*, 258101.
- (18) Alm, E.; Baker, D. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11305.
- (19) Munoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5872.
- (20) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
- (21) Makhatadze, G. I.; Lopez, M. M.; III, J. M. R.; Thomas, S. T. *Protein Sci.* **1998**, *7*, 689.
- (22) Ibarra-Molero, B.; Loladze, V. V.; Makhatadze, G. I.; Sanchez-Ruiz, J. M. *Biochemistry* **1999**, *38*, 8138.
- (23) Krantz, B. A.; Dothager, R. S.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *337*, 463.
- (24) Went, H. M.; Jackson, S. E. *Protein Eng. Des. Sel.* **2005**, *18*, 229.

- (25) Dastidar, S. G.; Mukhopadhyay, C. *Phys. Rev. E* **2005**, *72*, 051928.
- (26) Alonso, D. O. V.; Daggett, V. *Protein Sci.* **1998**, *7*, 860.
- (27) DeFlores, L. P.; Tokmakoff, A. *J. Am. Chem. Soc.* **2006**, *128*, 16520.
- (28) Harding, M. M.; Williams, D. H.; Woolfson, D. N. *Biochemistry* **1991**, *30*, 3120.
- (29) Stockman, B. J.; Euvrard, A.; Scahill, T. A. *J. Biomol. NMR* **1993**, *3*, 285.
- (30) Cordier, F.; Grzesiek, S. *Biochemistry* **2004**, *43*, 11295.
- (31) Cox, J. P. L.; Evans, P. A.; Packman, L. C.; Williams, D. H.; Woolfson, D. N. *J. Mol. Biol.* **1993**, *234*, 483.
- (32) Jourdan, M.; Searle, M. S. *Biochemistry* **2000**, *39*, 12355.

Appendix 8-A. Exact solution of the ME model¹⁷

In this appendix, an example will be given to see how exact calculation of the ME model works for a 4-residue protein ($N = 4$). According to Eq. (6) of Ref. ¹⁷, the partition function (Z) is given by the multiplication of the transfer matrices (Q) and is expressed by a polynomial of a dummy variable λ ,

$$Z(\lambda) = v_0^{0T} Q_1^0(\lambda) \dots Q_{N+1}^N(\lambda) v_0^{N+1} = \sum_{j=0}^N Z_j \lambda^j. \quad (8.A1)$$

where v_k^j is a $(j+1) \times 1$ column vector, the $(k+1)$ th element of which is 1. Other elements are 0. k ranges from 0 to j . The transfer matrix acts on this vector as

$$\begin{aligned} Q_{j+1}^j(\lambda) v_k^{j+1} &= v_{k-1}^j, & \text{for } k = 1, \dots, j+1 \\ Q_{j+1}^j(\lambda) v_0^{j+1} &= \sum_{k=0}^j \lambda^k w_{j,j+1-k} v_k^j. \end{aligned} \quad (8.A2)$$

Here, $w_{j,i}$ represents the weighting factor of a native stretch from residue i to j . Then, the partition function can be obtained as shown below.

$$\begin{aligned} Z(\lambda) &= v_0^{0T} Q_1^0 \dots Q_4^3 \left[Q_5^4 \equiv \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \lambda w_{4,4} & 0 & 0 & 0 & 0 & 0 \\ \lambda^2 w_{4,3} & 0 & 0 & 0 & 0 & 0 \\ \lambda^3 w_{4,2} & 0 & 0 & 0 & 0 & 0 \\ \lambda^4 w_{4,1} & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ &= v_0^{0T} Q_1^0 \dots Q_3^2 \left[Q_4^3 \equiv \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \lambda w_{3,3} & 0 & 0 & 0 & 0 \\ \lambda^2 w_{3,2} & 0 & 0 & 0 & 0 \\ \lambda^3 w_{3,1} & 0 & 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} 1 \\ \lambda w_{4,4} \\ \lambda^2 w_{4,3} \\ \lambda^3 w_{4,2} \\ \lambda^4 w_{4,1} \end{pmatrix} \\ &= v_0^{0T} Q_1^0 Q_2^1 \left[Q_3^2 \equiv \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ \lambda w_{2,2} & 0 & 0 & 0 \\ \lambda^2 w_{2,1} & 0 & 0 & 0 \end{pmatrix} \right] \left[\begin{pmatrix} 1 \\ \lambda w_{3,3} \\ \lambda^2 w_{3,2} \\ \lambda^3 w_{3,1} \end{pmatrix} + \begin{pmatrix} \lambda w_{4,4} \\ \lambda^2 w_{4,3} \\ \lambda^3 w_{4,2} \\ \lambda^4 w_{4,1} \end{pmatrix} \right] \end{aligned}$$

$$\begin{aligned}
&= v_0^{0T} Q_1^0 \left[Q_2^1 \equiv \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ \lambda w_{1,1} & 0 & 0 \end{pmatrix} \right] \cdot \left[(1 + \lambda w_{4,4}) \begin{pmatrix} 1 \\ \lambda w_{2,2} \\ \lambda^2 w_{2,1} \end{pmatrix} + \begin{pmatrix} \lambda w_{3,3} + \lambda^2 w_{4,3} \\ \lambda^2 w_{3,2} + \lambda^3 w_{4,2} \\ \lambda^3 w_{3,1} + \lambda^4 w_{4,1} \end{pmatrix} \right] \\
&= v_0^{0T} \left[Q_1^0 \equiv (0 \ 1) + (1 \ 0) \right] \cdot \left[(1 + \lambda w_{4,4} + \lambda w_{3,3} + \lambda^2 w_{4,3}) \begin{pmatrix} 1 \\ \lambda w_{1,1} \end{pmatrix} \right. \\
&\quad \left. + (1 + \lambda w_{4,4}) \begin{pmatrix} \lambda w_{2,2} \\ \lambda^2 w_{2,1} \end{pmatrix} + \begin{pmatrix} \lambda^2 w_{3,2} + \lambda^3 w_{4,2} \\ \lambda^3 w_{3,1} + \lambda^4 w_{4,1} \end{pmatrix} \right] \\
&= 1 + \lambda (w_{1,1} + w_{2,2} + w_{3,3} + w_{4,4}) + \lambda^2 (w_{2,1} + w_{3,2} + w_{4,3} + w_{1,1} w_{3,3} + w_{2,2} w_{4,4} + w_{1,1} w_{4,4}) \\
&\quad + \lambda^3 (w_{3,1} + w_{4,2} + w_{2,1} w_{4,4} + w_{1,1} w_{4,3}) + \lambda^4 (w_{4,1})
\end{aligned}$$

Appendix 8-B. Calculation of free energy surface

```

% exactMElp3Dcharge.m -----
% Calculation of free energy surface

clear all

num=76; % number of residues
coord=load('ubiqcoorda.txt'); % Ca coordinate
coordc=load('ubiqcoordcharge.txt'); % coordinates for the charged atoms
temperature=[48.2 53 57.9 62.7 67.4 72.2];
hpin1=1:17; hpin2=42:49; strand1=12:17; strand3=42:45; strand5=66:71;
nrcheck1(1:num)=0;nrcheck1(hpin1)=1; % check indices for reaction coordinate I-II
nrcheck2(1:num)=0;nrcheck2(hpin2)=1;nrcheck2(strand5)=1; % check indices for reaction
coordinate I-II
cthrlsd=7; % cutoff distance of native contact between Ca
R=8.3145;
ncont=contacta(coord,cthrlsd); % make a contact map
ee=-0.4765*4186.8; % interaction energy
ds1=-3.3*0.54*4.1868; % residue fixing entropy
ds(1:num)=ds1;
Cn0=6; % Flory characteristic ratio
for ii=2:num-2; % calculating loop entropy
    for jj=ii+1:num-1;
        dslp(ii,jj)=R*log(2*pi*(norm(coord(ii-1,:)-coord(jj+1,:)))^3/3/(2*pi*Cn0*(jj-ii)*3.6^2/3)^1.5);
    end
end
fe=ncont.*ee; % interaction energy map
fe=calcharge(fe,coordc); % interaction energy map modified by charge repulsion
dr=0.75; % increment of the radius of the folded sphere
dr0=6; % radius of the first sphere

for temper=1:length(temperature);
T=temperature(temper)+273.15;

% Calculation of native contact for exact ME model -----
ww1=zeros(num+1,num+1);
for temp=1:length(T); % calculating weighting factor of the native stretch from ii to jj
    for ii=1:num;
        for jj=ii:num;
            ie=sum(sum(fe(ii:jj,ii:jj)))/2-T(temp)*sum(ds(ii:jj));
            ww=exp(-(ie)/R/T(temp));
            ww1(jj,ii)=ww;
        end
        ww1(ii,ii+1)=1;
    end
end

vec2ns=zeros(num+1,2); % 1d partition - exact ME model -----
vec2ns(1,:)=1;
for m=2:num+1;
    clear vectt1;
    for k=1:m;
        qtt=zeros(num+1,1); % qtt is the transfer matrix element
        qtt(k)=ww1(m-1,m-k+1);
        vectt1(:,k)=conv(vec2ns(:,k),qtt);
    end
end

```

```

    end
    vectt2=sum(vectt1(1:num+1,:),2);    % vec2ns is the transfered row vector
    vec2ns=cat(2,vectt2,vec2ns);    % the column elements represent the polynomial of lambda
end
partfj(:,temper)=vec2ns(:,1);    % projected partition function in 1D

numnr1=sum(nrcheck1);    % 2d contact - exact ME model -----
numnr2=sum(nrcheck2);

vec2ns2=zeros(numnr1+1,numnr2+1,2);
vec2ns2(1,1,:)=1;
for m=2:num+1;
    clear vectt12;
    qtt2=zeros(numnr1+1,numnr2+1);
    qtt2(1,1)=1;
    vectt12(:,1)=conv2(vec2ns2(:,1),qtt2);
    for k=2:m;
        qtt2=zeros(numnr1+1,numnr2+1);
        qtt2(sum(nrcheck1(m-k+1:m-1))+1,sum(nrcheck2(m-k+1:m-1))+1)=ww1(m-1,m-k+1);
        vectt12(:,k)=conv2(vec2ns2(:,k),qtt2);
    end
    end
    vectt22=sum(vectt12(1:numnr1+1,1:numnr2+1,:),3); % third dimension of vec2ns represents
    vec2ns2=cat(3,vectt22,vec2ns2);    % the transfered row vector in 2D FE surface
end
partfj2=vec2ns2(:,1);    % projected partition function in 2D
partf2t(:,temper)=partfj2;

nr1s1=zeros(num,num);    % double stretch + loop -----
nr2s1=zeros(num,num);
ww1=zeros(num,num);

for jj=1:num;    % calculating weighting factor of the native stretch from jj to jj+ii-1
    for ii=1:num-jj+1;
        ie=sum(sum(fe(ii:ii+jj-1,ii:ii+jj-1)))/2-T*sum(ds(ii:ii+jj-1));
        ww=exp(-(ie)/R/T);
        ww1(jj,ii)=ww;    % average over j
        nr1s=sum(nrcheck1(ii:ii+jj-1));    % # of folded residues for coordinate 1 in the stretch
        nr2s=sum(nrcheck2(ii:ii+jj-1));    % # of folded residues for coordinate 2 in the stretch
        nr1s1(jj,ii)=nr1s;
        nr2s1(jj,ii)=nr2s;
    end
end
wsum2lp=zeros(num,1);
wwnr2lp(1:numnr1+1,1:numnr2+1)=0;
for jj=1:num;
    for LL=1:jj-1;
        for kk=1:num-LL-1;
            pp=jj-LL;
            for mm=LL+kk+2:num-pp+1;
                ieadd=sum(sum(fe(kk:LL+kk-1,mm:pp+mm-1))); % interaction between diff. stretches
                wwlp=ww1(LL,kk)*ww1(pp,mm)*(exp(-(ieadd-T*dslp(LL+kk,mm-1))/R/T));
                wsum2lp(jj)=wsum2lp(jj)+wwlp;
                nr1s=nr1s1(LL,kk)+nr1s1(pp,mm);
                nr2s=nr2s1(LL,kk)+nr2s1(pp,mm);
                wwnr2lp(nr1s+1,nr2s+1)=wwnr2lp(nr1s+1,nr2s+1)+wwlp;
            end
        end
    end
end

```

```

    end
  end
end
wsum2lps(:,temper)=wsum2lp;          % 1d partition function by DSA + loop
wwnr2lps(:,temper)=wwnr2lp;         % 2d partition function by DSA + loop

strange=1:num;                       % 3D contact -----
wsum=zeros(num,length(strange));
wsum2=zeros(numnr1+1,numnr2+1,length(strange));

for st=1:length(strange);             % folding nucleus
  for k=1:num;                         % determine the location the residues in the spheres
    distid(k)=ceil((norm(coord(k,:)-coord(strange(st,:),:))-dr0)/dr)+2;
  end
  distid(find(distid<2))=2;distid(st)=1;
  for k=1:max(distid);
    foldidold=(distid/k<1);
    fidold=find(distid<k);
    fidnew=find(distid==k);
    nfoldnew=length(fidnew);
    if (nfoldnew)
      for nn=1:2^(nfoldnew)-1; % permuting by the conversion from decimal to binary number
        folded=de2bi(nn);
        foldedm=zeros(1,nfoldnew);foldedm(1:length(folded))=folded;
        foldidnew=foldidold;
        foldidnew(fidnew)=foldedm;
        unfoldid=1-foldidnew;
        fid=find(foldidnew>0); % fid is indices of folded residues (1 - 76)
        looptest=unfoldid(fid(1):num-1).*unfoldid(fid(1)+1:num);
        loopid=fid(1)+find(abs(looptest(1:end-1)-looptest(2:end)));
        nloop=floor(length(loopid)/2);
        if (nloop>1)
          ie=sum(sum(fe(fid,fid)))/2-T*sum(ds(fid));
          for m=1:nloop;
            ie=ie-T*dslp(loopid(2*m-1),loopid(2*m));
          end
          ww=exp(-(ie)/R/T);
          wsum(length(fid),st)=wsum(length(fid),st)+ww;
          wsum2(sum(nrcheck1(fid))+1,sum(nrcheck2(fid))+1,st)=wsum2(sum
(nrcheck1(fid))+1,sum(nrcheck2(fid))+1,st)+ww;
        end
      end
    end
  end
end
wsum3ds(:,temper)=sum(wsum,2);
wwnr3ds(:,temper)=sum(wsum2,3);

end

save partfjcharge.dat partfj -ascii          % save projected partition functions
save partf2tcharge partf2t
save partfjlpcharge.dat wsum2lps -ascii
save partf2tlpcharge wwnr2lps
save partfjlp3dcharge.dat wsum3ds -ascii
save partf2tlp3dcharge wwnr3ds

```

Appendix 8-C. Calculation of folding probability on FE

```
% exactMElp3DchargeProbMap.m -----
% Calculation of the folding probability of residues for given locations on free energy surface
% -----

clear all

num=76;                % number of residues
coord=load('ubiqcoorda.txt');    % Ca coordinate
coordc=load('ubiqcoordcharge.txt');    % coordinates for the charged atoms

temperature=57.9;

hpin1=1:17; hpin2=42:49; strand1=12:17; strand3=42:45; strand5=66:71;

nrcheck1(1:num)=0;nrcheck1(hpin1)=1;    % check indices for reaction coordinate I-II
nrcheck2(1:num)=0;nrcheck2(hpin2)=1;nrcheck2(strand5)=1;    % check indices for reaction
coordinate I-II

checkres=cat(2,hpin1,hpin2,strand5);

checknum=8;            % location where the prob. will be checked (check pt.)
checkpt=[17,14;17,12;15,14;13,11;16,4;5,4;17,8;11,9];

cthrld=7;              % cutoff distance of native contact between Ca
R=8.3145;

ncont=contacta(coord,cthrld);    % make a contact map

ee=-0.4765*4186.8;      % interaction energy
ds1=-3.3*0.54*4.1868;   % residue fixing entropy
ds(1:num)=ds1;
Cn0=6;                 % Flory characteristic ratio
for ii=2:num-2;        % calculating loop entropy
    for jj=ii+1:num-1;
        dslp(ii,jj)=R*log(2*pi*(norm(coord(ii-1,:)-coord(jj+1,:)))^3/3/(2*pi*Cn0*(jj-ii)*3.6^2/3)^1.5);
    end
end

fe=ncont.*ee;          % interaction energy map
fe=calcharge(fe,coordc);    % interaction energy map modified by charge repulsion

dr=0.75;              % increment of the radius of the folded sphere
dr0=6;                % radius of the first shpere

for temper=1:length(temperature);
    temper
    T=temperature(temper)+273.15;

    % Calculation of native contact -----
    ww1=zeros(num+1,num+1);
    for temp=1:length(T);    % calculating weighting of the native stretch from ii to jj
        for ii=1:num;
            for jj=ii:num;
                ie=sum(sum(fe(ii:jj,ii:jj)))/2-T(temp)*sum(ds(ii:jj));
```



```

        ww=exp(-(ie)/R/T(temp));
        ww1(jj,ii)=ww;
    end
    ww1(ii,ii+1)=1;
end
end

% 1d partition -----
vec2ns=zeros(num+1,2);
vec2ns(1,:)=1;
for m=2:num+1;
    clear vectt1;
    for k=1:m;
        qtt=zeros(num+1,1);      % qtt is the transfer matrix element
        qtt(k)=ww1(m-1,m-k+1);
        vectt1(:,k)=conv(vec2ns(:,k),qtt);
    end
    vectt2=sum(vectt1(1:num+1,:),2); % vec2ns is the transfered row vector
    vec2ns=cat(2,vectt2,vec2ns); % the column elements represent the polynomial of lambda
end
partfj(:,temper)=vec2ns(:,1); % projected partition function in 1D

%figure;plot(0:num,-R.*T.*log(partfj)./4186.8);

% 2d contact - exact ME model -----
numnr1=sum(nrcheck1);
numnr2=sum(nrcheck2);

for resi=1:num; % Summation of weighting factors when residue# = resi is folded
    vec2ns2=zeros(numnr1+1,numnr2+1,2);
    vec2ns2(1,1,:)=1;
    for m=2:num+1;
        clear vectt12;
        qtt2=zeros(numnr1+1,numnr2+1);
        if m~=resi+1 % when m = resi+1, qtt2=0 (see Eq (8) of PRL, vol.88, 258101 (2002))
            qtt2(1,1)=1;
        end
        vectt12(:,1)=conv2(vec2ns2(:,1),qtt2);
        for k=2:m;
            qtt2=zeros(numnr1+1,numnr2+1);
            qtt2(sum(nrcheck1(m-k+1:m-1))+1,sum(nrcheck2(m-k+1:m-1))+1)=ww1(m-1,m-k+1);
            vectt12(:,k)=conv2(vec2ns2(:,k),qtt2);
        end
        vectt22=sum(vectt12(1:numnr1+1,1:numnr2+1,:),3); % the third dimension of vec2ns
        vec2ns2=cat(3,vectt22,vec2ns2); % represents the transfered row vector in 2D FE surface
    end
    partfj2res(:,resi)=vec2ns2(:,1); % projected weighting factors in 2D for residue # = resi
end

for mapk=1:checknum; % folding probability of residues for a given check pt. on FE
    resmap1s(:,mapk,temper)=permute(partfj2res(checkpt(mapk,1)+1,checkpt(mapk,2)+1,:),[3 1
2]);
end

% double stretch + loop -----
nr1s1=zeros(num,num);

```

```

nr2s1=zeros(num,num);
ww1=zeros(num,num);

for jj=1:num;          % calculating weighting factor of the native stretch from jj to jj+ii-1
  for ii=1:num-jj+1;
    ie=sum(sum(fe(ii:ii+jj-1,ii:ii+jj-1)))/2-T*sum(ds(ii:ii+jj-1));
    ww=exp(-(ie)/R/T);
    ww1(jj,ii)=ww;          % average over j
    nr1s=sum(nrcheck1(ii:ii+jj-1));      % # of folded residues for coordinate 1 in the stretch
    nr2s=sum(nrcheck2(ii:ii+jj-1));      % # of folded residues for coordinate 2 in the stretch
    nr1s1(jj,ii)=nr1s;
    nr2s1(jj,ii)=nr2s;
  end
end

resmap2lp=zeros(num,checknum);

for jj=1:num;
  for LL=1:jj-1;
    for kk=1:num-LL-1;
      pp=jj-LL;
      for mm=LL+kk+2:num-pp+1;
        nr1s=nr1s1(LL,kk)+nr1s1(pp,mm);
        nr2s=nr2s1(LL,kk)+nr2s1(pp,mm);
        for mapk=1:checknum;          % check if the microstate is located at check pt
          if nr1s==checkpt(mapk,1) & nr2s==checkpt(mapk,2);
            ieadd=sum(sum(fe(kk:LL+kk-1,mm:pp+mm-1))); % interaction energy between
diff. stretches
            wwlp=ww1(LL,kk)*ww1(pp,mm)*(exp(-ieadd-T*dslp(LL+kk,mm-1))/R/T);
            resmap2lp(kk:LL+kk-1,mapk)=resmap2lp(kk:LL+kk-1,mapk)+wwlp;
            resmap2lp(mm:pp+mm-1,mapk)=resmap2lp(mm:pp+mm-1,mapk)+wwlp;
          end
        end
      end
    end
  end
end
resmap2lps(:,:,temper)=resmap2lp;

% 3D contact -----
strange=1:num;
resmap3=zeros(num,checknum);

for st=1:length(strange);          % folding nucleus
  for k=1:num;          % determine the location the residues in the spheres
    distid(k)=ceil((norm(coord(k,:)-coord(strange(st),:))-dr0)/dr)+2;
  end
  distid(find(distid<2))=2;distid(st)=1;
  for k=1:max(distid);
    foldidold=(distid/k<1);
    fidold=find(distid<k);
    fidnew=find(distid==k);
    nfoldnew=length(fidnew);
    if (nfoldnew)
      for nn=1:2^(nfoldnew)-1; % permuting using conversion from decimal to binary number
        folded=de2bi(nn);
      end
    end
  end
end

```

```

foldedm=zeros(1,nfoldnew);foldedm(1:length(folded))=folded;
foldidnew=foldidold;
foldidnew(fidnew)=foldedm;
unfoldid=1-foldidnew;
fid=find(foldidnew>0);      % fid is indices of folded residues (1 - 76)
looptest=unfoldid(fid(1):num-1).*unfoldid(fid(1)+1:num);
loopid=fid(1)+find(abs(looptest(1:end-1)-looptest(2:end)));
nloop=floor(length(loopid)/2);
if (nloop>1)
    for mapk=1:checknum;      % check if the microstate is located at check pt
        if sum(nrcheck1(fid))==checkpt(mapk,1) & sum(nrcheck2(fid))==checkpt(mapk,2);
            ie=sum(sum(fe(fid,fid)))/2-T*sum(ds(fid));
            for m=1:nloop;
                ie=ie-T*dslp(loopid(2*m-1),loopid(2*m));
            end
            ww=exp(-(ie)/R/T);
            resmap3(fid,mapk)=resmap3(fid,mapk)+ww;
        end
    end
end
end
end
end
end
end
resmap3s(:,temper)=resmap3;

end

save resmap1charge.dat resmap1s -ascii
save resmap2charge.dat resmap2lps -ascii
save resmap3charge.dat resmap3s -ascii

```

Subroutine 1

```
% contacta.m -----  
% Build a contact map by comparing the Ca distance with cthrsld  
% -----
```

```
function ncont=contacta(coord,cthrslid)
```

```
[num ntrash]=size(coord);
```

```
ncont=zeros(num);
```

```
for k=1:num-3;
```

```
    for m=k+3:num;
```

```
        if (norm(coord(k,:)-coord(m,:))<cthrslid)
```

```
            ncont(k,m)=1;
```

```
        end
```

```
    end
```

```
end
```

```
ncont=ncont+ncont';
```

Subroutine 2

```
% calcharge.m -----
```

```
% Calculate the electrostatic repulsion energy between charged atoms
```

```
% -----
```

```
function fe=calcharge(fei,coordc);
```

```
perm=1/4/pi/8.854e-12/80*6.022e23*(1.602e-19)^2*1e10;
```

```
ctlength=3.6;
```

```
[num tt]=size(coordc);
```

```
fe=fe;
```

```
for k=1:num-1;
```

```
    for m=k+1:num;
```

```
        kk=coordc(k,1);
```

```
        mm=coordc(m,1);
```

```
        distance=norm(coordc(k,2:4)-coordc(m,2:4));
```

```
        if mm-kk > 2 & distance < ctlelength*(mm-kk)
```

```
            Erepel=perm*(1/distance-1/ctlength)/(mm-kk);
```

```
            fe(kk,mm)=fe(kk,mm)+Erepel;
```

```
            fe(mm,mm)=fe(mm,mm)+Erepel;
```

```
        end
```

```
    end
```

```
end
```

Chapter 9

Transition state of ubiquitin unfolding explored by mutation studies

9.1. Introduction

The preceding chapters have provided a detailed assessment of the free energy landscape for ubiquitin unfolding. Chapter 6 described evidence that the μs burst phase unfolding of ubiquitin from the transition state involved the unfolding of contacts within strands III-V of its β -sheet while the strand I-II hairpin remained largely intact. In Chapters 7 and 8, the details of this free energy surface were more carefully revealed through temperature-dependent T-jump experiments and statistical mechanical modeling. We interpreted the increased amplitude of non-exponential unfolding on the μs time scale at high temperatures as increased heterogeneity of the unfolding transition state.

In this chapter, the details of the ubiquitin unfolding free energy landscape are further investigated through site-specific mutation studies. Three ubiquitin mutants that to stabilize or destabilize interactions within strands III-V of the β -sheet or the α -helix are investigated with T-jump experiments, and the μs and ms relaxation behavior are

contrasted with wildtype ubiquitin. The analysis of this data draws on the methods described in Chapters 7 and 8.

9.1.1. Background: Φ value analysis

Point mutation of an amino acid residue in the protein has been widely used to map out the folding transition state of proteins, most commonly using Φ value analysis.¹⁻³ A mutant is designed to give a small perturbation that decreases the stability of the native protein and the transition state without changing the native structure.² The influence of this mutation on the transition state is characterized through a Φ value: the ratio of the change of the activation free energy of folding ($\Delta\Delta G_F^\ddagger$) to the change of the free energy of folding ($\Delta\Delta G_F$) for that mutation relative to the wildtype,

$$\Phi_m = \frac{\Delta\Delta G_F^\ddagger}{\Delta\Delta G_F}. \quad (9.1)$$

Usually, Φ_m varies between 0 and 1, the two cases shown in Fig. 9-1. The value $\Phi_m = 1$ indicates that a given mutation equally changes the stability of the transition state and the folded state, and that the mutated residue is fully structured at the transition state. When $\Phi_m = 0$, on the other hand, the mutated residue is not involved in the transition state at all. In most cases, Φ_m lies between 0 and 1, which can be interpreted either as the partially folded state of the residue or as the presence of the parallel routes.⁴ In the latter case, the mutated residue is involved in part of multiple transition states.

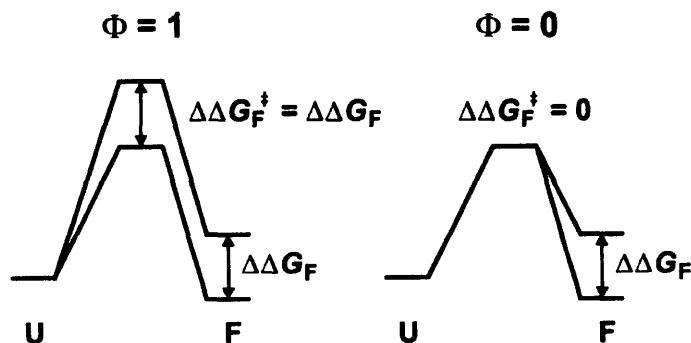


Figure 9-1. Two extreme cases of the Φ value analysis.

As an complementary method, the Ψ value analysis has been developed, which is useful for the investigating the role of pair-wise interactions in the folding transition state.⁵⁻⁷ In this method, two histidine mutations (bi-His) are introduced in such a way that the two mutated residues can be linked by coordination to a metal ion. Although the candidate of mutation is limited to the residues on the surface of the protein, the stability of the bound residues can be controlled continuously by varying the concentration of the transition metal ions such as Zn^{2+} , Co^{2+} , and Cu^{2+} . This method has proven effective for stabilizing two β strands or part of an α helix.

Both Φ and Ψ value analyses have been applied to ubiquitin folding. The Φ value analysis showed that the β strands I and II and the α helix are folded at the transition state,^{7,8} which is consistent with other results. A fragment study has reported that the N-terminal peptides (1-35) show a small population having a native-like structure while C-terminal peptides (36-76) form a nonnative structure.^{9,10} Also a partially folded structure of ubiquitin including strands I and II and the α helix has been observed in the organic solvent and water mixture.¹¹⁻¹³ A simulation study supports the result of the Φ value analysis as well.¹⁴ The transition state of the Ψ value analysis is broader than that of the Φ value analysis and includes the β strands V and III as well,^{6,7} which is supported by another simulation study.¹⁵ However, the propensity of overestimating the stability of the mutated residue in the Ψ value analysis is still under debate.¹⁶

Most of these mutation experiments have been performed at room temperature by the stopped flow method, the time scale of which is limited to the ms range. In this chapter, we combine the conventional Φ value analysis (Φ_m) and Φ_T value analysis shown in Chapter 7 to describe the transition state during the thermal unfolding process. Also, the origin of the fast unfolding phase observed on μs timescales is discussed.

9.2. Experimental

The ubiquitin mutants were kindly provided by Ali Shandiz from Tobin Sosnick's group at the University of Chicago. We use three of an extensive series of mutants that the Sosnick group has cloned in their studies of ubiquitin folding.^{6,7,17} Since the mutant samples were lyophilized from organic solvent after purification, sometimes, they did not

dissolve well in water. Sonicating at low pH (0.18% DCl, 1/200 dilution) was helpful for this problem. Also, to remove a trace of trifluoroacetic acid (TFA) added in the purification step, the sample was lyophilized under acidic conditions (0.18% DCl, 1/200 dilution). About 5 mg of sample was lyophilized twice more after dissolving in 500 – 800 mL of D₂O. The T-jump and the DVE probe experiment were performed at a sample concentration of 30 mg/mL in the 0.35% DCl solution. The remaining experimental procedures and conditions are the same as those described in Chapter 4 and 7.

9.2.1. Overview of the mutants

Three different mutants are used in this chapter, which are named according to the convention in Ref. 6. The locations of mutated residues are shown with blue-colored boxes in the projected β sheet in Fig. 9-2. Mutant *i* and *l* are bi-His mutants of the β sheet and α helix, respectively, and β strands in mutant *g* are covalently linked. In mutant *l*, two more mutations are located in the α helix towards the other side of the β sheet.

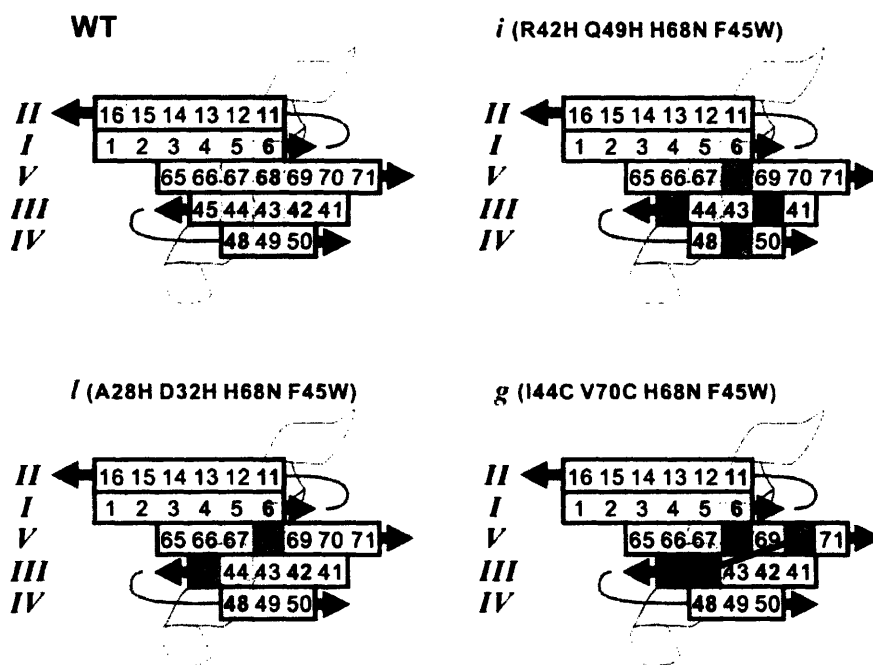


Figure 9-2. Projection of the β -sheet registries of ubiquitin wildtype and mutants. Each square box in the β sheet indicates an amino acid residue with the residue number. Charged residues are indicated with red figures. Blue-colored residues are mutation sites. For mutant *l*, mutations at residue 28 and 32 are located in the C-terminus (upper right part) of the α helix. Residue 44 and 70 in mutant *g* are covalently linked.

Since the experiments were performed in acidic condition (0.35% DCl) for reversible thermal unfolding, all side chain amine groups, indicated by red numbers in Fig. 9-2, are positively charged. In wildtype ubiquitin (WT), there are five positively charged residues in the β sheet. Since the side chains of Lys 6, Arg 42, and His 68 are oriented away from the hydrophobic core (out of the plane), while those of the residue Lys 11 and Lys 48 penetrate inward, side chain repulsion effects between strand I and II, and III and IV are small. However, at acidic condition, the repulsion between residues 6, 68, and 42 destabilizes the protein and lowers the melting temperature relative to neutral pH.

Mutant *i* and *l* are bi-His mutants originally designed for Ψ analysis.⁶ However, we cannot utilize the metal-ion-induced ligation of the mutated residues because of the lower pH than the pKa of His of 6.5. At this condition, the side chain of the His residue is protonated and cannot bind to the metal ion. Instead, the mutation affects the protein stability by charge-charge repulsion and disrupting structure-stabilizing hydrogen bonding networks for the water in the vicinity of those residues. In mutant *i*, the addition of a charge at residue 49 (residue 42 is already charged in WT) makes the protein unstable by about 2.1 kcal/mol calculated from the electrostatic interaction term in Eq. (8.2). Although the increase of the energy is mainly exerted on the newly charged residue, the long-range electrostatic repulsion also affects the stability of other charged residues slightly. In all mutants, Phe 45 is mutated to Trp to allow fluorescence detection of folding and His 68 is mutated to Gln to avoid an unwanted metal ion induced linkage in the original Sosnick experiments.^{6,7} The Trp mutation at residue 45 is known to destabilize the protein slightly by ~ 0.4 kcal/mol.¹⁸ On the other hand, removal of the charged His 68 stabilizes the protein by 1.9 kcal/mol. As a result, in mutant *i*, strand IV (residue 49) is destabilized in addition to Trp 45 in strand III while strand V is stabilized.

For mutant *l*, the addition of the two charges at residues 28 and 32 in the α helix decrease the stability of the protein by 2.5 kcal/mol but the destabilization between the two mutated residues is very small because of their proximity in the sequence.

Mutant *g* differs from *i* and *l* in that the β sheet is stabilized through a covalent linkage. In this mutant, residues 44 and 70 in strand III and V are mutated to Cys and cross-linked with dichloroacetone to form the a covalent -S-CH₂-CO-CH₂-S- bridge.¹⁹

Therefore, the stability of the strand III and V is greatly increased. A control experiment with the unlinked bi-Cys mutant has also been tried under excess dithiothreitol (DTT) to avoid dimerization by a disulfide linkage. However, thermal aggregation is too fast to obtain meaningful data. A quoted “mutant *g*” indicates the covalently linked one in this chapter.

9.3. Results

9.3.1. Equilibrium thermal unfolding

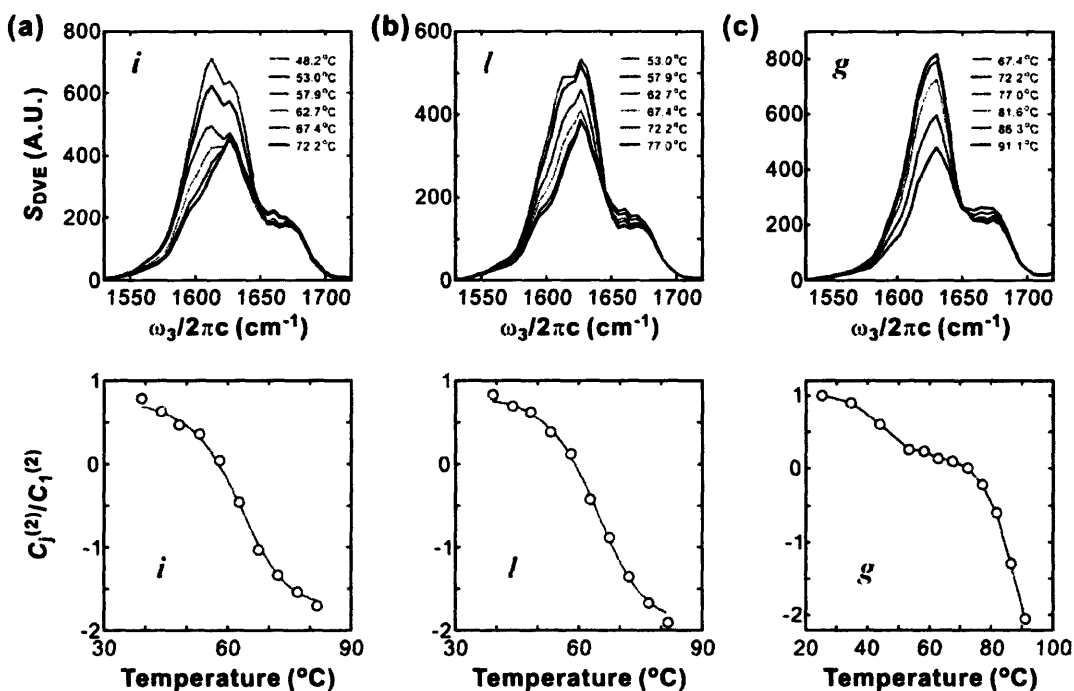


Figure 9-3. Equilibrium thermal unfolding of three mutants, (a) *i*, (b) *l*, and (c) *g*. Melting curves in the lower panels are constructed from the SVD 2nd components. Red curves are fits to the two-state thermodynamics.

The equilibrium thermal unfolding of the mutants monitored through their DVE spectra are shown with their melting curves in Fig. 9-3. The unfolding of mutant *i* and *l* is similar to that of WT. The spectral intensity in the ν_{\perp} region decreases and the peak shifts to the blue side as a result of the disruption of the β sheet. The melting curves constructed by the second SVD component show a sigmoid shape and the fitting to Eqs. (7.1) and

(7.2) gives the thermodynamic parameters summarized in Table 9-1. As mentioned in Section 9.2.1, the destabilization by the insertion of charges seems to be compensated by the stabilization by a removal of the charged His 68. The melting curve in Fig. 9-3c shows that the thermal unfolding of mutant *g* is not complete even at 91°C. Therefore, thermodynamic parameters cannot be obtained from mutant *g*. Nonetheless, the similarity of the DVE spectrum at 91°C to those of other mutants at the highest temperature implies a fair amount of unfolding of the protein.

	T_m (°C)	ΔH_m (kcal/mol)	ΔS_m (cal/mol·K)	ΔC_p (cal/mol·K)
WT	64.0	40.9	121	1264
<i>i</i>	62.6	37.5	112	813
<i>l</i>	64.1	37.7	112	693

Table 9-1. Thermodynamic parameters of wildtype and mutant ubiquitin.

9.3.2. Transient thermal unfolding

9.3.2.1. Overview

The T-jump ranges of mutant *i* and *l* are chosen similar to those of WT for comparison. Fig. 9-4 shows the transient DVE difference spectra of the three mutants. In all three cases, unfolding is faster than that of the wildtype. Also, the crossover to refolding kinetics at the balanced position ($k_u P_F = k_f P_U$) is reached at 2 – 3 ms, which is also earlier than the 4 – 6 ms in the wildtype. Both mutants *i* and *l* show the fast relaxation on the 10 – 100 μ s time scale in the v_{\perp} region. The monotonic non-exponential decay of the signal resembles the relaxation of WT at the highest temperature in Chapter 7, which shows the heterogeneous feature in the transition state. Compared to *i* and *l*, unfolding of mutant *g* does not show a fast μ s relaxation. The difference spectra are almost same until 20 μ s and show a slight decay at 100 μ s. However, the amplitude of the change at 3 ms is similar to those of mutant *i* and *l*, which implies the ms unfolding rate is

comparable. The absence of the μs unfolding phase indicates that unfolding is dominated by activated barrier crossing.

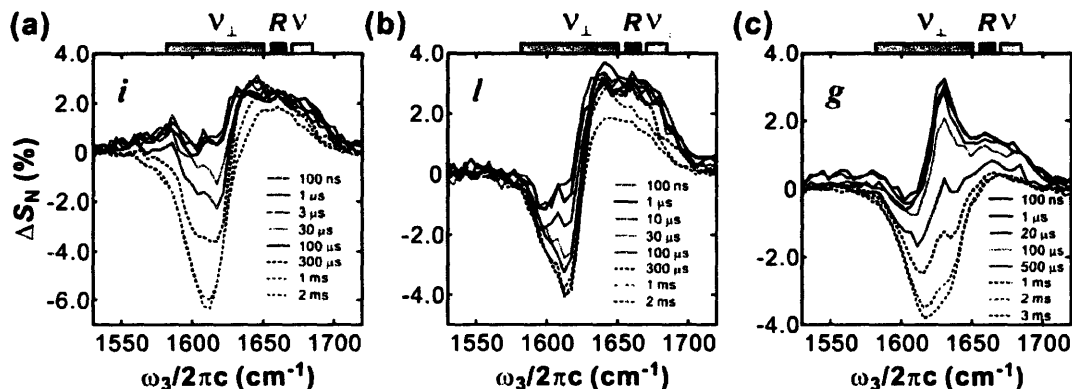


Figure 9-4. Transient DVE difference spectra of (a) mutant *i* ($57.9^{\circ}\text{C} \rightarrow 68.2^{\circ}\text{C}$), (b) mutant *l* ($62.7^{\circ}\text{C} \rightarrow 71.9^{\circ}\text{C}$), and (c) mutant *g* ($77.0^{\circ}\text{C} \rightarrow 78.6^{\circ}\text{C}$).

9.3.2.2. Mutant *i*

Using the analysis described in Section 7.3.4, the population relaxation of the folded species is reconstructed using the first SVD component for the v_{\perp} region. The results are presented in Fig. 9-5a. Relative to the minimum of the relaxation profile, the short time period (unfolding region) and long time period (refolding region) are analyzed separately. A combination of stretched exponential and exponential fit of the unfolding part gives the timescale and the amplitude of the μs unfolding as in Eq. (7.6),

$$C_j^{(1)} = A_{\mu\text{s}} \exp\left[-\left(\tau_j/\tau_{\mu\text{s}}\right)^{\beta}\right] + A_{\text{ms}} \exp\left[-\tau_j/\tau_{\text{ms}}\right], \quad (9.2)$$

where τ_j is the j th delay time ranging from 100 ns to the time when $C_j^{(1)}$ is its minimum. The relative amplitudes $A_{\mu\text{s},\text{rel}}$ and $A_{\text{ms},\text{rel}}$ for the μs and ms part are obtained using Eq. (7.7). Fitting results are summarized in Table 9-2. The burst phase unfolding time scales of 9 – 45 μs are somewhat longer than that of WT.

From the refolding transients for delays longer than 10 ms, the relaxation rates at three initial temperatures are obtained by an exponential fit. Thermodynamic parameters for the barrier are obtained by fitting the folding and unfolding rate constants to Eq. (7.8), which gives $E_0 = 6.7$ kcal/mol, $E_1 = -60$ cal/mol·K, and $E_2 = -0.90$ cal/mol·K².

$T_i - T_f$ (°C)	β	$\tau_{\mu s}$ (μs)	$A_{\mu s,rel}$ (%)	$A_{\mu s,rel} + A_{ms,rel}$ (%)	$P_{\mu s}/P_F$ (%)
53.0 – 63.3	0.61	9.3	4.3	13.2	1.8
57.9 – 68.2	0.57	45.3	6.3	19.8	3.8
62.7 – 74.5	0.83	20.7	11.9	22.6	9.4

Table 9-2. Relaxation parameters extracted from the SVD component of the v_{\perp} region (1582 – 1651 cm^{-1}) of mutant i . β and $\tau_{\mu s}$ are the stretched exponential ($\exp[-(t/\tau_{\mu s})^{\beta}]$) parameters for the μs relaxation part. $A_{\mu s,rel}$ and $A_{ms,rel}$ are the relative amplitudes of the μs and ms relaxation compared to the expected change when the temperature is kept constant at the final temperature of the T-jump. $P_{\mu s}/P_F$ is the relative population change during the μs relaxation to the folded population at the initial temperature.

Using the fast unfolding kinetics and the information on the barrier height, expected unfolding population changes are calculated by solving the differential equation given in Eq. (7.10). The results are compared with experimental values in Fig. 9-5a. Consistent with the WT data at $T_i = 67^{\circ}C$ in the previous section, the calculated population changes does not properly reproduce the experimental changes. Part of the reason for the deviation of the experimental and calculated relaxations is the inaccuracies in timescale and amplitude of the fast unfolding phase because of its merging with the slow unfolding phase. The global stretched exponential fit of the whole unfolding part is shown and compared with those of mutant l and WT in Table 9-4.

$$C_j^{(1)} = A \exp\left[-\left(\tau_j/\tau_{\mu s}\right)^{\beta}\right]. \quad (9.3)$$

Interestingly, the relaxation time of mutant i is similar to that of the WT.

$T_i - T_f$ (°C)	β_{\parallel}	τ_{\parallel} (ms)	β_r	τ_r (ms)
53.0 – 63.3	0.97	3.7	1.00	4.8
57.9 – 68.2	0.88	3.9	1.00	5.3
62.7 – 74.5	0.93	4.9	0.98	5.8

Table 9-3. Relaxation parameters extracted from the first SVD component of the v_{\parallel} (1670 – 1690 cm^{-1}) and random coil (1656 – 1666 cm^{-1}) regions of mutant i .

This similarity of the relaxations of mutant *i* and WT can also be found in other frequency regions. In Fig. 9-5b, relaxations of the random coil region and v_{\parallel} region are compared with the temperature relaxation. As expected, the relaxations of both regions are longer than the temperature relaxation. The relaxation parameters by a stretched exponential fit are summarized in Table 9-3. This longer relaxation time indicates positive changes of these regions as a result of unfolding of the protein.

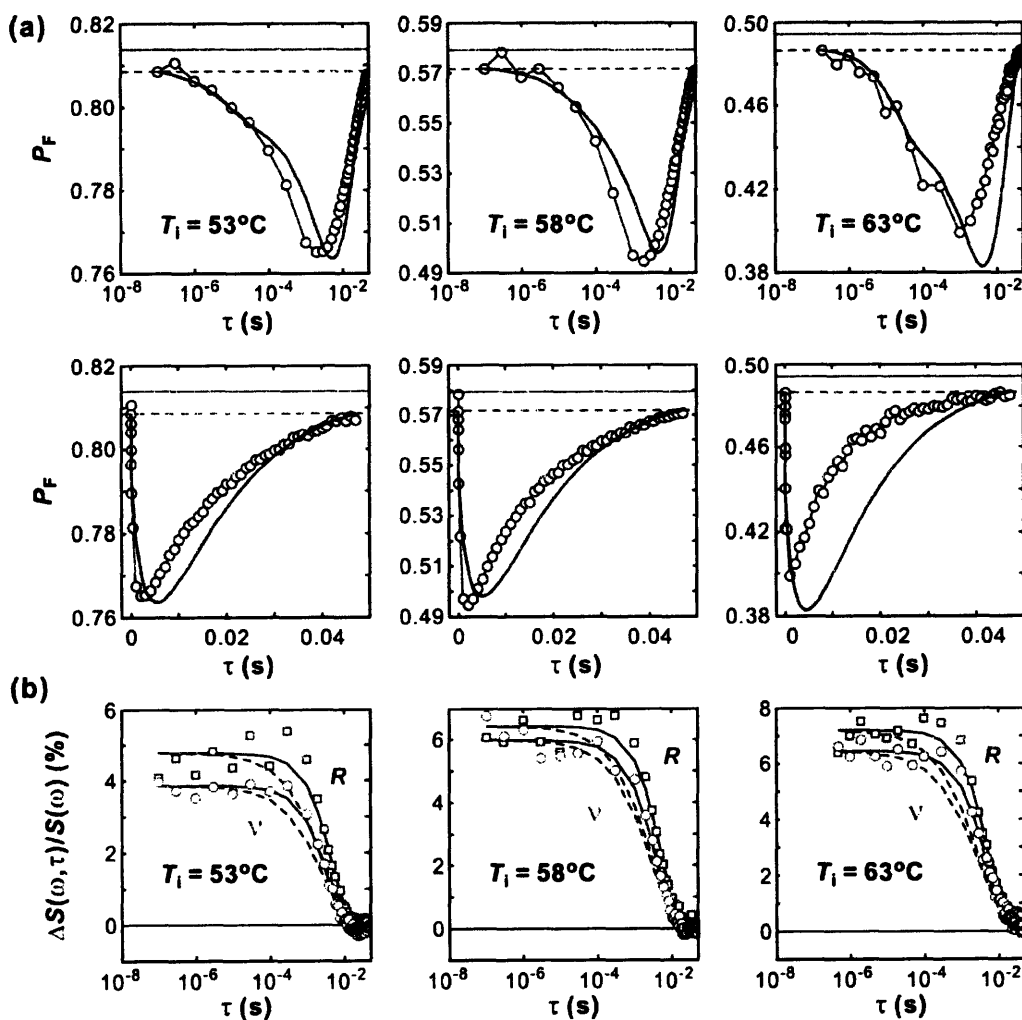


Figure 9-5. Relaxations of mutant *i*. (a) Population changes of the folded species after the T-jump are obtained from the experiment (open circle) and the calculation (red line) for the v_{\perp} region. The solid horizontal and the dashed horizontal lines represent the initial population and the steady state value, respectively. Same data are shown with a log (first row) and linear (second row) timescale. (b) Spectral relaxations of the v_{\parallel} (open circle) and random coil (open square) regions. Solid lines are stretched exponential fit and dashed lines are the temperature relaxation profile normalized to the amplitude of each of the fits.

Additionally, the temperature Φ value (Φ_T) is calculated using Eq. (7.11),

$$\Phi_T(T) = \frac{\partial \Delta G_F^\ddagger(T)/\partial T}{\partial \Delta G_F(T)/\partial T} = \frac{E_1 + \Delta S_m + (2E_2 + \Delta C_p/T_m) \cdot (T - T_m)}{\Delta S_m + (\Delta C_p/T_m) \cdot (T - T_m)}, \quad (9.4)$$

$$= \frac{51 + 0.62(T(^{\circ}\text{C}) - T_m(^{\circ}\text{C}))}{112 + 2.42(T(^{\circ}\text{C}) - T_m(^{\circ}\text{C}))}$$

which results in $\Phi_T = 0.51$ at $T_i = 53^{\circ}\text{C}$ and 0.46 at $T_i = 63^{\circ}\text{C}$. This decreasing trend with temperature is also similar to that of WT that is interpreted by the increase of the heterogeneity of the transition state.

All these relaxation features shown above indicate that the unfolding process of mutant *i* is governed by the heterogeneous process though it is unclear whether the mechanism is same as that of the WT at high temperature. This question will be addressed with calculations of the free energy surface in Section 9.4.1.

9.3.2.3. Mutant *l*

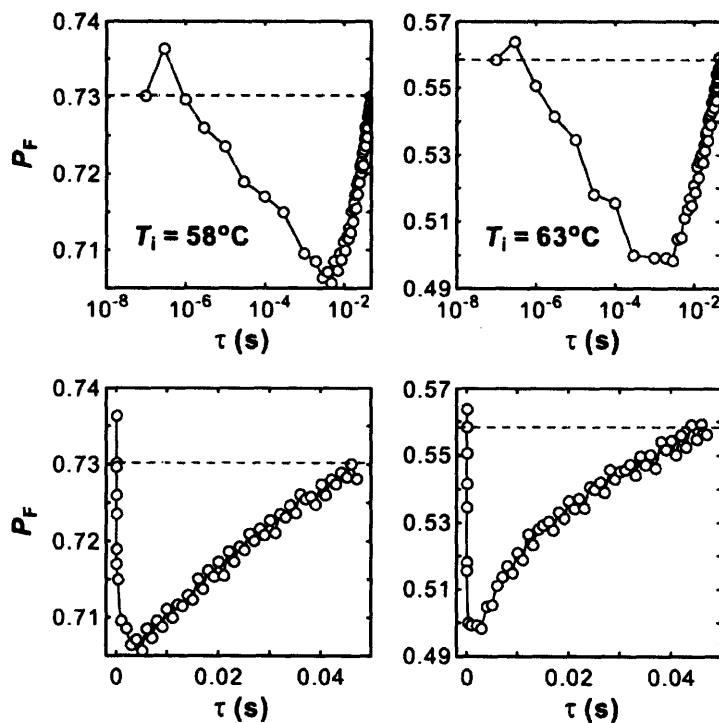


Figure 9-6. Experimentally obtained population relaxations of the folded species of mutant *l* are shown with a log and linear timescale. Dashed horizontal lines indicate the initial population.

The unfolding of mutant *l* shows more strange behavior both in the μs and ms timescale. The unfolding relaxation is highly heterogeneous ($\beta < 0.4$) as shown in Fig. 9-6 and Table 9-4. As for mutant *i*, a global stretched exponential fit is appropriate for the unfolding part. The unfolding rate is faster than those of WT at $T_i = 67^\circ\text{C}$ and mutant *i* though it is slower than the molecular timescale of 1 μs . On the other hand, the refolding part is opposite. The refolding relaxation (17 – 19 ms) is slower than those of WT (14 – 17 ms) and mutant *i* (12 – 15 ms). Since the melting temperatures of mutant *l* and WT are similar, the unfolding and refolding rate should be correlated with each other. However, the faster unfolding and slower refolding of mutant *l* are anti-correlated with those of WT, implying that the unfolding/refolding mechanism is different. It is clear that the slow refolding part should be governed by the barrier-crossing process. Also, from the timescale of tens of μs , unfolding of mutant *l* is expected to be a quasi-barrierless process. In other words, the unfolding phase is dominated by this barrierless process and the amount of the ms unfolding is very small. The small amplitude of the population change in Table 9-4 compared to those of mutant *i* and WT also indicates most of the relaxation came from the fast unfolding phase. Detailed interpretations will be followed in Section 9.3.2.

	$T_i - T_f$ ($^\circ\text{C}$)	β	$\tau_{\mu\text{s}}$ (μs)	A_{rel} (%)	$\Delta P_F/P_F$ (%)
<i>i</i>	53.0 – 63.3	0.57	227	12.7	5.3
	57.9 – 68.2	0.71	252	19.5	11.8
	62.7 – 74.5	0.47	85.3	24.4	19.3
<i>l</i>	57.9 – 67.3	0.29	76.9	8.9	3.2
	62.7 – 71.9	0.38	21.9	20.5	7.2
WT	67.4 – 77.9	0.49	277	19.6	16.8

Table 9-4. Global stretched exponential fit of the unfolding part. A_{rel} is the relative signal change normalized to the maximum changes expected when the temperature is fixed at the final temperature of the T-jump. ΔP_F is the population change calculated from A_{rel} .

9.3.2.4. Mutant *g*

Fig. 9-7 shows two T-jump relaxations of the v_{\perp} region of mutant *g*. As mentioned above, unfolding is not completed until 91°C and the thermodynamic parameters are not available. Although quantitative analyses are not possible, the T-jump relaxations show several trends. Firstly, the data show fast refolding relaxations of about 8 ms, which is twice as fast as that of WT. Because of lack of the equilibrium constant, separate comparisons of refolding and unfolding rates with those of WT are not possible. The fast relaxation can be attributed to the relaxation at high temperature. Also, the amplitude of the μs unfolding is small as expected from the structure of the protein. The ligation of strand III and V may block the sequential unfolding path, in the transition of which strands III-V are partially unfolded. Also, the reduced heterogeneity of the transition state may allow only the barrier crossing process.

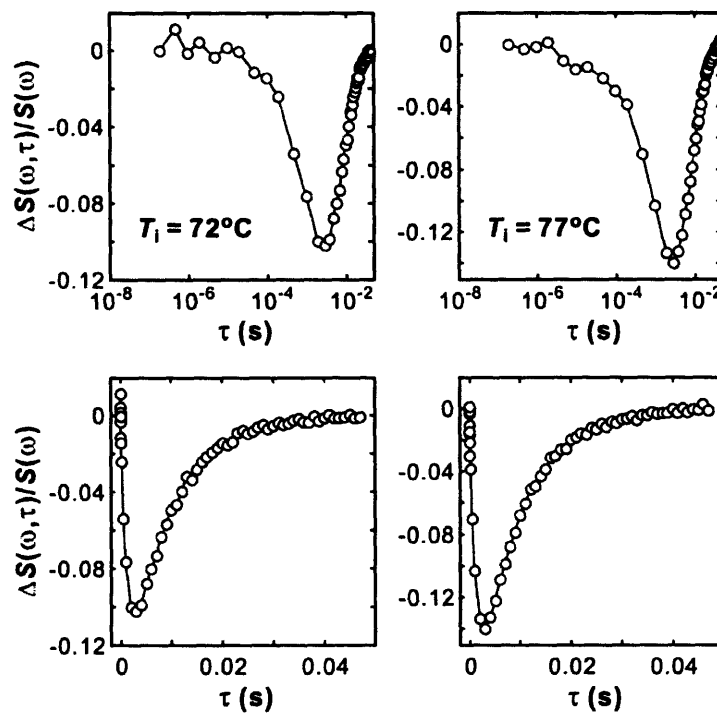


Figure 9-7. Relaxations of mutant *g*.

9.4. Discussion

9.4.1. Transition state from Φ value analysis

The transition state of ubiquitin folding has been known to be localized in the β hairpin by strand I and II and the α helix.⁸ The extension of the transition state to the strands V and III has been reported^{6,7} but it is still controversial.¹⁶ In Chapter 8, the transition state calculated by the ME model includes the α helix and β strands I and II with partially unfolded strands III-V. This more extended transition state in thermal unfolding can be rationalized by the Hammond postulate.²⁰ The postulate states that the structure of the transition state resembles the structure of a state that is energetically closer. In our experiment, as the folded state becomes less stable at higher temperature, it will be energetically closer to the transition state. Then, according to the postulate, the structure of the transition state at the melting temperature should resemble the folded structure more than that at room temperature at which most of other kinetic experiments have been performed.

	T (°C)	$\Delta\Delta G_F$ (cal/mol)	$\Delta\Delta G_F^\ddagger$ (cal/mol)	Φ_m
<i>I</i>	53.0	148	65	0.44
	57.9	173	-37	-0.21
	62.7	167	-151	-0.91
<i>L</i>	57.9	11	108	-
	62.7	-5.7	24	-

Table 9-5. Φ_m values for mutant *i* and *l*.

The Φ_m values of mutant *i* and *l* are calculated using Eq. (9.1) and summarized in Table 9-5. The high Φ_m values of mutant *l* cannot be trusted because of the very small difference of the free energy of folding ($\Delta\Delta G_F$) between WT and mutant *l*. Since the

minimum $\Delta\Delta G_F$ for the reasonable Φ_m value analysis is 0.6 kcal/mol,² even the Φ_m values of mutant i might not be trustworthy. Nevertheless, the changes of the activation free energy ($\Delta\Delta G_F^\ddagger$) itself that is obtained from the relaxation rate should be real. The difference of the relaxation rates by several ms is a significant amount of change compared to the experimental error. Also, the decrease of the melting temperature of mutant i is sufficient to confirm at least its destabilization. Then, one finds a contradictory behavior of Φ_m values, which is similar to that of Φ_T values. Although the folded state of mutant i is destabilized ($\Delta\Delta G_F > 0$), the transition state is stabilized according to the negative Φ_m values ($\Delta\Delta G_F^\ddagger < 0$). This behavior can be explained by the increase of the heterogeneity of the transition state as the case of Φ_T values of WT. The change in the barrier height by the mutation infers the involvement of the mutated residue in the transition state, which supports the extended transition state at high temperature.

For better interpretation on the molecular level, the free energy surfaces of mutants are calculated using the hybrid model ($Q_{EX+L+3D}$ in Eq. (8.11)). To introduce a destabilization at the charged sites, we include an electrostatic repulsion in the interaction energy term as described in Chapter 8,

$$F_{ji}' = \sum_{contacts} \varepsilon + \sum_{charged} \frac{e^2}{4\pi\epsilon_r\epsilon_0} \left(\frac{1}{r} - \frac{1}{l_c} \right) - T \sum_{k=i}^{i+j-1} \Delta s_k \quad (9.5)$$

$$w_{ji}' = \exp\left(-F_{ji}'/RT\right)$$

The effect of the mutation of Trp 45 is included by raising the interaction energy (ε) between residue 45 and its native contacts by 0.4 kcal/mol.¹⁸

Fig. 9-8 shows the calculated changes in the free energy surface by the T-jump. For mutant i , there is a big change in the transition state region compared to WT. Before the T-jump at $T_i = 58^\circ\text{C}$, the transition state is not the same as that of WT at $(N_F(\text{I-II}), N_F(\text{III-V})) = (17, 7)$. This region is significantly populated. This change pushes the transition state to $(N_F(\text{I-II}), N_F(\text{III-V})) = (13, 4)$ at which strands III to V are mostly destabilized and the β hairpin by strand I and II are partially denatured. On the other hand, as the temperature is raised, the transition state moves toward the folded side, $(N_F(\text{I-II}), N_F(\text{III-V})) = (17, 7)$. Therefore, a broad range of the structures between the two

regions will relax with the downhill unfolding manner. The relaxation of this broad distribution of structures will appear as a non-exponential form because the timescale of the relaxation would be dependent on the location on the free energy surface before the T-jump. For example, the unfolding from A to B involves the downhill unfolding of strands III to V that will occur on the molecular timescale of 1 μ s in WT experiment in Chapter 6 and 7. The relaxation from B to C will happen on the longer timescale because unfolding of strand I and II accompanies unfolding of the α helix as well. Because of this heterogeneity of the timescale, the unfolding part may be represented best by the global stretched relaxation. In addition, the shift of the barrier position to another transition state, $((N_F(I-II), N_F(III-V)) = (13, 11))$ is also observed as the case of WT, which explains the decreasing trend of the Φ_T values in Section 9.3.3.2.

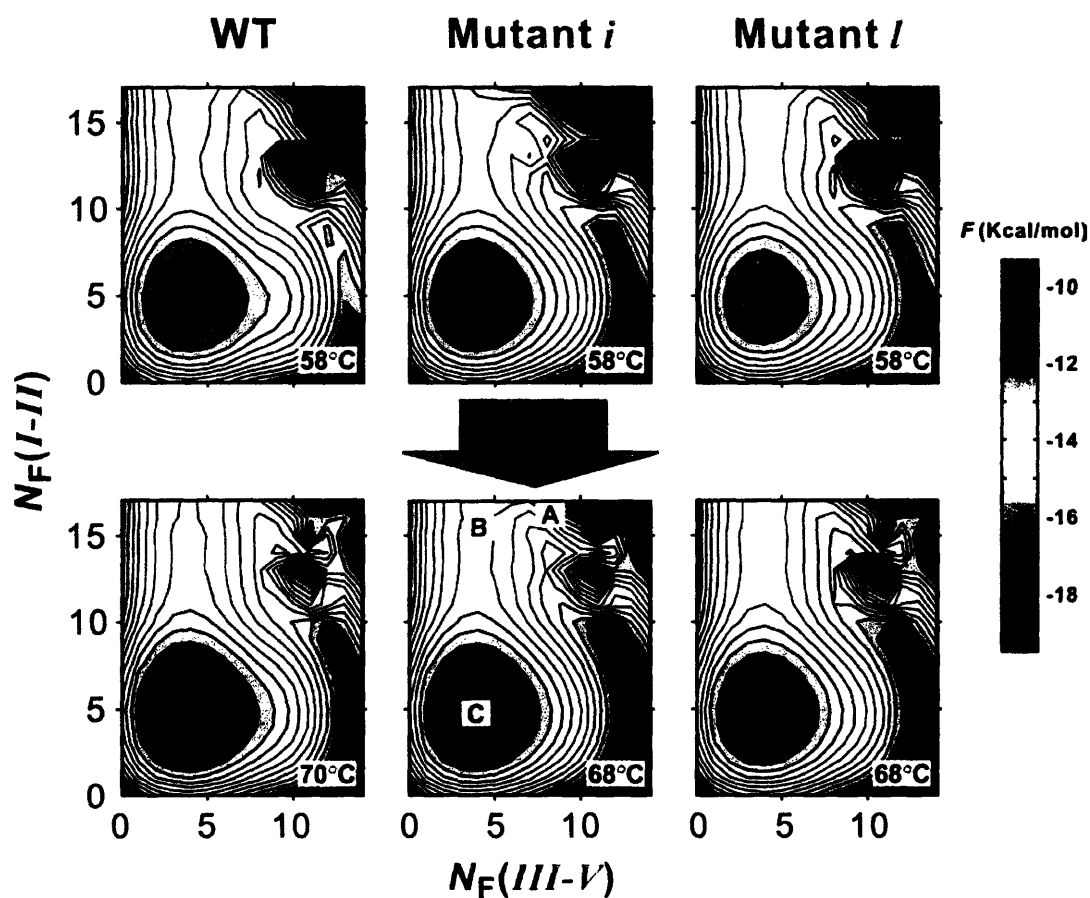


Figure 9-8. Comparison of the changes of the free energy surface of the wildtype and mutants by T-jump.

Compared to mutant *i*, the free energy of mutant *l* is almost same as that of WT. Actually, this is an expected result since the insertion of charges in the α helix will not induce local changes in the β -sheet stability. Also, the charge repulsion effect within the α helix is relatively small because of the small separation of 4 residues between the two charged residues. The two residues cannot be far apart from each other spatially even in the unfolded state. The 3D free energy surface calculation performed in Section 8.3.4 also supports that no partial unfolding of the α helix is present in the transition state. Then, where does the μ s unfolding relaxation come from? This will be discussed in the next section.

9.4.2. Fast unfolding phase of mutant *l*

A possible mechanism for the anti-correlated behavior of the timescales of the unfolding and refolding phases of mutant *l* can be proposed from the comparison with other mutants. First, mutant *g* does not show a fast μ s unfolding phase. This indicates that unfolding of strand III (or strand V as well), which is highly stabilized in mutant *g*, is necessary for the fast unfolding phase either in a partially unfolded or a fully unfolded form. Secondly, in mutant *i*, the amplitude of the fast phase is bigger than that of WT in spite of the stabilization of strand V by removing charged His 68. Therefore, the involvement of strand V in the fast unfolding phase would seem doubtful, unless we have overestimated the electrostatic repulsion effect of His 68. To connect structural changes to the time scale of the μ s unfolding, the 1 μ s barrierless unfolding of WT results from unfolding of the strand III, IV, and part of V. This is the transition from region A to B as labeled in Fig. 9-8, and explained in Chapter 6 and 7. The more heterogeneous unfolding of mutant *i* includes additional unfolding transitions such as from region B to C, which involves the unfolding of strands I and II and the α helix as discussed in the previous section. This additional process makes the observed unfolding timescale of mutant *i* increase to 100 – 300 μ s. However, the low barrier height of mutant *i* also induces a faster ms relaxation, the beginning of which merges with the tail of the μ s relaxation. This makes the apparent time scale of the μ s process longer in our global stretched exponential fit. Therefore, we estimate that the barrierless unfolding in the transition B to

C would occur on the timescale of 100 μs or less. This timescale is consistent with our previous measurement observed by the v_{\parallel} region as presented in Chapter 6. Unfortunately, the fast thermal aggregation of mutant *i* and *l* prevents the sufficient data collection, and the signal to noise ratio is not good enough for the analysis of the v_{\parallel} region where the signal intensity is low.

Based on these facts and the exclusion of unfolding through an orthogonal coordinate (the destabilized α helix) by free energy calculations as described in Section 8.3.4, the fast unfolding phase of mutant *l* in the tens of μs range should be related to the transition from B to C. For this process, there must be some population at B where a fair amount of residues in strand III to V are unfolded. In mutant *l*, only two mutated residues in this region are His 68 and Trp 45. In our calculation, mutation F45W destabilizes the protein by 0.4 kcal/mol as reported in the chemical denaturant experiment at room temperature¹⁸ but the stability at high temperature could be different. The side chain of this residue is partially exposed to the solvent. If the destabilization of F45W at room temperature results from the increase of the hydrophobicity of Trp, the protein will be less stable at higher temperature because the importance of the hydrophobic interaction increases with temperature. In this case, the free energy surface of mutant *l* becomes similar to that of mutant *i* and a small population fills the region B where strand III-V are fairly unfolded. After the μs process, ms unfolding will follow but this change is small and slow because of the relatively higher barrier compared to those of WT and mutant *i*.

9.5. Conclusion

The transition state of thermal unfolding of ubiquitin has been investigated in this chapter by site-specific mutations and temperature changes. Although most of the previous kinetics experiments, which have been performed at room temperature, limit the transition state to the α helix and β strand I and II, the thermal unfolding experiment of mutant *i* reveals that the transition state at the thermal melting temperature is extended more to the strand V, III, and IV. This transition state structure more closely resembles the folded state, which is consistent with the Hammond postulate. The involvement of each of the β strands has been discussed by comparing the relaxation time of the fast

component of mutants and WT and calculating free energy surface. We conclude that the barrierless unfolding process of strand III-V occurs on 1 μ s timescale while unfolding of the remaining hairpin consisting of strand I and II, which is also a downhill process, occurs on tens to 100 μ s timescale.

References

- (1) Matouscheck, A.; Jr, J. T. K.; Serrano, L.; Fersht, A. R. *Nature* **1989**, *340*, 122.
- (2) Fersht, A. R.; Sato, S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7976.
- (3) Pain, R. H. *Mechanisms of protein folding*, 2 ed.; Oxford University Press: Oxford, 2000.
- (4) Fersht, A. R.; Itzhaki, L. S.; ElMasry, N. F.; Matthews, J. M.; Otzen, D. E. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10426.
- (5) Krantz, B. A.; Sosnick, T. R. *Nat. Struct. Biol.* **2001**, *8*, 1042.
- (6) Krantz, B. A.; Dothager, R. S.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *337*, 463.
- (7) Sosnick, T. R.; Dothager, R. S.; Krantz, B. A. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17377.
- (8) Went, H. M.; Jackson, S. E. *Protein Eng. Des. Sel.* **2005**, *18*, 229.
- (9) Cox, J. P. L.; Evans, P. A.; Packman, L. C.; Williams, D. H.; Woolfson, D. N. *J. Mol. Biol.* **1993**, *234*, 483.
- (10) Jourdan, M.; Searle, M. S. *Biochemistry* **2000**, *39*, 12355.
- (11) Harding, M. M.; Williams, D. H.; Woolfson, D. N. *Biochemistry* **1991**, *30*, 3120.
- (12) Stockman, B. J.; Euvrard, A.; Scahill, T. A. *J. Biomol. NMR* **1993**, *3*, 285.
- (13) Cordier, F.; Grzesiek, S. *Biochemistry* **2004**, *43*, 11295.
- (14) Marianayagam, N. J.; Jackson, S. E. *Biophys. Chem.* **2004**, *111*, 159.
- (15) Dastidar, S. G.; Mukhopadhyay, C. *Phys. Rev. E* **2005**, *72*, 051928.
- (16) Fersht, A. R. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17327.
- (17) Shandiz, A.; Sosnick, T. R. *To be published* **2007**,
- (18) Khorasanizadeh, S.; Peters, I. D.; Butt, T. R.; Roder, H. *Biochemistry* **1993**, *32*, 7054.
- (19) Yin, L.; Krantz, B.; Russell, N. S.; Deshpande, S.; Wilkinson, K. D. *Biochemistry* **2000**, *39*, 10001.
- (20) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334.

Chapter 10

Transient 2D IR spectroscopy of ubiquitin unfolding

10.1. Introduction

In Chapter 6 through 9, the temperature-dependent unfolding dynamics of ubiquitin were probed with dispersed vibrational echo (DVE) spectroscopy after a T-jump. Although interesting unfolding dynamics related to the transition state are captured by DVE spectroscopy, there are two main ambiguities associated with the technique. First, a DVE spectrum is a projection of a two-dimensional infrared (2D IR) spectrum onto the ω_3 axis, and it cannot discern frequency dependent changes along the ω_1 axis. In other words, the time evolution of diagonal and cross peaks cannot be uniquely disentangled. Second, as an absolute value measurement, DVE cannot separate changes in the fundamental ($\nu = 0 \rightarrow 1$) and overtone ($\nu = 1 \rightarrow 2$) transitions. Even though the anharmonicity of the individual amide I vibrations is measured to be 16 cm^{-1} , the splitting between positive and negative peaks in the protein spectrum depends on couplings between individual oscillators and may change with the protein structure. Therefore, transient 2D IR spectroscopy, which is the most informative tool among the various

nonlinear IR spectroscopic techniques, is needed to remove ambiguities in DVE spectroscopy.

At equilibrium, 2D IR spectroscopy has been applied to watch vibrational dynamics in peptides,¹⁻⁸ proteins,⁹⁻¹¹ and DNA molecules.^{12,13} The unique sensitivity of amide I 2D IR spectroscopy to the two vibrations of the anti-parallel β -sheet structure (ν_{\perp} , ν_{\parallel}),^{14,15} has been utilized to monitor the thermal unfolding of RNase A,⁹ ubiquitin,¹⁰ and the trpzip2 peptide⁷. However, non-equilibrium, transient 2D IR spectroscopy is rare mainly due to technical difficulties. Only a couple of experiments have been reported using a UV pulse initiation.^{16,17} The realization of transient 2D IR spectroscopy after a T-jump is even more difficult because of the low repetition rate of the T-jump laser (20 Hz), which is 1/50 of that of the probe laser. Data collection at this significantly lower repetition rate requires accurate synchronization between the T-jump and probe laser and improvement of the signal-to-noise ratio. In addition, Fourier transform 2D IR spectroscopy is an interferometric technique and is extremely sensitive to the small refractive index change induced in the sample by a T-jump. Therefore, it is important to keep the relative phase of the third-order signal and the local oscillator field constant.

In this chapter, the realization of T-jump 2D IR spectroscopy is presented. As presented in Chapter 4, improvement of the signal-to-noise ratio using the balanced detection scheme and accurate control of the delay between the T-jump laser and the probe laser make it possible to take the transient 2D IR spectrum within a reasonable time (two hours). Transient unfolding of wildtype ubiquitin is monitored and the results are compared with those of the DVE experiment presented in Chapter 7. Transient DVE spectra are reconstructed by projecting 2D IR spectra, and changes in diagonal peak and cross peak regions are analyzed separately. Also, as a new observable, transient changes in the homogeneous linewidth are observed during unfolding and compared with other results.

10.2. Experimental

Although experimental details can be found in Chapter 4, the alignment of the infrared pulses through the sample and the double difference scheme used to obtain transient 2D IR spectra are summarized briefly in this section.

Since 2D IR spectroscopy is a heterodyne measurement, the stability of the relative phase between the third-order signal and the local oscillator (LO) is important. In the equilibrium measurements presented in Chapter 5 and 6, the LO does not pass through the sample. However, in the T-jump experiment, the refractive index of the sample varies with the temperature profile, which causes a drift of the relative phase between the LO and the third-order signal. Since the amplitude of the transient change is a small fraction of the equilibrium spectrum, a small phase shift mixes real and imaginary components and distorts the difference spectrum. Therefore, the LO is also sent with the other beams through the same T-jump point in the sample, to keep the relative phase between the LO and the third-order signal constant. Two alignment geometries of the LO and the three beams (a, b, and c) that generate the third-order signal are illustrated in Fig. 10-1a. In geometry A, the LO is focused on the same spot with the other three beams and the phase changes of all four beams are identical. In alignment geometry B, the LO is vertically displaced by 100 μm from the focal spot of the other three beams. Even in this case, the temperature at the two spots is similar because the focal spot size of the T-jump laser (500 μm in diam.) is much bigger than that of the 6 μm beams (100 μm in diam.). Geometry B was our first geometry designed to avoid an unwanted pump-probe signal that may be generated in geometry A with one of the three beams as a pump and the LO as a probe. This pump-probe signal would interfere with the heterodyned signal. Even in geometry A, however, it was discovered that a slight difference in the third-order signal and the LO path lengths between the focal point on the sample to the detector induces a time delay of 35 ps as mentioned in Chapter 4. In other words, when the LO and the third-order signal are temporally overlapped at the detector, there is a time delay of 35 ps between the LO and the other three beams at the focal spot on the sample. At this delay, a pump-probe signal is negligible and geometry A also works. In this chapter, results with both geometries are presented, but geometry A is preferred because the alignment is more straightforward.

Both the equilibrium and transient 2D IR spectra were obtained by undersampling with a τ_1 step of 14 fs from 0 – 2.1 ps and 0 – 1.2 ps for rephasing and non-rephasing spectra, respectively. Also, balanced detection is performed and the detected signals on the upper and lower stripes of a dual stripe array are subtracted to remove the LO intensity and homodyne signal to obtain heterodyned components, which reduces baseline fluctuations.

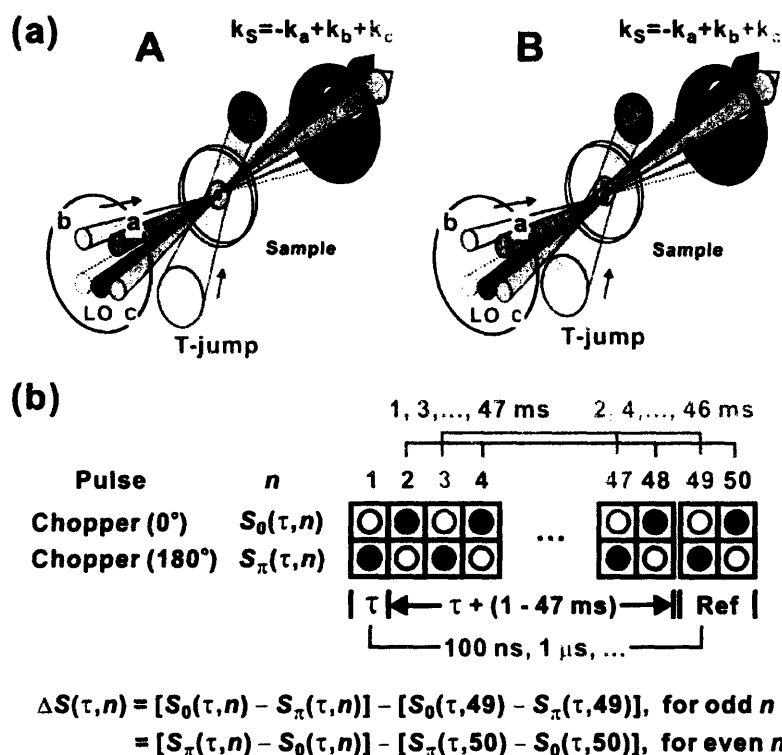


Figure 10-1. Beam alignment and data processing for transient 2D IR spectroscopy. (a) Alignment of beams in the sample region. In geometry A (Left), the LO hits the same spot with the other three beams, (a, b, and c) generating the third-order signal. In geometry B (Right), the LO is vertically displaced from the other three beams by 100 μ m. (b) Double difference scheme for 2D IR difference spectra. Difference spectra are obtained by subtracting the reference spectra by the 49th and 50th pulses from the odd- and even-numbered raw spectra. Odd-numbered spectra give difference spectra at τ (red), 2, 4, ..., 46 ms (orange) and even-numbered spectra give difference spectra at 1, 3, ..., 47 ms (green).

Transient 2D IR spectra are obtained by a double difference scheme

$$\begin{aligned}\Delta S(\tau, n) &= [S_0(\tau, n) - S_\pi(\tau, n)] - [S_0(\tau, 49) - S_\pi(\tau, 49)], \quad \text{for odd } n \\ &= [S_\pi(\tau, n) - S_0(\tau, n)] - [S_\pi(\tau, 50) - S_0(\tau, 50)], \quad \text{for even } n\end{aligned} \quad (10.1)$$

As shown in Fig. 10-1b, the 50 sets of infrared (6 μm) pulses after a T-jump pulse constitute one delay set which spans from the first variable delay τ ($< 1\text{ms}$) between a T-jump pulse and the first probe pulse to $\tau + 47\text{ms}$ ($\approx 47\text{ms}$. τ is much shorter than several ms delays). A delay between the T-jump pulse and the first infrared pulse, τ , is controlled electronically using a delay generator. Two reference spectra are obtained by the last two pulses (49th and 50th). Since beam c is chopped at 1 kHz, the combination of the third-order signal and the LO are detected at every other pulse as marked with an open circle. A closed circle represents just the LO. A raw 2D IR spectrum of each delay is constructed by subtracting the LO signal (closed circle) from the sum of the third-order signal and the LO (open circle) in the same column (same delay), each of which is collected at a different cycle of the chopper phase, 0° (S_0) or 180° (S_π). A 2D IR difference spectrum is obtained by subtracting the reference spectrum from the raw 2D IR spectrum at each delay. Here, two reference spectra by the 49th and 50th pulses are needed for odd- and even-numbered (n) raw spectra, respectively. Since the chopper phase is cycled with a period of 2.5 sec and 5 sec for the rephasing and non-rephasing spectra, respectively, a drift in the pulse energy of the 6 μm beams during this period makes the resulting difference spectrum noisy when an even-numbered spectrum is subtracted from an odd-numbered spectrum (or vice versa). Therefore, difference spectra at $\tau, 2, 4, \dots$, and 46 ms (odd n) are obtained using a reference spectrum by the 49th pulse. Difference spectra at the other delays (even n) are constructed using the reference spectrum by the 50th pulse.

The concentration of the ubiquitin sample is 30 mg/mL in the 0.35% (w/w) DCl solution.

10.3. Results and Discussion

10.3.1. Equilibrium differences

Fig. 10-2 shows equilibrium 2D IR spectra taken at 63°C and 72°C, which are the initial (T_i) and final (T_f) temperatures, respectively, of the T-jump experiment for ubiquitin unfolding in this chapter. Also shown are their differences. Because of inhomogeneous broadening, the two transitions of the β sheet (ν_{\perp} , ν_{\parallel}) are not clearly resolved but appear as a broad diagonal peak both for the fundamental transition (positive) and for the overtone transition (negative). Cross peaks in the off-diagonal region are small in the parallel polarization geometry (ZZZZ), but are enhanced and form a cross-peak ridge in the upper left corner in the perpendicular polarization geometry (ZZYY). Loss of negative intensity (a positive change) in the lower right corner also indicates the presence of the positive cross peak in this region.

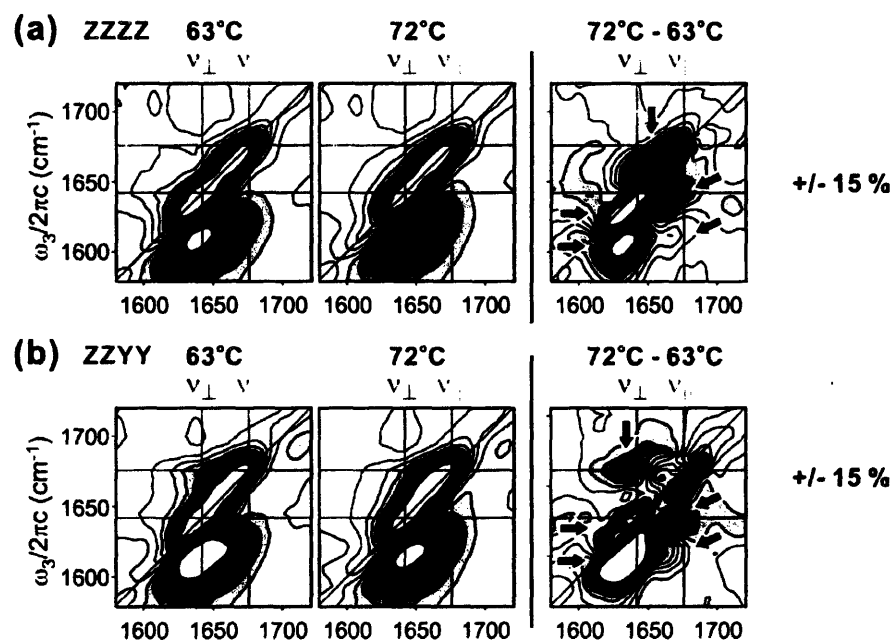


Figure 10-2. Equilibrium thermal unfolding of ubiquitin monitored by 2D IR spectroscopy in (a) parallel (ZZZZ) and (b) perpendicular (ZZYY) polarization geometries. Spectra are normalized to the maximum of the 63°C spectrum. 21 contours are plotted for +/- 60% of the spectra at 63°C and 72°C and for +/- 15% of the difference spectra. Positive and negative peaks are indicated by red and blue colors. In the difference spectra, red and blue arrows indicate the diagonal peaks in the red and blue side in the ω_3 axis, respectively. Green arrows represent cross peaks.

As the temperature is raised, diagonal intensity on the red side of the ν_{\perp} transition shifts to the blue side, which results in the formation of a doublet of negative and positive peaks for the fundamental and overtone transitions, respectively (marked with red arrows). This decrease in the diagonal intensity and blue shift indicates the disruption of the β sheet structure. Concomitantly, in the off-diagonal region (lower right corner), the negative and positive doublet marked with green arrows appear indicating the depletion of the cross-peak intensity. In the blue side of the ZZZZ difference spectrum, a positive diagonal peak is present because of an increase of the intensity of the random coil region (blue arrow). On the other hand, in the ZZYY difference spectrum, this region is dominated by the depletion of the cross peak intensity (green arrow) resulting from the enhancement of the cross peak intensity in this polarization geometry. The spectral features of these two difference spectra help in the interpretation of the transient difference spectra in the T-jump experiments.

10.3.2. Ultrafast responses

As observed in the transient DVE experiment, there are ultrafast responses that arise from solvated regions of the protein. (see Section 7.3.2) The ultrafast responses in the DVE spectra of NMA ($T_i = 58^{\circ}\text{C}$) and ubiquitin ($T_i = 25^{\circ}\text{C}$) induced by a T-jump are characterized by two features: the increase of the signal due to increased solvent transmission at $6\ \mu\text{m}$ and the depletion of the intensity on the red side of the spectrum ($\sim 1600\ \text{cm}^{-1}$) due to decreased hydrogen bonding with the solvent. 2D IR versions of these responses during the T-jump for NMA ($58^{\circ}\text{C} \rightarrow 68^{\circ}\text{C}$) and ubiquitin ($25^{\circ}\text{C} \rightarrow 35^{\circ}\text{C}$) are shown in Fig. 10-3.

The reference spectrum of NMA consists of two peaks from the 0 – 1 (positive) and 1 – 2 (negative) transitions of a single oscillator. In the transient 2D IR difference spectrum at 200 ns, both the positive and negative diagonal peaks become stronger because of the transmission increase of the probe beams and the LO. (Since the DVE spectrum is related to the squared absolute value of the projection of the complex 2D IR spectrum onto the ω_3 axis, increases in the magnitudes of both the positive and negative peaks in the 2D IR spectrum result in positive changes of the DVE signal.) At the same

time, both peaks are blue shifted because of the weakening of solvent hydrogen bonds. The loss of intensity of the diagonal region on the red side (marked with a red ellipse) corresponds to the depletion of the DVE intensity around the 1600 cm^{-1} region. The spectral shape of the difference spectrum at 1 ms is almost the same as that of the 200 ns spectrum except for the decrease in the magnitude of the signal. This lack of changes in the spectral shape is expected because no structural change is possible in NMA. The change is mostly due to the modulation of the transmission of the solvent by a temperature re-equilibration.

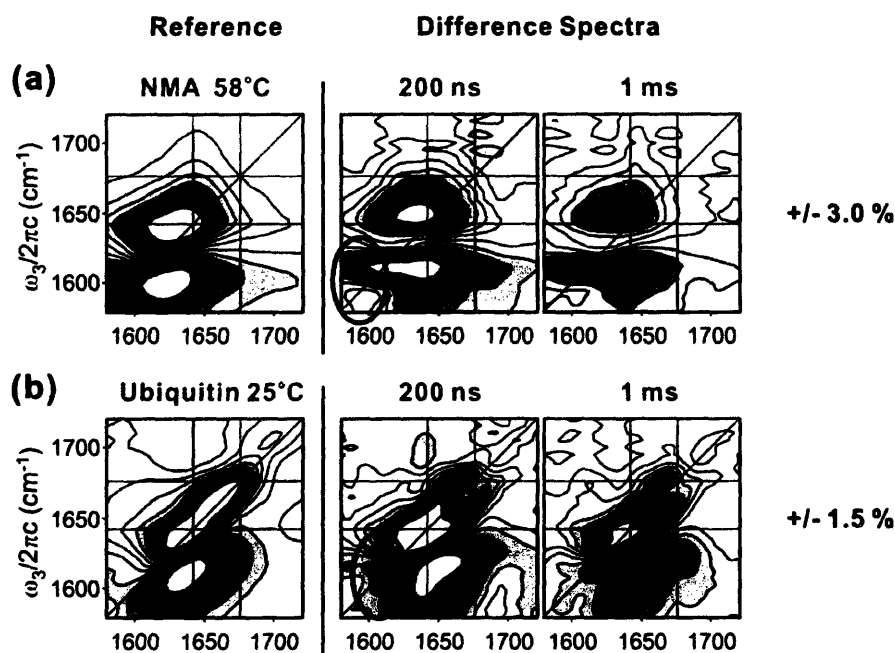


Figure 10-3. Transient 2D IR spectra of (a) NMA ($58^{\circ}\text{C} \rightarrow 68^{\circ}\text{C}$) and (b) ubiquitin ($25^{\circ}\text{C} \rightarrow 35^{\circ}\text{C}$). Left panels are the reference 2D IR spectra of NMA (58°C) and ubiquitin (25°C) in the parallel polarization geometry (ZZZZ). A reference spectrum is the average of the two spectra obtained from the 49th and 50th pulses (Fig. 10-1b). 21 contours are plotted in $\pm 60\%$ of the maximum. Transient 2D IR difference spectra ($\Delta S(\tau)$) are shown in the right panels. $\pm 3.0\%$ (NMA) and $\pm 1.5\%$ (ubiquitin) of the maximum of the reference spectra are plotted with 21 contours. Depletion of the signal on the red side of the diagonal regions at 200 ns is marked with red ellipses for both proteins. The guidelines parallel to the two frequency axes (ω_1 and ω_3) mark the two resonances of the β -sheet in ubiquitin. The same lines are used for NMA for comparison. Data is collected in geometry A shown in Fig. 10-1a.

Similar changes are observed in the transient difference spectra of ubiquitin in Fig. 10-3b. At this low temperature ($T_i = 25^\circ\text{C}$ to $T_f = 35^\circ\text{C}$), significant structural changes are not expected. Therefore, the increase in amplitude both in the positive and negative peaks and the depletion of the signal in the diagonal region on the red side result from the transmission increase and hydrogen bond breaking, respectively. The similar spectral shapes of the two transient difference spectra at 200 ns and 1 ms also mirror NMA. However, there is a slight difference between the reference spectrum and transient difference spectra. The anti-diagonal width of the positive peak of transient spectra is broader than that of the reference spectrum. The anti-diagonal width of a 2D IR spectrum represents the homogeneous linewidth.¹⁸ Therefore, this change indicates the increase in homogeneous broadening with temperature. Quantitative analysis of this effect follows in Section 10.3.6.

10.3.3. Transient thermal unfolding of ubiquitin

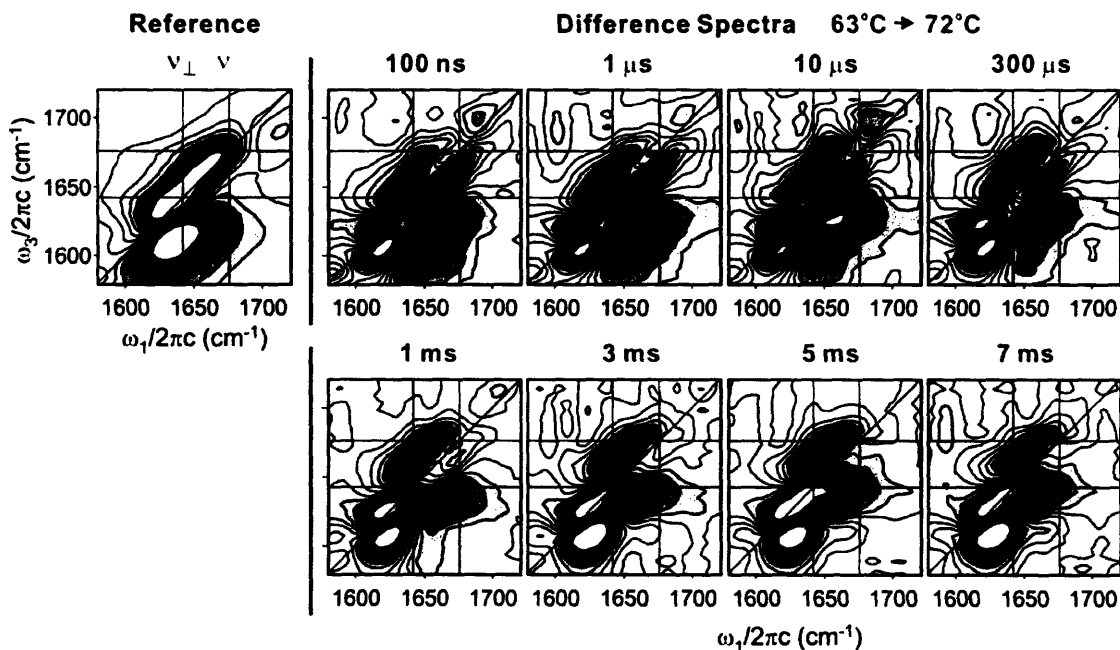


Figure 10-4. Transient 2D IR spectra after initiated by a T-jump from 63°C to 72°C . For the reference spectrum (ZZZZ), $\pm 60\%$ of the maximum at 63°C are plotted with 21 contours. Transient difference spectra are plotted as a function of delay τ . 21 contours are plotted in $\pm 1.5\%$ of the maximum of the reference spectrum. The spectra from 1 – 7 ms are obtained from the same data set as the 100 ns spectrum. Data is collected in geometry B shown in Fig. 10-1a.

Thermal unfolding of ubiquitin is monitored by transient 2D IR spectroscopy (ZZZZ) following a T-jump from 63°C to 72°C. 2D IR difference spectra are plotted in Fig. 10-4 as a function of delay τ following the T-jump. Also, frequency-dependent changes are shown by plotting slices for constant ω_1 in Fig. 10-5. Initial changes at $\tau = 100$ ns are similar to those observed for the $T_i = 25^\circ\text{C}$ T-jump in Fig. 10-3b. The difference spectrum shows an increase in the amplitude of the broad positive and negative diagonal peaks and depletion on the red side of the diagonal region. Also, there is increased broadening in the anti-diagonal width. The similarity of this to NMA and $T_i = 25^\circ\text{C}$ ubiquitin spectra indicates that the changes on this time scale result from residues of the solvent exposed regions. However, time-dependent spectral changes observed in the subsequent spectra are different due to thermal unfolding of the protein. This is apparent from the similarity between the transient difference spectrum at 7 ms and the equilibrium difference spectrum in Fig. 10-2a.

The transient spectral changes can be divided into three different frequency blocks. Changes in the red and blue sides of the ω_3 axis are separated as DVE results, and the red side is further divided into diagonal and off-diagonal regions to observe the behaviors of diagonal and cross peaks separately. First, in the low-frequency diagonal region (lower left), the amplitude of the negative/positive doublet, which is initially caused by the disruption of the hydrogen bonding with the solvent, rises and shifts to the blue along the diagonal axis. The blue shift of this ν_{\perp} region indicates disruption of the β sheet. The depletion and blue shift of this region are captured by vertical slices through the transient 2D IR spectra at $\omega_1 = 1620$ and 1642 cm^{-1} . (Fig. 10-5a and b, respectively) The positions of the positive and negative peaks (diagonal peaks) of the slice at $\omega_1 = 1620\text{ cm}^{-1}$ (red arrows) do not change along the ω_3 axis but the amplitude is slightly increasing. This change is represented by a temporal profile at $(\omega_1, \omega_3) = (1620, 1607)$ (Fig. 10-5d) showing an instantaneous change at 100 ns which is followed by a small increase until 3 ms and decrease because of the temperature re-equilibration.

The slice at $\omega_1 = 1642\text{ cm}^{-1}$, which is the peak frequency of the ν_{\perp} transition in the FTIR spectrum, shows a clear blue shift in the 2D spectrum. The difference signal in

the diagonal region, $(\omega_1, \omega_3) = (1642, 1639)$ is positive initially but becomes negative due to the blue-shift of the spectrum as unfolding of the β sheet proceeds. As expected, the temporal profile of this region (Fig. 10-5e) is similar to the relaxation of the ν_{\perp} region of DVE results in Chapter 7. A small μs relaxation is followed by a larger ms relaxation and subsequent refolding due to the temperature re-equilibration. Therefore, the response in this region has two components. Responses at $\omega_1 = 1620 \text{ cm}^{-1}$ due to hydrogen bond disruption are instantaneous but time-independent afterwards, whereas those at $\omega_1 = 1642 \text{ cm}^{-1}$ show big changes during unfolding of the β sheet. These two different responses are related to the two negative diagonal peaks on the red side of the equilibrium difference spectrum in Fig. 10-2, which are hardly resolvable in DVE spectroscopy.

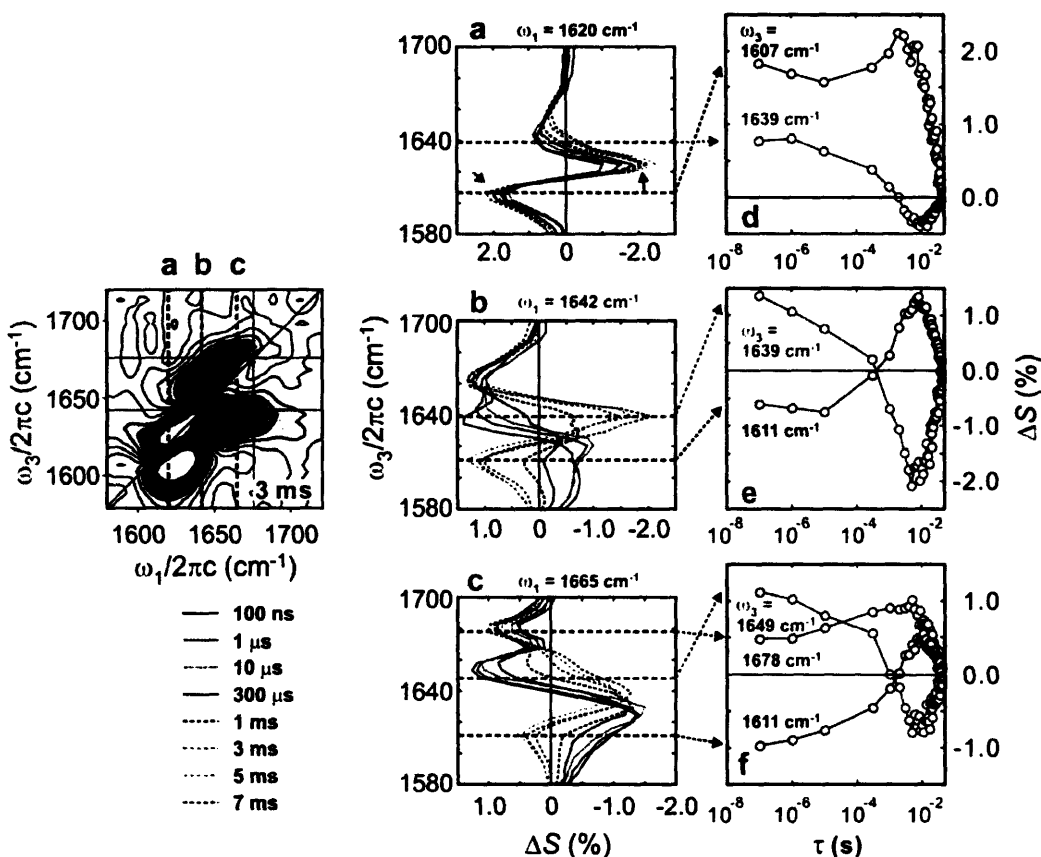


Figure 10-5. Transient changes monitored by slices at (a) $\omega_1 = 1620 \text{ cm}^{-1}$, (b) $\omega_1 = 1642 \text{ cm}^{-1}$, and (c) $\omega_1 = 1665 \text{ cm}^{-1}$. (d) – (f) Semi-log relaxation profiles are constructed by subsequent ω_3 slices.

The disruption of the β sheet can also be monitored by the cross peak intensity. Initially, the off-diagonal region for $\omega_1 > \omega_3$ (lower right) is dominated by a negative peak originating in the D₂O transmission increase. However, with increased delay, the negative and positive cross peak doublet is formed and pushes the negative diagonal peak to the blue. The formation of cross peaks of opposite signs indicates that the decrease in the cross peak intensity and disruption of the vibrational couplings are due to unfolding of the β sheet. The slice at $\omega_1 = 1665 \text{ cm}^{-1}$ shows a blue shift in this region. However, these cross peaks behave differently from the diagonal peaks. The relaxation profiles of the two cross-peak regions at $(\omega_1, \omega_3) = (1665, 1611)$ and $(\omega_1, \omega_3) = (1665, 1649)$ mirror each other, except for the sign. The both have μs and ms components as does the diagonal region of the ν_{\perp} mode, $(\omega_1, \omega_3) = (1642, 1639)$. This mirror image implies that changes in the positive and negative cross peaks are correlated. On the other hand, the responses of the diagonal peaks in the red side at $(\omega_1, \omega_3) = (1620, 1607)$ and $(\omega_1, \omega_3) = (1642, 1611)$, which are the responses of the solvent exposed regions, do not have the μs component. These differences in profiles of diagonal and off-diagonal regions imply that changes to the vibrational dynamics of the solvent exposed residues are not strongly correlated with those of the β sheet.

For high detection frequencies (upper part), the changes are less dramatic than the red side of the spectrum. The transmission increase induces a positive change in this region and the increase of the random coil component by thermal unfolding also appears as a positive change. As a result, a positive lobe is always present in this region. However, changes in the anti-diagonal width of this region provide similar information as that of the red side of the spectrum and this will be discussed in Section 10.3.6.

10.3.4. Comparison with DVE results

As shown in Chapter 3, a DVE spectrum is mathematically identical to the squared absolute value of the projection of a complex 2D IR spectrum onto the ω_3 axis as

$$\tilde{\mathbf{S}}_{\text{DVE}}(\tau_1 = 0, \tau_2, \omega_3) \propto \left| \int_{-\infty}^{\infty} \tilde{\mathbf{S}}'(\omega_1, \tau_2, \omega_3) + i\tilde{\mathbf{S}}''(\omega_1, \tau_2, \omega_3) d\omega_1 \right|^2. \quad (10.2)$$

Therefore, the result of the T-jump DVE experiment in Chapter 7 can be reproduced from 2D IR data in this chapter by integrating them over the ω_1 axis. Fig. 10-6a and b show the relative changes in the measured and reconstructed DVE spectra. Although the spectral shape and the amplitude of changes are somewhat different due to the different spectral shape of the femtosecond infrared pulses, their time-evolutions describe the same dynamics. At 100 ns, both spectra show depletion of the intensity at about 1600 cm^{-1} . Also, a decrease and blue shift in the ν_{\perp} region due to the unfolding of the protein occurs on the μs to ms time scale.

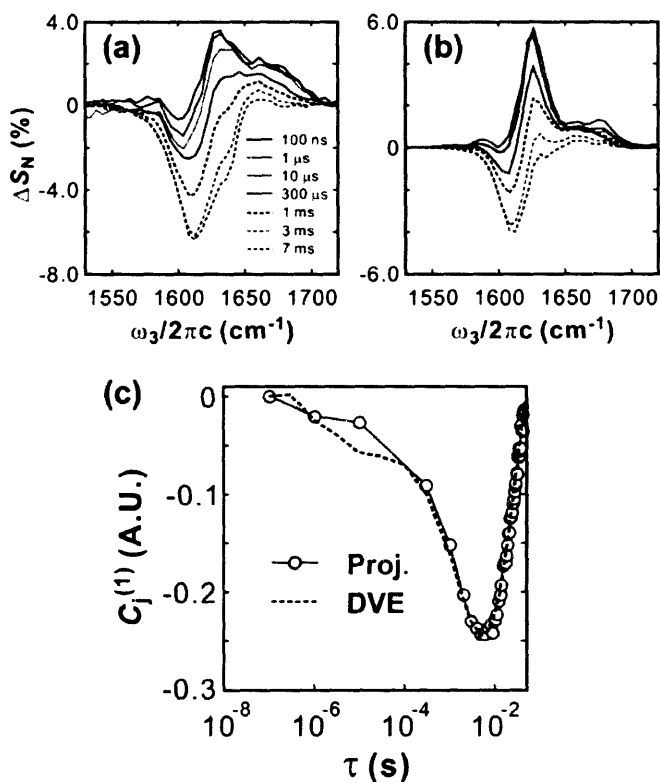


Figure 10-6. Comparison of DVE spectra measured and reconstructed from 2D IR spectra. (a) Plotted are DVE difference spectra obtained from the T-jump at $T_i = 63^\circ\text{C}$ presented in Chapter 7. Each spectrum is divided by the maximum of the reference DVE spectrum. (b) DVE difference spectra are constructed from the absolute value square of the projection of complex 2D IR spectra onto ω_3 axis. (c) Relaxation profiles constructed by the first SVD component of ν_{\perp} region ($1577 - 1651\text{ cm}^{-1}$) are compared.

For better comparison, we applied SVD to the v_{\perp} region after correcting for the temperature relaxation effect (Eq. (7.5) in Section 7.3.4),

$$\Delta S_{2D}'(\tau) = \Delta S_{2D}(\tau) - \Delta S_{2D}(\tau = 100 \text{ ns}) \cdot \exp[-(\tau/\tau_r)^\beta], \quad (10.3)$$

where $\Delta S_{2D}(\tau)$ is the normalized 2D IR difference spectrum at delay τ . Temperature relaxation is given by a stretched exponential with $\beta = 0.68$ and $\tau_r = 2.7$ ms. The relaxation profiles are obtained from coefficients of the first SVD component and compared in Fig. 10-6c. As expected, the two relaxation curves match each other well. This successful reproduction of DVE spectra from 2D IR spectra confirms that the two different measurements are probing the same dynamics, which provides a very important basis for the comparisons made in further analyses shown below.

10.3.5. Dynamics monitored by diagonal and off-diagonal regions

We have observed changes on two different timescales in the v_{\perp} region of DVE measurements in previous chapters and interpreted them as unfolding from two slightly different regions in the folded well on the free energy surface. The basic assumption in this interpretation is that both the μs and ms relaxations result from the same unfolding pathway in which unfolding of strands III-V precedes unfolding of strand I and II. In this case, the two relaxations have the same spectral characteristics. However, DVE measurements cannot resolve the dynamics appearing at two different regions in the ω_1 dimension due to the projection and interference along the ω_1 axis. For example, if we simply divide the ω_1 dimension into two regions, diagonal and off-diagonal, two possible cases can give the same result. The first case is that both relaxations in the diagonal and off-diagonal regions have two components, which corresponds to our assumption. The second case is that each region shows a single-timescale change but the timescales are different. These two distinct cases will appear as the same relaxation in the one-dimensional projection. To remove this ambiguity of DVE spectroscopy and confirm our assumption, we performed the SVD analysis separately for the diagonal and off-diagonal regions of the v_{\perp} ($1580 - 1654 \text{ cm}^{-1}$) region in the ω_3 axis.

The SVD results are summarized in Fig. 10-7. The first SVD component spectra are shown in the left panels. The spectrum A for all the ω_1 regions resembles the transient difference spectrum at 7 ms, which indicates the SVD analysis captures the spectral changes of unfolding. Also, the relaxation profile matches very well with that from the DVE measurement as shown in the right panel. Separate SVD analyses for the diagonal part and off-diagonal part result in the two spectra in the lower left panel (B and C). Surprisingly, the original component spectrum (A) can be reconstructed simply by patching these two component spectra (B and C) together. Both relaxation profiles also follow the DVE result as well.

The biggest spectral change in the diagonal region and the off-diagonal region is the blue shift of peaks due to the blue shift of the ν_{\perp} transition upon unfolding of the β sheet. The similarity of the relaxation profiles in the two separate analyses indicates that the blue shift of the diagonal peaks and the cross peaks occur coincidentally. This confirms our assumption of similar multi-component relaxation across the spectrum in interpreting the DVE measurement.

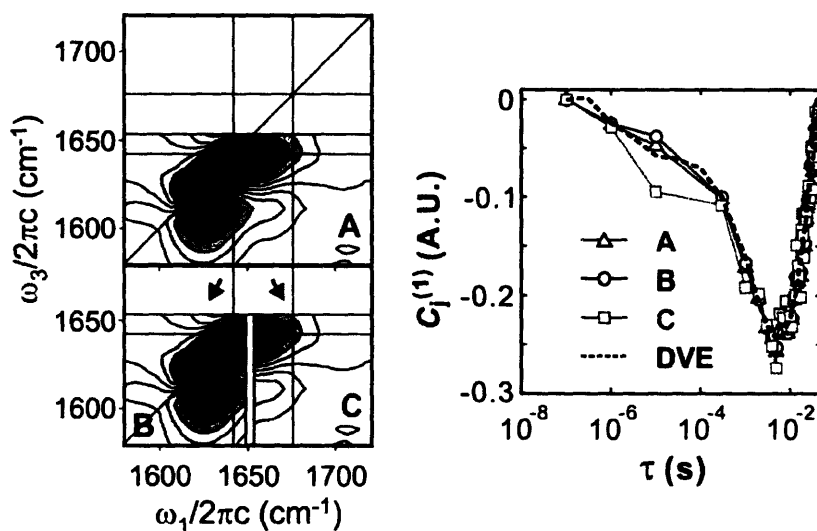


Figure 10-7. SVD analyses for different frequency blocks of the transient 2D IR difference spectra. The first SVD component spectra are shown in the left panels for the red side in the ω_3 axis ($1580 - 1654 \text{ cm}^{-1}$) and (A) $1500 - 1800 \text{ cm}^{-1}$, (B) $1500 - 1651 \text{ cm}^{-1}$, (C) $1651 - 1800 \text{ cm}^{-1}$ in the ω_1 axis, respectively. The first SVD component coefficients show temporal profiles. Also shown with dashed line is the profile from the DVE measurement.

10.3.6. Homogeneous broadening during thermal unfolding of protein

For proteins consisting of many amide I oscillators, the lineshape and linewidth of a linear infrared spectrum strongly depends on the static distribution of site energies of amide I oscillators. Because of this broad, inhomogeneous character, the homogeneous part of broadening cannot be independently observed in the spectrum. The homogeneous and inhomogeneous parts can be separated in the 2D lineshape.¹⁸ Since the diagonal and anti-diagonal frequency axes reflect the distribution of the amide I transitions and homogeneous broadening, a frequency-dependent homogeneous linewidth can be obtained from the anti-diagonal slice.

In addition to vibrational population relaxation, homogeneous line-broadening of amide I vibrations of proteins have two origins: solvent perturbations and structural fluctuations of the protein. The solvent perturbation induces site energy modulations while structural fluctuations induce variations in coupling strength between oscillators. The relaxation time of the frequency-frequency correlation function of NMA has been calculated to be about 1 ps, which is the timescale of the solvent modulation.¹⁹ The additional broadening by structural fluctuation gives a homogeneous linewidth of 10 – 15 cm^{-1} for proteins. This homogeneous broadening is relatively independent of the site frequency. However, it depends on the structural state of the protein because conformational fluctuations and solvent penetration contribute more in the unfolded state than in the folded state. Even in the folded state, it will differ by secondary structural motifs depending on their flexibility.

The relative change of the homogeneous linewidth ($\Delta\Gamma_r$) during a T-jump can be defined as

$$\Delta\Gamma_r(\tau, \omega_{AD}) = \frac{\Gamma(\tau, \omega_{AD}) - \Gamma(0, \omega_{AD})}{\Gamma(0, \omega_{AD})}, \quad (10.4)$$

where $\Gamma(\tau, \omega_{AD})$ is the full width at half maximum of the peak along the anti-diagonal slice ($\omega_{AD} = (\omega_1 + \omega_3)/2$) of a transient 2D IR spectrum at delay τ .

In Fig. 10-8, time-dependent changes of the homogeneous width for the anti-diagonal slices above the diagonal axis (positive peak) are plotted for the three frequency

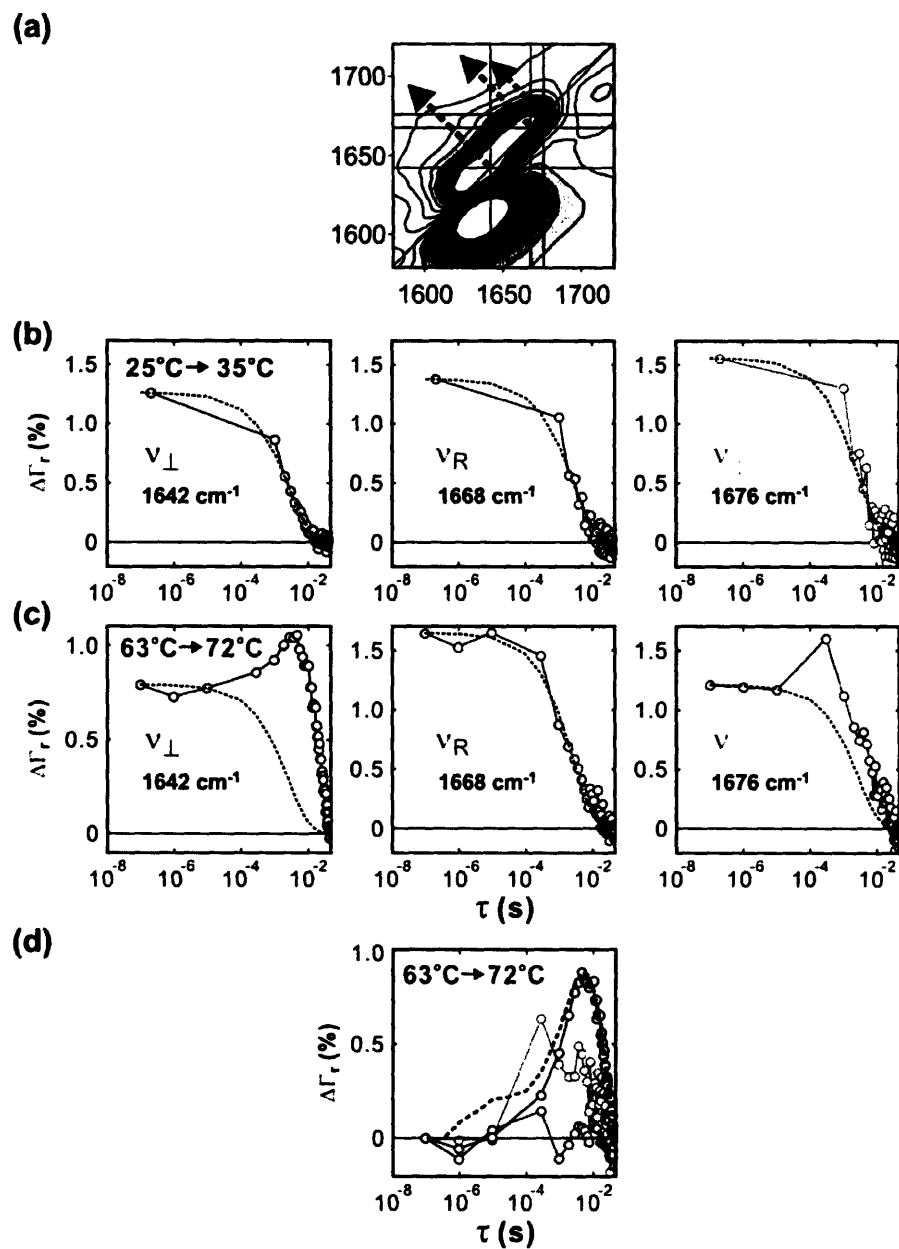


Figure 10-8. Changes in anti-diagonal width. (a) Anti-diagonal slices for three frequency components are marked with dashed arrows: ν_{\perp} mode (blue), random coil (red), and ν_{\parallel} mode (green). (b) The relative changes in the anti-diagonal width for a T-jump from $T_i = 25^{\circ}\text{C}$ are plotted with a temperature relaxation profile normalized to the initial change (dashed line). (c) Changes in anti-diagonal width are plotted for a T-jump from $T_i = 63^{\circ}\text{C}$. (d) Temperature corrected profiles of three frequency components. The DVE relaxation profile for the ν_{\perp} region is also shown with a dashed line (magenta) for comparison.

components: the ν_{\perp} mode ($\omega_{AD} = \omega_{\perp} = 1642 \text{ cm}^{-1}$), the random coil region ($\omega_{AD} = \omega_R = 1668 \text{ cm}^{-1}$) and the ν_{\parallel} mode ($\omega_{AD} = \omega_{\parallel} = 1676 \text{ cm}^{-1}$). For the T-jump from $T_i = 25^{\circ}\text{C}$ in Fig. 10-8(b), the changes of all three frequency components tracks the temperature relaxation profile. Since the magnitude of the change is very small ($< 2\%$), width changes can be assumed to be linear with temperature within the relatively small T-jump range of 10°C . Therefore, the increase in linewidth at low temperature is solely a result of the increased thermal fluctuations of the protein structure or solvent perturbations without a structural change.

However, in the T-jump from $T_i = 63^{\circ}\text{C}$, deviations from the temperature profile are observed depending on ω_{AD} . For the ν_{\perp} mode ($\omega_{AD} = 1642 \text{ cm}^{-1}$) in Fig. 10-8c, the initial increase of 0.8% does not relax with temperature but increases to about 1.1% until 6 ms. By subtracting the temperature profile, the increase due to unfolding of the protein can be obtained as shown in Fig. 10-8d. A stretched exponential ($= \exp[-(\tau/\tau_r)^{\beta}]$) fit for the increasing part gives a time scale of $\tau_r = 1.0 \text{ ms}$ with $\beta = 0.66$. Compared to the DVE relaxation curve of the ν_{\perp} region, the relaxation is slower on the μs timescale, but almost the same on the ms timescale.

The homogeneous width for the ν_{\parallel} mode ($\omega_{AD} = 1676 \text{ cm}^{-1}$) also shows an additional increase on top of the temperature profile after $10 \mu\text{s}$. Although the outlier at $300 \mu\text{s}$ is probably experimental error, the following data points confirm that the relaxation is similar to that of the ν_{\perp} mode.

The changes at $\omega_{AD} = 1668 \text{ cm}^{-1}$, which represents the random coil region, tracks the temperature relaxation as those at the low temperature T-jump do. The random coil region is already flexible and solvent exposed in the folded state and so additional broadening by unfolding is small. This conclusion is supported by the broad linewidth of this region in the equilibrium spectrum. The anti-diagonal widths for three frequency components of the equilibrium ZZZZ spectrum at $T = 63^{\circ}\text{C}$ in Fig. 10-2a are 10.6, 11.0, and 12.7 cm^{-1} , for the ν_{\perp} mode, the ν_{\parallel} mode, and the random coil region, respectively. The largest linewidth of the random coil region suggests its largest flexibility in structure before the T-jump. Also, their increases from 63°C to 72°C are 17%, 15%, and 7%,

respectively. The increase of the random coil region is the smallest as in the transient case. The overall increase of widths in the equilibrium measurement is much larger than those in the transient measurement because the transient unfolding is not completed (about 15% of the equilibrium change) due to the fast temperature re-equilibration.

10.3.7. Polarization effect

The relative intensity of the diagonal and cross peaks is determined by the polarization of the incident beams and the third-order signal. For ZZYY and ZZZZ polarization geometries, the ratio of the cross peak intensity to the diagonal peak intensity is always greater in the ZZYY spectrum than in the ZZZZ spectrum. This enhancement of the cross peaks is shown in the equilibrium difference spectra in Fig. 10-2. The decrease in the cross peak intensity observed in the upper left corner of the ZZYY difference spectrum and does not show up in the ZZZZ difference spectrum because the stronger positive diagonal peak obscures the small difference.

Similar features are observed in the transient spectra in Fig. 10-9. The behaviors of the red side on the ω_3 axis are similar for both polarization geometries. There is an initial depletion in the red side of the diagonal region at $\tau = 200$ ns, and these changes are followed by a blue shift of the diagonal peaks. Also, a cross peak doublet is formed in the off-diagonal region (lower right). The only difference is that the change of the cross peak intensity is larger than the change of the diagonal peak intensity, as expected. In addition, this enhancement is helpful to interpret the change in the other cross peak region. In the off-diagonal region in the upper left corner of the ZZYY difference spectra, the left side of the positive peak decreases with increasing delay time as marked with red arrows. As shown in the equilibrium difference spectrum (left panel), this location is the cross peak region. Therefore, the negative change of the cross peak intensity by unfolding of the β sheet interferes with the positive diagonal peak, which changes its spectral shape. However, because of the small magnitude of changes in the transient spectra, the cross peak change does not dominate the positive peak as it does in the equilibrium difference. The loss of the cross peak is illustrated more clearly in the plots of the slices at $\omega_1 = 1620$ cm^{-1} in Fig. 10-9c. ZZYY slices show a depletion at about $\omega_3 = 1675$ cm^{-1} after 1 ms.

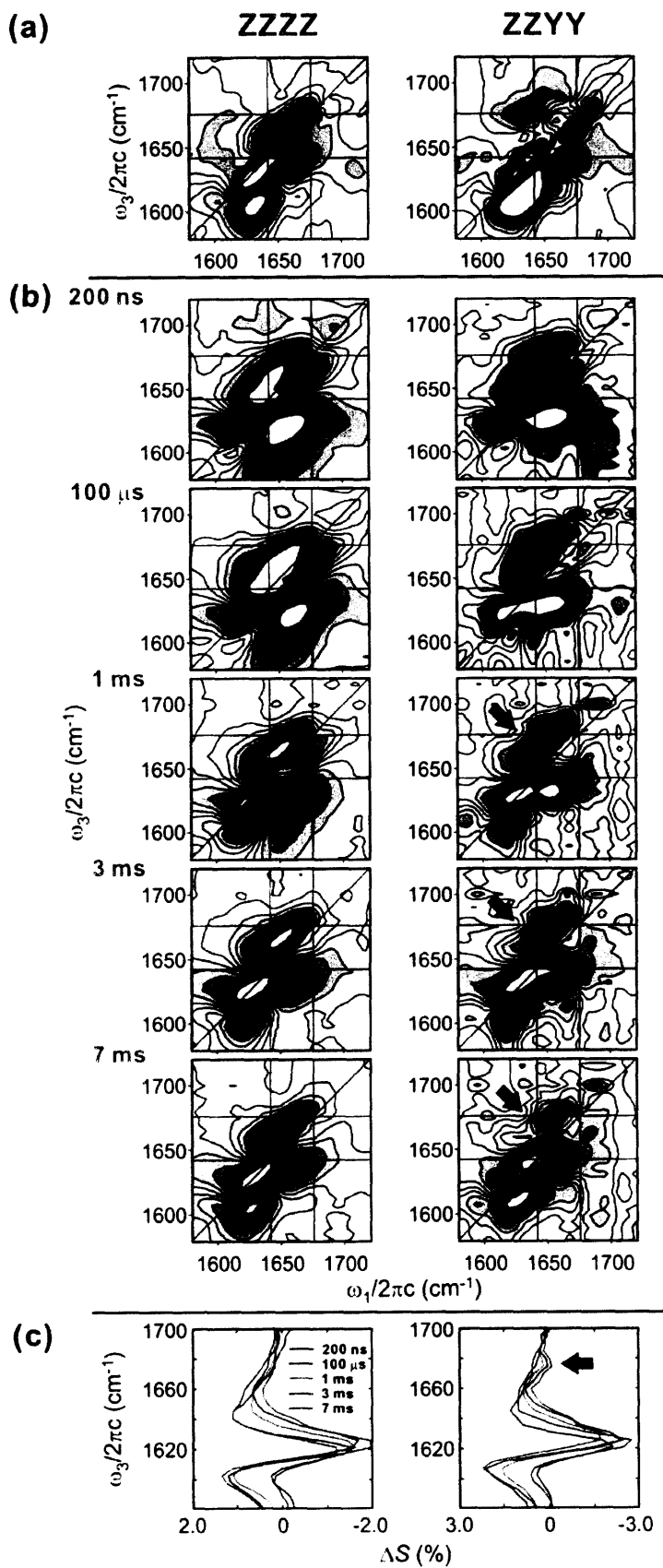


Figure 10-9. Comparison of different polarization geometries. (Left) Parallel (Right) Perpendicular. (a) Equilibrium difference spectra are the same as those in Fig. 10-2. (b) Transient difference spectra are plotted with 21 contours for $\pm 1.5\%$ (ZZZZ) and $\pm 2.0\%$ (ZZYY) of the maximum of the reference spectrum of each polarization geometry at $T_i = 63^\circ\text{C}$. Red arrows in ZZYY difference spectra indicate the decrease in signal for the cross peak region. (c) ω_1 slices are plotted at $\omega_1 = 1620 \text{ cm}^{-1}$. Red arrows indicate the depletion of the intensity in the off-diagonal region due to the loss of the cross peak.

10.4. Concluding Remarks

In this chapter, transient 2D IR spectroscopy is used to monitor transient thermal unfolding of ubiquitin. A great enhancement of the signal to noise ratio through the balanced detection scheme using a dual-stripe array detector and undersampling makes it possible to take a 2D IR surface within a reasonable time, which seemed previously impossible because of the low repetition rate of the T-jump laser.

By introducing a second dimension (ω_1) to the spectrum, the transient 2D IR measurements offer more degrees of freedom in the analysis of previous measurements and allows for new analyses. First, using the projection relationship with DVE spectroscopy, the transient changes in DVE measurement are reproduced from 2D IR spectra. The changes in relaxation profiles prove that both techniques probe the same dynamics. Second, transient 2D IR spectroscopy helps remove ambiguities in the one-dimensional DVE spectroscopy. Almost identical results in the diagonal and off-diagonal regions of the transient spectrum confirm that the relaxation in these two regions is coincident and that our previous interpretation based on this assumption is valid. Finally, as a new aspect, changes in the anti-diagonal width are investigated as a measure of the flexibility of protein. The anti-diagonal width of the 2D IR spectrum reflects the homogeneous broadening process. The width increases with temperature because structural fluctuations of the protein grow with the increased flexibility of the unfolded state. The frequency-dependent responses in the homogeneous width are interpreted as variation in the flexibility of different structural components.

Although various features in the transient 2D IR data help in interpreting our previous results and offer new analytical measures, the technique is still qualitative compared to DVE spectroscopy. However, with further improvements, it promises to become an independent measurement technique rather than a complementary technique.

References

- (1) Woutersen, S.; Mu, Y.; Stock, G.; Hamm, P. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11254.
- (2) Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *114*, 2727.
- (3) Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *115*, 7737.
- (4) Mukherjee, P.; Krummel, A. T.; Fulmer, E. C.; Kass, I.; Arkin, I. T.; Zanni, M. T. *J. Chem. Phys.* **2004**, *120*, 10215.
- (5) Fang, C.; Wang, J.; Kim, Y. S.; Charnley, A. K.; Barber-Armstrong, W.; Smith, A. B., III; Decatur, S. M.; Hochstrasser, R. M. *J. Phys. Chem. B* **2004**, *108*, 10415.
- (6) Kim, Y. S.; Wang, J.; Hochstrasser, R. M. *J. Phys. Chem. B* **2005**, *109*, 7511.
- (7) Smith, A. W.; Chung, H. S.; Ganim, Z.; Tokmakoff, A. *Journal of Physical Chemistry B* **2005**, *109*, 17025.
- (8) Wang, J. P.; Chen, J. X.; Hochstrasser, R. M. *J. Phys. Chem. B* **2006**, *110*, 7545.
- (9) Chung, H. S.; Khalil, M.; Tokmakoff, A. *J. Phys. Chem. B* **2004**, *108*, 15332.
- (10) Chung, H. S.; Khalil, M.; Smith, A. W.; Ganim, Z.; Tokmakoff, A. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 612.
- (11) Mukherjee, P.; Kass, I.; Arkin, I.; Zanni, M. T. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3528.
- (12) Krummel, A. T.; Mukherjee, P.; Zanni, M. T. *Journal of Physical Chemistry B* **2003**, *107*, 9165.
- (13) Krummel, A. T.; Zanni, M. T. *J. Phys. Chem. B* **2006**, *110*, 13991.
- (14) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *Journal of Chemical Physics* **2004**, *120*, 8201.
- (15) Demirdöven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. *J. Am. Chem. Soc.* **2004**, *126*, 7981.
- (16) Bredenbeck, J.; Helbing, J.; Hamm, P. *J. Am. Chem. Soc.* **2004**, *126*, 990.
- (17) Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. *Nature* **2006**, *444*, 469.
- (18) Tokmakoff, A. *J. Phys. Chem. A* **2000**, *104*, 4247.
- (19) DeCamp, M. F.; DeFlores, L. P.; McCracken, J. M.; Tokmakoff, A.; Kwac, K.; Cho, M. *J. Phys. Chem. B* **2005**, *109*, 11016.

Hoi Sung Chung

hsc00001@mit.edu

Business Address

77 Massachusetts Avenue, 6-030
Cambridge, MA 02139
(617) 253-7372 / 312-3416 (cell)

Home Address

70 Pacific Street, #212
Cambridge, MA 02139
(617) 312-3416

Education **Massachusetts Institute of Technology** Cambridge, MA
Ph.D. in Physical Chemistry, February 2007.
Thesis: Thermal unfolding dynamics of proteins probed by nonlinear infrared spectroscopy.

Seoul National University Seoul, South Korea
M.S. in Chemistry, August 2000. Thesis: Application of a molecular lens formed by a focused nonresonant infrared laser pulse to benzene and carbon disulfide molecular beams.

Seoul National University Seoul, South Korea
B.S. in Chemistry, February 1998.

Awards Lester Wolfe Graduate Fellowship (2003 – 2004, 2006); Young Scientist Grant, Optical Society of America (2004); Asan Undergraduate Scholarship (1996-1997)

Research **MIT, Department of Chemistry** Cambridge, MA
Experience Advisor: Andrei Tokmakoff (November 2001 – February 2007)

Seoul National University, Department of Chemistry Seoul, South Korea
Research associate (September 2000 – July 2001)
Supervisor: Doo Soo Chung
Graduate Student Research Assistant (M.S., March 1998 – August 2000)
Advisor: Doo Soo Chung

Teaching **MIT, Department of Chemistry** Cambridge, MA
Experience *Thermodynamics and Kinetics*. Teaching Assistant. (February 2002 – May 2002)
Principles of chemical science. Teaching Assistant (September 2001 – December 2001)
Seoul National University Seoul, South Korea
General Chemistry. Teaching Assistant (March 2000 – June 2000)
Instrumental Analysis Laboratory. Teaching Assistant (March 1998 – June 1998)

Selected
Publications “Visualization and characterization of the infrared active amide I vibrations of proteins,” **H. S. Chung** and A. Tokmakoff, *J. Phys. Chem. B*, **110**, 2888 (2006).

“Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin,” **H. S. Chung**, M. Khalil, A. W. Smith, Z. Ganim, and A. Tokmakoff, *Proc. Natl. Acad. Sci. USA*, **102**, 612 (2005).

“Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding,” **H. S. Chung**, M. Khalil, and A. Tokmakoff, *J. Phys. Chem. B*, **108**, 15332 (2004).

“Molecular Lens Applied to Benzene and Carbon Disulfide Molecular Beams,” **H. S. Chung**, B. S. Zhao, S. H. Lee, K. Cho, S. Hwang, S. H. Shim, S. M. Lim, W. K. Kang, and D. S. Chung, *J. Chem. Phys.* **114**, 8293 (2001).