Stony Brook University

Probing the Protein Folding Mechanism by Simulation of Dynamics and Nonlinear Infrared Spectroscopy

A Dissertation Presented

by

Christopher Nicholas Marai

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

May 2010

Stony Brook University

The Graduate School

Christopher Nicholas Marai

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Jin Wang – Dissertation Advisor
Professor, Department of Chemistry and Department of Physics

John Reinitz – Chairperson of the Defense
Professor, Applied Mathematics and Statistics

David F. Green
Assistant Professor, Department of Applied Mathematics and Statistics

Daniel P. Raleigh
Professor, Department of Chemistry

Robert C. Rizzo
Assistant Professor, Department of Applied Mathematics and Statistics

Dale G. Drueckhammer
Department of Chemistry, Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Probing the Protein Folding Mechanism by Simulation of Dynamics and Nonlinear Infrared Spectroscopy

by

Christopher Nicholas Marai

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

2010

This dissertation presents simulation based approaches towards the study of the dynamics and nonlinear spectroscopy of protein folding. The folding mechanisms of two model protein systems, the N-terminal domain of ribosomal protein L9 (NTL9) and the synthetic Beta3s mini-protein were investigated. All-atom molecular dynamics (MD) simulations and two-dimensional infrared spectroscopy (2DIR) computations were employed in the investigation of the folding mechanisms of these model systems. In this work, the folding mechanism and transition state ensemble (TSE) of the 56-residue N-terminal domain of L9 (NTL9) was probed. The TSE was identified from high temperature unfolding all-atom MD simulations in conjunction with experimentally determined Φ-values. The TSE ensemble of NTL9 was found to be largely native in composition, with a well defined secondary structure. In the progression to folding after crossing the TSE our data suggests that much of the drive towards the native state of NTL9 is spent optimizing electrostatic interactions between stable secondary structure elements. This work also proposes the use two-dimensional infrared spectroscopy (2DIR) to characterize the folding mechanism of the mini-protein Beta3s. In this study Beta3s was folded by MD simulation and intermediate conformational ensembles were identified. The two-dimensional infrared spectrum was calculated for the intermediate and native states of the mini-protein. A direct structure-spectra relationship was determined by analysis of conformational properties and specific residue contributions. The structural origins of diagonal and off-diagonal peaks in the 2DIR spectrum were identified for the native and intermediate conformational ensembles in the folding mechanism. This work supports the implementation of computational techniques in conjunction with experimental 2DIR to study the folding mechanism of proteins. In addition to exploring the folding mechanism the work presented here may also be applied in combination with experiment to refine and validate current molecular dynamics force fields.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to acknowledge the following for their helpful discussions and support of my work. Firstly, The Wang Lab, specifically Ronaldo Olivera, Dr. Qiang Lu, Bo Han, and Dr. Haidong Feng who have been helpful in discussions of my work and provided valuable feedback and support on a number of topics. The Mukamel Group, including Brit Highland and Dr. Wei Zhuang who provided helpful tips, insight and technical dialogue as we refined their SPECTRON software. Professor Shaul Mukamel, who provided helpful thoughts about my work and who is largely responsible for the years of progress made to allow 2DIR spectra to be simulated. Professor Dan Raleigh and Dr. Jae-Hyun Cho who have taught me much about the experimentalist view of science and contributions to the work on NTL9. I would like to thank my committee Professors, David Green, Professor, Dan Raleigh, John Reinitz and Rob Rizzo and my outside member Dale Drueckhammer. I am indebted for their helpful insights, criticisms, and support, that have not only helped me build this dissertation but also build my character. Finally, I would also like to thank my advisor Professor Jin Wang who has been supportive in this process and also provided inspiration to continually explore science from a top down perspective.

Introduction

This thesis begins by discussing the protein folding problem in Chapter 1. The techniques applied in some of this work towards understanding the folding problem experimentally are described in Chapter 2. These methods were implemented in the derivation of data for the N-terminal domain of ribosomal protein L9 (NTL9) studies in Chapter 5. In Chapter 3 the underlying concepts and practical considerations concerning the use of molecular dynamics simulations, the primary technique used in this thesis, are reviewed. In Chapter 4 the two-dimensional infrared (2DIR) experimental technique is described. This section also highlights the fundamental principles and practical steps required to simulate 2DIR signals. In Chapter 5 the identification and characterization of the transition state ensemble (TSE) of NTL9 is described through a synergy of experimentally and computationally derived data. The TSE was identified from high-temperature implicit-solvent all-atom unfolding simulations in conjunction with experimentally determined Φ-values. Chapter 6 describes the calculations of the 2DIR spectrum of the protein folding mechanism of the Beta3s protein. The structural origin of 2DIR signals for the native and intermediate conformational ensembles in the folding mechanism of Beta3s were identified. Chapter 7 builds on the results and methodologies of Chapter 6, summarizing the implementation of simulated folding pathways and the calculation of 2DIR spectra to characterize protein folding at an atomic level of resolution. Finally, Chapter 7 suggests new protocols that may be implemented with current reduced representation folding models. These procedures are expected to be useful for model refinement and expanding our understanding of protein folding through comparisons with experimental data.

1 Protein Folding

1.1 Proteins

Proteins are biological polymers of amino acids that are essential to all living organisms. Information about the function of a protein is encoded in their amino acid sequence which is derived from genetic material [1, 2]. Proteins function in several capacities including the signaling processes, catalytic activity and structural components critical to life [3]. Ultimately, the function of a protein is dependent on its amino acid sequence folding into its final folded three-dimensional shape. Their malfunction is often at the root of disease, and as such proteins are the primary target of pharmaceuticals developed for the treatment of human disease. Specifically, the activity of misfolded proteins has been implicated in diseases including Alzheimer's, Diabetes, Parkinson's disease, many cancers and cancer-related syndromes [3, 4]. Thus, an understanding of the protein folding mechanism is of importance for pharmaceutical design and molecular biology [3, 5-9]. In the last several decades the mechanism of protein folding has been the subject of extensive experimental and theoretical study [10-14]. In this dissertation the folding mechanisms of two model protein systems are characterized in a synergy between experiment and theory.

1.2 The Protein Folding Problem

The current study of the protein folding mechanism is based on the early experiments by Anfinsen et al. [1, 2, 15]. Their studies revealed that an unfolded protein will spontaneously refold into its native state [15]. This suggested that all the information necessary to fold proteins is contained in the amino acid sequence. However, the question of how proteins find their 3D structure from a sequence of amino acids coded by the genome, still remains. Anfinsen's "thermodynamic hypothesis" of protein folding suggests that an amino acid sequence folds spontaneously to a conformation with the lowest free energy [2]. Yet, kinetically, the question remains, how does a protein fold in experimentally observed folding timescales. It is known from Levinthal's paradox that a random search of all possible geometries would take longer than the age of the universe to fold the average protein, consequently another process must facilitate the process [16]. Thus, the study of protein folding has focused on finding a physically based underlying mechanism responsible for guiding the process thermodynamically and kinetically [10-14].

1.3 Energy Landscape Theory

The Energy Landscape Theory of protein folding provides a statistical description based approach to the folding problem. It proposes that proteins fold in an ordered way, through an ensemble of structures under an energy biased folding pathway [17, 18]. (Figure 1-1) The energy bias, or relative stability of the folded state versus the unfolded state, facilitates folding on the observed timescales [17, 18]. Under this model, a protein made up of random residues would produce an energy surface that is rough and relatively unfunneled (unbiased), also called frustrated. On the other hand, natural proteins are not random and thus have evolved to incorporate a native state bias that drives the kinetics of folding in an energetic funnel. In this model the free energy of folding is composed of three components, the entropy, the native driving force and the roughness contribution of the energy pathway [17-22]. As the protein folds, the interactions between the protein and solvent as well as those within the protein facilitate the native state bias and funnel [17]. In this respect the evolution of such interactions during the folding of process can be likened to a reaction coordinate in a chemical reaction [17]. Experimentally, kinetics and contacts between residues have been mapped through the folding process. Similarly, with a theoretical model of sufficient detail, these reaction coordinates and others may also be monitored.

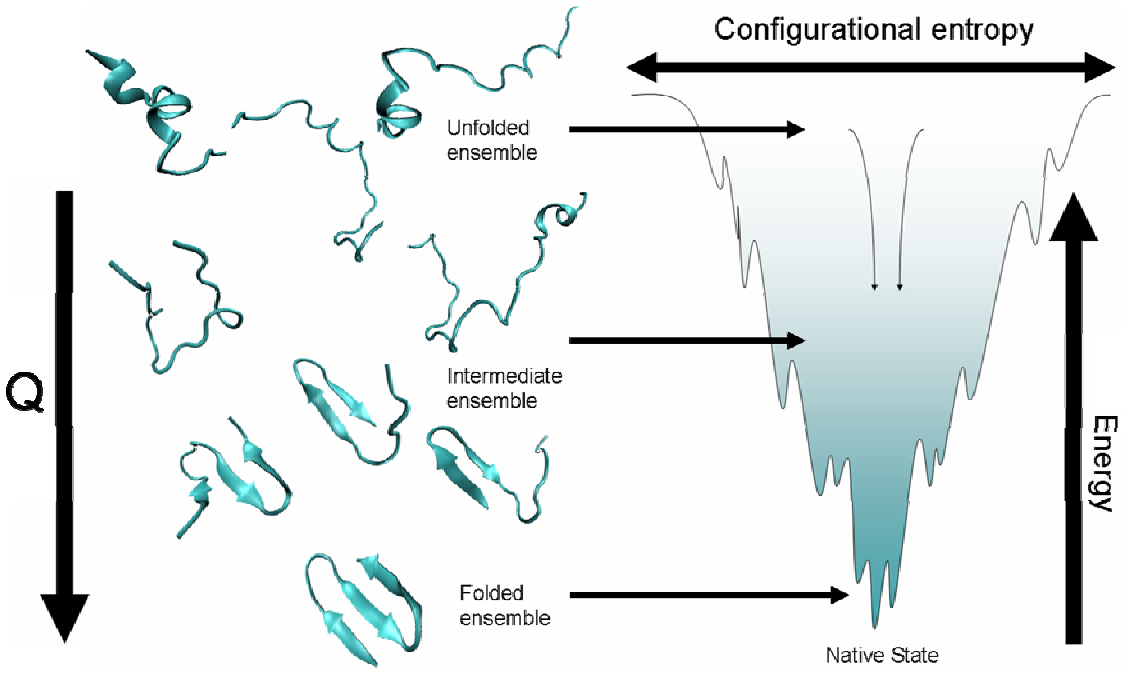


Figure 1-1: The energy landscape of protein folding.
The energy landscape depiction of protein folding. Folding occurs through the progressive organization of ensembles of structures (left) through a funnel-shaped energy landscape (right). Conformational entropy loss during folding is compensated for by the energy as more native interactions form. Adapted from reference [23].

1.4 Experimental Approaches for Probing Protein Folding

Traditionally, the majority of experimental data on protein folding has been obtained through kinetics experiments [21, 24]. The degree of foldedness of a protein is determined in kinetics experiments by conformationally sensitive techniques including fluorescence, circular dichroism and other probes [21, 24]. In these experiments the protein's general structure is monitored as it is folded or unfolded in the presence of chemical or thermal denaturants. The stopped flow method [25] is a specific technique employed to modify the protein's chemical environment and induce folding or unfolding. The data characterizing NTL9 in Chapter 5 was derived by this method.

Experimental data on protein folding obtained through kinetics experiments does not report on atomic level structural changes. Consequently these experiments alone are not sufficient to fully describe the detailed dynamics that characterize the folding process. Moreover, experimental methods with structural resolution have lacked the temporal resolution necessary to observe ultra-fast folding processes. Although, progress has been made by advanced nuclear magnetic resonance (NMR) techniques, lengthy folding processes have yet to be explored [26]. Recent advances in two-dimensional infrared spectroscopy (2DIR), using techniques derived from NMR, are shedding new light on the mechanism of protein folding [27-31], these will be discussed in detail in Chapter 4.

1.5 Computational Approaches for Probing Protein Folding

The computer simulation of protein dynamics offers a detailed atomic-level structural and temporal understanding of the protein folding process. Several different variations of simulation exist to probe a variety of aspects of the process including kinetics and thermodynamics [32]. Additionally, simulation can probe atomic level changes and the importance of solvent in the folding process [33]. Molecular dynamics simulation, discussed further in Chapter 3, is now a common method used in probing the folding and unfolding free energy surface of molecules [33, 34]. (Figure 1-2) In this dissertation Chapter 5 and 6 reveal that when combined with experimental data, synergies between simulation and experiment reveal greater knowledge about the folding mechanisms of proteins [11, 13, 35-42].

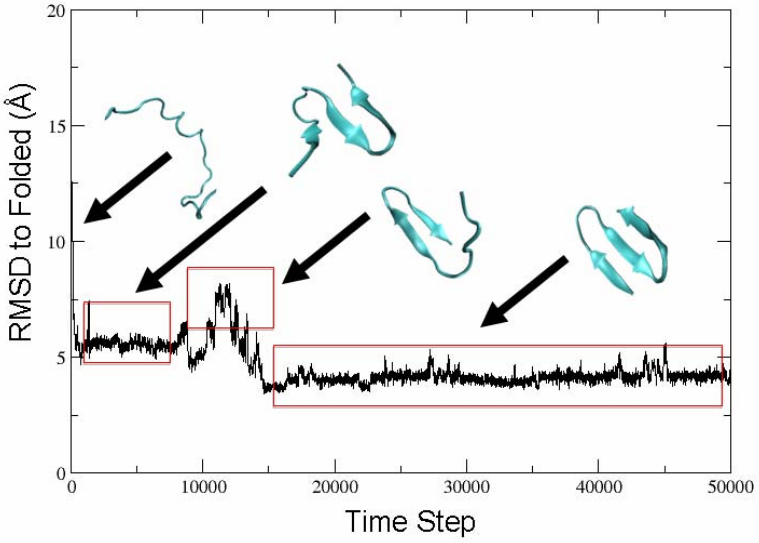


Figure 1-2: A typical protein folding trajectory plotted as RMSD versus time step.
In this trajectory for the folding of mini-protein Beta3s (in cyan) the intermediate conformations configurations were assigned to the time and RMSD region along the folding pathway. This highlights the utility for atomic resolution structural determination of folding pathways by all atom simulations.

2 The Transition State Ensemble and Two-State Folding

Determination of the transition state ensemble (TSE) of proteins is a necessary component of the description of the folding process because these conformations are central to folding and unfolding kinetics [43]. In the work presented in Chapter 4 the TSE of N-terminal domain of ribosomal protein L9 (NTL9) is identified and characterized. Energetically, the TSE represents an ensemble of structures that occupies the highest free energy in the folding pathway [33, 44-46]. (Figure 2-1) In the terms of energy landscape theory the TSE is a bottleneck through which all folding paths must cross on route to the native state [17, 18]. This concept is most easily understood in terms of a two-state protein where the denatured and native states are separated by a single high free energy barrier, known as the transition state ensemble. (Figure 2-1)

Transition state ensembles are inherently difficult to characterize because they are both transient, high energy and partially unstructured. Unlike in simple chemical reactions, the TSE in proteins is not localized to an individual bond. This is because of the large number of non-covalent interactions in protein structure that are broken and formed in the TSE. Due to the complexity of the TSE, single-residue mutation techniques, such as Φ-value analysis, are employed to explore two-state dynamics and the folding pathway of small proteins and peptides [37, 47-64].

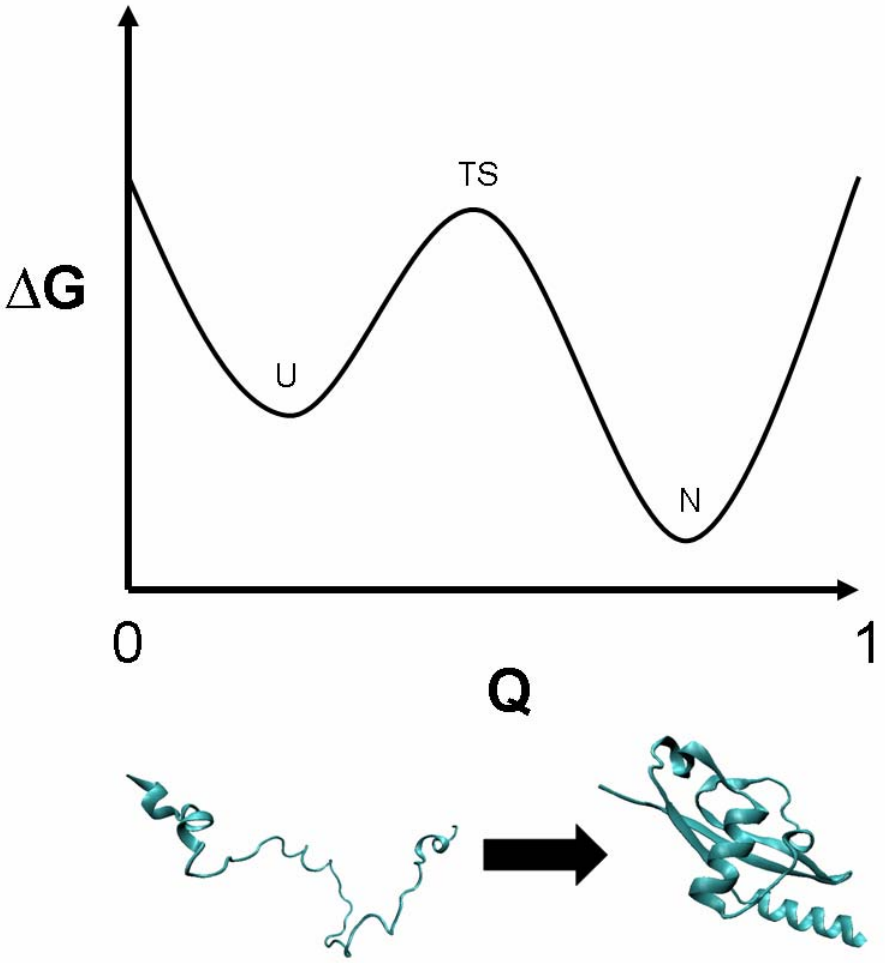


Figure 2-1: The two-state protein folding model.
The free energy ΔG versus reaction coordinate (Q) is plotted for a two state protein. The unfolded state ensemble exists in the well labeled U. The folded ensemble exists in the well labeled N. The barrier between the folded and unfolded state represents the transition state ensemble, labeled TS.

2.1 Φ-Value Analysis

The Φ-value analysis technique is a tool used by experimentalists to study the structure of the folding transition state ensemble in small two-state proteins and peptides [56, 59, 65, 66]. Traditional structural elucidation methods including protein nuclear magnetic resonance (NMR) and x-ray diffraction (XRD) are dependent on well-defined, long-lived ensembles, making them unsuitable for the study of transition states. Φ-value analysis compares, through folding kinetics, the conformational folding stability of the wild-type protein with that of proteins containing one or more point mutations. This analysis results in a so called Φ-value that is related to the mutated residue's energetic contribution to the folding TSE. (Figure 2-2) The Φ-value represents the ratio of the energetic destabilization introduced by the mutation to the transition state versus that introduced to the native folded state [59, 64-66]. Mathematically the Φ-value is defined as:

$$\Phi = \frac{\Delta G_W^{TS-D} - \Delta G_M^{TS-D}}{\Delta G_W^{N-D} - \Delta G_M^{N-D}} \qquad = \qquad \Phi = \frac{\Delta\Delta G^{TS-D}}{\Delta\Delta G^{N-D}} \tag{2-1}$$

Where $\Delta G^{TS\text{-}D}{}_W$ - $\Delta G^{TS\text{-}D}{}_M$ is the free energy difference between the transition state ensemble and the denatured state ensemble for the wild-type (W) and mutant (M) protein and $\Delta G^{N\text{-}D}{}_W$ - $\Delta G^{N\text{-}D}{}_M$ is the difference between the native and denatured state ensembles [59, 64-67].

The Φ-value calculation ideally results in a number from 0 to 1, where 0 suggests the mutation has no effect on the structure of the TSE. This is illustrated in the bottom of

Figure 2-2. In terms of folding this indicates that the mutated residue (and its surrounding region) in the wild-type protein is not critical in the folding transition state and thus not critical in the folding mechanism. On the contrary a Φ-value of 1 suggests the mutation has a significant effect on the folding pathway (top Figure 2-2) and therefore its surrounding region and intermolecular interactions are critical in the folding pathway of the protein.

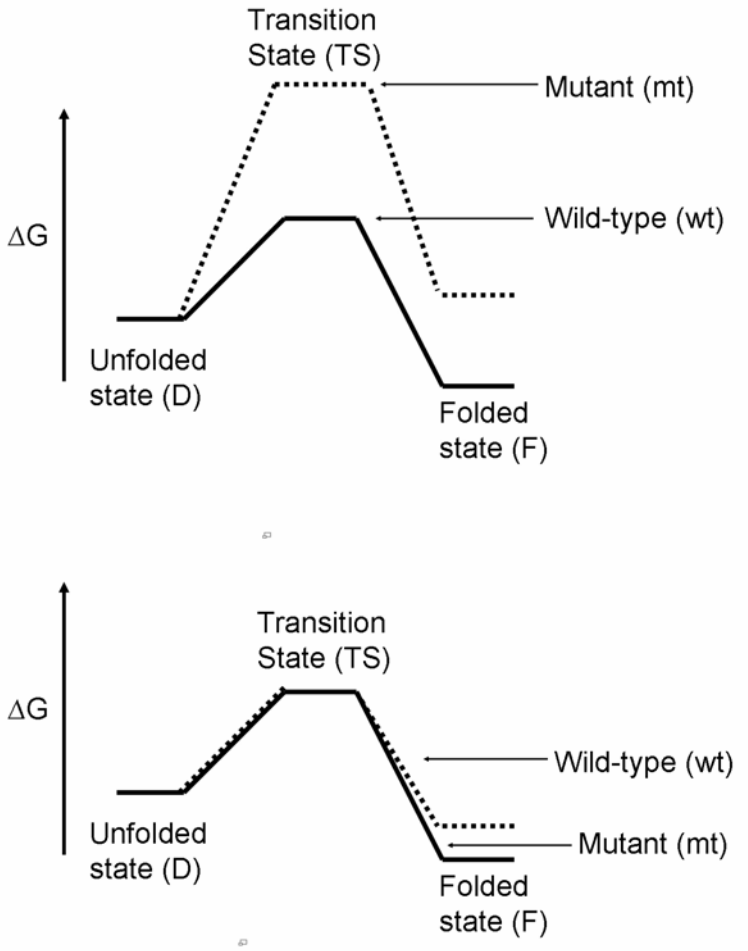


Figure 2-2: Φ-value energy scheme.
Energy diagram displaying the energy of the denatured state, D, the transition state, TS, and the native state, F for both the mutant and wild-type proteins. Top: Illustrates the case when the Φ-value is 1. Bottom: Illustrates the case when the Φ-values is 0.

2.2 Computational Determination of Φ-Values

In addition to experiment, molecular dynamics simulation can aid in the determination of the transition state and Φ-values for particular residues in a protein folding pathway [36, 38, 68-70]. The calculation of Φ-values from simulated structures is generally not attempted directly by calculation of the free energy because of the extensive computational effort required to obtain computational free energies. Alternatively, computational investigation of either unfolding or folding simulations provides detailed structures on the folding pathway. Since structures are available, calculated Φ-values (Φ_{calc}) involve N^{Conf}, the number of contacts for a particular residue in the conformation of interest and N^{Folded} the number of native contacts in the folded native state ensemble of the protein. This is a suitable approximation since experimental Φ-value analysis is interested in the general structure around a particular mutated residue which is ultimately related to atomic contacts. The calculated Φ-value (Φ_{calc}) can be defined:

$$\Phi_i^{calc} = \frac{N_i^{Conf}}{N_i^{Folded}} \qquad (2\text{-}2)$$

Where N_i^{Conf} is the number of contacts for a particular residue i in the conformation of interest and N_i^{Folded} is the number of native contacts for i in the folded native state ensemble of the protein. Calculated Φ-values should theoretically fall between 0 and 1, like their experimental counter parts, however if more contacts exist in a non-native state than in the native state Φ_{calc} derived may be greater than 1.

The calculated Φ-value can be used computationally in many ways including the determination of the transition state ensemble. This is accomplished by computing the

Φ-values for a series of structures along the folding or unfolding trajectory and comparing them to experimental Φ-values [38]. This methodology is implemented in the study of NTL9 described in Chapter 5. Alternatively, it is possible to calculate Φ-values *ab initio* from a set of pre-calculated transition state ensemble and folded states [36, 71]. Finally, experimental Φ-values can be used as constraints to drive folding calculations towards the folded state and along folding pathways [36, 71].

2.3 Assumptions of Φ-Value Analysis

Φ-value analysis assumes several properties of the system investigated, these conditions need to be understood when implementing the method [67, 72]. Specifically, Φ-value analysis assumes the energy landscape contains a deep global minimum that is well defined. This implies that the intermediate conformation closely correlates with the energy of the structure, a fact necessary to structurally evaluate destabilizations introduced by mutations. The method also assumes that the folding pathway is not altered significantly by the introduced mutations. In general this is avoided by employing conservative mutations to maintain the predominant aspects of the folding landscape. Finally, Φ-values are restricted to the 0 to 1 range which means that a mutation will not increase stability and thus not lower the energy of the native or transition state for the mutant relative to the wild type. This assumes that interactions that stabilize the folding transition state are native like, which is not always true. Recent studies have shown that non-native interactions may facilitate folding and stabilize the transition state ensemble to speed up folding [48, 73, 74].

2.4 Corrected Φ-values

The corrected Φ-value calculation attempts to address the fact that at times mutations may affect a non-native interaction that participates in folding [48, 59, 75]. In this calculation denatured state effects are included so that Φ is defined below:

$$\Phi = \frac{(\Delta G^{TS}_{mt} - \Delta G^{TS}_{wt}) - (\Delta G^{D}_{mt} - \Delta G^{D}_{wt})}{(\Delta G^{N}_{mt} - \Delta G^{N}_{wt}) - (\Delta G^{D}_{mt} - \Delta G^{D}_{wt})} \qquad (2\text{-}3)$$

Where the ΔG^{D}_{Mutant} - $\Delta G^{D}_{Wild\text{-}type}$ terms take into account the denatured state energy contribution of the selected residues. Figure 2-3 describes, graphically, the consequence to the Φ-value of taking into account the denatured state effects.

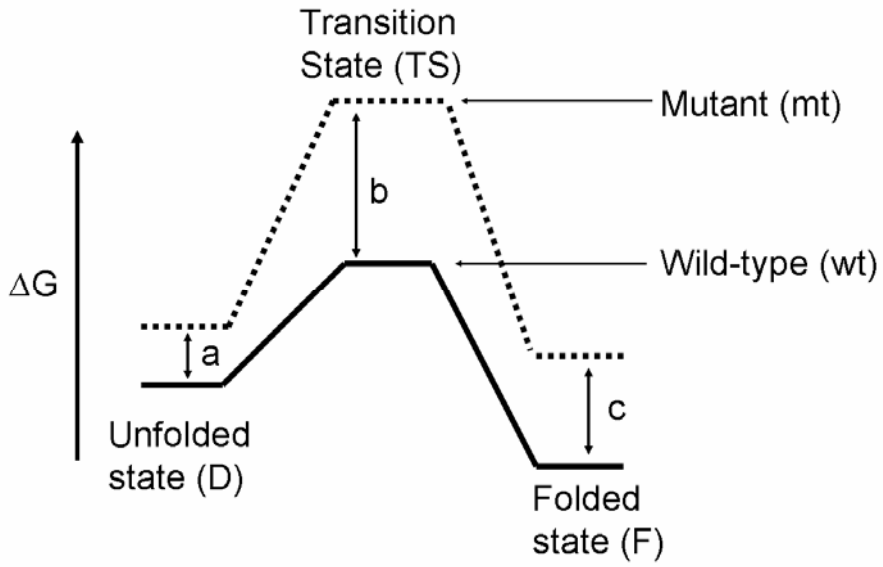


Figure 2-3: Energy diagram for the corrected Φ-value analysis method.
The free energy profiles are drawn for the wild-type (solid) and mutant (dotted) protein configurations. The unfolded state (D) is on the left and the folded state (F) is on the right separated by the higher line representing the transition state (TS). In the corrected method a is calculated represents the $\Delta\Delta G^{D}_{Mutant}$ - $\Delta\Delta G^{D}_{Wild\text{-}type}$ term in equation X. The terms b and c are common to the original Φ-value analysis method and correspond to $\Delta\Delta G^{TS}_{Mutant}$ - $\Delta\Delta G^{TS}_{Wild\text{-}type}$ and $\Delta\Delta G^{F}_{Mutant}$ - $\Delta\Delta G^{F}_{Wild\text{-}type}$ respectively.

2.5 Summary

In Chapter 5 the experimental and computational Φ-values discussed here are employed in the study of the folding mechanism of NTL9. Specifically, this work models the implementation of experimental Φ-values and molecular dynamics simulation towards the structural identification and characterization of the transition state ensemble a critical component of the folding pathway. The primary tool facilitating an atomic level description used in this study was molecular dynamics simulation the fundamentals of which are discussed in Chapter 3.

3 Computational Techniques

This chapter describes the computational methods and data analysis implemented in generating results for this dissertation. The emphasis is on molecular dynamics (MD) calculations because they constitute the primary tool for obtaining data in this work. An explanation of the sampling techniques for data derived from MD simulation is also included. The two-dimensional photon-echo spectroscopy calculations, also important to this work, are treated separately in the theoretical background of the 2DIR in Chapter 4.

3.1 Molecular Dynamics Simulations

Molecular Dynamics (MD) simulations compute the motion and interactions of molecules in time according to the laws of Newtonian Physics [76]. MD simulations vary widely in scope and implementation, as such they have utility in exploring the properties and dynamics of chemical systems from biological molecules to material sciences [77]. In the past two decades MD simulations have become one of the principal tools in the theoretical study of biological molecules [78]. Additionally, MD simulations are also implemented in the determination of structures from XRD and NMR experiments [79-84].

Molecular Dynamics simulations calculate of series of configurations versus time that reveal time correlated motions of molecules, through the sequential integration of Newton's equations of motion [76, 78, 85]. This means that MD simulations are theoretically deterministic unlike other popular stochastic simulation techniques such as Monte Carlo simulation. Mathematically simulation is accomplished using Newton's law, $F = ma$. Initially, the force on an atom F_i is determined by the negative gradient of the potential energy function V which is a function of molecular positions r. This is expressed as equation 3-1:

$$F_i = -\frac{\partial V}{\partial \mathbf{r}_i} \tag{3-1}$$

The movement of atoms is then simulated by in time (t) solving the following equations relating the velocity v_i, position r_i and force F_i of atom i:

$$\frac{dv_i(t)}{dt} = \frac{F_i(t)}{m_i} \tag{3-2}$$

Where the velocity v_i can be rearranged in terms of atomic positions in time, in equation 3-3 where r is the atomic coordinate.

$$\frac{dr_i(t)}{dt} = v_i(t) \tag{3-3}$$

It is then necessary to represent this as a function of time where equations 3-2 and 3-3 can be represented for small changes in time [76, 85]. For small time steps the leap-frog algorithm [85] integrates these laws of motion to update the position of atoms after Δt, where:

$$r_i(t + \Delta t/2) = r_i(t) + v_i(t + \Delta t/2)\Delta t \tag{3-4}$$

and

$$v_i(t + \Delta t/2) = v_i(t + \Delta t/2) + \frac{F_i(t)}{m_i}\Delta t \tag{3-5}$$

Following these calculations the force on atoms can be recalculated and the process computed iteratively for a desired simulation time. Due to the computational expense in calculating these equations there is a practical limit on the number of time steps and atoms simulated.

3.1.1 The Potential Energy Function

The calculation of the forces during molecular dynamics simulations requires knowledge of the potential energy of the system [85]. The potential energy is represented as a function of the bonds, angles, torsions and interactions between atoms comprising bonds and it is frequently represented by a Lennard-Jones potential. The potential energy of the system is described in Equation 3-6, where it is shown that V, the potential energy

function is differentiable with respect to atomic coordinates to give the force acting on an atom [85]. In MD calculations the potential energy function can be broken down into two practical components, the set of equations to describe the components of the potential energy and the parameters used in solving these equations. Equation 3-6, below represents the all-atom MD force field equation applied in the AMBER [86], package. Different packages, CHARMM [87], GROMACS [88] implement slight variations of this equation.

$$V(r^N) = \sum_{bonds} k_b (l - l_0)^2 + \sum_{angles} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \varepsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\} \quad (3\text{-}6)$$

In Equation 3-6, the first term represents the energy attributable to covalently bonded atoms. The second term represents the energy due to the angle position of atoms involved in covalent bonding. The third term gives the torsion energy, a result of bond twisting. Finally the fourth term represents the non-bonded energy between all atom pairs which is broken down into van der Waals interactions (first term) and electrostatic energies (second term). In theory the last term should be calculated for all possible pairs of atoms in the system, this is however not practical considering after a particular distance, interactions will become energetically negligible. As such cutoff distances are specified in MD calculations to determine when to ignore such calculations because atoms are far enough apart that their energetic contribution is negligible.

3.1.2 Molecular Dynamics Force Fields

In order to implement a force field, it is necessary to have values for the parameters contained within, including force constants, equilibrium bond lengths (l) and angles (θ), charges (q). These values can be variable depending upon the environment but are ultimately approximated by parameter sets taking into account, to varying degrees these conditions. Moreover, different force fields exist and thus necessitate different force field parameters. In our simulations we used the AMBER ff99 [89], AMBER ff03 [90], CHARMM19 [87] and CHARM22 [91] force fields and parameter sets. Several review articles evaluate the differences between each of the commonly implemented force fields [92-96].

3.2 Sampling and Comparison to Experiments

The comparison of molecular dynamics simulations to experiment is facilitated by principles of statistical mechanics. Experimental data is typically derived from a sample that contains a very large number of molecules in many states, a macroscopic system. Molecular dynamics simulations, however, typically occur on one molecule in only one state (at multiple times), a microscopic system. Statistical mechanics facilitates the relationship between the two through the Ergodic hypothesis [97, 98]. This is expressed as:

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} \qquad (3\text{-}7)$$

Where $<A>$ is the average value of an observable A. This is useful when comparing MD simulations to experiment because experimental values are ensemble averages while averages defined by MD are time averages.

The relationship between the ensemble average and time average suggests that for a very long simulation time the entire phase space, (all states of some observable A) are fully sampled. Unfortunately, this requires a significant amount of simulation time which is not always possible due to practical computational limitations. An alternative to generating a very lengthy trajectory is to employ a sampling technique that attempts to cover as many of the different states of a system in as little amount of time as possible [99-103].

In this work the Ergodic hypothesis is the basis for our primary sampling method which involved running many simulations from different starting conditions. This is illustrated in Figure 3-1, which displays the trajectories of many simulations started from an initial random seed. This method allowed us to run a very long simulation, as several

separate trajectories on several different machines. Additionally, in the works presented in Chapter 5 and Chapter 6 increased sampling was also facilitated by running longer simulations at reduced computational cost through the implementation of implicit solvation models.

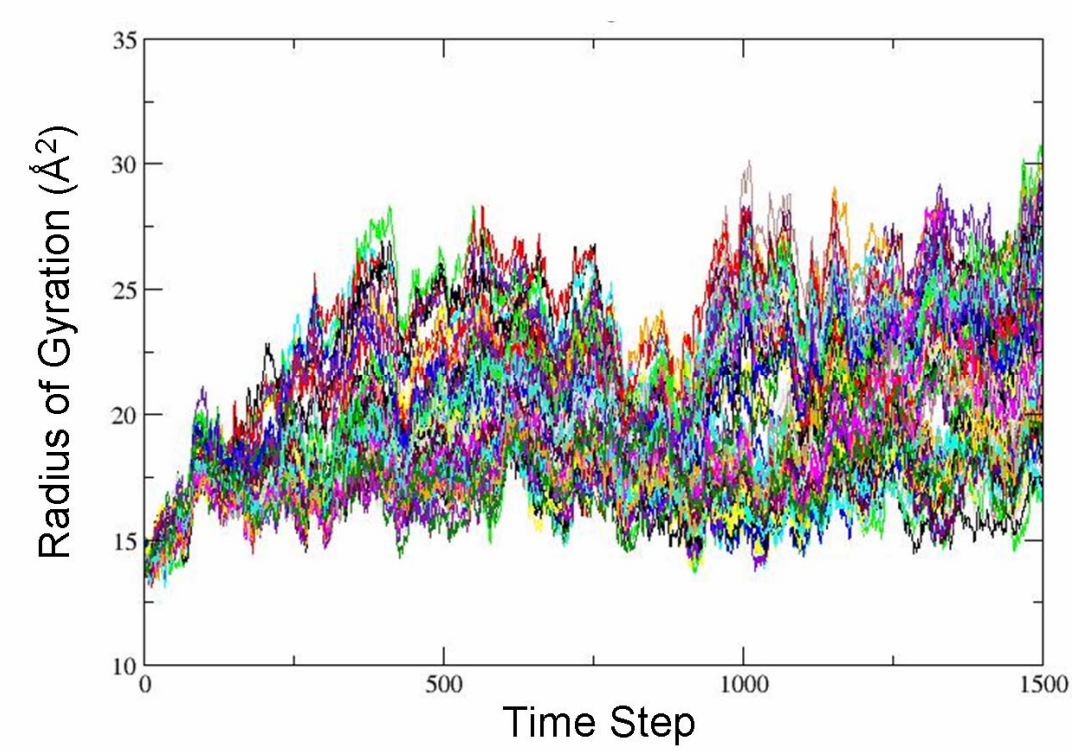


Figure 3-1: Molecular dynamics sampling scheme for unfolding of a protein.
One hundred unfolding trajectories are started with random seeds defining initial simulation velocities. Resultant simulations diverge and sample a greater proportion of the potential energy space compared to a single simulation. A reaction coordinate, the radius of gyration, is monitored versus time for 100 unfolding trajectories, each trajectory represented by a different color line.

3.3 Solvent Models

Protein conformation and dynamics are known to be dependent upon their solvating environment [104]. Consequently, it is important to consider solvent in MD simulations of proteins. Solvent can be considered in two main categories, explicit and implicit [104-107]. Explicit model simulations include solvent molecules in a box surrounding the protein and calculate the forces between these molecules and other molecules in the simulation. This method adds to the number of atomic positions over which to integrate and therefore requires more computational effort, but a more realistic simulation. Implicit models attempt to limit computational time used by approximating the solvent as a continuous medium in either considering accessible surfaces area (ASA) [108], and or electrostatic interactions with an external electrostatic continuum [109]. In general implicit models will simulate folding faster because the have fewer atom positions to integrate as well as this lack of solvent effect. In comparison to experiment and explicit models of solvent, implicit solvent models have some drawbacks. Specifically because interactions with solvent molecules are not included the friction and thus reduced speed of dynamic processes is not included.

4 Structural Elucidation of Proteins by Infrared Spectroscopy

4.1 Introduction

An understanding of the protein folding mechanism necessitates characterization of transient ensembles of structures over short time scales. Typically on the order of picoseconds to seconds, the short timescales involved in protein folding make it difficult to detect structural changes with atomic level resolution [24, 110-112]. Traditionally the majority of experimental data on protein folding has been obtained through kinetics experiments that cannot fully report on atomic level structural changes and thus the structural dynamics that characterize the process [21, 24]. Additionally, experimental methods with atomic level structural resolution have lacked the temporal resolution necessary to observe ultra-fast folding processes. These methods, including nuclear magnetic resonance (NMR) and x-ray diffraction (XRD) have resulted in indirect or time-averaged information about structures [113]. Although progress has been made by advanced NMR and XRD techniques the ability to extract temporal or dynamic information by these methods is limited to milliseconds to seconds time scales [4, 26, 114-122]. Consequently, these procedures cannot sample the conformational dynamics involved in the folding of proteins nor can they sample the process in a sequential order at the timescales necessary to follow the protein folding process.

4.2 Infrared Spectroscopy

Optical techniques including fluorescence [4, 123, 124], infrared Raman [112, 125-127], and UV circular dichroism [128], utilizing high frequency probes have offered the ability to understand processes at fast timescales, although they traditionally have lacked structural resolution [129, 130]. Among them, infrared (IR) spectroscopy has emerged as among the most versatile of the optical techniques in the study of protein structure and dynamics [123, 131, 132]. Unlike other optical techniques, IR methods can be photochemically inert and do not require chemical modification to samples. Additionally, the utility of IR spectroscopy for structural elucidation of molecules from 1D absorption experiments is well known [133].

IR spectroscopy is sensitive to protein structure by the excitation of vibrational resonances in proteins associated with the bonds in the peptide backbone or side chains [134-137]. Specifically, the amide-I band of the protein backbone absorbs from 1600-1700 cm^{-1}. It is the most widely investigated band because its shape is sensitive to backbone structure and solvation but not side chains and it is a mode common to all proteins [138, 139]. Specifically, the amide-I band consists of the carbonyl stretch of the protein backbone and displays structural sensitivity due to coupling between in-phase bending of N-H and stretching of C-H bonds [139, 140]. Structure-frequency correlations for the amide-I band are well defined from experiment and include β-sheet absorption at 1630–1640 cm^{-1} and a weaker band above 1680 cm^{-1}. The random coil and α-helix structures are also known to result in absorptions at 1640–1650 cm^{-1} and 1640–1660 cm^{-1}, respectively. (Figure 4-1) Finally, although 1DIR is a useful technique to

probe molecular structure and environment the ensemble-averaged 1D signals do not report on information about vibrational coupling and structural fluctuations.

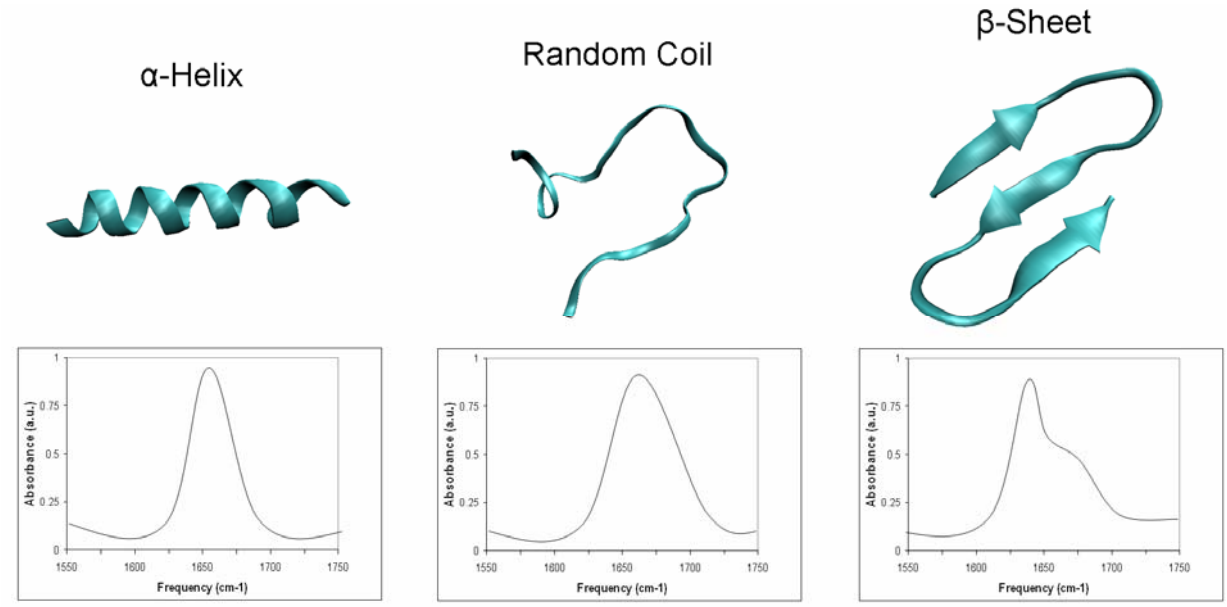


Figure 4-1: The Amide-I absorbance for different protein secondary structure.
1DIR spectra of α-helix, β-sheet and random coil secondary structure.

4.3 2DIR Spectroscopy

2DIR correlation spectroscopy, using techniques derived from NMR, improves information revealed by 1DIR, ultimately shedding new light on the mechanism of protein folding [27-31, 126]. Using advanced multiple pulse coherent spectroscopy techniques like 2DIR, transitions can be spread across two axes revealing vibrational couplings resulting from three-dimensional structural contacts [141-146]. Coupled with ultrafast optical technologies, the 2DIR methods result in a structure-based tool that is responsive at the ultrafast timescales (femtoseconds and longer) present in the folding mechanism.

Two predominant ultra-fast 2DIR correlation spectroscopic techniques exist; they differ in the way they probe the molecular system and the data they reveal. The simplest is the pump-probe technique that collects information in the time domain, a method common in UV-Visible spectroscopy, and pioneered in the IR by Hamm and Hochstrasser [147]. In this technique an initial frequency specific pulse (the pump pulse) is applied to the system. After a waiting time, a broadband frequency probe pulse is incident on the system. A 2D correlation plot of the system's response to the two pulses is derived by plotting the probe frequency at each pump frequency [147]. The second method, called the infrared photon-echo technique, collects data in the frequency domain instead of the time domain and is interpreted in the frequency domain by Fourier transform. The 2D photon-echo experiment involves interacting the sample with three pulses that result in a photon-echo being emitted [148]. The echo signal depends on the timing sequence and directionality of the three pulses. The photon-echo technique will be described in the following section and is then simulated in Chapter 6.

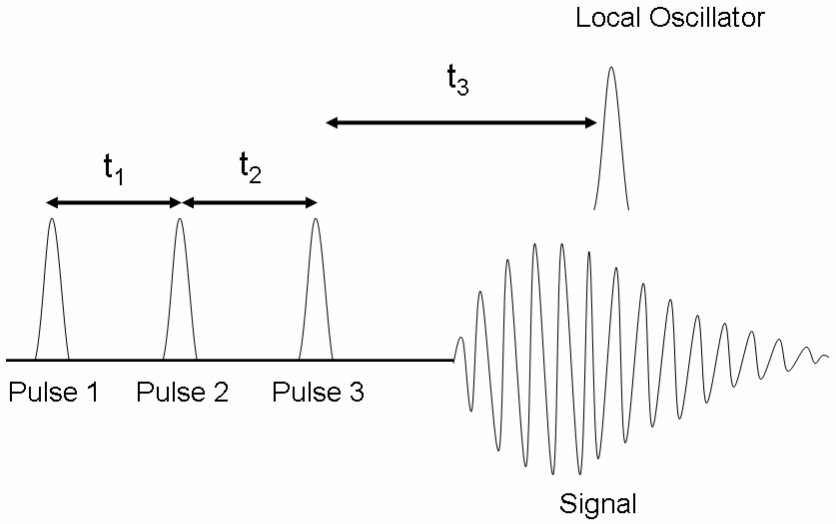


Figure 4-2: The pulse configuration for a photon-echo experiment with heterodyning detection.
The signal is recorded as a function of time delays, t_1, t_2, t_3. Pulse 1,2 and 3 are incident on the sample and the resultant signal emitted is probed by a local oscillator to extract phase and intensity information.

4.4 The 2DIR Photon-Echo Technique

Two-dimensional infrared photon-echo experiments involve probing the sample with three time- and phase-controlled laser pulses. The three phase-locked IR pulses, labeled 1, 2, and 3, with wavevectors $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ interact with the sample at time intervals τ, between 1 and 2, and t, between 2 and 3. (Figure 4-2 and 4-3) The first pulse creates coherences between vibrational modes. The second pulse creates a population at a specific frequency which is determined in the time domain by the delay between the first and second pulse and is called τ. After a variable time delay, T, the third pulse interrogates the sample once more, which results in a signal being emitted. The signal is generated in a phase matched direction dependent on the incident pulse directions and for the so called $\mathbf{k}_I$ experiments, includes incident pulses with wave vectors $-\mathbf{k}_1$, $+\mathbf{k}_2$, $+\mathbf{k}_3$. (Figure 4-3) The phase matching pulse polarization guarantees that the echo-signal will be separate from other non-linear process. Finally, a heterodyne pulse is interacted with the signal to extract phase and amplitude information. The time interval between this pulse and the third pulse represents the second dimension in the experiment and carries information about the final coherent state of the sample [148].

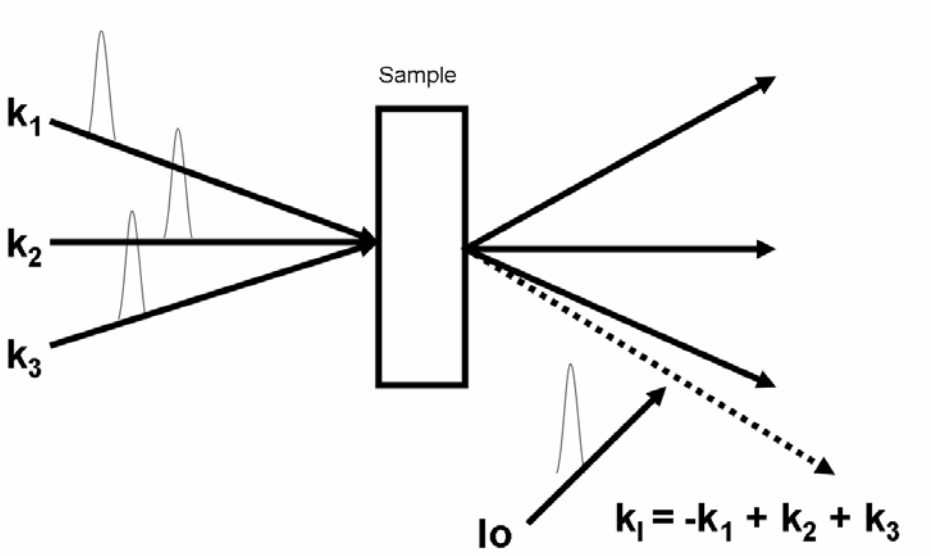


Figure 4-3: The photon-echo experimental setup.
The pulse sequence and geometry used in the generation of a three pulse for a photo-echo signal. Pulses are denoted by wave vectors, $-\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$ and resultant signal $\mathbf{k}_l$. Local oscillator is denoted lo and incident on the signal.

4.5 2DIR Photon-Echo Experimental Methods

In practice 2DIR photon-echo experiments involve a complex setup of controlled pulses that ultimately interrogate the sample [148]. The incident IR pulses employed are weak (on the order of 30nJ) to prevent photochemical reactions. Data is collected as a function of τ and *t* which is controlled by implementing timing delays on pulses. *T*, the time between pulse 2 and 3 is held fixed and may be 0 fs or longer depending on interest in the evolution of population state with time, *T* of 0 fs is sufficient for investigating static structural ensembles. Additional information about energy transfer in the system can be extracted depending on the experiment for different values of *T*. The time frame *t* is held fixed and τ is stepped over the interested range in small increments (0-1500 fs in about 18 fs steps). The signal resulting from the incident pulses is then heterodyned by an additional pulse to extract not just its amplitude but also phase. The signal is then recorded for t from 0 to 1500 fs in 18 fs steps in a process that takes about 35 minutes. This can also be determined interferometrically in one shot rather than by scanning *t*. Since the experiment, and thus the signal, occurs in the time domain it is Fourier transformed experimentally by a monochromator resulting in a spectrum with the frequency axes common to IR techniques. The two frequency dimensions ω_τ and ω_t of the 2D correlation spectra $S(\omega_\tau,T;\omega_t)$ correspond to the double Fourier transforms of the time traces of the signal $S(\tau,T;t)$ described.

4.6 The Characteristics of the 2DIR Spectra of Proteins

The diagonal peaks in a 2DIR photon-echo experiment correspond to transitions observed in traditional 1DIR experiments. (Figure 4-4) The specific 1DIR peak locations are useful since they correspond to common protein backbone conformations as noted in 4.1. In the 2D these peaks also contain additional information, specifically diagonal elongation of the peaks (inhomogeneous broadening) that represent the instantaneous distribution of frequencies in the transition. Additionally, anti-diagonal peak widening (perpendicular to the diagonal) is a result of homogenous contributions and related to the homogenous broadening parameter (Γ) and the time between pulse 2 and 3 (t) in the experiment. Furthermore, 2D spectra also contain two signals of opposite sign along the diagonal separated by the vibrational anharmonicity. These peaks correspond to the absorption of a second vibrational quantum. The signal on the diagonal corresponds to the 0 ->1 (denoted blue on the spectra) vibrational transition and the shifted peak to the 1 -> 2 transition (denoted red on spectra). (Figure 4-4)

In addition to the supplementary diagonal information, 2DIR spectra contain off-diagonal peaks that provide detailed structural information regarding mode coupling [146, 148]. The off-diagonal cross-peaks appear when the population of one mode is dependent on that of other modes. This is an indication of coupling which results from structural contacts and thus provides information about the 2D and 3D structure of the protein investigated. Specifically, the cross-peak positions in relation to the diagonal in the β-sheets of this spectra create a characteristic "Z" shape that can be used to track the extent of the β-sheet [149]. These cross-peaks are a result of coupling between the low and high frequency signals that originate from the β-sheet amide-I conformation [150].

Additionally, further structural information can be derived from cross-peaks if the particular residue in a protein is known or isotopically labeled [127, 130]. In such cases the vibrational contribution to the spectra from an isotopically labeled residue can be determined and compared to other similarly labeled residues providing specific structural information.

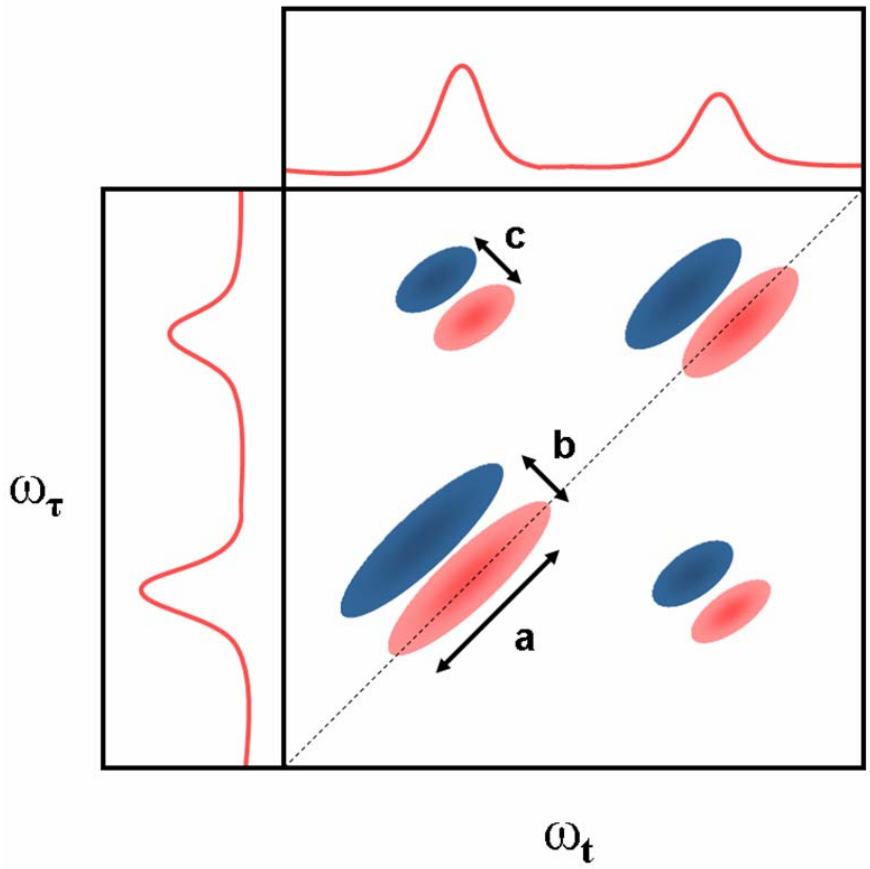


Figure 4-4: Characteristics of a 2DIR spectrum. 1D linear spectra on the left and top axis. The red (positive) signals corresponding to the 1→0 and the IR excited state emissions and the blue (negative) to the 1→2 transitions absorptions. The distance a represents peak broadening which represents the distribution of states for the mode. The distance b represents the vibrational anharmonicity between the states probed. The distance c, splitting between the positive and negative cross peaks can be used to extract the magnitude of coupling between the coupled modes.

4.7 Simulating 2DIR of the Amide-I

Modeling of the 2D coherent vibrational response is complicated but it has recently made possible with state-of-the-art computational techniques [151-158]. A practical description of the 2DIR signal calculation process for protein structures is outlined below [151], and treated in-depth in a recent review article [158].

1. Initial Structures

The initial step in simulating 2DIR spectra of proteins involves obtaining protein and solvent configurations. Molecular dynamics simulations containing at minimum a protein backbone and explicit solvent molecules are typically employed to generate an ensemble of configurations.

2. Building the Hamiltonian

Once a suitable structural ensemble is available, a fluctuating anharmonic vibrational-exciton Hamiltonian and transition dipole matrix is calculated for one and two-exciton states of each protein configuration. This Hamiltonian consists of the local amide unit composed of amide bond geometries and nearest covalent neighbors as well as non-covalent interactions including protein backbone, side chains and solvent regions near the C=O modes. MD calculations are insufficient in this application because they only model low-frequency modes, thus quantum mechanical calculations are implemented in this stage.

3. Preparing the Response Function

Following creation of the Hamiltonian, the four-point correlation functions of transition dipoles are calculated to describe relevant motions and fluctuations of the system [159]. The response of any system to an external perturbation can be expressed in

terms of a correlation function. The correlation function describes the time-dependent behavior or spontaneous fluctuation in the dipole moment in absence of a field. At a structural level the function contains information on states of the system and broadening due to relaxation. In simulating a 2DIR photon-echo experiment a four-point function is utilized because there are four states we are interested in including when t = 0, after waiting time one, after waiting time 2 and after waiting time 3. In 1DIR calculations this is a two-point correlation function that represents the system at t = 0 and t = 1.

4. Calculating the Response Function

Finally, the response functions, representing response of the four-point correlation functions to the electric field are calculated [160].

5. Calculating the Signal

The optical signal in the experiment is related to the polarization induced in the system by the external field interaction, which is third order [161]. This requires the response function, which is dependent upon the four-point correlation function determined previously, to be calculated in presence of the external fields. Since the experiment takes place in the time domain and we are interested in the frequency domain the calculated signal is Fourier transformed.

4.8 SPECTRON 2DIR Simulation Software and Analysis

SPECTRON is a software developed by the Mukamel group implements the described above following generation of structures [151]. The user is expected to define the local amide modes of interest, thus classifying the system and bath for creation of the local amide Hamiltonian. Generation of the local amide Hamiltonian requires information about the vibrational modes of the system as computed from quantum mechanical calculations, these are provided in the software as a parameter set. The frequency ranges and polarizations of pulses are then specified according the experimental setup desired. Calculations are then relatively straight forward and result in the signal which can then be plotted and analyzed.

4.8.1 Analysis of Simulated 2DIR Spectra

The simulation of 2DIR spectra is important to understanding the structural origin of specific peaks. Unlike in NMR, the origin of signals in 2DIR is complicated by the variety of infrared transitions in large structures [41, 129, 130]. Consequently, specific spectral peaks that describe detailed structures from 2DIR spectra have yet to be fully discerned. Isotopic labeling can be helpful in this endeavor but it is limited to species for which isotopic labels exist [130]. When the structure is known, as in the case of simulation, approximate techniques can be used to make peak-structure correlations. Specifically, normal mode decomposition (NMD) is a method that provides a simple analysis of the Local Amide Hamiltonian revealing strengths of cross peak interactions and mode excitation values.

4.8.2 Normal Mode Decomposition

In NMD analysis the excitonic Hamiltonian is diagonalized to obtain eigenvalues and eigenvectors for each of the residues in the system. The diagonalized matrix consists of Eigenvectors $C_{i,}$ the magnitude of which is the Eigenvalues E_i, where $HC_i=E_iC_i$. The elements in the vector C_i are can be denoted $c_{ij,}$ and the wave function is represented by,

$$|\phi_i\rangle = \sum_{j=1}^{N} c_{ij} |\psi_j\rangle \qquad (4\text{-}7)$$

where,

$$\sum_{j=1}^{N} c_{ij}^{\ 2} = 1 \qquad (4\text{-}8)$$

The Eigenvalues extracted from this diagonalization represent the frequency of each mode (residue). Consequently the $c_{ij}^{\ 2}$ value related to the Eigenvectors represents the contribution of mode (residue) *i* to mode (residue) *j*. The contribution from each *i* and *j* is what is analyzed in NMD analysis. NMD analysis is presented in the peak-structure correlations for the Beta3s protein in Chapter 6.

5 Identification and Characterization of the Transition State Ensemble of the N-terminal Domain of L9

In this chapter the folding pathway and transition state ensemble (TSE) of the 56-residue N-terminal domain of L9 (NTL9) is investigated using molecular dynamics (MD) simulation. The TSE of NTL9 is characterized at atomic resolution. The TSE was identified from high-temperature unfolding implicit-solvent all-atom MD simulations in conjunction with experimentally determined Φ-values. The TSE ensemble of NTL9 was found to be largely native in composition with the major components of secondary structure well defined. However, in the TSE these secondary structure elements do not make optimal interactions between other secondary structure elements. This is particularly evident in the N-terminal nucleus region which exhibits fewer native contacts and a greater solvent accessible surface area than the other regions. The data suggests that much of the drive towards the native state after passing the TSE is spent optimizing electrostatic interactions between stable secondary structure elements.

> The work presented in this chapter, "Identification and Characterization of the Transition State Ensemble of the N-terminal Domain of L9" was developed in collaboration between Christopher N. J. Marai, Jae-Hyun Cho, Daniel Raleigh and Jin Wang.

5.1 Introduction

Determination of the transition state ensemble (TSE) of a protein is a necessary component of the description of the folding process because these conformations are central to folding and unfolding kinetics [162]. Analysis of the transition state ensemble provides significant insight into how the primary amino acid sequence finds it final folded, and functional, form. Determination of the transition state ensemble for a folding or unfolding process is inherently complicated, however. Unlike in simple chemical reactions the transition state of proteins is not localized to an individual bond. This is because of the large number of non-covalent interactions in protein structure that are broken and formed in the TSE. Energetically, this ensemble of structures occupies the highest free energy in the folding pathway and many possible conformations [33, 44-46].

Advances in the study of proteins are making it possible to characterize the folding process and the intermediates along the folding path [163]. Experimentally, kinetics provides the only way to obtain information about the transition state ensemble. The experimental approach most commonly employed is the protein engineering method [164], called Φ-value analysis, in which mutations are made throughout a protein and the relative effects upon stability and folding rates are measured. Additional information about these methods was presented in Chapter 2.

In addition to experimentally derived data, computational simulations can be employed to provide structural and dynamic information about protein folding pathways at atomic resolution [165, 166]. Determination of the folding pathway and its free energy

is difficult, by all-atom molecular dynamics simulations with current technologies. Moreover, the low-dimensional reaction coordinates that are often employed in understanding molecular dynamics simulation data are imperfect at capturing the many degrees of freedom and conformational changes involved in protein folding. Consequently, structural properties backed by experimental evidence, are commonly employed synergistically to approximate the transition state ensemble (TSE) [163, 167-170]. Such complementarily is important in the study of TSE because their characterization is particularly challenging but required for a comprehensive description of the folding pathway [171].

In this study, determination of the NTL9 TSE was made based upon the identification of transition states from 100 unfolding simulations. The TS was located primarily by looking at a single reaction coordinate, native contacts, in the calculation of Φ-RMSD. The five identified TS structures are believed to represent some of the transition state structures of NTL9. The transition state of a two-state protein is however, better represented by an ensemble of structures located at the highest energy intermediate between the folded and unfolded state. Since many different states of a protein can be energetically degenerate the ensemble may be diverse. Thus, a TSE in reality exists as an ensemble of microscopically defined states along many reaction coordinates and represents several different structures of the same energy.

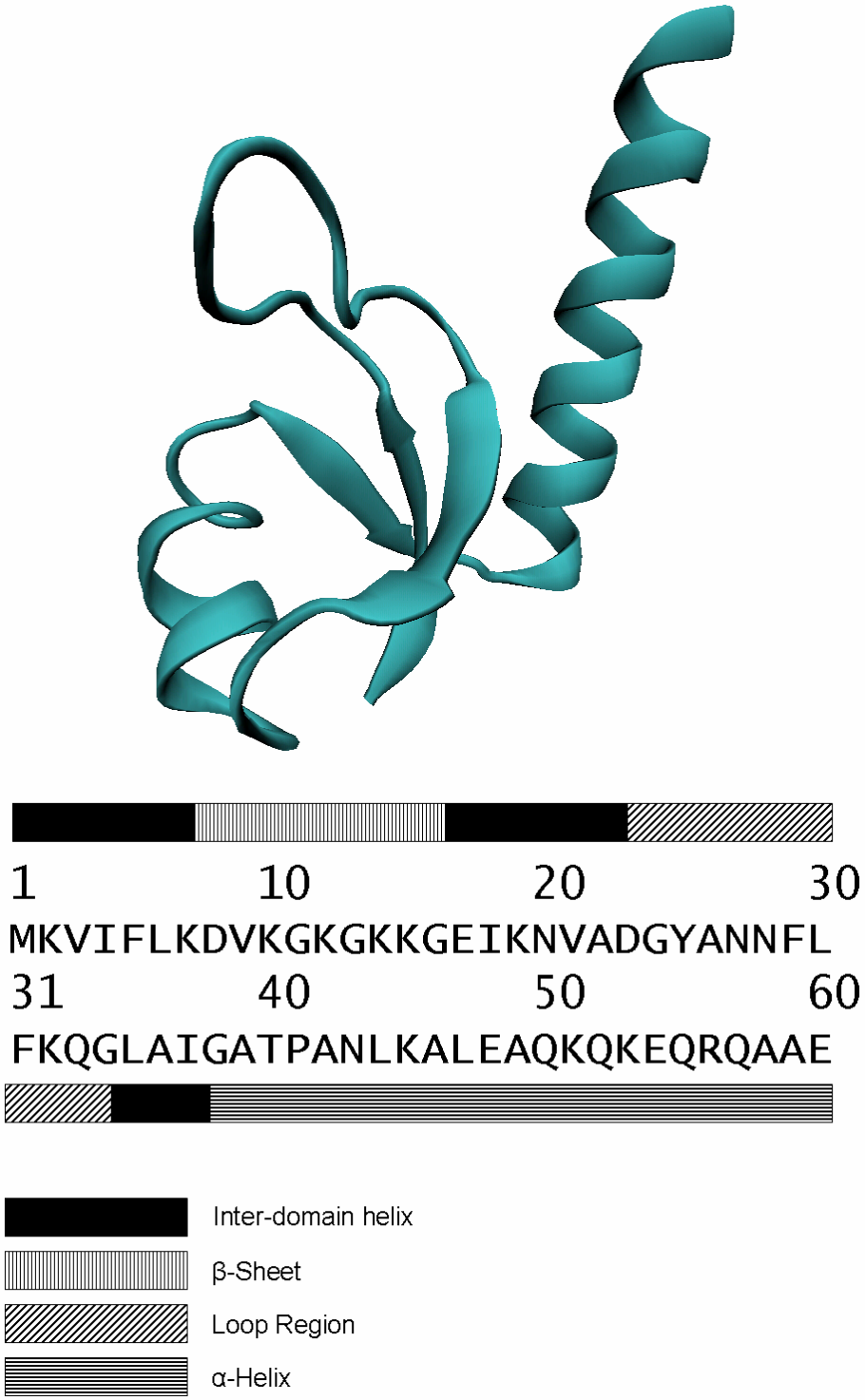


Figure 5-1: Native structure of NTL9 (PDB ID code 1DIV 1-60).
The structure is composed of (from N-terminus) alternating β strands residues 1-7 and 14-20 this is connected to another β-strand, residues 33-39, separated by an α helix residues 21-33 and finally terminated by an inter-domain connector helix.

In this work we investigate the folding pathway and transition state ensemble of the 56-residue N-terminal domain of L9 (NTL9) using molecular dynamics (MD) simulation. NTL9 was chosen for study for several reasons. It is a small, 56-residue protein that folds under a wide range of conditions, independently of the C-terminal domain of the L9 structure and without the aid of chaperones or metal binding ions [172]. NTL9 is among the smallest and simplest examples of an ABCD motif or the split β-α-β fold, which is found in a large number of proteins [173-175]. (Figure 5-1) The protein is composed of a three-stranded anti-parallel β-sheet flanked by two α-helices, one of which constitutes part of the inter-domain helical linker of the full L9 protein [173-175]. It is known experimentally to fold by a reversible two-state mechanism on the millisecond timescale, with a single rate-determining transition state [172, 176-179]. While the two-state nature eliminates the need to study an intermediate, characterization of the transition state is critical to understanding the folding mechanism since other partially folded intermediates are unobservable [24]. Finally, extensive work has been preformed to understand the nature of the folding pathway of NTL9 yet little is known about the TSE [172, 174, 176-198]. Detailed characterization of NTL9's TSE would represent significant progress towards understanding the folding pathway of this protein and complement the significant experimental research already accomplished on the protein.

5.1.1 Experimentally Derived Structures of NTL9 1DIV and 2HBB

In this study the 1DIV PDB structure of L9 was used to derive the 1-56 residues of NTL9 investigated [175]. 1DIV was determined by XRD at 2.8Å for the full 149 residues of ribosomal protein L9 [175]. The N-terminal domain of the 1DIV structure

was found to have poor electron density believed to be a result of crystal packing effects and central helix flexibility [175]. Specifically, and relevant to this work, the loop region residues 10-17 were not visible in the electron density map of 1DIV. At the time of the work presented here the 1DIV structure was the best available structure of NTL9. Since then, however, the 2HBB structure containing residues 1-51 of L9 was resolved by XDR at 1.9Å. The new structure provides detail about the unresolved loop region and as such a comparison was made to the 1DIV and 2HBB structures studied.

The figures and contact map comparison of 1DIV and 2HBB are illustrated below (Figure 5-2, 5-3 and 5-4). It is clear from the contact map and figures that the turn region immediately before residues 14, 15 and 16 the second β-sheet structure exhibits no contacts with residues 47-51 in the new 2HBB structure. In this work data about residue 14 in the Φ-value analysis would be expected to be affected by the additional contacts present in the native structure 1DIV. Specifically, due to these extra native state contacts, the calculated Φ-values for residue 14 could potentially be artificially low. Since native state contacts about residue 17 were not particularly reliable the Φ calculations for this residue could possibly be affected.

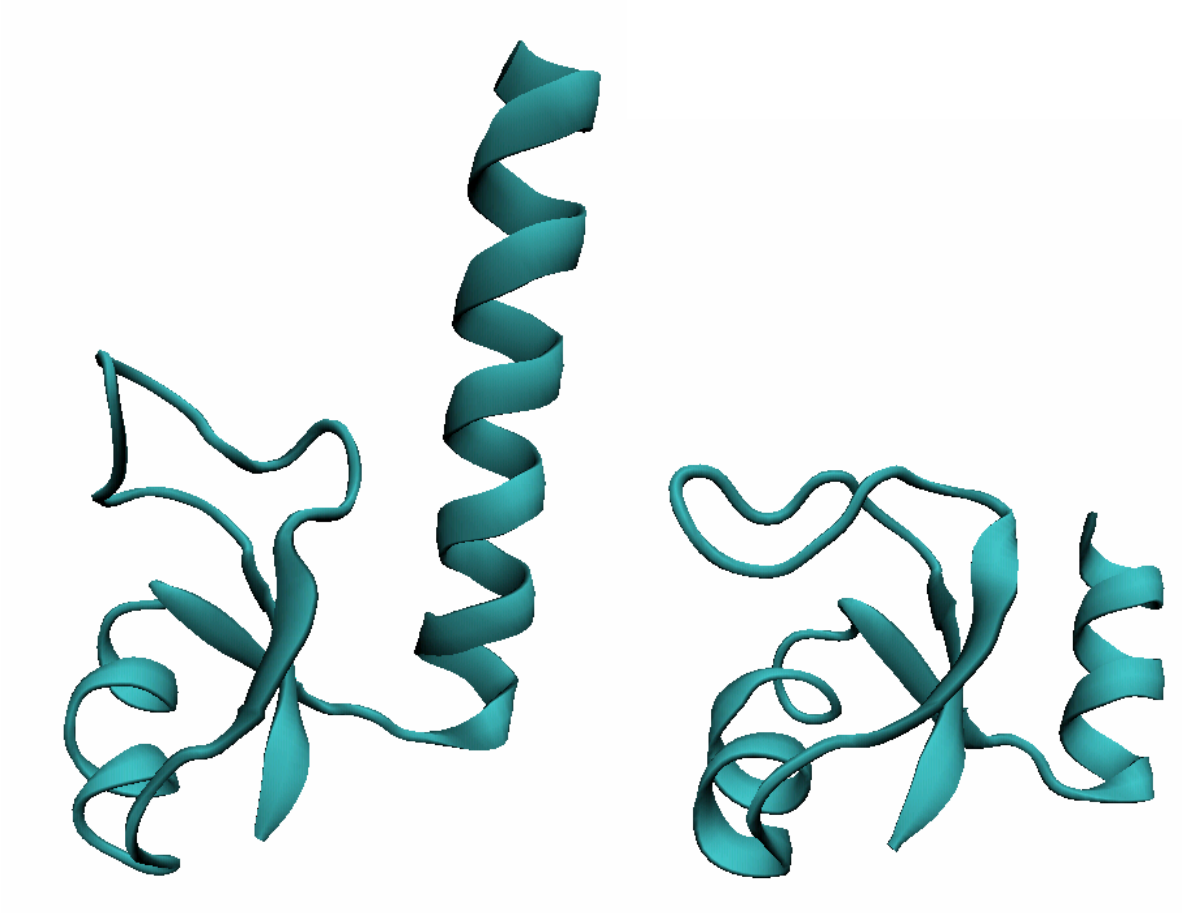


Figure 5-2: Cartoon representation of 1DIV (left) and 2HBB (right).

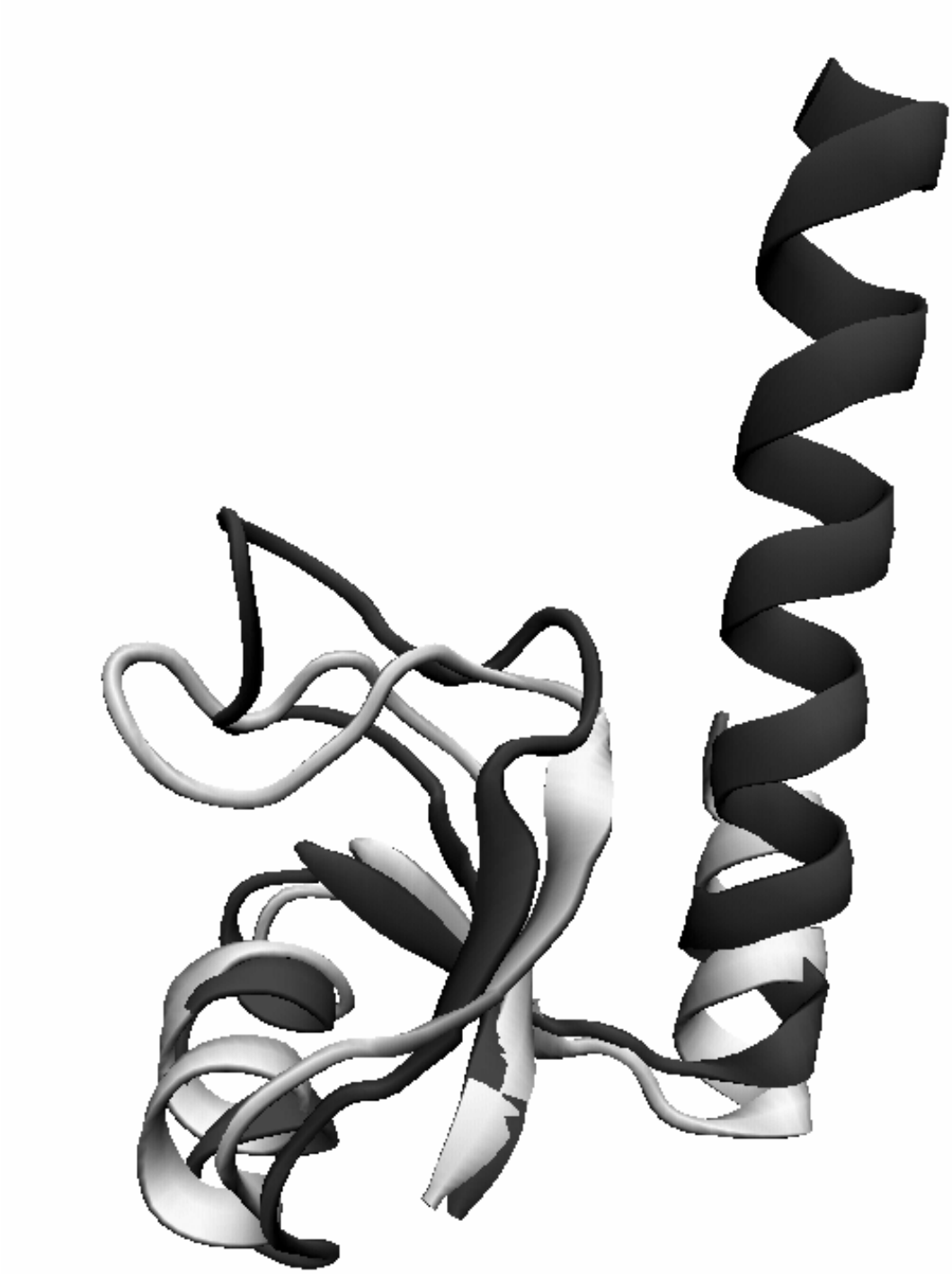


Figure 5-3: 1DIV and 2HBB over laid structures. 1DIV crystal structure derived at 2.6Å and 2HBB derived at 1.9Å.

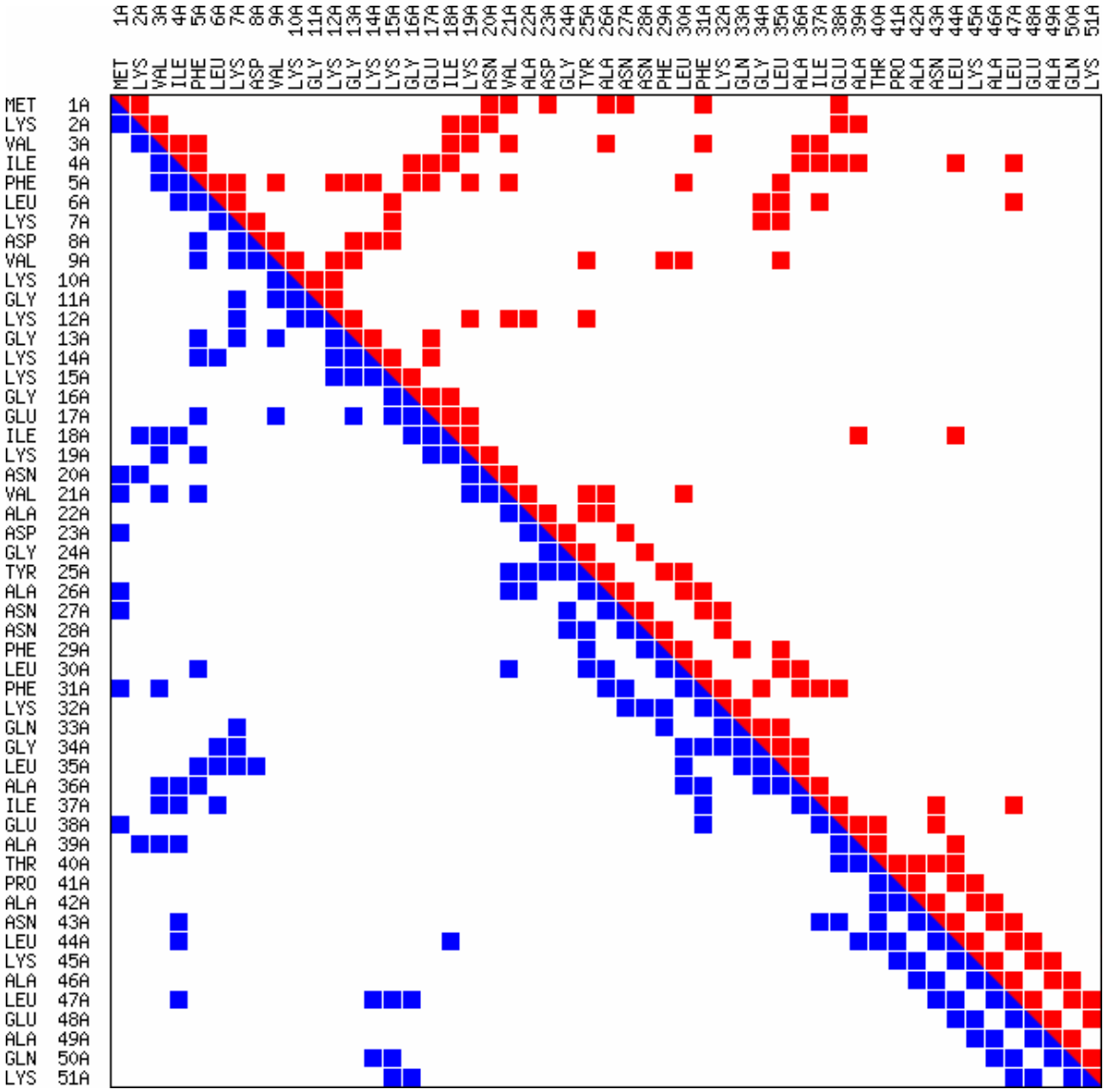


Figure 5-4 Contact map for 1DIV 1-52 and 2HBB structures of NTL9.
1DIV shown in dark color and 2HBB in light color. Contacts defined within 4.2Å cutoff. Significant differences are noted for contacts between residues 10-17 and 46-51, present in 1DIV and not in 2HBB.

5.2 Methods

5.2.1 Identification of the TSE

A TSE of unfolding made up of 5 TS conformations was identified from an ensemble of 100 unfolding simulations. The 5 TS conformations were determined by a protocol involving the Φ-RMSD method [199, 200], described in more detail in section 5.2.5. Initially, the 100 unfolding simulations were screened for the lowest Φ-RMSD structures, a Φ-RMSD of 0.35 or less was considered acceptable as is common in the literature [199, 200]. Following identification of those structures with a correlation of 0.78 or better correlation of Φ_{calc} and Φ_{exp} were chosen as representative structures. (Figure 5-5) These structures appear in the upper left corner of Figure 5-5. The remaining 95 had lower correlations when looking at their lowest Φ-RMSD value and thus a lower probability of being a member of the TSE.

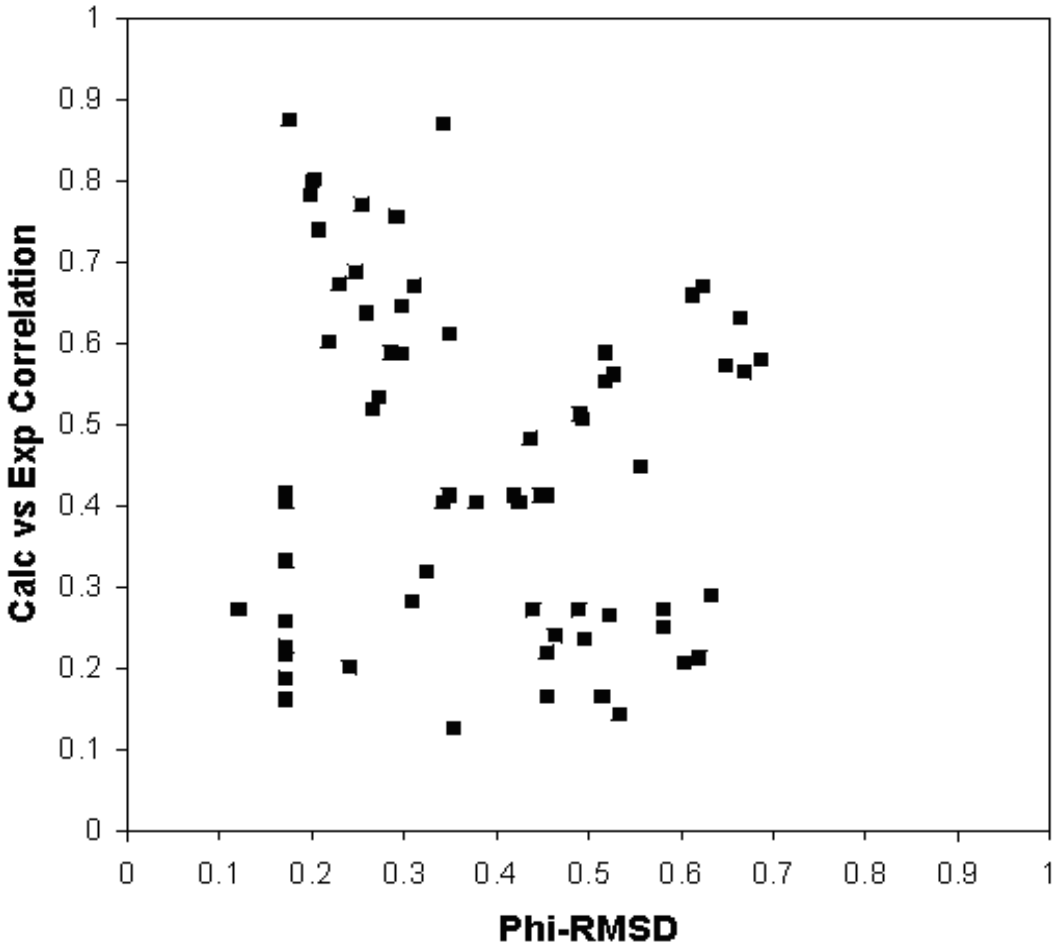


Figure 5-5: Correlation between calculated and experimental Φ-values versus Φ-RMSD for all positive values.
Each point represents the best structure (as determined by lowest Φ-RMSD value) from each of the 100 unfolding trajectories. Structures populating the upper-left region of the space represent best possible transition state conformations.

Following identification of the TSE by these methods structures were confirmed by calculated β_T values [201-203], and correlation of calculated Φ-values versus experimental Φ-values. Although, p_{fold} calculations [204, 205] would be ideal for confirming TSE, performing a significant number of the simulation in order to obtain the results would be computationally prohibitive considering the size of NTL9.

5.2.2 Structure Setup and Equilibration Procedures

The structure of NTL9 was derived from the L9 protein 2.6Å crystal structure in the protein data bank (code 1DIV) [174]. The L9 crystal structure was truncated at residue 60 and used as the starting structure for the simulations. NTL9 is formally defined as residues 1-56, however because residue 56 is in the middle of the α-helix linker we used 1-60. This structure extended by four residues from NTL9 (1-56) allowed the C-terminal domain helix to fray slightly in simulation while exerting minimal impact on the representation residues in the 1-56 region. The more recent and higher resolution NMR models of NTL9 (code 2HBB) [180], would not have facilitated this. The MD simulations were preformed with the AMBER 8 MD Package [206], using the AMBER03 Force Field and the Generalized Born [207, 208], implicit solvent model. The modified bondi radii, mbondi2 was used to improve GB performance [208]. The initial coordinates were prepared by first removing the hydrogen atoms and building the system in the leap module in AMBER. Prior to production simulations the coordinates were equilibrated to 300K in the GB solvent environment. Equilibration was achieved by an alternating series of restrained 250 cycle minimizations of increasingly weak restraints and 100 steps of molecular dynamics at increasing temperatures. The equilibration was ended with a final 100 steps of unrestrained MD simulation. Production unfolding MD simulations for analysis consisted of simulation after this point at 350K.

5.2.3 Unfolding Simulations

A series of simulations at a variety of temperatures were performed to determine the optimal temperature for unfolding. The optimal conditions were defined as the lowest

temperature to induce unfolding that still provided rapid sampling of the unfolding pathway yet retained sufficient resolution of conformation in the unfolding trajectory. Unfolding was tested by MD simulation on the equilibrated structure at 200K, 250K, 300K, 350K, 400K, 450K and 500K. (Figure 5-63) These initial high temperature MD simulations were performed for 0.8ns until significant unfolding was observed. It should be noted that the unfolding time noted in the MD simulations should not be interpreted as a realistic measure of actual unfolding time because implementation of the GB implicit solvent model neglects the effects of solvent viscosity, and so conformations are sampled much more rapidly. The degree of unfolding was determined by RMSD of the simulation trajectory to the equilibrated native conformation. From these simulations it was determined that 350K provided the ideal compromise between speed of sampling the unfolding pathway and resolution of detail in the unfolding process. Moreover, the unfolding temperature chosen, 350K also represented a realistic unfolding temperature of NTL9 which is not always the case with temperatures used in unfolding simulation [178].

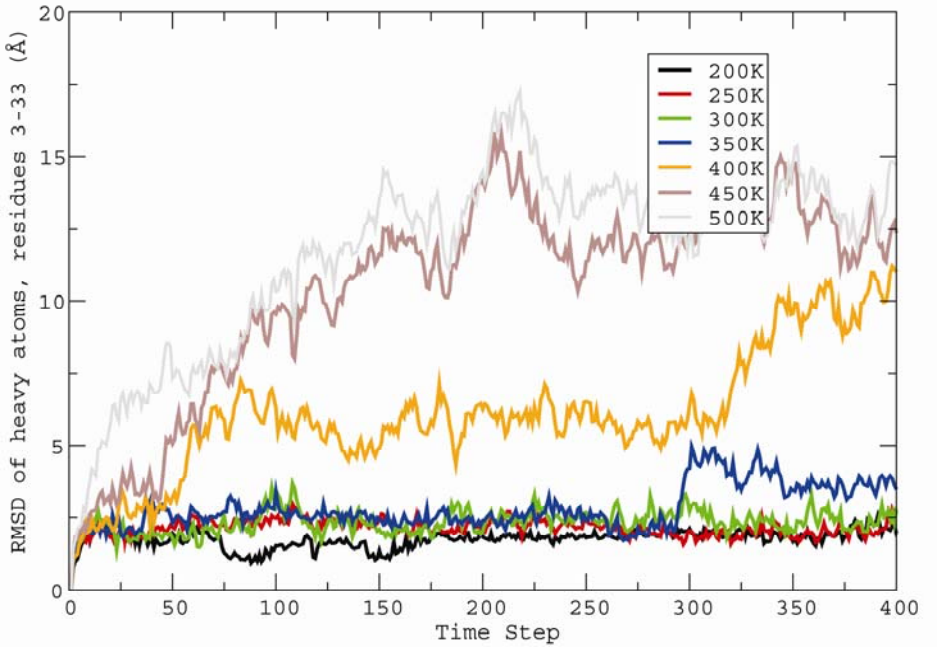


Figure 5-6: Determination of the optimal unfolding temperature for NTL9.
The RMSD of heavy atoms for NTL9 residues 3-33 versus time for different temperature simulations, 200K-500K are represented.

5.2.4 Unfolding Ensemble

The unfolding ensemble was created by sampling 100 independent MD simulations at 350K of 4ns each. The 100 simulations were run from different starting velocities as assigned by the random number generator function in AMBER. Each unfolding simulation was run in GB at 350K in implicit solvent with a 0.2fs time step with a 10Å van der Waals cutoff. High temperature unfolding was proposed in this study as a rapid and cost effective sampling method as is common in the literature [36, 38, 50, 58]. High temperature and implicit solvent (i.e.: limiting frictional forces) increased the rate of sampling of conformations on the PES and some but not every simulation in the ensemble was expected to sample the 300K TSE conformations.

In addition to identifying the five TS structures described, their coordinates were used to cluster all structures in our 100 unfolding trajectories. Clustering was based on Cα RMSD of residues 3-35 to each of the 5 TS structures identified. Structures within our unfolding ensemble of more than 200000 structures were considered part of a TS cluster if they had a RMSD of less than 1.5Å to any of the 5 previously identified states. Clustering revealed 114 structures, the properties of which are listed in Table 5-3.

5.2.5 Φ-Values and Φ-RMSD

The apparent deviation between the experimentally determined TSE and the simulation structure was monitored by calculating the root-mean-square deviation (RMSD) between the experimental Φ-values and calculated Φ-values for each structure in the unfolding trajectory [199, 200]. Experimental Φ-values employed in this work were derived from corrected Φ-value calculations that measured the change in free energy of the TS versus the native state due to a mutation at a particular residue. In the corrected calculation the denatured state effects are included so that Φ is defined as in Equation 5-1.

$$\Phi = \frac{(\Delta\Delta G_{mt}^{TS} - \Delta\Delta G_{wt}^{TS}) - (\Delta\Delta G_{mt}^{D} - \Delta\Delta G_{wt}^{D})}{(\Delta\Delta G_{mt}^{N} - \Delta\Delta G_{wt}^{N}) - (\Delta\Delta G_{mt}^{D} - \Delta\Delta G_{wt}^{D})} \qquad (5\text{-}1)$$

This is analogous to the computational case where all the denatured state contacts are included.

Computationally we defined Φ-values (Φ_{calc}) in equation 5-2, where N^{Conf} is the number of contacts for a particular residue in the conformation of interest while N^{Folded} is the number of native contacts in the folded native state of the protein.

$$\Phi_i^{calc} = \frac{N_i^{Conf}}{N_i^{Folded}} \tag{5-2}$$

Calculated Φ-values should theoretically fall between 0 and 1, however should more contacts be made in a non-native state than in the native state the calculated Φ-value may be greater than 1. The Φ-RMSD was defined as described in Equation 5-3 where M_Φ is the total number of residues being reported.

$$\Phi - RMSD = \sqrt{\frac{1}{M_\Phi}\sum_I \left[\Phi_i^{calc} - \Phi_i^{\exp}\right]^2} \tag{5-3}$$

Φ-values for 14 of 19 residues with experimentally reported Φ-values were included in our calculations (Table 5-1). Residues V9, A22, A39, T40 and A42 were not included in the calculations because they did not form enough contacts in the native state calculations. Native contacts for the calculated Φ-values were determined to be those contacts that occurred in 2/3 of the contacts averaged from five, 1ns, 300K simulations. A lack of contacts for V9, A22, A39, T40 and A42 may have been observed because these residues exist in more flexible regions of the protein or are small in size possibly limiting the number of contacts observed in MD simulation. V9 (not included) for instance resides on the edge of a turn, while A22 has a small side chain. A39, T40 and

A42, all reside at the loop region between the C-terminal α-helix and the folded N-terminal structure. Contacts were counted with the contact.pl script from the MMTSB [209], protein modeling tool kit software. By this method residues are considered in contact when the minimum inter-residue distance of all pairs of heavy atoms is less than 4.2Å.

Table 5-1: Calculated Φ-Values

Residue	Experiment	TS1	TS2	TS3	TS4	TS5
1	*0.62*	0.60	1.20	0.60	0.60	0.60
3	*0.60*	0.75	0.94	0.56	0.56	0.75
4	*0.61*	0.80	0.53	0.80	0.80	0.80
6	*0.43*	0.59	0.59	0.59	0.59	0.59
7	*0.05*	0.00	0.00	0.00	0.00	0.00
8	*0.18*	0.00	0.00	0.00	0.00	0.00
14	*0.10*	0.00	0.00	0.00	0.00	0.00
17	*0.18*	0.56	0.00	0.00	0.56	0.56
18	*0.51*	0.57	0.57	0.57	0.57	0.57
21	*0.63*	0.62	1.24	0.62	0.62	0.62
35	*0.59*	0.54	0.54	0.54	0.54	0.54
36	*0.54*	0.56	1.12	0.56	0.56	0.56
37	*0.51*	1.01	1.01	1.01	1.01	1.01

Table 5-1: Calculated versus experimental Φ-values of NTL9 residues investigated. Calculate Φ-values greater than 1 represent situations where more residual contacts are noted in the TSE than in the native state.

5.2.6 SASA and β_T Analysis

Calculation of the β-Tanford (β_T) value is one way to validate potential TS conformations [201, 210]. It has been shown that m-values from the kinetic and equilibrium experiments correlate with solvent accessible surface area (SASA) and can be combined to generate a dimensionless parameter (β_T) that reports on the relative compactness of the transition state ensemble [201, 203]. Mathematically, β_T can be represented by:

$$\beta_T = \frac{m_{fold}}{m_{fold} - m_{unfold}} \tag{5-4}$$

For a two-state model the m-values from experiment can provide a value for β_T where:

$$\beta_T = 1 - \frac{m_{off}}{m_{on} + m_{off}} \tag{5-5}$$

Due to the correlation between SASA and m-values we can arrange an estimated β_T value as follows:

$$\beta_T = \frac{SASA_{Unfolded} - SASA_{TSE}}{SASA_{Unfolded} - SASA_{Folded}} \tag{5-6}$$

We determined the SASA of the unfolded state by taking an average of the SASA of the structure at the end of each of the 100 trajectories. $SASA_{TSE}$ is the SASA value in angstroms squared ($Å^2$) as calculated by the naccess program [211], of any structure along the trajectory. The naccess default radii were used in this calculation. $SASA_{Folded}$ was derived from the SASA of the initial equilibrated simulation structure from which all 100 unfolding trajectories started.

5.3 Results and Discussion

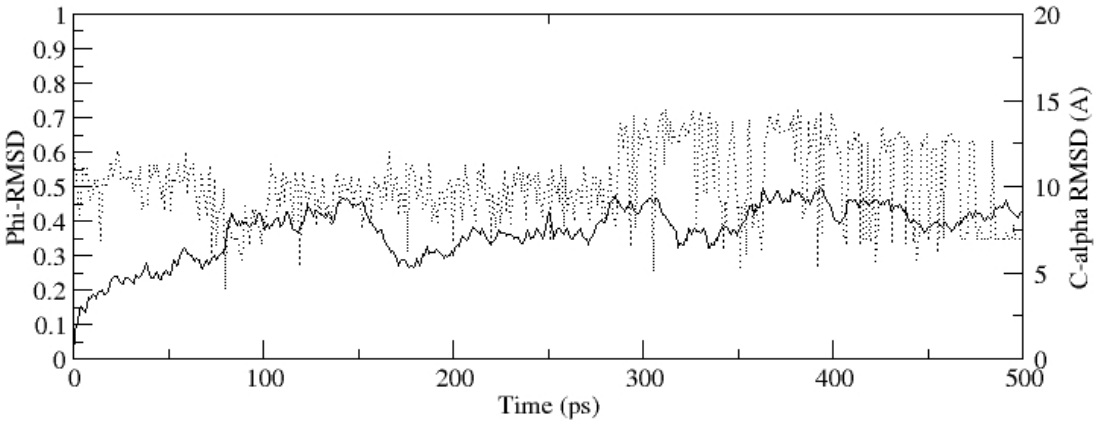
5.3.1 The TSE of Unfolding

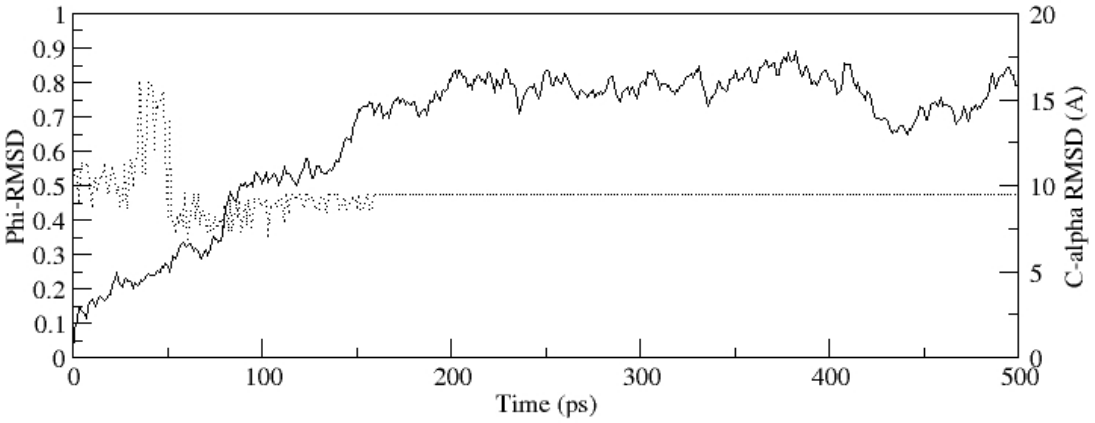
A TSE of unfolding constituting 5 TS conformations was identified from an ensemble of 100 unfolding simulations. The five TS conformations were determined by examining the ensemble of unfolding simulations for the lowest Φ-RMSD structures with the best correlation of Φ_{calc} and Φ_{exp}. To verify that identified conformations were in fact transition state structures, we analyzed several parameters in comparison to experimentally known values. Each of the 5 TS structures will be denoted as TS1, TS2, TS3, TS4 and TS5.

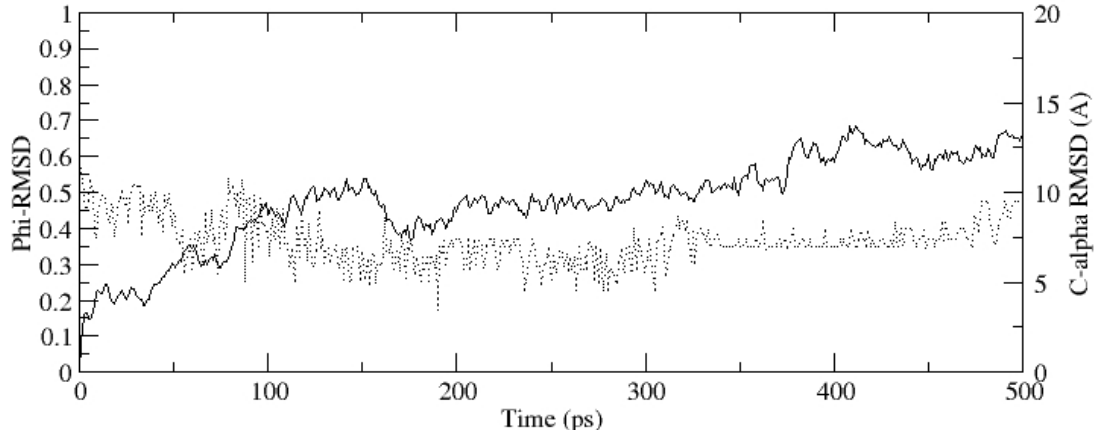
The structures of the proposed TS conformations in the unfolding trajectories were found to be consistent with what is known about transition states in general. All TS structures identified occurred rapidly in the unfolding pathway which is also consistent with what is known about the location of transition states on folding pathways [163, 172, 177]. This is an expected characteristic of the TS because in general the major transition state of folding/unfolding is believed to be a distorted form of the native state [212]. Moreover, for NTL9 this is expected since it is known that the TS state contains significant of secondary character [193].

Examination of the C_α-RMSD versus time data shows the locations of the identified TS correspond to large changes in RMSD. This is consistent with large structural motions which are expected to occur around the TS. (Figure 5-4) Moreover, the largest C_α-RMSD changes correspond well with the lowest Φ-RMSD. The C_α-RMSD changes were particularly evident in TS1, TS2 and TS3. (Figure 5-9) The changes in C_α-

RMSD in the TS4 and TS5 were present but of a much lower magnitude than the others. This smaller change may have been a result of their slightly later position in the unfolding pathway relative to the other TS conformations we identified. RMSD changes that occur later in the trajectory would not reflect the same magnitude in change due to the compounded changes in which other components of the protein unfolded adding to the RMSD data.


Phi-RMSD
C-alpha RMSD (A)
Time (ps)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
0
100
200
300
400
500


Phi-RMSD
C-alpha RMSD (A)
Time (ps)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
0
100
200
300
400
500


Phi-RMSD
C-alpha RMSD (A)
Time (ps)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
0
100
200
300
400
500

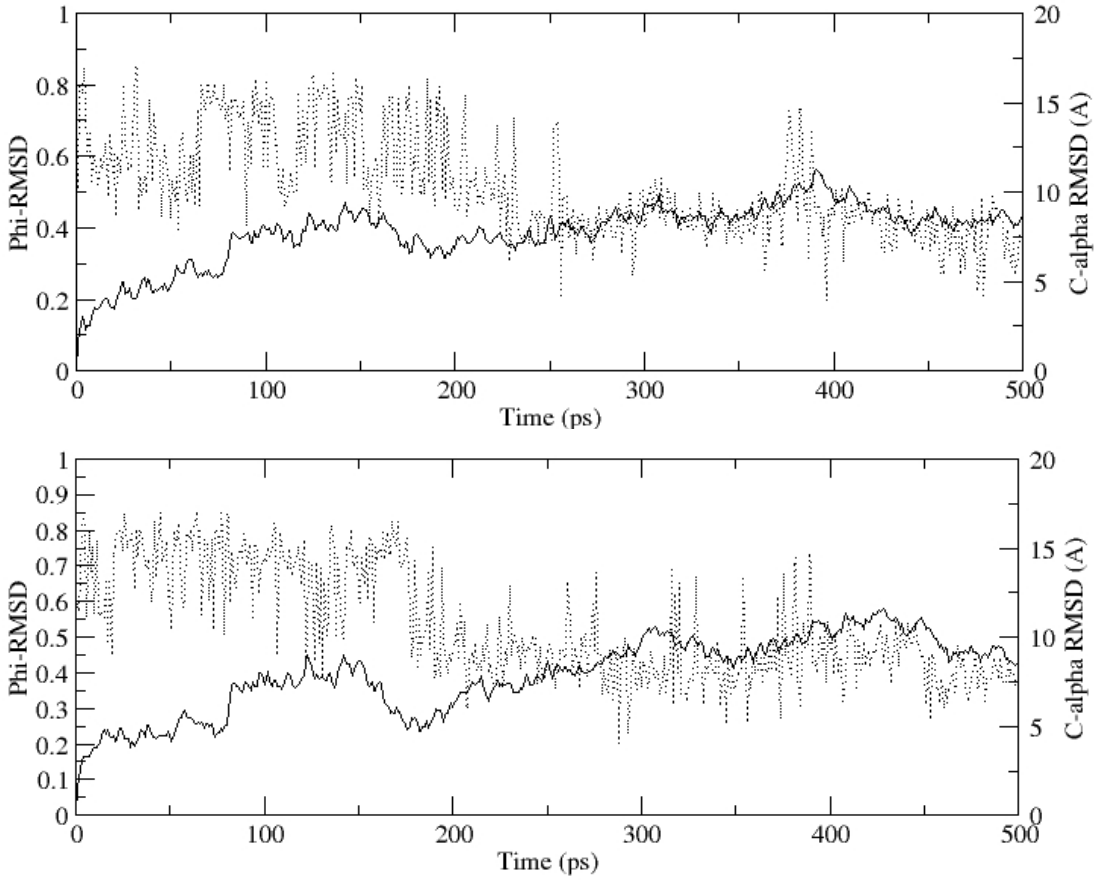
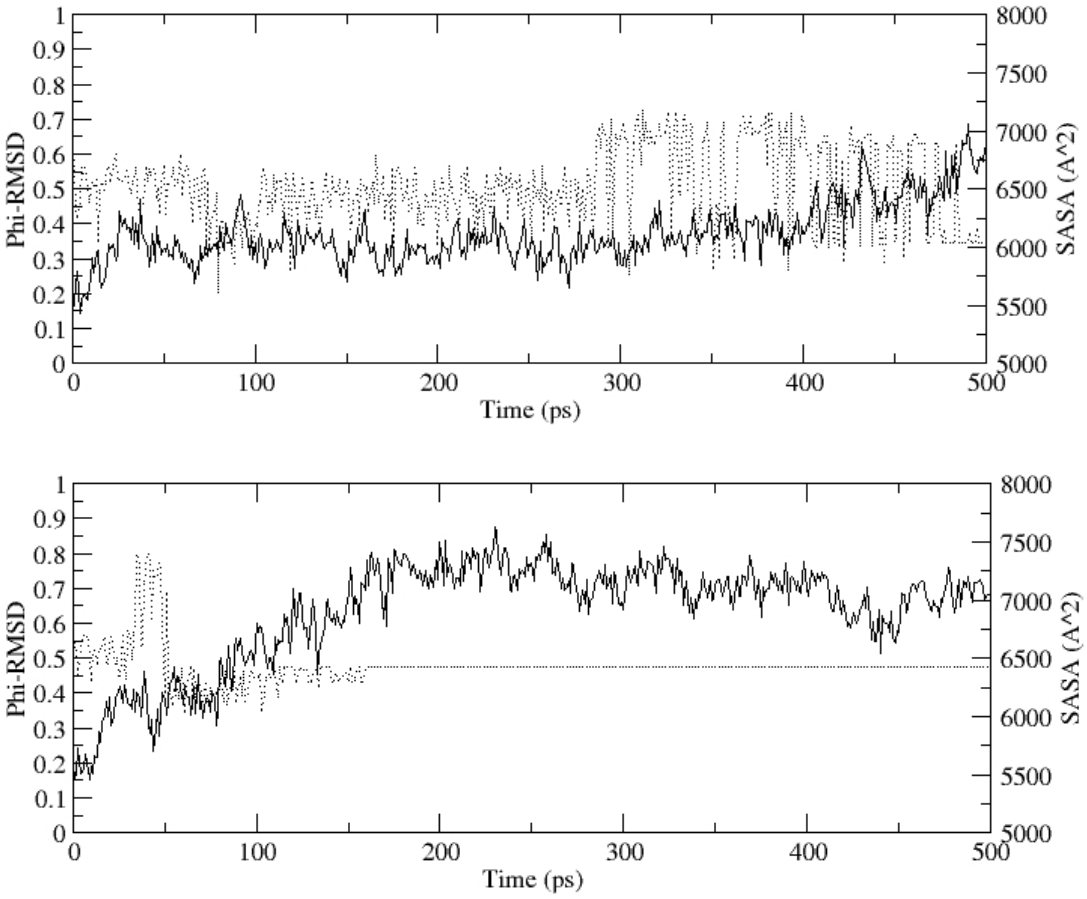


Figure 5-7: Identification of the members of the TSE along trajectories at 350K by Φ-RMSD versus structural RMSD.
The plots represent Φ-RMSD, light lines, (see methods) versus structural RMSD, dark lines, of the C-α relative to the initial simulation structure. (*Top to Bottom* TS1, TS2, TS3, TS4, TS5)

The change in solvent accessible surface area (SASA) was also monitored for each of the unfolding trajectories and plotted against Φ-RMSD. (Figure 5-8) In this analysis it was expected that an increase in SASA corresponded to the protein structure change towards unfolding by loss of contacts made in the TS. The changes in SASA were consistent for all identified TS conformations at which point small changes in SASA between 300Å^2 and 800Å^2 were observed around the identified TS. (Figure 5-8) This observation is inline with other experimental work suggesting the TS structure of NTL9 to be relatively compact and contain significant secondary structure [182, 193].

Small changes in SASA are generally attributed to the dissolution of non-covalent interactions stabilizing the transition state, where as large changes are commonly caused by changes in hydrophobic interactions. This observation is also in agreement with the previous work suggesting that electrostatic interactions are only weakly formed in the TS of NTL9 [189].



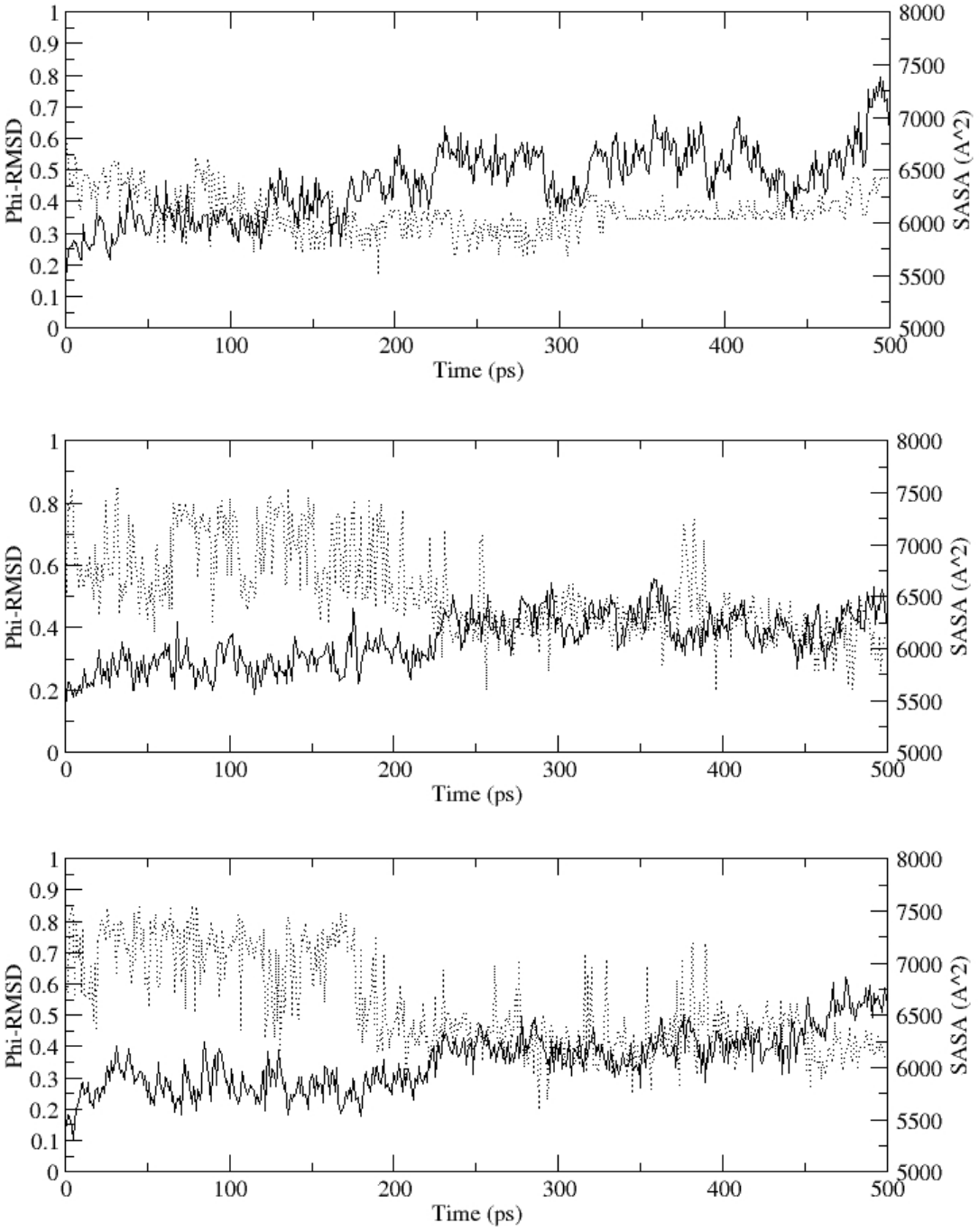


Figure 5-8: Identification of the members of the TSE of NTL9 along unfolding trajectories at 350K by Φ-RMSD versus SASA.
The plots represent the Φ-RMSD, light lines, verse the absolute SASA (Å^2) dark lines. (*Top to Bottom* TS1, TS2, TS3, TS4, TS5)

5.3.2 Validation of the TSE

Validation of the identified TS conformations is supported by the correlation between the Φ_{calc} and Φ_{exp} for the 14 residues analyzed. The average correlation among the 5 structures was 0.83, with a range of between 0.78 and 0.88 using the Pearson linear correlation method. (Table 5-2) Additionally, R^2 analysis of the experimentally determined versus calculated Φ-values revealed an acceptable range of between 0.64 and 0.77. (Figure 5-11 and Table 5-2) Moreover, the trends in calculated Φ-values were consistent across each residue for all of the identified TS structures. (Figure 5-9) The calculated TS β_T values were similar to the experimental value of between 0.6 an 0.7 [213], and had an average of 0.60. (Table 5-2) This method further supports the selected TS conformations since the β_T value derived from experiment and calculation good agreement. In addition to these properties, the properties of an ensemble of 114 structures created through clustering of the original five structures, highlights variation in the TSE. (Table 5-3)

Table 5-2: Properties of the TS Conformations

TS	Time (ps)	P	R^2	SASA (Å^2)	β_T	Radius of Gyration (Å)	C_α RMSD (Å)
TS1	81	0.80	0.64	5925	0.69	16.93	6.7
TS2	62	0.87	0.76	6274	0.58	16.56	6.5
TS3	191	0.88	0.77	6505	0.50	17.11	8.4
TS4	397	0.78	0.61	6083	0.64	18.59	10.3
TS5	289	0.80	0.64	6274	0.58	16.29	9.0
Average	204	0.83	0.68	6212	0.60	17.10	8.2

Table 5-2: Properties of the TSE conformations of NTL9.
The average results determined from average values for each of the clusters for each TS identified.

Table 5-3: Transition State Ensemble Properties

	Radius of Gyration (Å)	SASA (Å^2)	C_α RMSD (Å)	Fraction of Native Contacts
Average	16.82	6167	9.0	0.29
Min	14.50	5817	4.7	0.20
Max	19.27	6574	11.9	0.41
Std Deviation	1.22	160	1.7	0.05

Table 5-3: Transition State Ensemble Properties of NTL9. The average results determined from average values for ensembles of 114 TS structures.

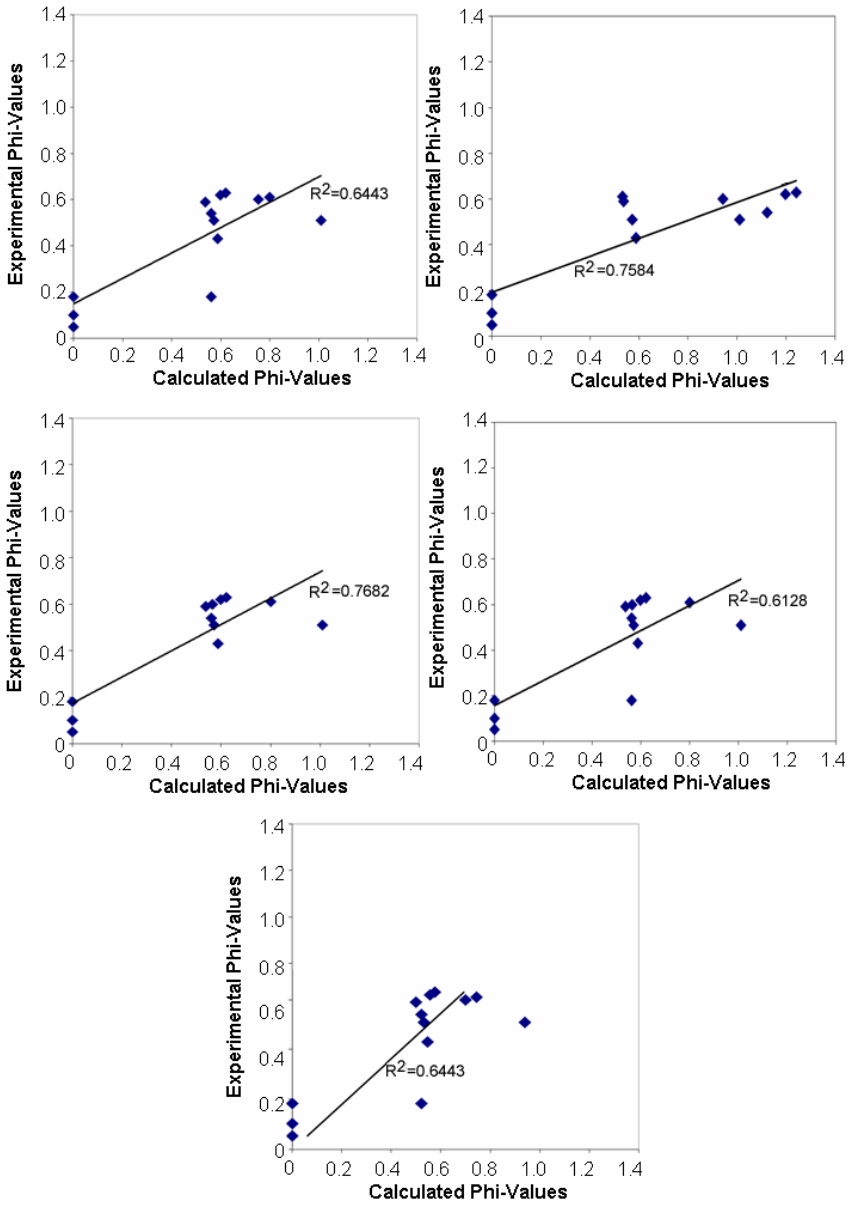


Figure 5-9: The R^2 analysis of experiment versus calculated Φ-values for the TSE members of NTL9.
Top to Bottom TS1, TS2, TS3, TS4, TS5.

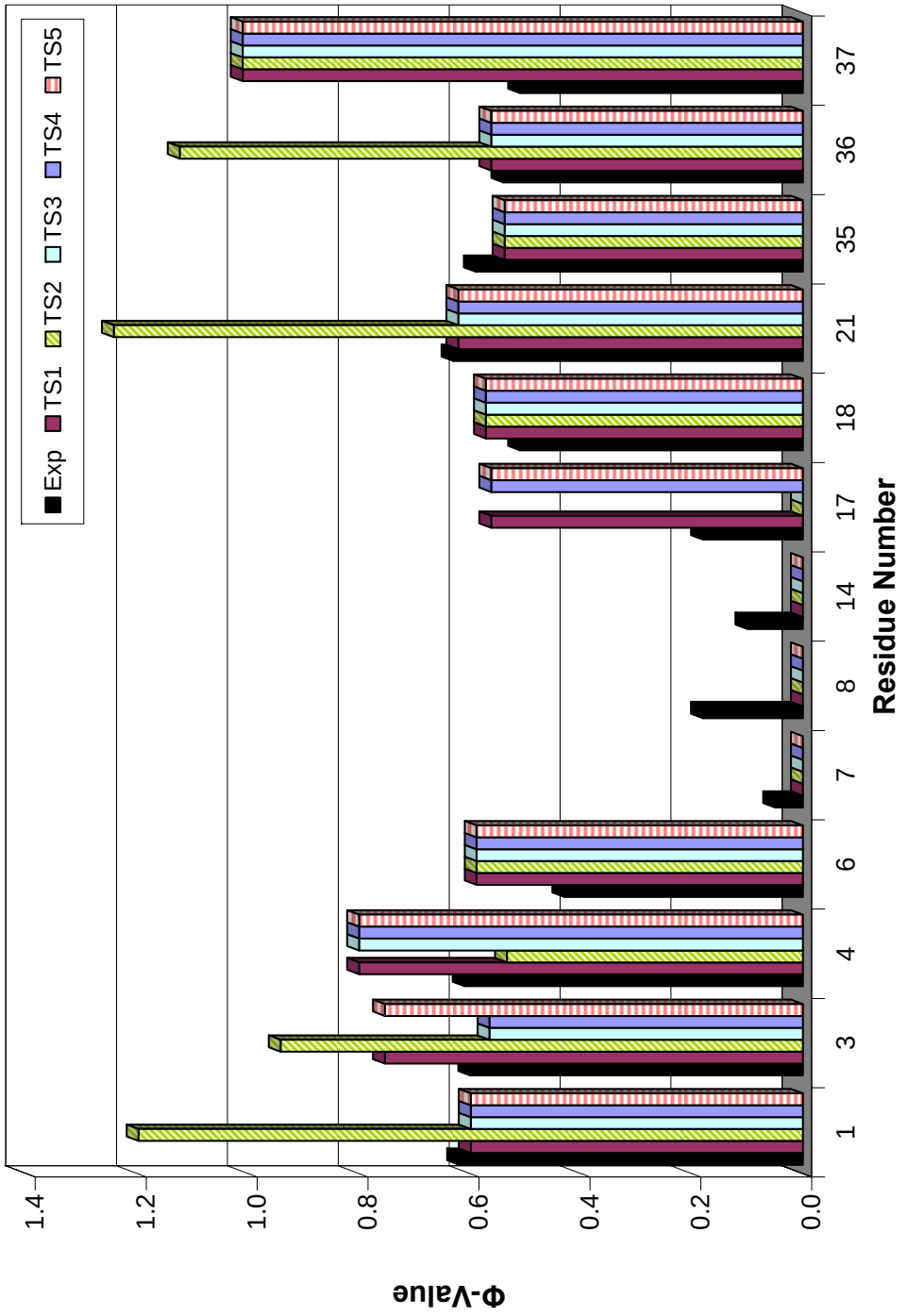


Figure 5-10: Experimental versus calculated Φ-values for NTL9.
The experimental Φ-values represented by the empty circles and are connected by a line for easier tracking by eye. The calculated Φ-values for each of the 14 residues identified are as follows TS1 triangles, TS2 asterisks, TS3 plus sign, TS4 squares, TS5 full circles (these overlap with TS1).

5.3.3 Properties and Characterization of the TSE

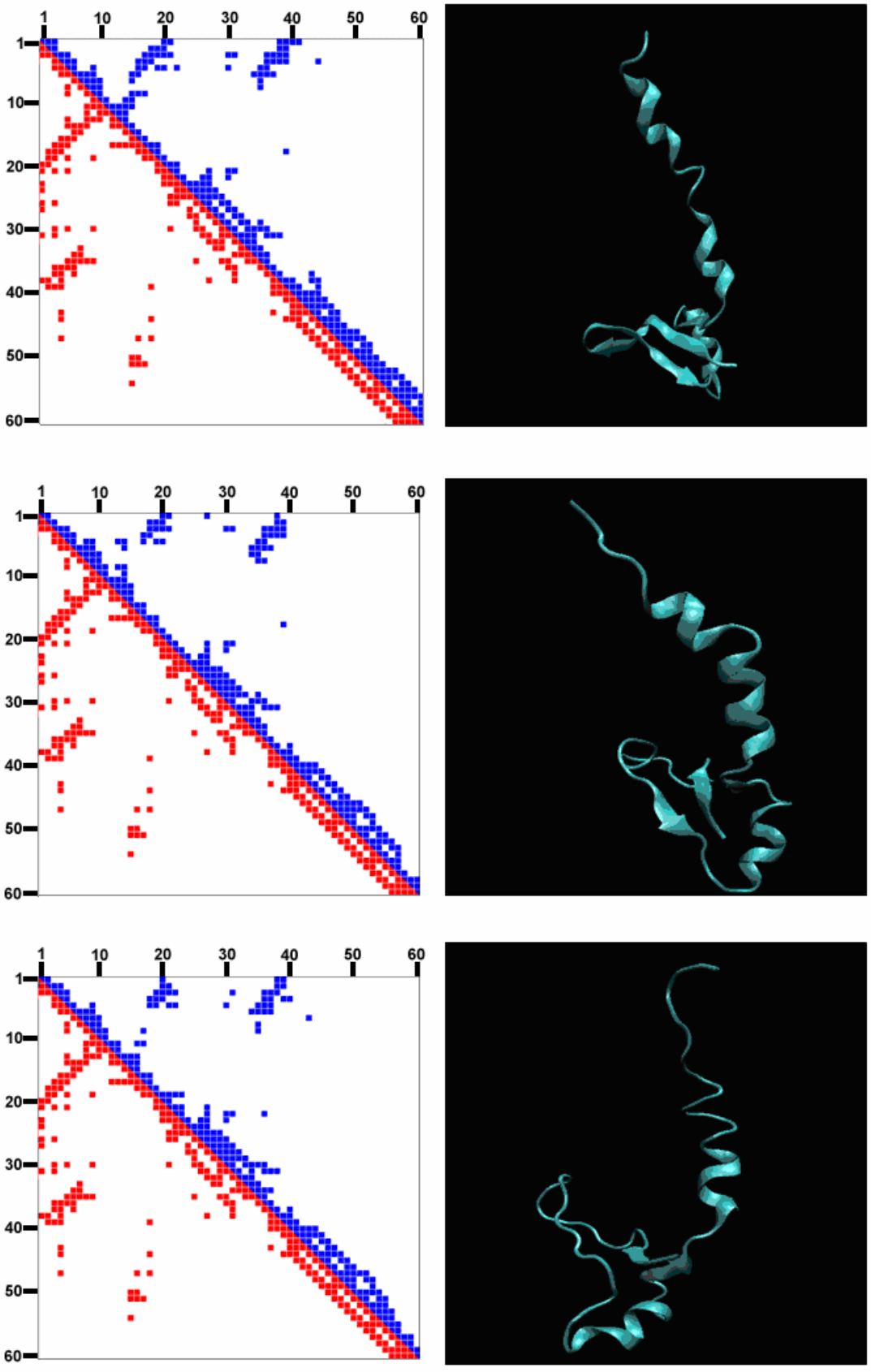
The properties of the 5 identified TS conformations are summarized in Table 5-2. Additionally, because these 5 structures do not represent an absolute description of the TS, a TS is better thought of as an ensemble of structures, average properties of the structures are also listed.

Following the identification of the TSE we examined and characterized the TSE relative to the native state of NTL9. The folded state was broken down into its β-sheet, loop region and helical elements. (Figure 5-1) These elements were analyzed for each of the TS structures.

5.3.4 β-strand 1

The residues on the first β-strand (Residues 1 to 7) in the NTL9 native structure form several contacts with residues of the second β-strand (Residues 16 to 21). Additionally, sitting between two β-sheet structures, residues 1 to 6 make contact with residues of the third β-strand (Residues 33 to 39). (Figure 5-1 and 5-11) The first β-strand is preserved in the TS containing nearly complete β-character according to secondary structure analysis by DSSP [214]. (Figure 5-12) In the case of TS1 this β-strand character was extended beyond residue 6 and until residue 12. Native contacts between this β-strand 1 and β-strand 2 are well maintained in the TSE. In the TS structures native contacts were lost between residues 1 and residues 22-24 which make up the beginning of the α-helix. (Figure 5-11) However, it is noted that in the TSE non-native contacts exist between residues 3 and 5 and residue 22 near the region between the α-helix and end of β-strand 2. These non-native contacts appear in TS4, however here

they occur with residue 21. (Figure 5-11) The novel contacts noted in the TSE may help support the inclusion of the final N-terminal region in the final folded state. Furthermore, this supports the high experimental Φ-value of residue 22 which was not included in by the calculated Φ-values.


1
10
20
30
40
50
60

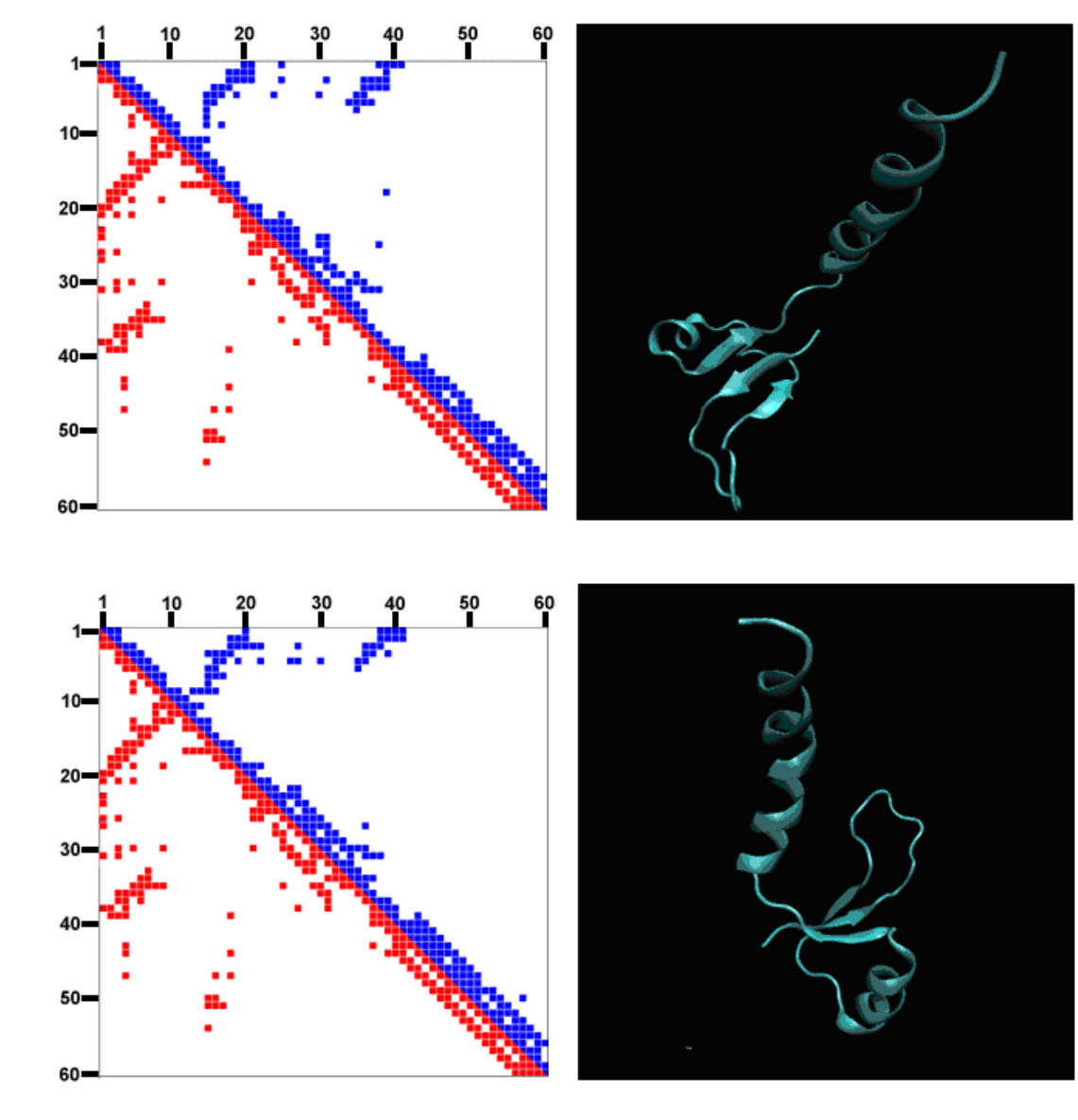
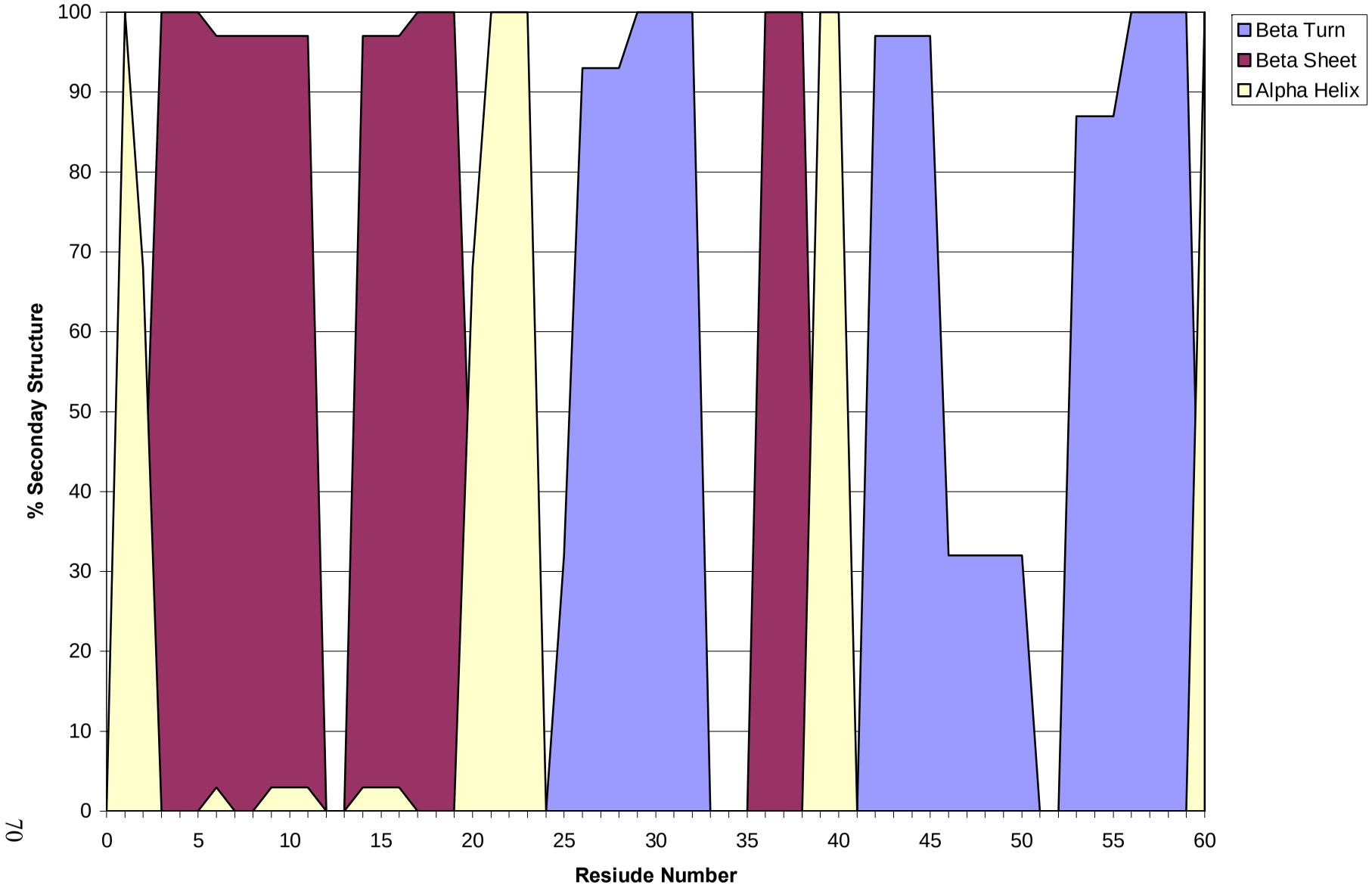
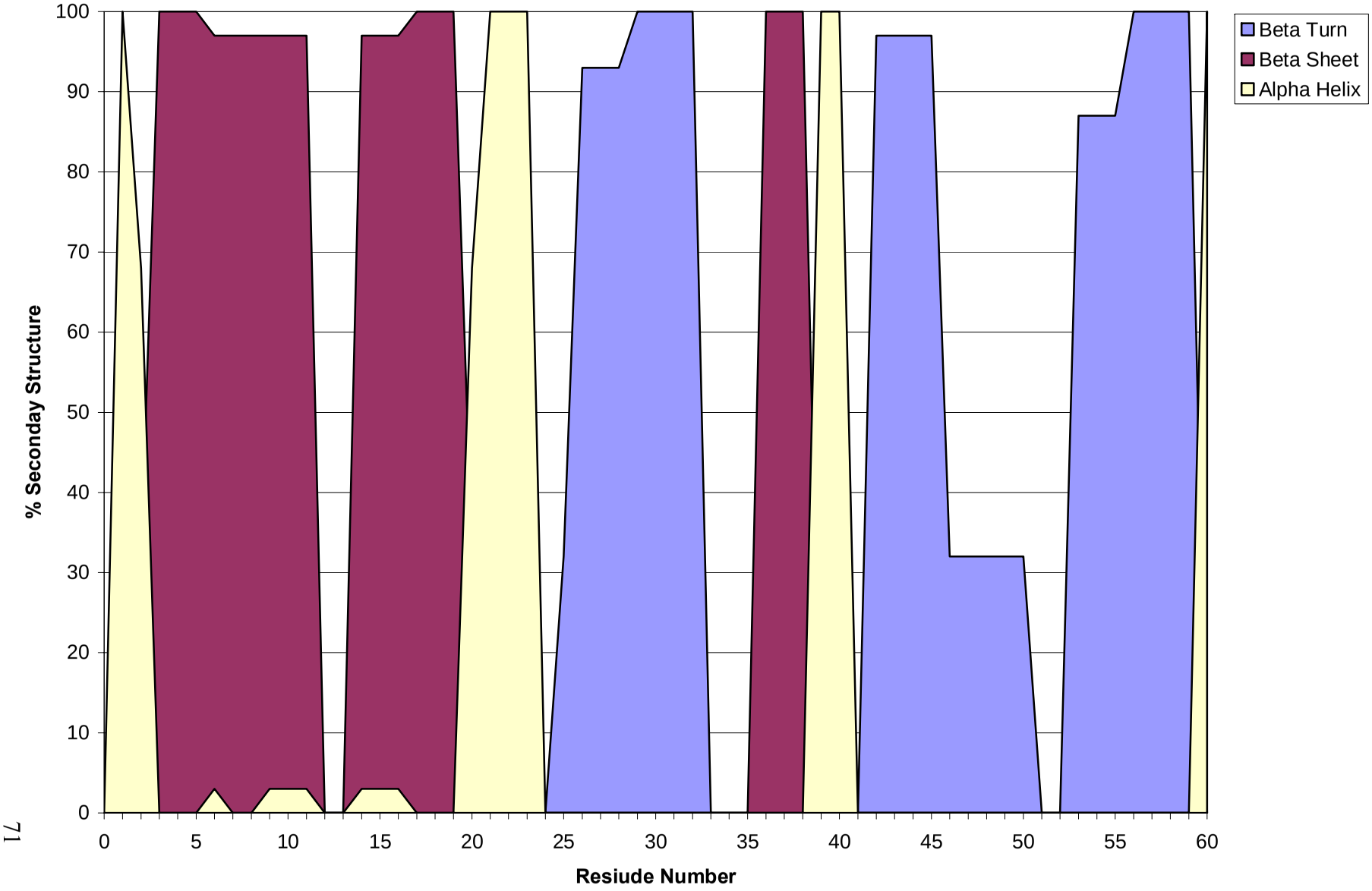
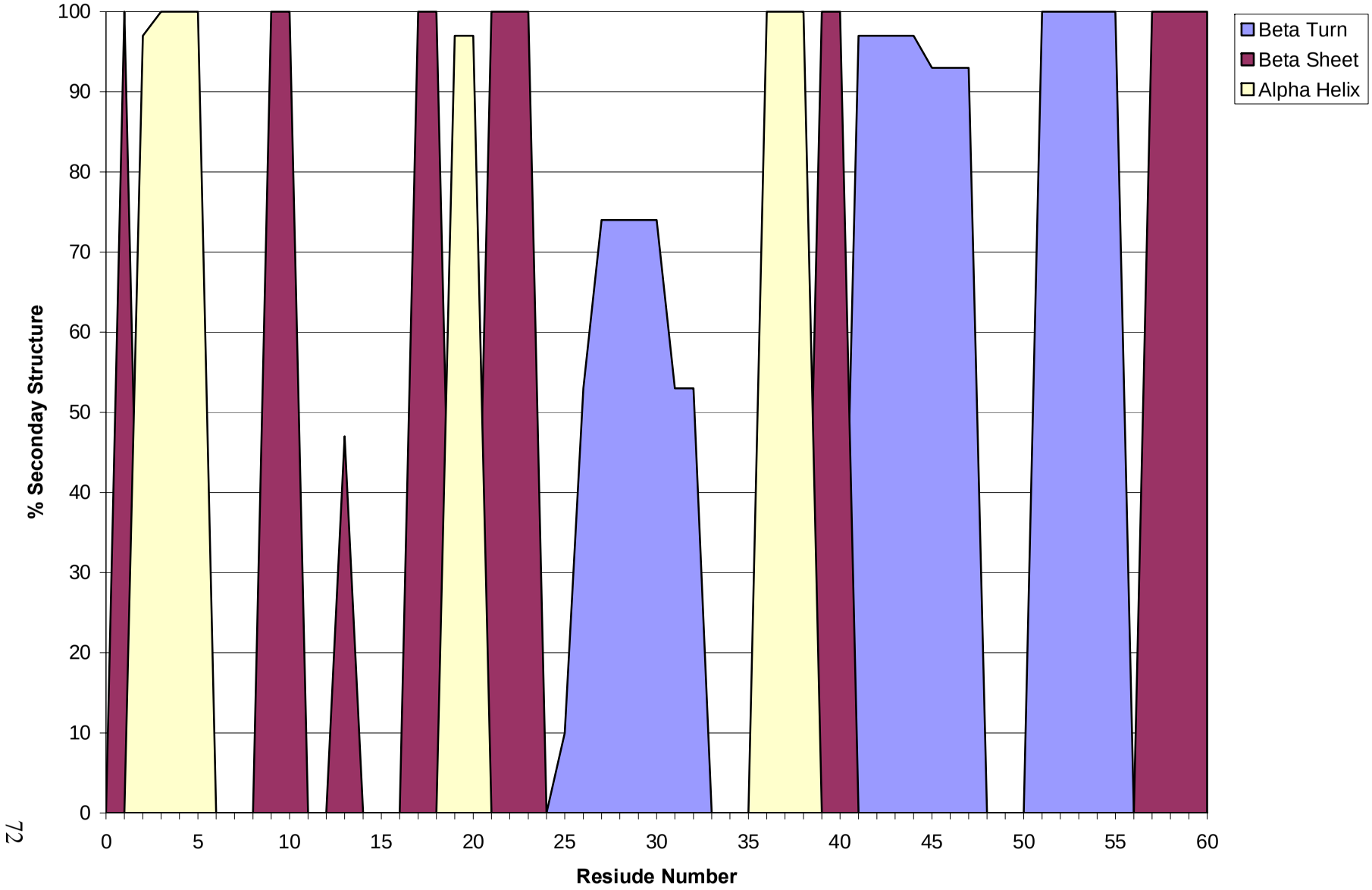
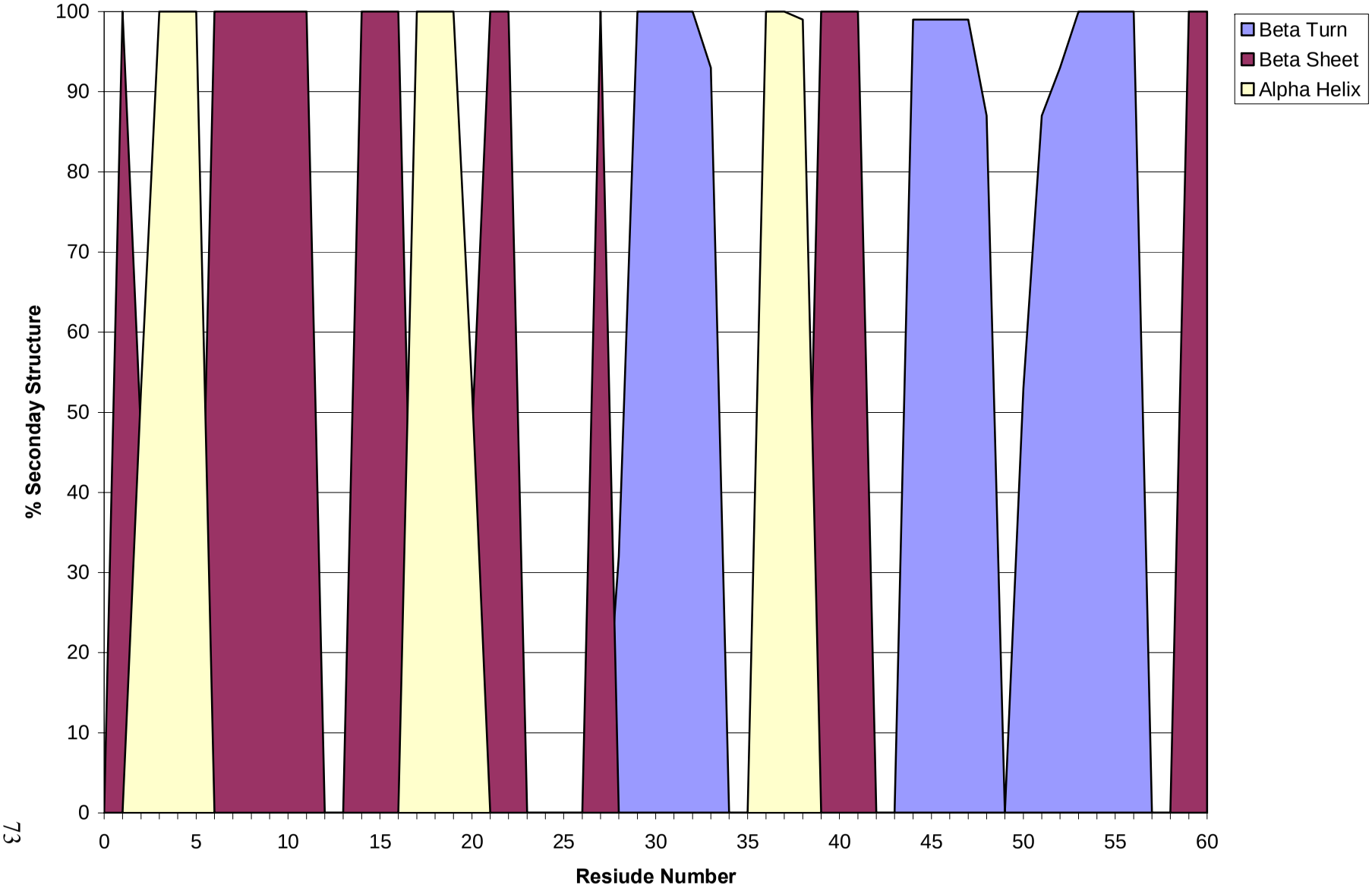


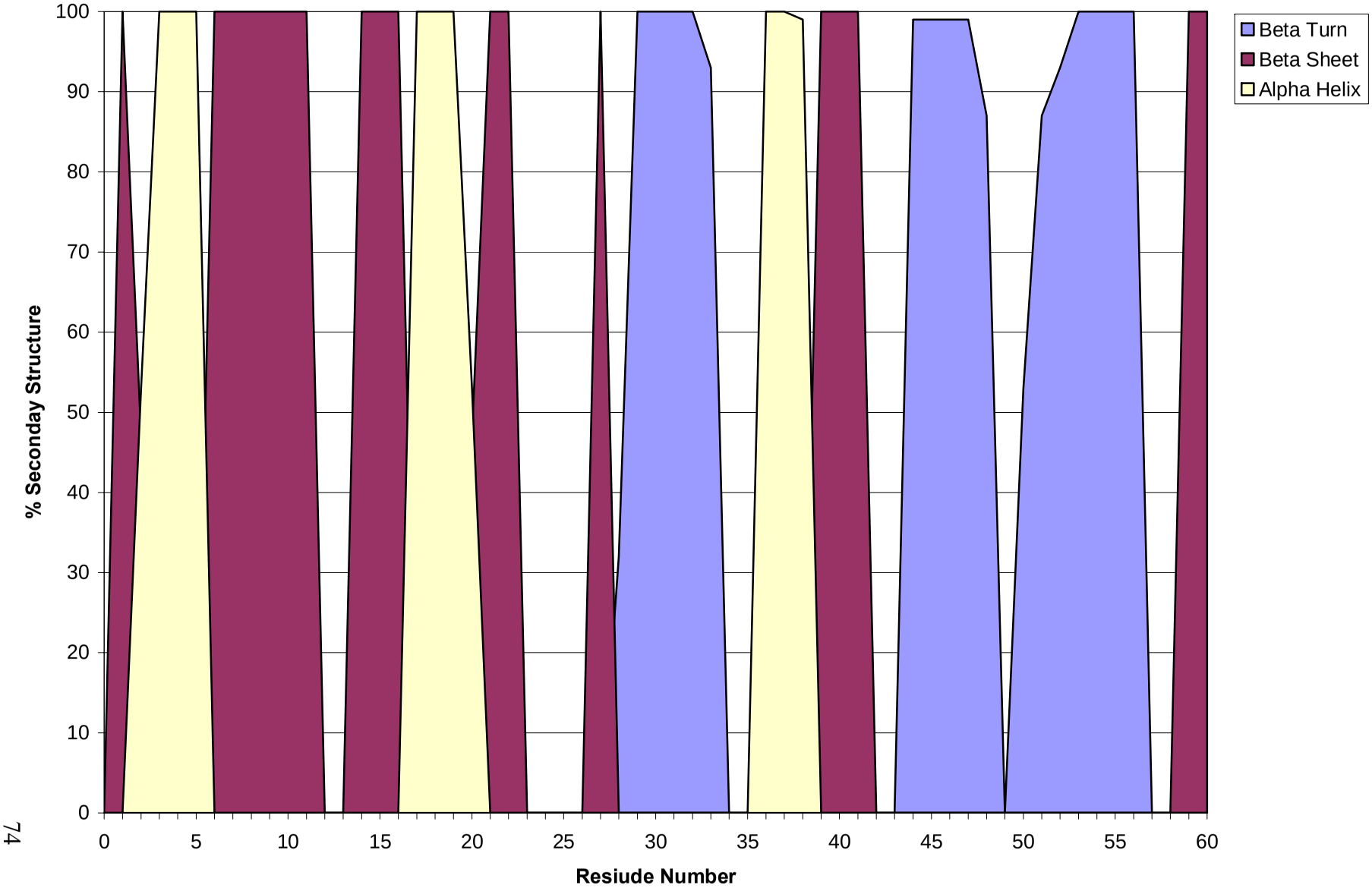
Figure 5-11: The contact maps for each of the TS structures in the TSE of NTL9. *Left* Contact map darker color TS structure relative to the lighter color, native state. *Right* A plot of the 3D structure of the TS. (*Top to Bottom* TS1, TS2, TS3, TS4, TS5)


Beta Turn
Beta Sheet
Alpha Helix
% Seconday Structure
Resiude Number
0
10
20
30
40
50
60
70
80
90
100
0
5
10
15
20
25
30
35
40
45
50
55
60


Beta Turn
Beta Sheet
Alpha Helix
% Seconday Structure
Resiude Number
0
10
20
30
40
50
60
70
80
90
100
0
5
10
15
20
25
30
35
40
45
50
55
60


Beta Turn
Beta Sheet
Alpha Helix
% Seconday Structure
Resiude Number
0
10
20
30
40
50
60
70
80
90
100
0
5
10
15
20
25
30
35
40
45
50
55
60


100
90
80
70
60
50
40
30
20
10
0
% Seconday Structure
0
5
10
15
20
25
30
35
40
45
50
55
60
Resiude Number
Beta Turn
Beta Sheet
Alpha Helix


100
90
80
70
60
50
40
30
20
10
0
% Seconday Structure
0
5
10
15
20
25
30
35
40
45
50
55
60
Resiude Number
Beta Turn
Beta Sheet
Alpha Helix

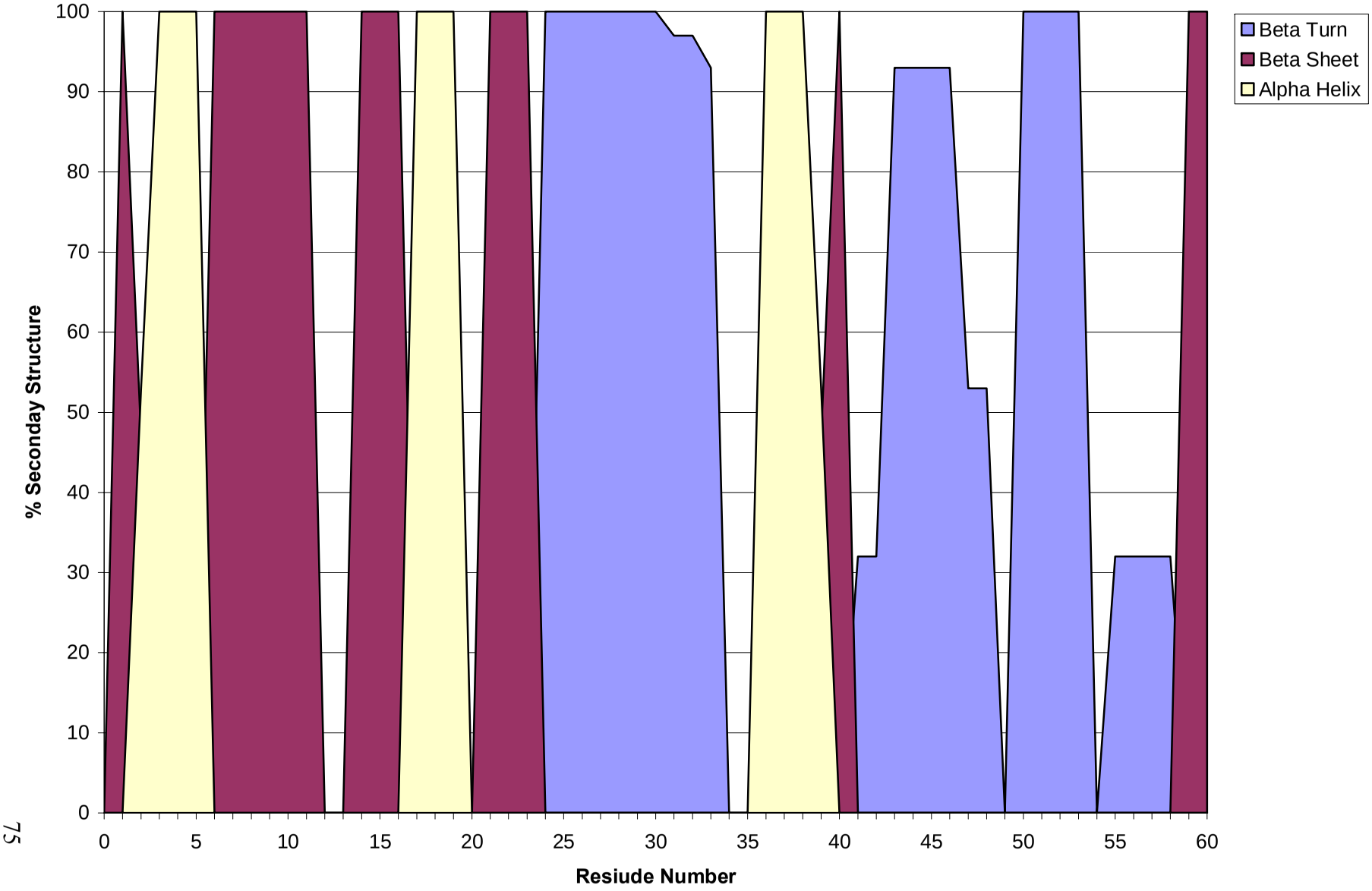

Beta Turn
Beta Sheet
Alpha Helix
% Seconday Structure
Resiude Number
0
10
20
30
40
50
60
70
80
90
100
0
5
10
15
20
25
30
35
40
45
50
55
60

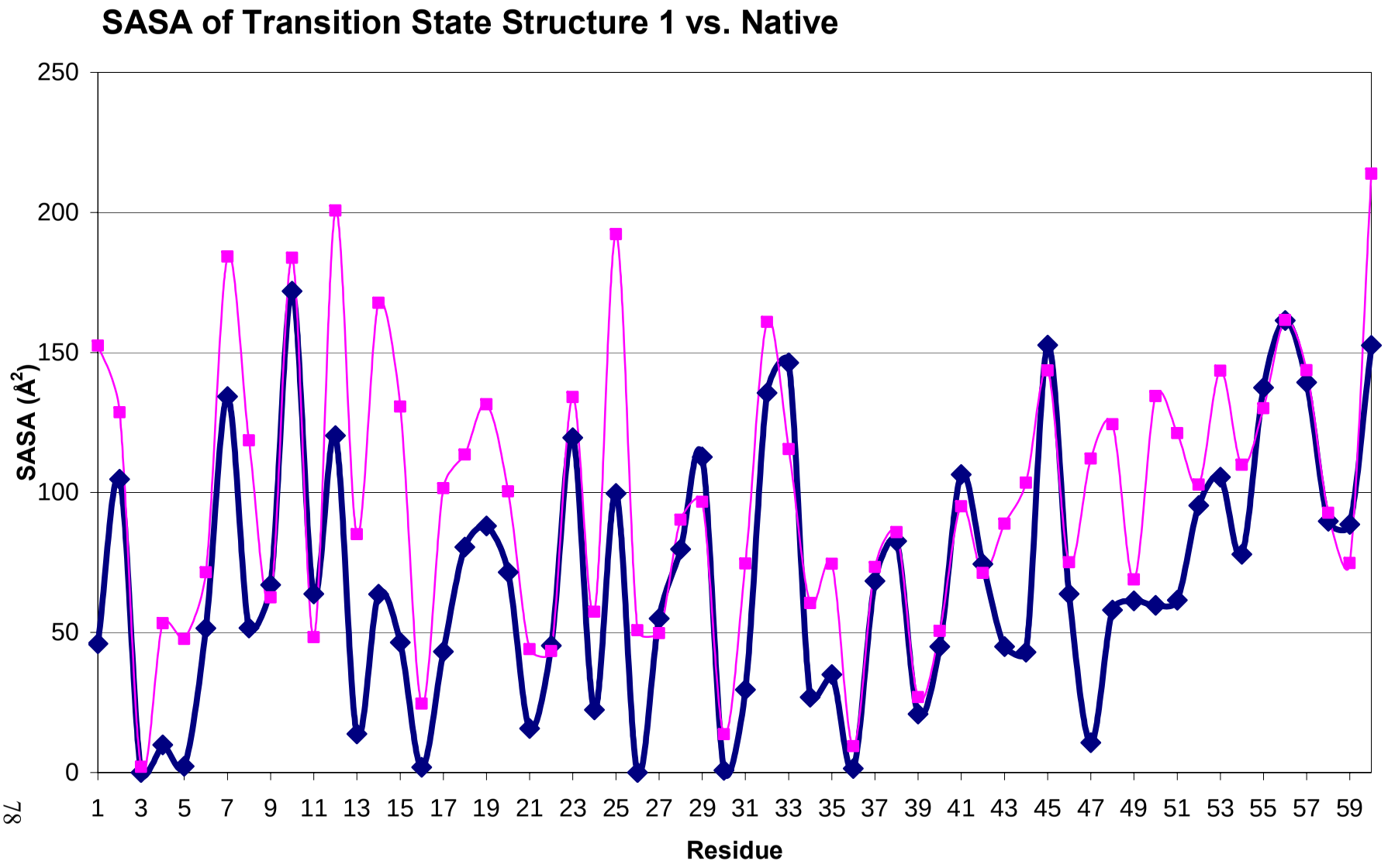
Figure 5-12: The DSSP analysis of the residues in the native state and TSE structures of NTL9.
Identical DSSP results exist for native state and TS1 as well as TS3 and TS4 (*Top to Bottom:* NATIVE STATE, TS1, TS2, TS3, TS4, TS5)

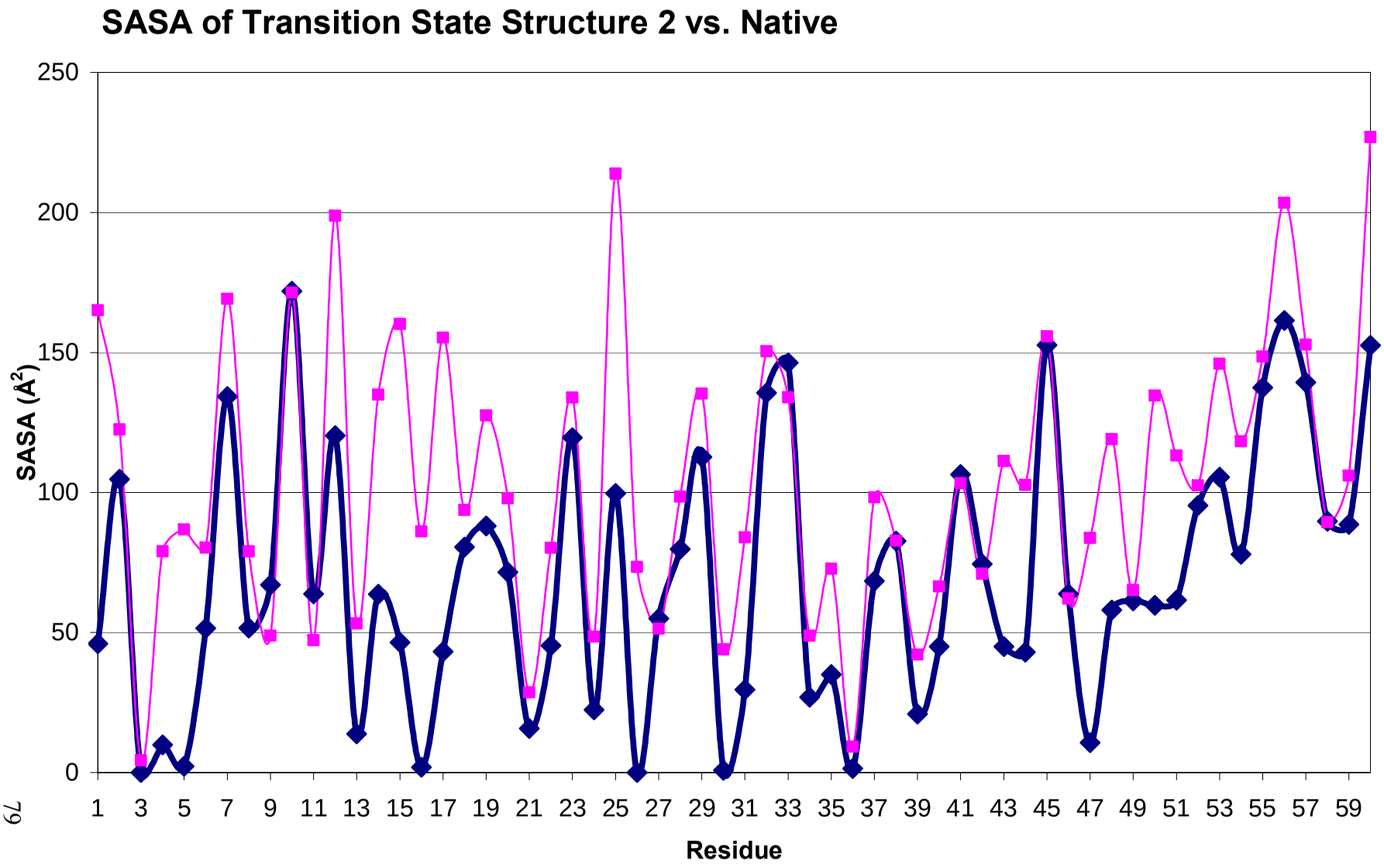
Contacts with the third β-strand structure in the TSE of NTL9 are well persevered. Additionally, non-native contacts are noted in the TSE between the N-terminal region and the end region of β-sheet 3. This is the region constituting the beginning of the inter-domain helix (Residues 39-41). This region may act as scaffolding in the folding pathway to facilitate the incorporation of the N-terminus in the native state. Moreover, loss of contacts between residue 3 and the 43-44 residue region of the helix are noted in the TSE. It suggests the importance of simulating an appropriate length of inter-domain helix. However, it is difficult to come to any conclusions about these contacts because they are interacting with the N-termini which may be particularly flexible in our simulations in part due to lack of friction with water molecules because of the implicit solvent model employed. Furthermore, native structure contacts between the helix and the most N-terminal residues are not seen in the more recent NMR models of NTL9 [180].

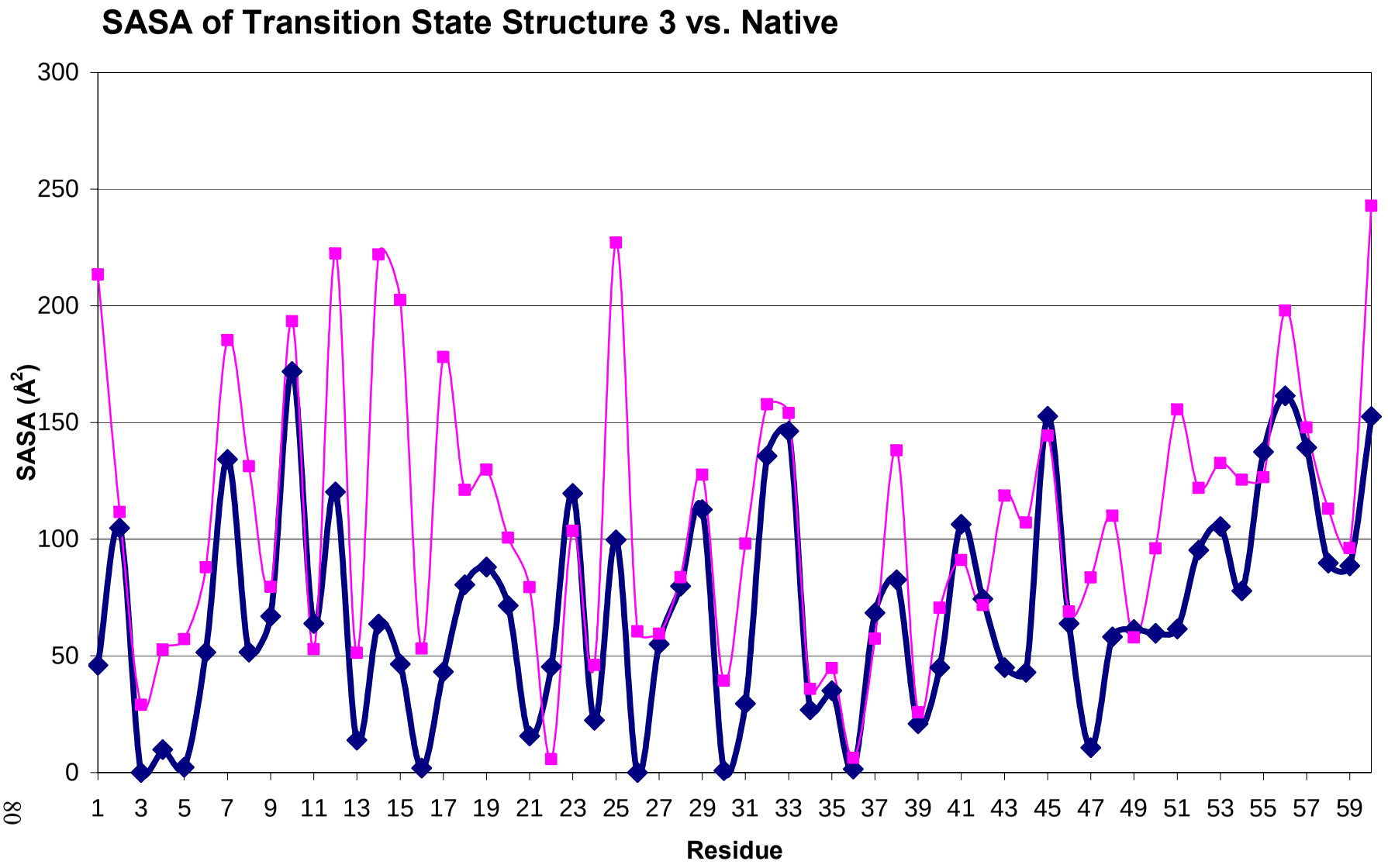
5.3.5 Turn – Loop Region and β-strand 2

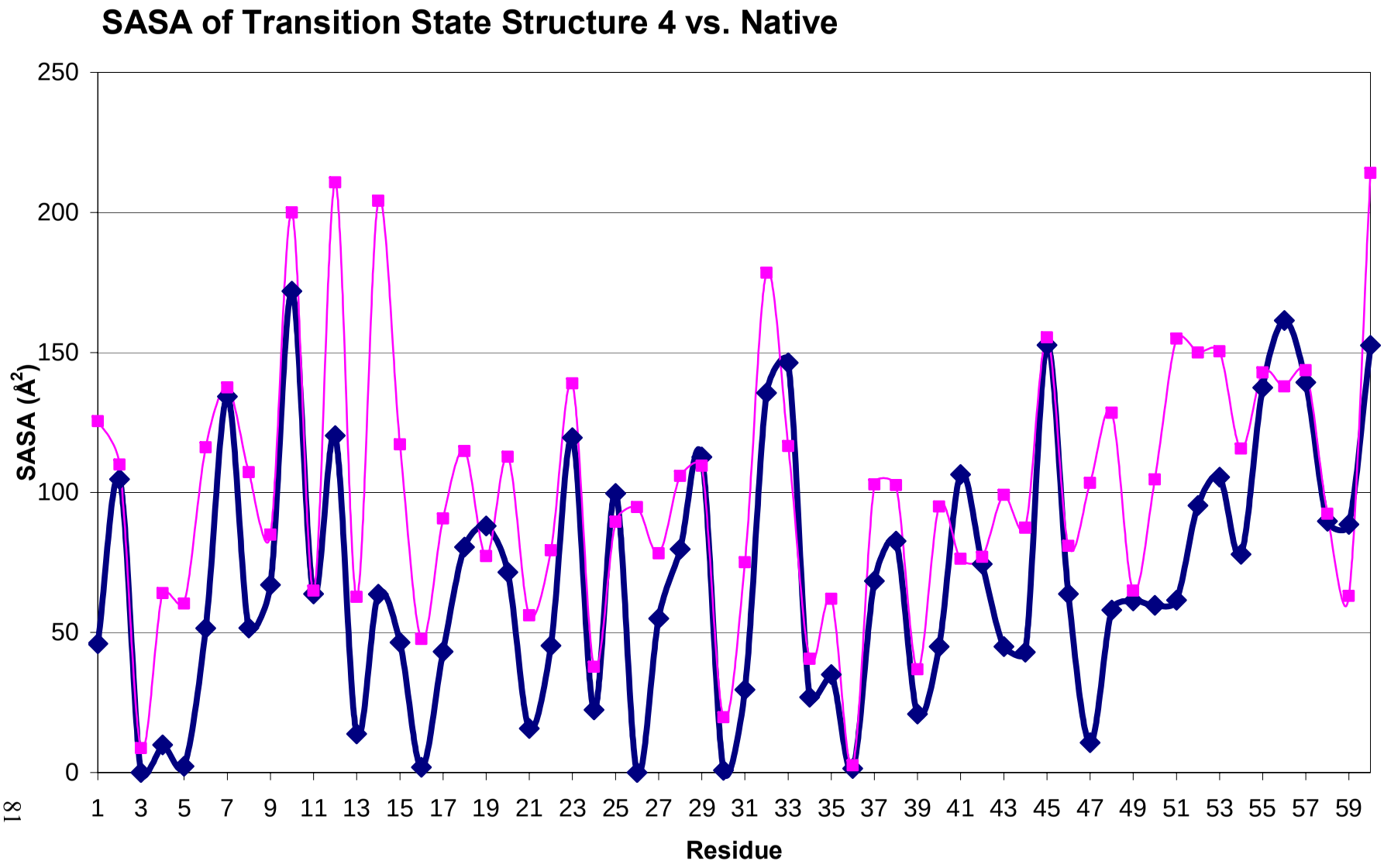
The turn or loop region, is defined by the residues between the first and second β-sheets (Residues 7-16). (Figure 5-1) Contacts in the native state suggest that this loop region is compact, tucking the majority of turn and end residues into the region under β-

sheet 2. (Figure 5-11) In the unfolding simulations the loop region was the first region to lose native form after which followed a loss of interactions between the secondary structural elements (viewed from trajectory). In the native state the beginning of this region including parts of the end of β-strand 1, (Residues 6-9) make contacts with the other end of the loop region, residues 11-14. DSSP analysis identified strong (90-96%) loop regions for these residues in the native state. (Figure 5-12) In the TS, DSSP analysis of this region yielded a mix of results, including very strong (100%) loop character for all but residue 13 in TS5 to nearly no loop character in TS1. TS1 also displayed non-native contacts between residues 9-11 and 13-14. Evident in all conformers of the TSE is that the loop structure adjacent to residue 11 is not fully formed. Further analysis of contacts between residues in this region suggests significant loss of the loop structure with native contacts missing between the residues of the beginning and end of this region. (Figure 5-11) Contacts between residue 17 of β-strand 2 and members of the loop region (Residues 7-16) are missing in all conformations of the TSE. This is consistent with the low experimental Φ-value for residue 17. Further analysis of the SASA for these residues also points to observations from unfolding trajectories that this turn region is not fully tucked towards the second β-sheet structure. (Figure 5-13) In the native state these contacts exist and form electrostatic interactions. However, in the TS the conformation of this loop does yet not facilitate the formation of electrostatic interactions, consistent with experiment [189].


SASA of Transition State Structure 1 vs. Native
SASA (Å²)
250
200
150
100
50
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Residue


SASA of Transition State Structure 2 vs. Native
SASA (Å²)
250
200
150
100
50
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Residue


SASA of Transition State Structure 3 vs. Native
SASA (Å²)
300
250
200
150
100
50
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Residue


SASA of Transition State Structure 4 vs. Native
SASA (Å²)
250
200
150
100
50
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Residue

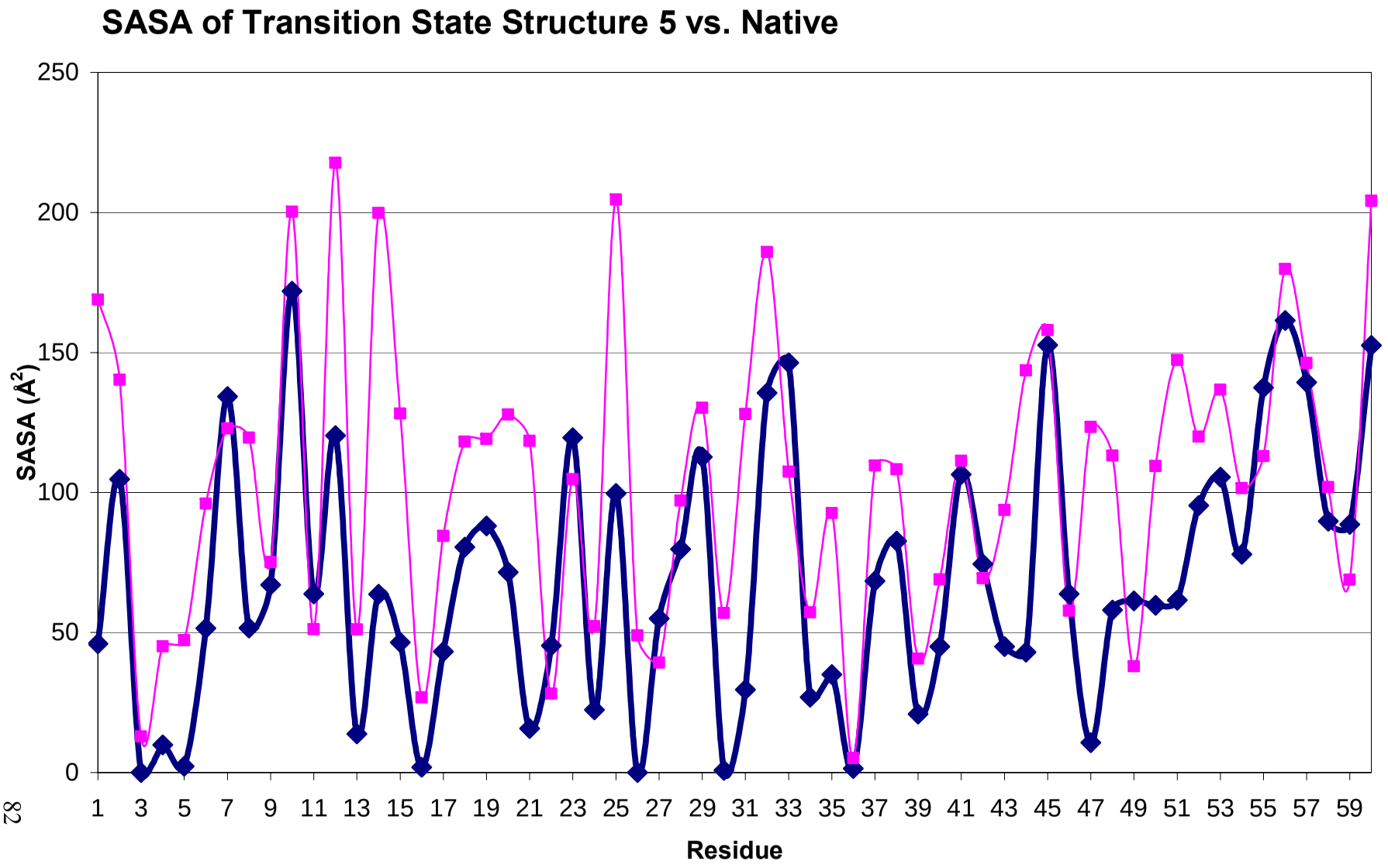

SASA of Transition State Structure 5 vs. Native
SASA (Å²)
250
200
150
100
50
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Residue

Figure 5-13: The SASA (Å^2) computed for each residue in each of the NTL9 TS structures identified.
These values are compared to the native state 1DIV experimental structure SASA (Å^2) values (dark line and diamonds).

Finally in the native state Lys-14 and Lys-15 of the loop region make close contacts with the α-helix at Gln-50 Lys-51 and Glu-54 [174]. Due to its location at the beginning of β-strand 2 this may be an important contact in the final form. Experimental Φ-values for such residues are low, moreover more recent structures [188], do not display this interaction and so it is possible that these observed interactions are artifacts of the structural elucidation process. In our analysis the TSE contact maps as well as DSSP analysis suggest that this region is not fully formed in the transition state of NTL9.

In addition to the native state contacts already mentioned, β-strand 2 also makes contacts with the inter-domain α-helix. (Figure 5-11) These native contacts were observed between residue 18 and residues 39, 44 and 47. In the TSE only the first of these interactions with residue 39 was observed and this occurred in 3 of 5 structures. DSSP analysis for the native state revealed strong β-sheet character from residues 17-20, but in 3 of the 5 TS conformations (TS2, TS3 and TS5) of the ensemble this β-sheet structure was not present or was reduced. (Figure 5-12) In the remaining 2 TS structures this β-sheet character was extended beyond the native state region around the loop.

Non-native contacts appeared consistently in the TSE between residue 3 and the end region of β-strand 2 structure, residue 22. This region is susceptible to considerable

flexibility being near the N-terminus of NTL9. This particular contact does not appear in the native state because the C_α distance is just under the cutoff for the contact map analysis. Thus, we suggest that this observation is an artifact of the analysis method and highlights the sensitivity of a reaction coordinate like residue-residue contacts.

5.3.6 α-Helix

A short 4 residue α-helix connects the second and third β-strands. The helix is made up of residues 25 to 29 according to DSSP analysis. (Figure 5-12) Primary contacts within this region are between components of the helix itself. In our TSE it was observed that the helix region was extended beyond residue 29 to residue 33 and in some cases even further. This extended helix structure is thought to be slightly more flexible in the TSE as slightly fewer intra-helix contacts were observed. (Figure 5-11) Additionally, the C_α RMSD of these residues for each of the TSE structures exhibit greater variation than other residues suggesting they have more flexibility. (Figure 5-14) This extended helix may play a role as a molecular scaffold upon which the final form of the nucleus region rests prior to sampling the conformational degrees of freedom prior to finding the native state. The fact that it was more flexible suggests that the helix has also not fully settled into its native conformation and as such in the TS the electrostatic interactions have not fully formed. (Figure 5-10, 5-11 and 5-14) Additionally, new contacts between residues 22 and 27 are noted in the TSE. It is thought that these may be transient contacts further supporting the idea that the final structure linking β-strand 2 and the helix is not fully developed in the TSE.

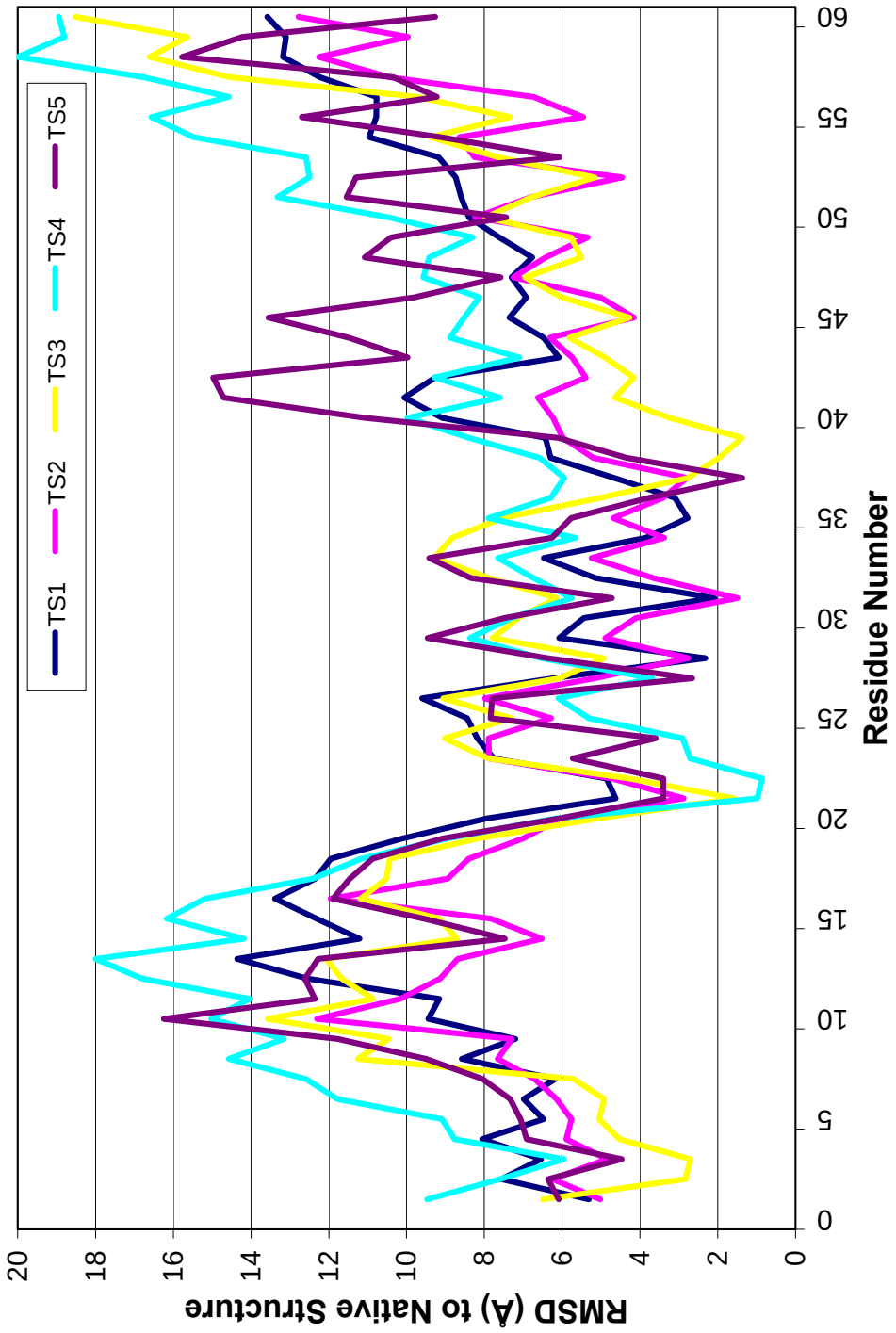


Figure 5-14: Residue C_α RMSD (Å) from the native state conformation of NTL9. The structural C_α RMSD (Å) per residue for each of the structures in the TSE. The RMSD was calculated relative to the native state conformation. (*Top to Bottom* TS1, TS2, TS3, TS4, TS5) Variation in C_α RMSD in the TSE indicates flexibility in the conformation of the residue in the TS.

5.3.7 β-strand 3

In the TSE the N-terminal region of the third β-strand does not make contacts with the beginning of the loop region as seen in the native state. (Figure 5-11) Investigation of this region by DSSP reveals a lack of fully formed secondary structure in adjacent structures. Specifically in the end region of β-strand 1 and the loop region connecting to β-strand 2 have not fully formed. (Figure 5-12)

5.3.8 Inter-domain Helix

The inter-domain helix in the transition state is largely formed. (Figure 5-11 to 5-13) Throughout the unfolding simulations the helix is largely structured. The C-terminal end displayed some fraying. DSSP analysis reveals that in the TSE the helix region is split into 2 parts divided at residue 49. (Figure 5-12) Later in the unfolding simulations the helix looses integrity in this region. Contact map analysis reveals that in the TSE fewer contacts are made between residues within the helix, similar to the other helix structure. This may point to greater helix flexibility in the TSE. (Figure 5-13) As mentioned earlier, the N-terminal region of L9 residues 1 and 2 make contact with residues 40 and 41 of the helix. No significant non-native contacts were found in the TSE. (Figure 5-13) Our observations of the inter-domain helix may be misleading, however, as it has been shown that helices are over stabilized in the GB implicit solvent model [215]. That said, it is particularly interesting that the helix loses integrity at

residue 49 since this would be expected to be even more pronounced under conditions that do not over stabilize the helix.

5.4 Conclusions

In this work, macroscopically derived Φ-values were utilized to determine the TS of a microscopic system. Since single reaction coordinate investigations can only represent some fraction of the entire ensemble, our five transition states represent a few transition states and not a statistically significant ensemble. In order to better describe the ensemble the five identified TS structures were the basis of clustering algorithms to generate a more representative TSE of NTL9. Ideally, we would have liked additional experimental data to build a more statistically significant TSE based on several reaction coordinates. Unfortunately, the study of transition states has proven difficult experimentally limiting data available. Nonetheless, the work here goes beyond the common methodology of identifying and characterizing transition states using only single or very few (< 10) unfolding simulations [163, 167-170].

In this paper we have identified and subsequently characterized the TSE of NTL9 at atomic resolution. This was accomplished through the Φ-RMSD results driven by experimental knowledge of Φ and β_T values. Given the great difficulties in locating the TSE we believe that this represents the best picture of the NTL9 transition state to date. We have determined that the TSE ensemble of NTL9 is largely native in composition with all major parts of secondary structure well defined. Our results are consistent with experimental knowledge which suggests that the TS is largely native in structure [216]. Moreover, in atomic level description of the TSE, secondary structure elements do not make optimal interactions between other units in the N-terminal nucleus region. Such optimal interactions are likely facilitated by electrostatic interactions and thus this

observation is also in agreement with past experimental work on the TS [188]. In the folding path of NTL9 after reaching the TSE results suggest that much of the drive towards the native state is spent optimizing contacts between stable secondary structure elements and forming contacts in loop region.

6 Probing the Folding of mini-protein Beta3s by Two Dimensional Infrared Spectroscopy; Simulation Study

In this chapter we use of infrared coherent two-dimensional infrared spectroscopy (2DIR) to characterize the folding mechanism of the mini-protein Beta3s is proposed. In this study Beta3s was folded by molecular dynamics (MD) simulation and intermediate conformational ensembles were identified. The one and two-dimensional correlation spectrum was calculated for the intermediate and native states of the mini-protein. A direct structure-spectra relationship was determined by analysis of conformational properties and specific residue contributions. We identified the structural origin of diagonal and off-diagonal peaks in the 2DIR spectra for the native and intermediate conformational ensembles in the folding mechanism. This work supports the implementation of computational techniques in conjunction with experimental 2DIR to study the folding mechanism of proteins. In addition to exploring the folding mechanism the work presented here can be applied in combination with experiment to refine and validate current molecular dynamics force fields.

> The work presented in this chapter, "Probing the Folding of mini-protein Beta3s by Two Dimensional Infrared Spectroscopy; Simulation Study" was developed in collaboration between Christopher N. J. Marai, Jin Wang and Shaul Mukamel and submitted for review and publication to PMC Biophysics.

6.1 Introduction

The biological activities of proteins are determined by the specific three-dimensional structure and dynamical properties of the molecule. The activity of misfolded proteins has been implicated in diseases including Alzheimer's, Diabetes, Parkinson's disease, many cancers and cancer-related syndromes; consequently an understanding of the protein folding mechanism is of importance to pharmaceutical design and molecular biology [5-9]. Our understanding of protein folding has largely remained elusive due to the short timescales at which folding occurs and the large number of degrees of freedom involved in tracking such mechanisms [111, 112, 217]. Energy landscape theory provides a novel framework for understanding the global principles of protein folding in terms of funnels [9]. A greater understanding of this process can be facilitated by further insight into both the structural and dynamical changes that occur during the folding process. A combination of experiment and calculations has recently been developed to monitor these changes in tandem.

6.1.1 Protein Folding

Traditionally the majority of experimental data on protein folding has been obtained through kinetics experiments which do not report on atomic level structural changes that characterize the folding process. Additionally, until recently, experimental methods with structural resolution have lacked the temporal resolution necessary to observe ultra-fast folding processes. These methods, including 1DIR, fluorescence, NMR and XRD, have resulted in indirect or time-averaged information about the

structure and the energy surface of proteins along the folding path. Although progress has been made by advanced NMR techniques combined with simulation, lengthy folding processes have yet to be explored [26]. Recent advances in 2DIR correlation spectroscopy, using techniques derived from NMR, are shedding new light on the mechanism of protein folding [27-31].

6.1.2 Infrared Spectroscopy

Similar to 1DIR absorption spectroscopy, 2DIR uses infrared wavelengths to probe the Amide-I and other vibrational bands present in protein structure [135]. The Amide-I band, consisting of the carbonyl stretch of the protein backbone is commonly probed by IR based techniques because it displays structural sensitivity due to coupling between in-phase bending of N-H and stretching of C-H bonds [140]. In proteins these bands respond to coupling between amide units and delocalization of vibrational states, thus reporting on the size and secondary structure of proteins. Using advanced, multiple-pulse coherent spectroscopy techniques like 2DIR, transitions can be spread across two axes revealing vibrational couplings resulting from three-dimensional structural contacts [141-145]. Coupled with ultrafast optical techniques, 2DCS IR spectroscopy results in a structure-based tool that is responsive at the ultrafast timescales present in the folding mechanism. Accordingly 2DIR spectroscopy is now being used to study unfolding processes in T-jump experiments.

6.1.3 Molecular Dynamics Simulation and Simulated 2DIR

It has been proposed that simulation of 2DIR spectra coupled with MD calculations can provide additional insight into the folding pathway, particularly when compared with experiment [218, 219]. The accurate simulation of one-dimensional absorption spectra of proteins by incorporating a dipole-dipole coupling scheme is well established [220]. Although the simulation of 2DIR presents greater challenges, it is now increasingly possible to reliably calculate the 2DIR spectra for a variety of small protein structures [39, 145, 151, 221, 222]. In these calculations a Local Amide Hamiltonian is generated with structural coordinates derived from MD simulations [151]. The signal is then simulated by a third-order response function dependent on all of the one and two-exciton states and their coupling to a thermal bath by the sum over states (SOS) method or nonlinear exciton equations (NEE). Spectra are calculated for structures, particularly intermediates, in a MD simulation of protein folding. The calculated spectra can then be compared to experimental 2DIR to elucidate new structural information about the folding mechanism. Additionally, structural information from 2DIR folding experiments coupled with MD and simulated 2DIR shows promise as a tool to validate and refine of MD force fields [141-145, 223].

In this study, simulated 2DIR was used to characterize the folding mechanism of the 20 residue β-sheet mini-protein Beta3s. (Figure 6-1 and Figure 6-2) Beta3s, a de-novo three-stranded β-sheet mini-protein, contains common protein structural motifs including a β-hairpin and an anti-parallel β-sheet [224, 225]. The structure and folding of

Beta3s has been probed extensively by NMR and through numerous computational studies, however much remains unknown about its folding mechanism [142, 226-231]. NOE and chemical shift data suggests that Beta3s exists in a single structured antiparallel beta-sheet conformation that is in equilibrium with a random coil conformation [142]. The sheet conformation exists with turns at Gly-6-Ser-7 and Gly14-Ser-15 and a population of 13-33% according to NOE intensities and 30-55% according to chemical shift data. Additionally, equilibrium sedimentation and NMR dilution experiments showed Beta3s to be monomeric in solution [142]. This data, together with the synergies of MD and 2DIR calculations, provide a unique opportunity to reveal spectra-structure correlations and explore the folding mechanism of Beta3s from a structural perspective.

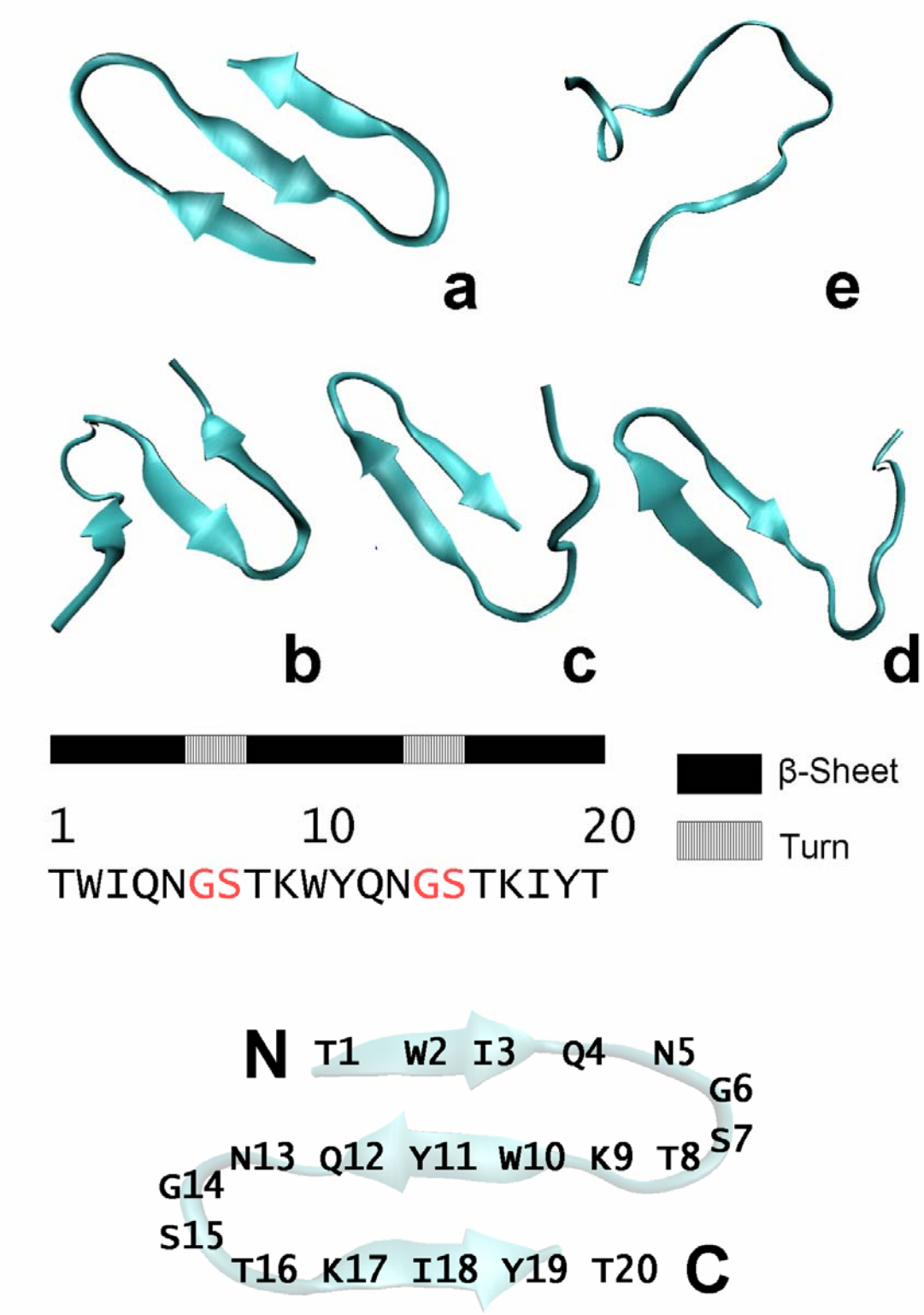


Figure 6-1: The residue composition and investigated conformational states of Beta3s.
Top: Residues in Beta3s structure shown in native conformation. Bottom: Cartoon rendering of Beta3s in a) native b) Ns c) Cs d) Ch-Curl conformations e) 6-12 Helix.

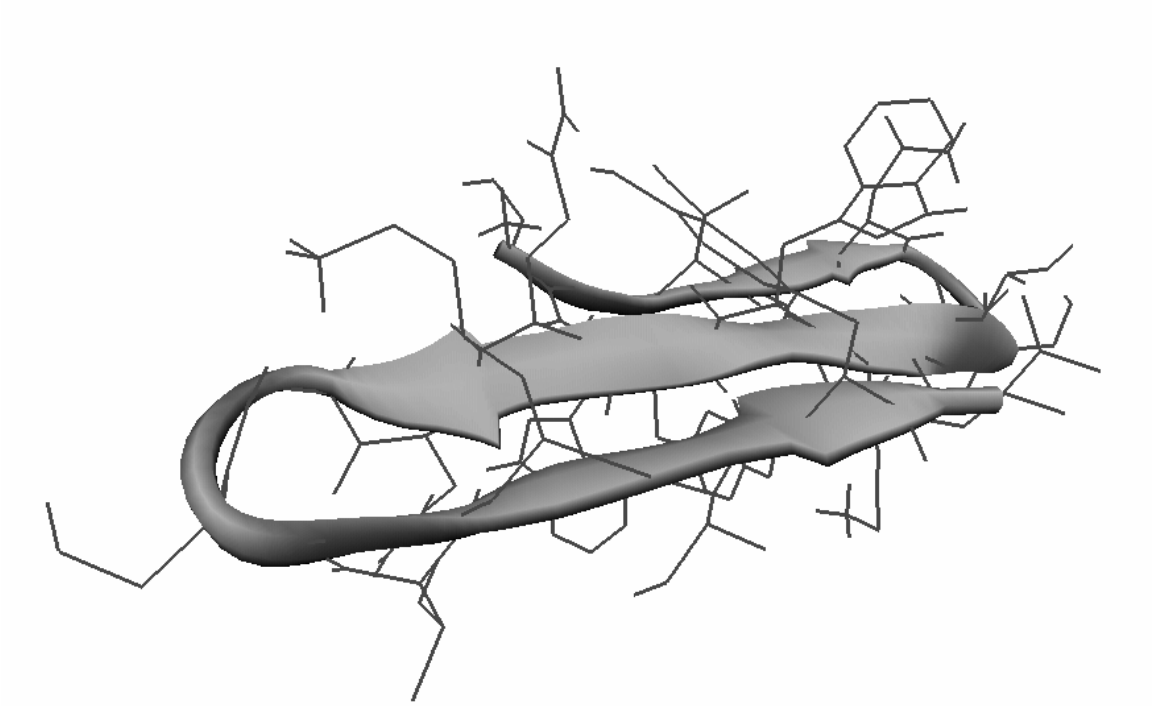


Figure 6-2: Beta3s native structure with cartoon backbone and stick representations of side chains.

6.2 Methods

6.2.1 Molecular Dynamics Simulations of Beta3s

Beta3s, a synthetic 20-residue peptide (Thr1-Trp2-Ile3-Gln4-Asn5-Gly6-Ser7-Thr8-Lys9-Trp10-Tyr11-Gln12-Asn13-Gly14-Ser15-Thr16-Lys17-Ile18-Tyr19-Thr20), [142], was modeled with the CHARMM PARAM19 force field [87]. This force field explicitly models all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms. The folding simulations were performed with the program CHARMM [87, 232], and analysis was carried out with the MMTSB [209] package as well as our own code. Solvent interactions were taken into account by an implicit model based on solvent-accessible surface area (SASA) [233]. Ten simulations of 200 ns each, starting from a linear structure and random seed were performed under default electrostatic cutoffs (7.5Å) and 330 K to sample the folding pathway of Beta3s. This procedure has been used extensively by Caflisch et al. to reversibly fold Beta3s into its NMR conformation and efficiently sample its folding landscape on several occasions [226-231]. Folding to the native conformation was confirmed by fraction of native contacts (Q-score) analysis. The peptide was considered folded to the known folded NMR conformation, detailed in reference 18, when at least 25 of 26 Nuclear Overhauser Effect (NOE) constraints were satisfied, a Q-score of greater than 0.95.

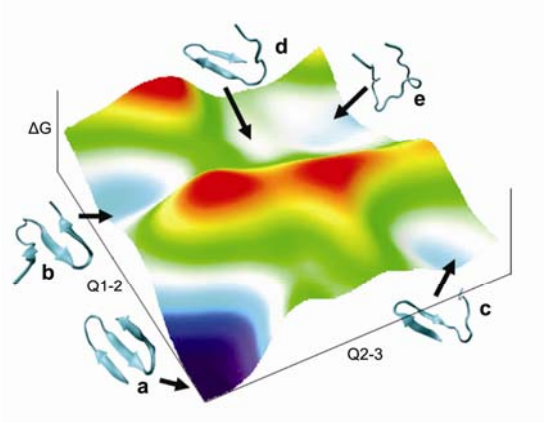


Figure 6-3: The free energy surface of Beta3s and assigned conformational states. Beta3s free energy surface approximated from data in previous works by Caflisch et al. [227, 228]. Axis ΔG, Q1-2 and Q2-3 representing, free energy and fraction of native contacts in each of the strands N-terminal (strand 1), central, (strand 2) and C-terminal (Strand 3) respectively. Blue indicates low and red indicates high free energy values on a free energy scale from 0-4 kcal/mol. Conformations, a) native b) Ns c) Cs d) Ch-Curl conformations e) 6-12 Helix, are assigned to local basins in accordance free energies described in previous work [227, 228].

6.2.2 Conformation Identification

Significant exploration of the conformational space of Beta3s has already been accomplished by a variety of rigorous methodologies [226-231]. (Figure 6-3) The most current work at the time of this study had identified several statistically significant conformations in the folding of Beta3s [227]. The most populated structures in the folding pathway include the mostly helical "6-12 helix" (Figure 6-1e), a curled structure "Ch-curl" (Figure 6-1d), a native like structure with the C-terminus out of register "Cs-or" (Figure 6-1c), a native like with the N-terminus out of register "Ns-or" (Figure 6-1b), and native structure (Figure 6-1a). These structures listed in table 6-1 and displayed in figure 6-1 were defined by Karplus and Caflisch et al. using the DSSP [234] backbone configuration and investigated in this study [227]. (Figure 6-1, Table 6-1) The DSSP configuration identifies residue conformations based upon the φ, ψ backbone angle [234]. In our study each of the 5 conformations studies were identified by comparison of the

DSSP calculated backbone configuration at each time step in the trajectories to the published backbone configurations for each intermediate. A total of five structures were chosen for each conformation investigated. Since many structures were identified and only a few were required with some heterogeneity only the middle structure of each quintile in a distribution of structures was chosen for investigation. The quintile distribution was formed from analysis of all matching structures by clustering backbone RMSD to the centroid. By this process we were able to account for slight structural variances in the backbone configurations that are more consistent with an ensemble of structures in a conformation.

Table 6-1: Conformational Properties of Beta3s

Conformation	Q	Radius of Gyration (Å)	SASA (Å^2)	Backbone DSSP String
612 Helix	0.59	7.98	2084	~HHHHHHHHHHS~~~~~
Ns-or	0.99	7.45	1993	EEEESTTEEEEESSEEEE
Cs-or	0.89	7.96	2004	EEEESSEEEEESSSEEEE
Ch-Curl	0.89	7.78	1951	~~SSGGG~~~EESSEETT
Native	1.00	7.50	1783	EEEESSEEEEESSEEEE

Table 6-1: Conformational Properties of Beta3s.
The averaged conformational properties of investigated Beta3s conformations, DSSP assigned backbone configuration, Q, fraction of native contacts, radius of gyration (Å), and SASA (Å^2), solvent accessible surface area.

It has been well documented that solvation and arrangement of solvent plays an important role in the 1D and 2DIR spectra of the amide bond [39, 235, 236]. Consequently the identified conformations from the CHARMM PARAM19 force field were solvated and all hydrogen atoms were added. The solvation process involved an initial minimization of the solvent around a constrained protein backbone followed by 20 ps of backbone constrained molecular dynamics to allow for adjustment of the protein-water interface. The CHARMM PARAM22 [39, 91], all atom force field was

implemented in this process. This procedure resulted in a total of 100 unique solvent environments for the 5 unique structures of each of the 5 conformations investigated.

6.2.3 Amide I Spectral Calculations 1D and 2D IR Spectra: SPECTRON

An ensemble of 500 structures for each conformation was used in the Amide-I spectral calculations. Simulation of the 1D and 2DIR spectra was carried out according to an approach described by Zhuang and Mukamel [151, 237-239], as implemented in the SPECTRON [151], software package. The Local Amide Hamiltonian (LAH) approach was applied to describe the peptide structures in our calculations [152, 240-243]. The Hamiltonian parameters were provided by Mukamel et al. as implemented in SPECTRON [151]. Vibrational couplings of different amide modes were calculated by the *ab initio* maps of Torri and Tasumi [220, 244, 245]. The three-pulse coherent four-wave mixing technique was simulated, where three incoming pulses with wave vectors $\mathbf{k_1}$, $\mathbf{k_2}$, $\mathbf{k_3}$ interact with the protein to generate a signal in the direction $\mathbf{k_I} = -\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}$. The sum over states (SOS) technique with a Lorentzian lineshape with varying FWHM values was used to simulate the 2DIR signal [151]. Details on the methods used in SPECTRON have been presented in reference 28 and 47. These parameters have been shown to provide reliable 1D and 2D spectra for both α-helical and β-content peptides that are suitable for comparison with experiment [235, 238, 246, 247]. The $\mathbf{k}_I$ signal is displayed in figures 6-4 and 6-5 by transforming the response function $S_{v,\gamma\beta\alpha}(t_3,t_2,t_1)$ as described in reference 47, to the frequency domain with first t_1 and third t_3 time delays:

$$W^{(3)}_{I,\nu,\gamma\beta\alpha}(\omega_1,t_2,\omega_3) = \left| \int_0^\infty \int_0^\infty dt_1 dt_3 S^{(3)}_{I,\nu,\gamma\beta\alpha}(t_3,t_2,t_1)\exp(i\omega_1 t_1 + i\omega_3 t_3) \right| \qquad (6\text{-}1)$$

where ω_1 and ω_3 are the Fourier conjugates to the t_1 and t_3 respectively where $t_2 = 0$.

6.2.4 Test of 2DIR computational protocols

The computational protocols involved in the calculation of 2DIR were tested prior to this work to confirm the correct setup of the simulation software SPECTRON [151]. As a simulation software for 2DIR spectra of proteins, SPECTRON has undergone significant testing and reliable results have been produced for β-sheet containing proteins [151, 237-239]. However, since no current 2DIR spectrum of Beta3s exists it is important to ensure our model produces reasonable results with other known systems. In this test the trpzip2 protein derived from protein databank (PDB) structure, 1LE1 [248] was equilibrated simulated by MD simulation, solvated and an ensemble generated according to the procedure outlined in the methods section for Beta3s. The ensemble was then subjected to 1D and 2DIR calculations of the Amide-I band using the $\mathbf{k_I}$ pulse orientation in SPECTRON. The resulting spectra produced are displayed in Figure 6-6 and are in good agreement with past experimental work and theoretical calculations on trpzip2 surveyed in detail by Ganim and Tokmakoff [39]. Specifically, in the 1DIR spectra the peak positions and relative intensities are qualitatively similar of those surveyed in Figure 3 in the Ganim and Tokmakoff review [39]. The 2DIR spectra is derived from the same initial structure input as the 1D spectrum, this input must be supplied correctly for useful results [151, 237-239]. Thus, by its dependence on the initial structure and good agreement between the simulated and experimental 1DIR, the 2DIR should be accurately simulated by these methods. Nonetheless, the 2DIR spectrum

was also simulated and compared to previous work [39]. The simulated 2DIR spectrum revealed peak characteristics and locations similar to those of experiment and the more accurate simulations shown in Figure 3 of reference [39]. Consequently, these results suggest that the methods we have employed in the simulation of Beta3s 2DIR are well executed and capable of providing reliable results.

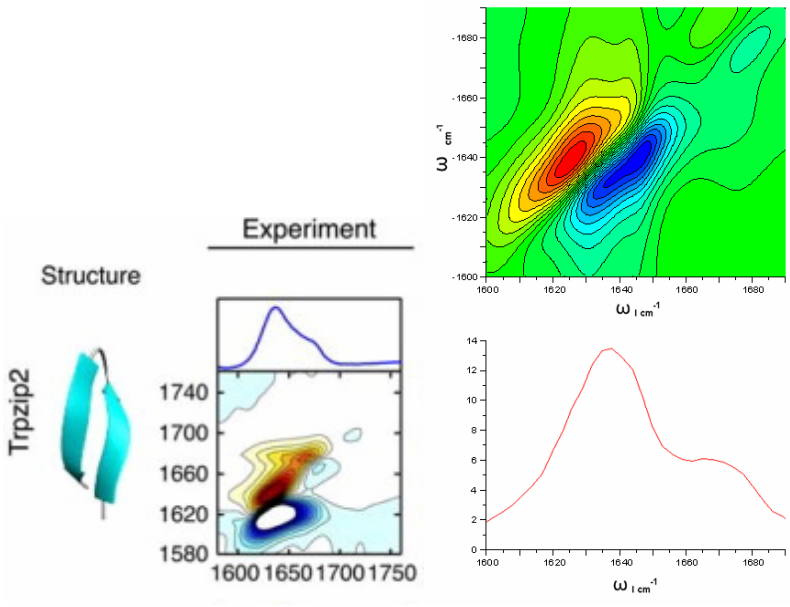


Figure 6-4: 1D and 2DIR of trpzip2.
Top: 2DIR spectra for trpzip2 as generated by SPECTRON. Bottom: 1DIR linear spectrum of trpzip2. Both spectra are in good agreement with previously studies of trpzip2.

6.2.5 Analyzing Residue Contributions Normal Mode Decomposition

Assessment of the spectra-structure correlations requires assignment of specific residues to peaks and cross peaks. Here we applied two methods, conformational difference analysis and normal mode decomposition (NMD). Conformational difference analysis coupled with two-dimensional contact maps involves qualitative examination of spectra relative to the structure of the native state and the other intermediate conformations. Since conformations have different structures the disappearance or presence of a particular peak relative to the native state and other states can reveal structure-peak correlations.

Normal mode decomposition (NMD) provides a simple analysis of the Local Amide Hamiltonian for cross peak interactions. In NMD analysis the excitonic Hamiltonian is diagonalized to obtain eigenvalues and eigenvectors for each of the residue. The eigenvalues provide a quantitative approximation of the IR frequency of each residue, although this does not fully take into account overlap between adjacent residues. The magnitude of the eigenvectors produced for each residue-residue interaction is squared providing an approximate contribution of the residue-residue interaction for each of the 19 x 19 interactions. The strength of this coupling is displayed in the bottom left corner of figure 6-7, stronger couplings are indicated by red peaks and weaker by blue peaks. Consequently, residue coupling contributions from non-adjacent residues, a result of secondary interactions, can be observed. Further description of NMD analysis follows.

6.2.6 Normal Mode Decomposition Method

Normal mode decomposition (NMD) is a method that provides a simple analysis of the Local Amide Hamiltonian revealing strengths of cross peak interactions and mode excitation values. In NMD analysis the excitonic Hamiltonian is diagonalized to obtain eigenvalues and eigenvectors for each of the residues in the system. The diagonalized matrix consists of Eigenvectors C_i the magnitude of which is the Eigenvalue E_i, where $HC_i=E_iC_i$. The elements in the vector C_i are can be denoted c_{ij} the wave function is represented by,

$$\left|\phi_i\right\rangle = \sum_{j=1}^{N} c_{ij} \left|\psi_j\right\rangle \tag{6-2}$$

where,

$$\sum_{j=1}^{N} {c_{ij}}^2 = 1 \tag{6-3}$$

The Eigenvalues extracted from this diagonalization represent the frequency of each mode (residue). Consequently the c_{ij}^2 value related to the Eigenvectors represents the contribution of mode (residue) *i* to mode (residue) *j*. The contribution from each *i* and *j* is what is plotted in the NMD analysis.

6.3 Results and Discussion

6.3.1 1DIR Spectra of Beta3s

The one-dimensional IR spectrum of each of the Beta3s conformational ensembles was calculated revealing significant differentiation between structures. (Figure 6-4) In the 1DIR, Amide-I band absorption at both high and low frequency peaks distinguish different secondary structure conformations and thus can be used to detect the degree of folding or point in the folding mechanism. Common Amide-I bands in proteins originating from the protein backbone configuration include, β structures absorbing at 1610-1640 cm^{-1} ($\nu\perp$) and 1680-1690 cm^{-1} ($\nu\parallel$), α-helix at 1640-1650 cm^{-1} and the 1650-1660 cm^{-1} random coil regions [140]. The low frequency ($\nu\perp$) β absorption is a result coupling perpendicular to the β-strand while the high frequency β absorption is ($\nu\parallel$) a result of in plane coupling [249]. The 1DIR data presented in figure 6-4 and table 6-2 has been calculated for explicitly solvated conformational ensembles with a homogenous broadening parameter (Γ) of 1.0 (solid line), 5.0 (dashed line) and 10.0 cm^{-1} (dotted line).

Table 6-2: 2DIR Diagonal Peak Characteristics

	Absorption Region (cm^{-1})				Splitting (cm^{-1})		
Conformation	**Beta Low**	**Alpha Low**	**Alpha High**	**Beta High**	**Beta-Alpha**	**Alpha-Alpha**	**Alpha Beta**
Nat	1630	1645	1665	1678	23	38	26
Ns	1638	N/A	1678	1692	42	N/A	18
Cs	1648	N/A	1678	1692	42	N/A	18
Ch	1632	N/A	1672	1688	66	N/A	16
612	1632	1662	N/A	1685	42	N/A	32

Table 6-2: 2DIR Diagonal Peak Characteristics.
Diagonal peak characteristics calculated from 2DIR at 5cm^{-1} line width parameter (Γ), position and relative distance, for Beta3s conformations.

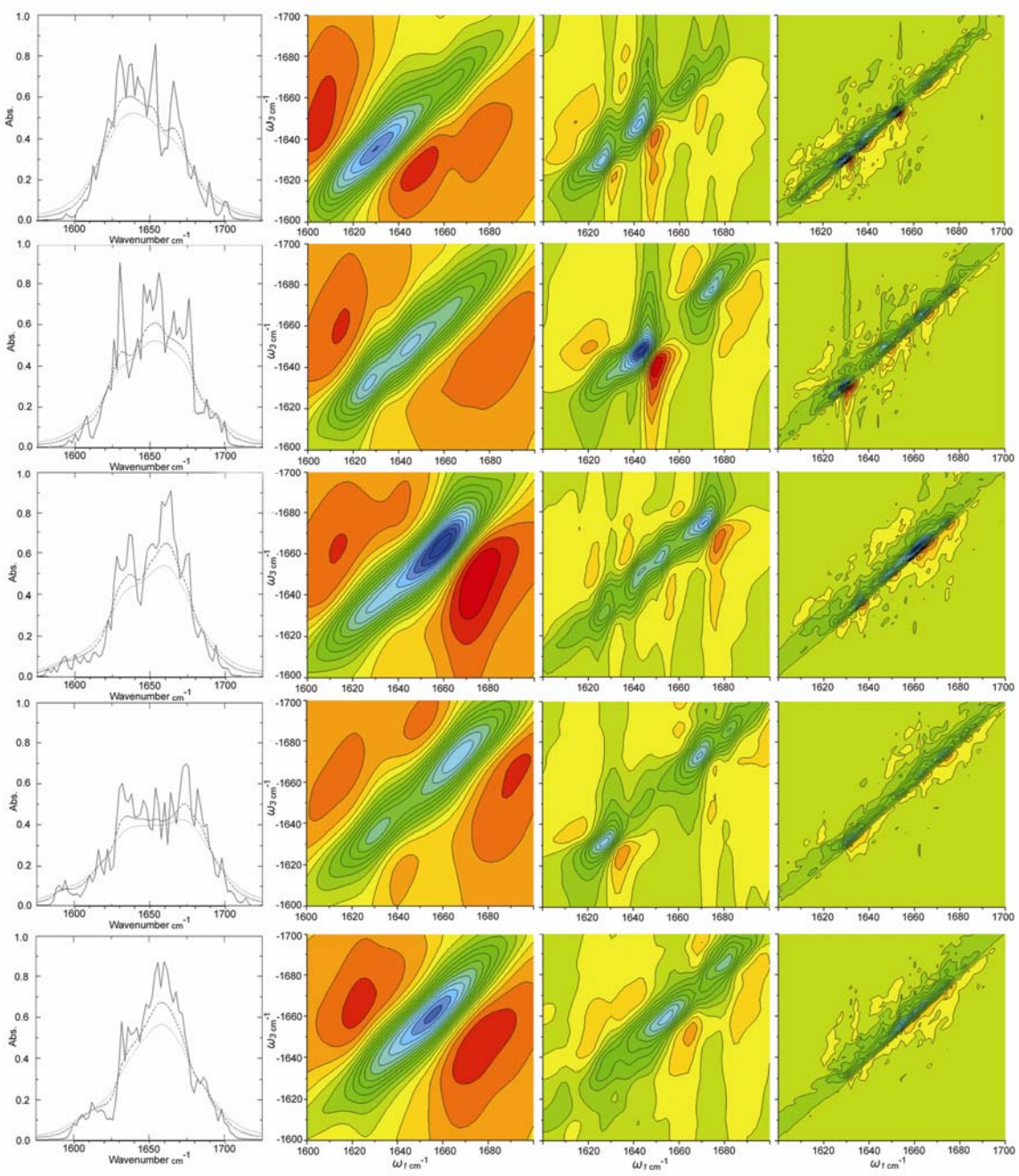


Figure 6-5: The simulated 2DIR spectrum of Beta3s.
Right: The linear absorption spectra for investigated conformations of Beta3s calculated with homogenous broadening parameters 1 cm^{-1} (solid line), 5 cm^{-1} (dashed line) and 10 cm^{-1} (dotted lines). Bottom axis is absorbance in cm^{-1} and side axis is absorbance. a) native b) 6-12 Helix c) Cs d) Ns and e) Ch-Curl. Left: The 2DIR spectra for the k_I experimental parameters for each conformation of Beta3s. Columns left to right represent line width parameters (Γ) 1 cm^{-1} (left), 5 cm^{-1} (middle) and 10 cm^{-1} (right). Rows indicate data for each of the conformations top to bottom native, Ns, Cs, Ch-Curl, 6-12 Helix.

The native conformation 1DIR spectrum of Beta3s at Γ of 10 cm^{-1} shows a maximum peak at 1636 cm^{-1} which is consistent with the low frequency $\nu\perp$ mode for a β-sheet structure resulting from oscillations in phase perpendicular to the β-strands (Figure 6-4, Table 6-2) [140]. In the 10 cm^{-1} data the dominant high frequency peak occurs at 1666 cm^{-1} which is lower than the typical 1680 cm^{-1} characteristic high frequency peak $\nu\parallel$ for β-sheets. Investigation of the data obtained at Γ of 1 cm^{-1} reveals the expected high frequency β-sheet $\nu\parallel$ mode at 1680 cm^{-1}. The data at 1 cm^{-1} also revealed a significant peak at 1654 cm^{-1}, typically associated with random coil and α-helical secondary structure [140]. The presence of the 1666 cm^{-1} and 1654 cm^{-1} peaks result from the turn region residues in the native conformation, although the absorption intensity for these modes is higher than expected. High absorption intensity is likely attributable to overlapping contributions from very similar modes occurring as a result of the structural homogeneity of the turn region versus the β-sheet regions in the ensemble. A comparison of the turn region versus β-sheet regions is described in the table 6-3 data and shows the β-sheet regions of Beta3s to be less homogenous in structure than the turn sections. (Table 6-3)

Table 6-3: Native Beta Regions RMSD

	RMSD (Å)				
Conformation	**Sheet 1**	**Turn 1**	**Sheet 2**	**Turn 2**	**Sheet 3**
Residue Range	**1-5**	**6-8**	**9-13**	**14-16**	**17-20**
Structure 1	2.28	0.85	1.82	1.39	1.79
Structure 2	1.94	0.55	2.34	1.12	1.24
Structure 3	1.56	0.91	1.63	0.88	2.89
Structure 4	2.10	0.80	0.84	0.99	2.34
Structure 5	1.82	0.89	2.28	1.15	1.94
Average RMSD (Å)	**1.94**	**1.11**	**1.78**	**0.80**	**2.04**

Table 6-3: Native Beta Regions RMSD.
The RMSD (Å) values for particular residue ranges (beta sheet ranges) for the native structure of Beta3s. The data in the table highlights the variance in conformation for each of the 5 native conformation structures analyzed structure (1-5) around the sheet regions by comparison to the turn regions. The data suggests turns are more rigid (and thus structurally homogenous) than sheet regions.

The major peak in the Ns conformation originated at 1653 cm^{-1} and was surrounded by a shoulder at 1630 cm^{-1} and another at 1671 cm^{-1} at Γ of 10 cm^{-1}. (Figure 6-4, Table 6-2) The right shoulders correspond to the high $\nu_{\parallel}$ frequency β-sheet absorption while the central peak and left peaks result from increased random coil character and the $\nu_{\perp}$ mode of the β-sheet in this conformation [140]. Interesting, the low frequency $\nu_{\perp}$, β-sheet peak shifted approximately 6 cm^{-1} lower compared to the native state. A decreasing distance between the high and low frequency Amide-I bands of the β-sheets has been shown to correspond to decreasing β-sheet content [39]. This is expected because the Ns conformation lacks the fully formed N-terminal β-sheet, and thus contains less β-sheet content when compared to the Native state.

The Cs conformation incorporating an out of register C-terminal region contains a predominate peak at 1660 cm^{-1} surrounded by a large shoulder at 1636 cm^{-1} and a weak shoulder at 1676 cm^{-1}. (Figure 6-4, Table 6-2) The 1DIR spectra at 1 cm^{-1} Γ further resolves these peaks particularly the weak shoulder at 1676 cm^{-1}. Relative to the Native structure a 4 cm^{-1} decrease in width between the $\nu_{\perp}$ (1636 cm^{-1}) and $\nu_{\parallel}$ (1676 cm^{-1}) modes of the β-sheet modes was observed indicative of the decrease in β-sheet structure in this conformation. Additionally, the evolution of a significant random coil peak as a result of the less structured C-terminal region was observed.

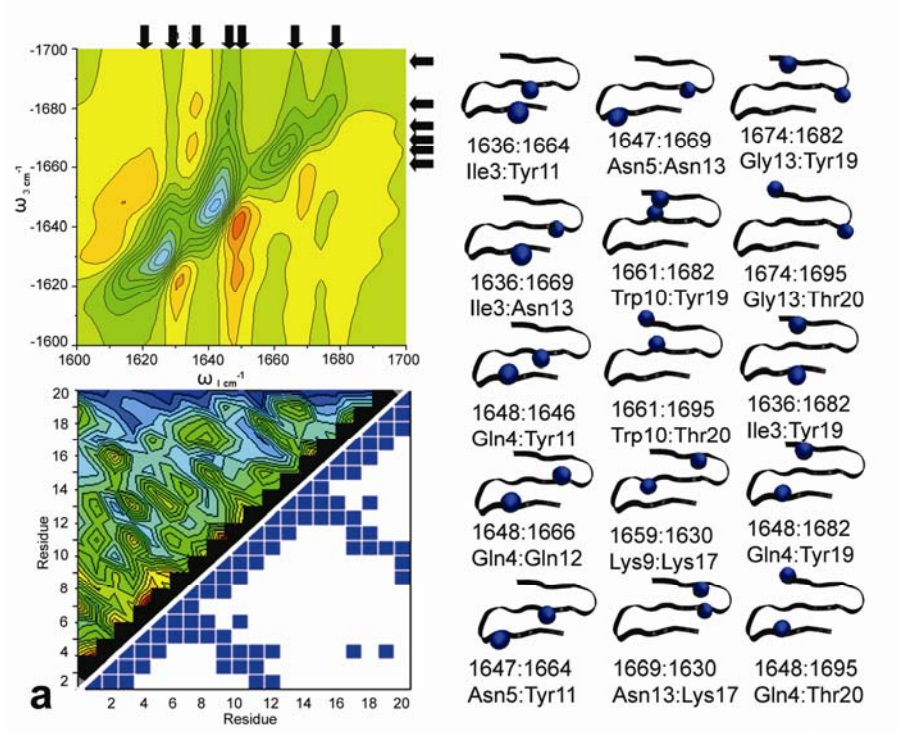
The Ch-Curl conformation is most similar to the Ns conformation containing a well ordered C-terminal β-sheet structure and disrupted N-terminal region. In the Ch-Curl structure however, the C-terminus turn is inverted. (Figure 6-1d) The 1DIR of Ch-Curl consists of 2 peaks at 1638 cm^{-1} and 1671 cm^{-1} in the 10 cm^{-1} Γ regime, additional resolution reveals multiple strong peaks from 1630 cm^{-1} to 1688 cm^{-1}. (Figure 6-4, Table 6-2) Critical β-sheet peaks are present as expected with the well structured C-terminus in this conformation. Variation and multiple strong peaks in the range of ~1645-1660 cm^{-1} appear to result from the mostly unstructured N-terminal part of the structure, providing the "Curl" component of this ensemble. The β-sheet peak occurred at 1638 cm^{-1}, 6 cm^{-1} higher than in the Ns conformation, likely a result of the different interactions between adjacent β-sheets due to the inversion of the C-terminus sheet [39].

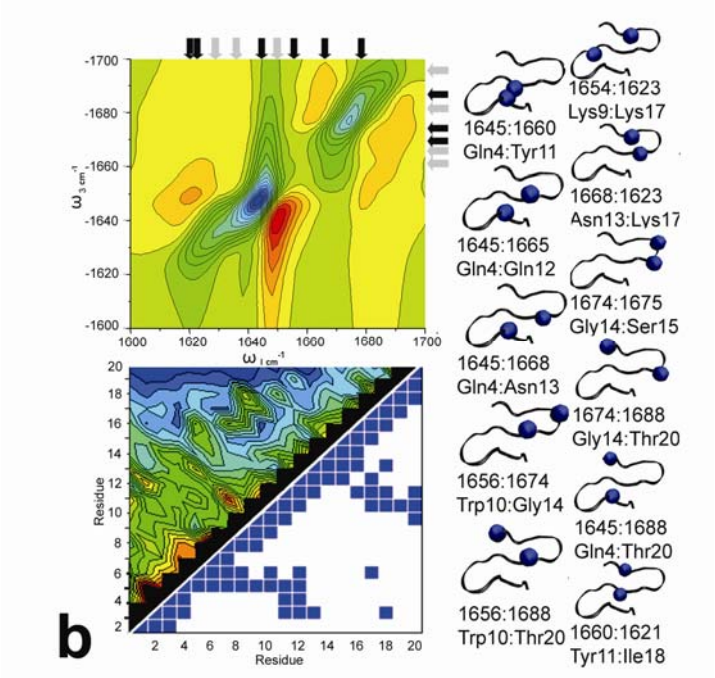
The 1DIR of the 6-12 helical conformation contains a single strong peak at 1658 cm^{-1} which corresponds the an α-helix or random coil Amide-I absorption [140]. (Figure 6-4, Table 6-2) Further resolution at Γ 1 cm^{-1} shows a splitting of the main peak into a 1656 cm^{-1} and 1660 cm^{-1} peak which likely corresponds to the α-helix structure absorption from residues 6 to 12 and the remainder of the structure which is largely in a random coil configuration.

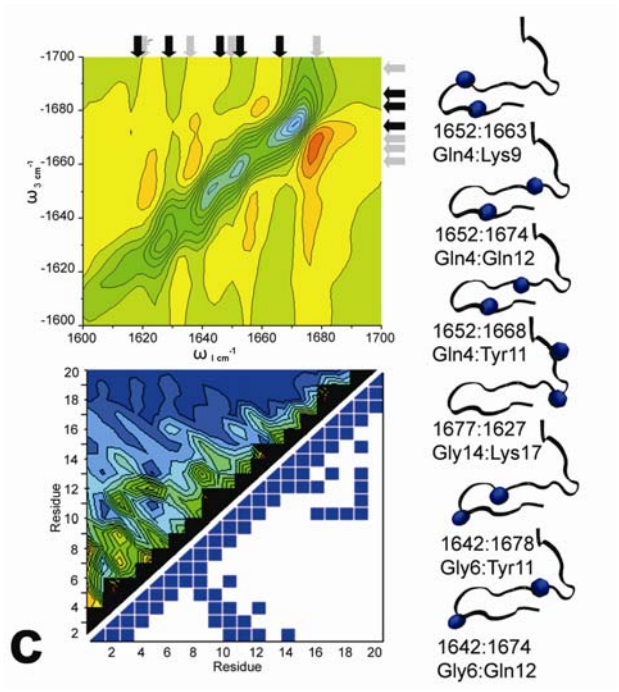
6.3.2 2DIR Spectra Beta3s

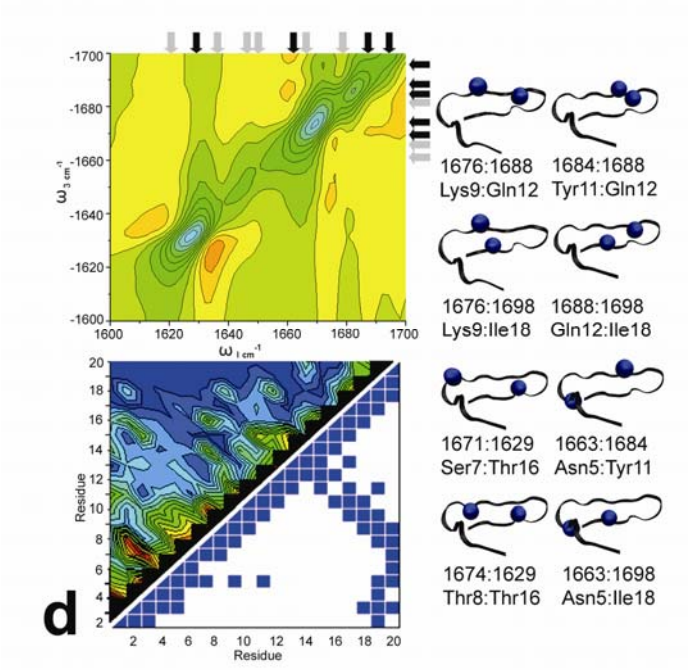
Simulated two-dimensional IR correlation spectroscopy (2DIR) reveals three-dimensional structural information about protein structure by reporting on vibration couplings and correlations between vibrations contacts. Although 1DIR appears to be

sufficient to distinguish the different conformations in the folding mechanism of Beta3s it lacks ability to reveal coupling between specific residues observed in the off-diagonal peaks of the 2D spectrum. Since 2DIR spectra are calculated with the same Local Amide Hamiltonian as the 1DIR spectra diagonal peak locations are identical for similar Γ parameters. The 2DIR as displayed in figure 3 was calculated for 3 different homogenous broadening parameters (Γ) of 1.0 (right), 5.0 (middle) and 10.0 cm^{-1} (left). In figure 6-4 each 2DIR spectrum is split into two regions, the upper left corner (blue peaks) and lower right corner (red peaks), containing signals originating from the 0->1 and 1->2 IR transitions respectively.


a
Residue
Residue
1636:1664
Ile3:Tyr11
1647:1669
Asn5:Asn13
1674:1682
Gly13:Tyr19
1636:1669
Ile3:Asn13
1661:1682
Trp10:Tyr19
1674:1695
Gly13:Thr20
1648:1646
Gln4:Tyr11
1661:1695
Trp10:Thr20
1636:1682
Ile3:Tyr19
1648:1666
Gln4:Gln12
1659:1630
Lys9:Lys17
1648:1682
Gln4:Tyr19
1647:1664
Asn5:Tyr11
1669:1630
Asn13:Lys17
1648:1695
Gln4:Thr20


b
Residue
Residue
1654:1623
Lys9:Lys17
1645:1660
Gln4:Tyr11
1668:1623
Asn13:Lys17
1645:1665
Gln4:Gln12
1674:1675
Gly14:Ser15
1645:1668
Gln4:Asn13
1674:1688
Gly14:Thr20
1656:1674
Trp10:Gly14
1645:1688
Gln4:Thr20
1656:1688
Trp10:Thr20
1660:1621
Tyr11:Ile18


c
1652:1663
Gln4:Lys9
1652:1674
Gln4:Gln12
1652:1668
Gln4:Tyr11
1677:1627
Gly14:Lys17
1642:1678
Gly6:Tyr11
1642:1674
Gly6:Gln12
Residue
Residue


d
1676:1688
Lys9:Gln12
1684:1688
Tyr11:Gln12
1676:1698
Lys9:Ile18
1688:1698
Gln12:Ile18
1671:1629
Ser7:Thr16
1663:1684
Asn5:Tyr11
1674:1629
Thr8:Thr16
1663:1698
Asn5:Ile18
Residue
Residue

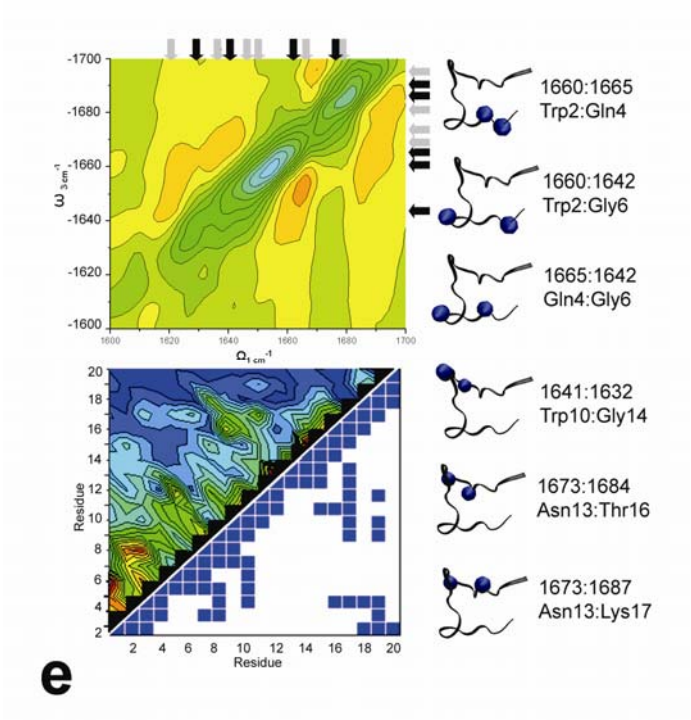


Figure 6-6: The structure-spectral correlation for investigated conformations of Beta3s. Letter designations a) native b) Ns c) Cs d) Ch-Curl conformations e) 6-12 Helix
Top Left: The 2DIR plot at Γ of 5 cm^{-1} for selected conformation. Arrows help highlight peak position. Black arrows highlight current conformation peaks while gray arrows highlight the original native conformation peak locations. Bottom Left: The Normal Mode Decomposition plot of residue coupling intensities (upper left) and structural contact map calculated at 6.7Å cutoff values. Right: The structural representation of the cross peak position and assigned residue. The blue spheres in the structure image represent the residues causing the cross peaks.

Table 6-4: Beta3s Native Peak Assignment

Assigned Peak/Residue											
ω_1	$-\omega_3$	Nat		Ns		Cs		Ch		612	
1636	1664	Ile3	Tyr11	NA	NA	NA	NA	NA	NA	NA	NA
1636	1669	Ile3	Asn13	NA	NA	NA	NA	NA	NA	NA	NA
1648	1664	Gln4	Tyr11	Gln4	Tyr11	Gln4	Lys9	NA	NA	NA	NA
1648	1666	Gln4	Gln12	Gln4	Gln12	Gln4	Gln12	NA	NA	NA	NA
1647	1664	Asn5	Tyr11	NA	NA	Gln4	Tyr11	NA	NA	NA	NA
1647	1669	Asn5	Asn13	Gln4	Asn13	NA	NA	NA	NA	NA	NA
1661	1682	Trp10	Tyr19	Trp10	Gly14	NA	NA	Lys9	Gln12	NA	NA
1661	1695	Trp10	Thr20	Trp10	Thr20	NA	NA	Lys9	Ile18	NA	NA
1659	1630	Lys9	Lys17	Lys9	Lys17	NA	NA	Ser7	Thr16	NA	NA
1669	1630	Asn13	Lys17	Asn13	Lys17	Gly14	Lys17	Thr8	Thr16	NA	NA
1674	1682	Gly14	Tyr19	Gly14	Ser15	NA	NA	Lys9	Gln12	NA	NA
1674	1695	Gly14	Thr20	Gly14	Thr20	NA	NA	Gln12	17	NA	NA
1636	1682	Ile3	Tyr19	NA	NA	NA	NA	NA	NA	NA	NA
1648	1682	Ile4	Tyr19	NA	NA	NA	NA	NA	NA	NA	NA
1648	1695	Ile4	Thr20	Ile4	Thr20	NA	NA	NA	NA	NA	NA

Table 6-4: Beta3s Native Peak Assignment.
Specific peak assignment of mode contributions from NMA for the native conformation of Beta3s. Non-native conformation modes noted when applicable. Mode-peak assignment for non-native conformations with slight variations in mode are presented when a peak is found in the same position as in the native conformation. White background represents modes of N-Terminal Sheet, Shaded background represents modes of C-Terminal Sheet. The complete table including conformation specific peak locations is located in table 6-7 in the appendix.

The native state 2DIR spectrum of Beta3s contains a full complement of cross peak interactions of the folded protein and was used as a point of reference for the other conformations. The native state exhibited 15 distinguishable off-diagonal cross peaks in the 2DIR spectra calculated at a line width of 5 cm^{-1} as noted in table 6-4 and figure 6-4. Three specific regions of the 2DIR spectra report on the general conformation of Beta3s, the 1620-1630 cm^{-1} and 1650-1680 cm^{-1} region contain signals resulting from the C-terminal β-structure. The peaks at 1636-1650 cm^{-1} in the middle of the spectra correspond to the N-terminal sheet, while the 1675 cm^{-1} to 1700 cm^{-1} region from 1620-1650 cm^{-1} reports on long-range coupling between C and N-terminal chains through the

central β-sheet. The Ns 2DIR spectra contained 10 of 15 native cross peaks and a novel peak at 1621 cm^{-1} and 1660 cm^{-1} (Figure 6-6b, Table 6-4 and 6-6) The Cs conformation included four of the native cross peaks and two unique signals. (Figure 6-6c, Table 6-4) The additional cross peaks were noted in the Cs conformation between the in 1642 cm^{-1}'s at 1674 cm^{-1} and 1678 cm^{-1}. (Figure, 6-6c, Table 6-4) The Ch conformation exhibited the six of 15 native cross peak interactions likely all from the C-terminal sheet. The Ch configuration also exhibited 2 novel peaks the in 1660 cm^{-1}'s at 1684 cm^{-1} and 1698 cm^{-1} range. (Figure, 6-6d, Table 6-4 and 6-6) Finally, the 6-12 helix displayed no native peaks and 5 new cross peaks. (Figure 6-6e, Table 6-5)

Table 6-5: Non-Native Peak Assignment

Assigned Peak/Residue							
ω_1	$-\omega_3$	Ns		ω_1	$-\omega_3$	Cs	
1660	1621	Tyr11	*Ile18*	1642	1678	Gly6	Lys9
				1642	1674	Gly6	Gln12

ω_1	$-\omega_3$	Ch		ω_1	$-\omega_3$	612	
1663	1684	Asn5	Tyr11	Trp2	Gln4	1660	1665
1663	1698	Asn5	Ile18	Trp2	Gly6	1660	1642
				Gln4	Gly6	1665	1642
				Trp11	Gly14	1641	1632
				Asn13	Thr16	1673	1684
				Asn13	Lys17	1673	1687

Table 6-5: Non-Native Peak Assignment.
Mode-peak assignments for non-native conformations are presented for non-native peaks derived from NMA analysis.

Table 6-6: Complete Peak Assignment for Beta3s

ω_1	$-\omega_3$	Nat		ω_1	$-\omega_3$	Ns		ω_1	$-\omega_3$	Cs	
1636	1664	Ile3	Tyr11	NA	NA	NA	NA	NA	NA	NA	NA
1636	1669	Ile3	Asn13	NA	NA	NA	NA	NA	NA	NA	NA
1648	1664	Gln4	Tyr11	1645	1660	Gln4	Tyr11	1652	1663	Gln4	Lys9
1648	1666	Gln4	Gln12	1645	1665	Gln4	Gln12	1652	1674	Gln4	Gln12
1647	1664	Asn5	Tyr11	NA	NA	NA	NA	1652	1668	Gln4	Tyr11
1647	1669	Asn5	Asn13	1645	1668	Gln4	Asn13	NA	NA	NA	NA
1661	1682	Trp10	Tyr19	1656	1674	Trp10	Gly14	NA	NA	NA	NA
1661	1695	Trp10	Thr20	1656	1688	Trp10	Thr20	NA	NA	NA	NA

1659	1630	Lys9	Lys17	1654	1623	Lys9	Lys17	NA	NA	NA	NA
1669	1630	Asn13	Lys17	1668	1623	Asn13	Lys17	1677	1627	Gly14	Lys17
1674	1682	Gly14	Tyr19	1674	1675	Gly14	Ser15	NA	NA	NA	NA
1674	1695	Gly14	Thr20	1674	1688	Gly14	Thr20	NA	NA	NA	NA
1636	1682	Ile3	Tyr19	NA	NA	NA	NA	NA	NA	NA	NA
1648	1682	Ile4	Tyr19	NA	NA	NA	NA	NA	NA	NA	NA
1648	1695	Ile4	Thr20	1645	1688	Ile4	Thr20	NA	NA	NA	NA

ω_1	$-\omega_3$	Ch		ω_1	$-\omega_3$	612	
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
1676	1688	Lys9	Gln12	NA	NA	NA	NA
1676	1698	Lys9	Ile18	NA	NA	NA	NA
1671	1629	Ser7	Thr16	NA	NA	NA	NA
1674	1629	Thr8	Thr16	NA	NA	NA	NA
1684	1688	Lys9	Gln12	NA	NA	NA	NA
1688	1698	Gln12	Ile18	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA

Table 6-6: Full Native Peak Assignment for Beta3s.
The table highlights the exact residue contact and peak location as derived from NMD analysis. Assigned residues may differ slightly from the residues for similar peaks in the native state. This is reported here because uncertainty peak assignment due to NMD in which delocalization is not accounted for. White background represents modes of N-Terminal sheet, dark shaded background represents modes of C-Terminal sheet, light shaded region indicated peaks due to C-Terminal and N-Terminal coupling through the central sheet structure.

6.3.3 Residue Contributions and Peak Assignment

One of the primary goals of 2DIR is the assignment of specific cross peaks to particular residue interactions to reveal the three-dimensional structure of proteins [135]. This has also been among the most challenging tasks for the 2DIR spectroscopists. In this work the small size of Beta3s along with the multiple conformations of known structure helps facilitate peak assignment. This is an ideal situation, unique to

computation, because experimentalists normally do not have atomic conformations for comparison or work with very large proteins that complicate structure assignment and necessitate isotopic labels to isolate specific peaks [135, 250].

Structure-peak assignment of Beta3s was determined by two methods, conformational difference analysis and normal mode decomposition (NMD). Examination of the Native structure ensemble alone suggests that Beta3s contains two distinct cross peak contributors originating from residues on the C-terminal and on the N-terminal β-strands. This, however, does not address which residues contribute to each peak. NMD analysis of the Native conformation of Beta3s allowed us to ascertain 15 readily identifiable residues that contribute to 15 peaks in the 2DIR spectra. (Table 6-4, Figure 6-6a) The eigenvalues from NMD allowed us to approximate the residue carbonyl group origin of peaks by assigning an IR absorption frequency to each residue. (Figure 6-6a, Table 6-4) The eigenvalues assigned to each residue by NMD are shown in Table 6-6. It is important to remember that NMD analysis provides absorption frequencies for each residue but that these do not fully incorporate delocalization and overlap between the modes of the residues.

Table 6-7: NMD Residue Eigenvalues

Residue	Nat	Ns	Cs	Ch	612
Thr2	1653	1646	1647	1660	1660
Ile3	1636	1640	1646	1657	1661
Gln4	1648	1645	1652	1656	1665
Asn5	1647	1644	1654	1663	1664
Gly6	1643	1642	1642	1648	1666
Ser7	1643	1650	1638	1671	1656
Thr8	1642	1640	1661	1674	1656
Lys9	1659	1654	1663	1676	1651
Trp10	1661	1656	1667	1678	1649
Tyr11	1664	1660	1668	1684	1641
Gln12	1666	1665	1674	1688	1642
Asn13	1669	1668	1674	1639	1673
Gly14	1674	1674	1677	1633	1632
Ser15	1636	1675	1632	1631	1630
Thr16	1667	1624	1631	1629	1684
Lys17	1630	1623	1627	1623	1687
Ile18	1627	1621	1684	1698	1693
Tyr19	1682	1620	1617	1613	1612
Thr20	1695	1688	1593	1594	1610

Table 6-7: NMD Residue Eigenvalues.
The NMD derived Eigenvalues (in frequency cm^{-1}) for each residue studied in Beta3s. The conformational structure (columns 2-6) associated residue (first column) and frequency in cm^{-1}. Note that NMD analysis provides frequency for each residue but that these do not fully incorporate delocalization and overlap between the modes of the residues.

NMD analysis showed six peaks originated from N-terminal strand residue interaction with the central strand and six peaks originating from C-terminal to central strand interaction. (Figure 6-4a, bottom left) Interestingly, NMD also revealed peaks resulting from long range coupling between C-terminal and N-terminal residues (coupled through the central strand) which provides an indicator of degree of Native structure.

NMD analysis assigns the peak at 1636 cm^{-1} to Ile3 and a peak at 1647 cm^{-1} to Asn5, both of which couple to central-strand Tyr11 and Asn13 at 1664 cm^{-1} and 1669 cm^{-1} respectively. (Table 6-4) This is consistent with our 1DIR data that under initial conformational analysis suggested a 1636 cm^{-1} peak originates on the N-terminal strand. Gln4 absorbing at 1648 cm^{-1} also interacted strongly with Tyr11 and Gln12

producing additional cross peaks at 1664 cm^{-1} and 1666 cm^{-1}. Peaks originating from Gln4 and Asn5 are established by coupling in the NMD plots in figure 6-6a (bottom left) but less discernable in the 2DIR spectra. These nearly identical peak locations are likely a result of strong coupling between nearby residues, suggesting significant electronic delocalization. Together the peaks from 1636 cm^{-1} to 1647 cm^{-1} provide a spectral region representative of the degree native-ness of the N-terminal sheet. Conformational analysis supports these peaks as an indicator of the N-terminal sheet since it was also found that these peaks are not present when the N-terminal region was interrupted, similar to the Ns conformation.

The native conformation of Beta3s exhibited 6 peaks consistent with residue interactions on the C-terminal β-strand. Specifically, Trp10 and Gly14 were found to interact with Tyr19 and Thr20 producing cross peaks from $\nu_{\parallel}$ interactions at 1661 cm^{-1}, 1674 cm^{-1} and 1682 cm^{-1} and a 1695 cm^{-1} respectively. (Figure 6-6a) Additionally, $\nu_{\perp}$ interactions were noted between residue Lys17 at 1630 cm^{-1} in the C-terminal sheet and Lys9 and Asn13 at 1659 cm^{-1} at 1669 cm^{-1}. Conformational analysis further supports the NMD results, since the majority of C-terminal strand residues do not produce cross peaks when interrupted in the Cs conformation. (Figures 6-6a)

In the Ns conformation, the interactions involving Trp2 and Asn5 were not noted corresponding to the Ns out-of-register structural disruption. (Table 6-4, Figure 6-6b) Gln4 however, exhibited an interaction with residues Tyr11, Gln12 and Asn13 forming a cross peak between 1645 cm^{-1} and 1660 cm^{-1} as well as 1665 cm^{-1} and 1668 cm^{-1}. Residues on the C-terminal strand produced similar peaks to that of the native structure.

A blue shift of ~5 cm^{-1} relative to the native conformation was noted for the 1656 cm^{-1} peak resulting from coupling to Trp10 on the central β-strand. The blue shift has been observed in prior work and occurs as the Amide modes localize as a result of β-sheet unfolding [249]. The majority of the long-range interactions between the C and N-terminal β-sheets were disrupted in this conformation as anticipated. (Figure 6-6b, Table 6-4)

The Cs conformation was noted to contain similar peaks to the native state with the exception of anticipated disruption in the C-terminal strand. (Figure 6-6c, Table 6-4) A single C-terminal interaction was noted between Gly14 and Lys17 producing the peak at 1627 cm^{-1} and 1677 cm^{-1}. Half of the N-terminal peaks were noted with peaks between Gln4 at 1652 cm^{-1} and Lys9, Tyr11 and Gln12. (Figure 6-6c) Cross peaks at the N-terminal region as a result of Ile3 interacting with Tyr11 and Asn13 were not present in the Cs conformation. Additionally, N-terminal signals due to the Asn5 to Asn13 interaction were not noted in the spectra or by the NMD. Interestingly, new peaks evolved at 1642 cm^{-1} and 1674 cm^{-1} as well as 1678 cm^{-1} a result of Gly6 interacting with Lys9 and Gln12 respectively. These spectral signatures reflect that structurally the N-terminal domain is perturbed in the Cs conformation as the C-terminal domain falls out of register. No blue shifts were noted relative to the native conformation in the Cs state. The contact map at the bottom of figure 6-6c displays the disruption of overall secondary structure. This is unsurprising, since it has been suggested that folding of Beta3s prefers to first fold a structurally stable C-terminal domain that forms a scaffolding to facilitate folding of the N-terminal domain [226-231]. In the spectra of Cs such structural changes are ultimately reflected in the red-shifting of noted peaks relative to the native state

caused by the peptide backbone shifting from a β-sheet (1630-1640 cm^{-1}) to a more α-helical conformation (1650-1660 cm^{-1}).

The Ch-curled conformation exhibited complete disruption of the peaks corresponding to the N-terminal region. (Figure 6-6d, Table 6-4) Two additional interactions near the N-terminus indicated by peaks at 1663 cm^{-1}, 1684 cm^{-1} and 1698 cm^{-1} attributed to Asn5 interacting with Tyr11 and Ile18. (Table 6-5) The remainder of the N-terminal domain did not produce cross peaks in this conformation. In the Ch-curl structure the C-terminal region is inverted relative to the native state and as such a new set of contacts cause the spectral cross peaks. The new peaks associated with the C-terminal sheet are red shifted as much as 10 cm^{-1} but are associated with residues similar to or close to those resulting in the C-terminal peaks of the native state. (Table 6-4 and 6-5) Ser7 and Thr8 interact with Thr16 at the same location as in the native state at 1629 cm^{-1} but the cross peak is far shifted Ser7 and Thr8 locations. Largely this is a result of the conformation of residues 7 and 8 which become helical in the turn region because of N-terminal domain disruption.

Most different from the native structure, the 6-12 helical conformation of Beta3s exhibited none of the native cross peaks. (Figure 6-6e, Table 6-4 and 6-5) NMD revealed that although some peaks and modal interactions appear similar to those in the Native state they originate from interactions of different residues. Six new interactions were noted as described in table 6-5. Local backbone interactions in the 1660 cm^{-1}'s between Trp2 and Gln4 and Gly6 as well as those between Asn13, Thr16 and Lys17 are typical of the random coil and α-helical structure noted here. Additionally, the NMD plot reveals

significant coupled interactions along the diagonal indicating strong coupling to nearby N+1. . .N+3 residues [140]. (Figure 6-6e, bottom left) Such residue interactions are highly local and thus largely in the diagonal of the 2DIR spectra due to their very similar absorption frequencies.

6.3.4 The Folding Mechanisms

Computational studies by Caflisch et al. have suggested Beta3s folds through two possible pathways [227, 228]. The main folding pathway of Beta3s starts with the formation of the C-terminal side chain contacts followed by the N-terminal contacts. (Figure 6-2 and Figure 6-4) Additionally, the reverse was also found to be possible where the N-terminal structure forms first followed by the C-terminal ones. Contacts in the turn regions were also found to form first [227, 228]. The proposed folding pathways have been examined numerous times by many methodologies computationally but never experimentally [226-231]. The 2DIR data presented here can be coupled with 2DIR experiments to investigate the intermediates and the order in which they are sampled during the folding of Beta3s.

6.3.5 Spectral Signatures of Folding

In this study, specific spectra-structure correlations have been established that can provide unambiguous indicators of conformational identity during a folding experiment. Specifically, during folding Beta3s may sample the 6-12 helical conformation on its way to the native state. Transition from the 6-12 helix conformation to a more native like

structure (β-sheet) is well described by the diagonal signals which decrease from a 1660 cm^{-1} centered primary peak to one in the 1640 cm^{-1}. Moreover, evolution of folding along the primary pathway, sampling the Ns conformation before the native state, is indicated by the evolution of 1648 cm^{-1} cross peaks a result of Gln4 coupling as well as signals from the full compliment of C-terminal peaks. (Figure 6-6b, Table 6-4) The existence of the Gln4 peaks also differentiates the Ns conformation spectrally from the Ch-Curl conformation which lacks this structure but contains a similar C-terminal spectral signature. (Figure 6-6d) Finally, in the alternate folding pathway the N-terminal region forms first, thus sampling the Cs conformation prior to the native state. 2DIR spectral analysis along this folding path would include the Gln4 associated cross peaks and nearly none of the C-terminal associated peaks in the 1666 cm^{-1}-1680 cm^{-1} range. (Figure 6-6c, Table 6-4)

6.3.6 Isotopic Labeling Experiments

Isotope labeling can be used to manipulate 2DIR signals by enhancing desired spectral features as demonstrated by Hochstrasser [251]. ^{13}C and ^{18}O labeling of peptides can induce 65 cm^{-1} red shift of Amide-I bands providing detailed structural constraints. In this work the different conformations of Beta3s were known and so isotopic labeling is of minimal help in identifying structure-peak correlations because they can be determined by conformational analysis alone, however because the information is not available in experiments, labels may aid in tracking folding. In practice labeling of Gln4 and Tyr19 (both commercially available) could provide insight into coupling of the N-terminal sheet and C-terminal sheet respectively.

6.4 Conclusions

We have shown that 2DCS IR spectra of proteins coupled with conformational sampling though folding calculations can reveal significant structural information about the ensemble evolution in the folding mechanism. This coupled with experiment can help to provide unprecedented information about the folding process including structurally resolved folding kinetics. It is clear, even from the 1DIR, spectrum that the intermediates sampled in the folding mechanism of Beta3s exhibit distinct spectroscopic characteristics. Although 1DIR provides some insight into the specific conformation of the peptide the 2DIR method allows tracking of specific cross peaks and associated atomic contacts that provide critical indicators of the folding mechanism. The results presented here suggest that further insight into the folding pathway of Beta3s can be obtained from experimental work similar to T-jump experiments performed on Ubiquitin and Amyloids [219, 252, 253]. Additionally, considering the extent to which the Beta3s peptide has been studied computationally by a number of models, theses results can also be applied to validation of molecular dynamics force fields [226-231].

6.5 Acknowledgments

We would also like to thank Dr. Wei Zhuang and Brit Hyland for technical support and helpful discussions regarding data analysis. S.M. gratefully acknowledges the support of the National Institutes of Health (Grant GM59230), and the National Science Foundation (Grant CHE-0745892) This work was supported in part by J.W.'s 2005 NSF Career Award.

7 Summary

7.1 Characterizing the Protein Folding Mechanism

The work presented in this dissertation provides insights into the protein folding mechanism through the NTL9 and Beta3s model systems. In an approach that synergistically utilized experimental data and computation the TSE of NTL9 is characterized at atomic resolution. Results suggest that initial stages in the folding pathway involve building basic elements of secondary structure and hydrophobic collapse. After crossing the TSE, folding consists of optimizing interactions between the initially formed secondary structures. This study constitutes new knowledge of not only a necessary stage on of the folding pathway of NTL9 but also the TSE the of the split β-α-β fold motif it contains, which is also found in a large number of proteins. The 2DIR spectrum of the folding pathway of Beta3s has been simulated and analyzed. This represents the first time an all-atom folding pathway of a protein has been explored by 2DIR simulation. The structural origin of diagonal and off-diagonal peaks in the spectra for the native and intermediate conformational ensembles in the folding mechanism was revealed. Specific indicators of folding are revealed in the 2DIR spectrum of the folding pathway that will be useful in comparison to experiment. As such, the results presented suggest that further insight into the folding pathway of Beta3s can be obtained from experimental work similar to T-jump experiments in the past [252, 253]. Additionally, considering the extent to which the Beta3s peptide has been studied computationally by a number of models, theses results can also be applied to the validation of molecular dynamics force fields [226-231].

7.2 Procedures for Simulating Folding and 2DIR Spectra

The 2DIR spectrum of peptides and proteins is dependent upon the protein backbone configuration as well as the local environment including solvation. Unfortunately, at current there are few proteins that can be simulated by all-atom simulations. As such the work on Beta3s can demonstrate a generalized methodology for obtaining 2DIR spectra for the folding pathways using reduced representation models. The simulation of the 2DIR spectrum of a protein begins with a structural starting point, most often an ensemble. In the case of folding this requires an ensemble of structures at different locations on the folding pathways. In the case of large proteins the ideal all-atom folding simulations are not currently available due to the extensive computational cost. Reduced representation models (i.e.: lattice and Go-type models) however, offer a possible solution to investigate the folding pathways of proteins too large to currently simulate by all-atom simulation. These models are helpful because the folding pathways are well sampled with limited computational effort. Unfortunately, reduced representation models, often neglect components critical to accurate simulation of the 2DIR spectrum of proteins. In the work presented on Beta3s in Chapter 6, a reduced representation model was expanded to include relevant components. In this model, the solvent and non-polar hydrogen atoms were not explicitly solvated but rather added to the simulation ensembles prior to calculation of the 2DIR spectrum.

The procedure employed in the work on Beta3s and the general underling concept is expanded here to allow the calculation of 2DIR for folding pathways of larger proteins.

In addition to expanding the scope of 2DIR calculations this method, in conjunction with experimental spectra can explore the relevance of the models used to derive such pathways. Details of this general procedure are explained in the steps below and in Figure 7-1:

1) A folding or unfolding simulation containing a reduced representation model is run to sample the folding landscape. A significant number of such models exist with a reduced representation of either the solvent and protein system or both.
2) Trajectories from step one are analyzed for relevant intermediate and transition state ensembles. Numerous techniques already exist to conduct this step.
3) The models from step 2 are "scaled up" to include all-atom components relevant to the simulation of the 2DIR spectrum. In the most basic models the system is built out to include all the atoms of the backbone and side chains of the protein as well as the solvent. Since only relevant ensembles need to be investigated the system is now limited in scope which means computational requirements are reduced.
4) The built out system is subjected to an equilibration to remove bad contacts that may have been introduced in the last step. Several protocols exist to treat this same problem and are typically addressed when using an experimentally derived structure in all-atom molecular dynamics simulations. In essence these should involve a constrained minimization of the protein under investigation. Constraints should be placed on the backbone configurations and the remaining structure should be subject to short minimization and potentially molecular dynamics. Following this the structure should be solvated with explicit solvent.

The solvent should then also be subjected to a short molecular dynamics simulation with the proteins system constrained to position it correctly relative to the structure.

5) The ensemble structures including all-atoms and solvent are ready to be included in simulation 2DIR simulation software.

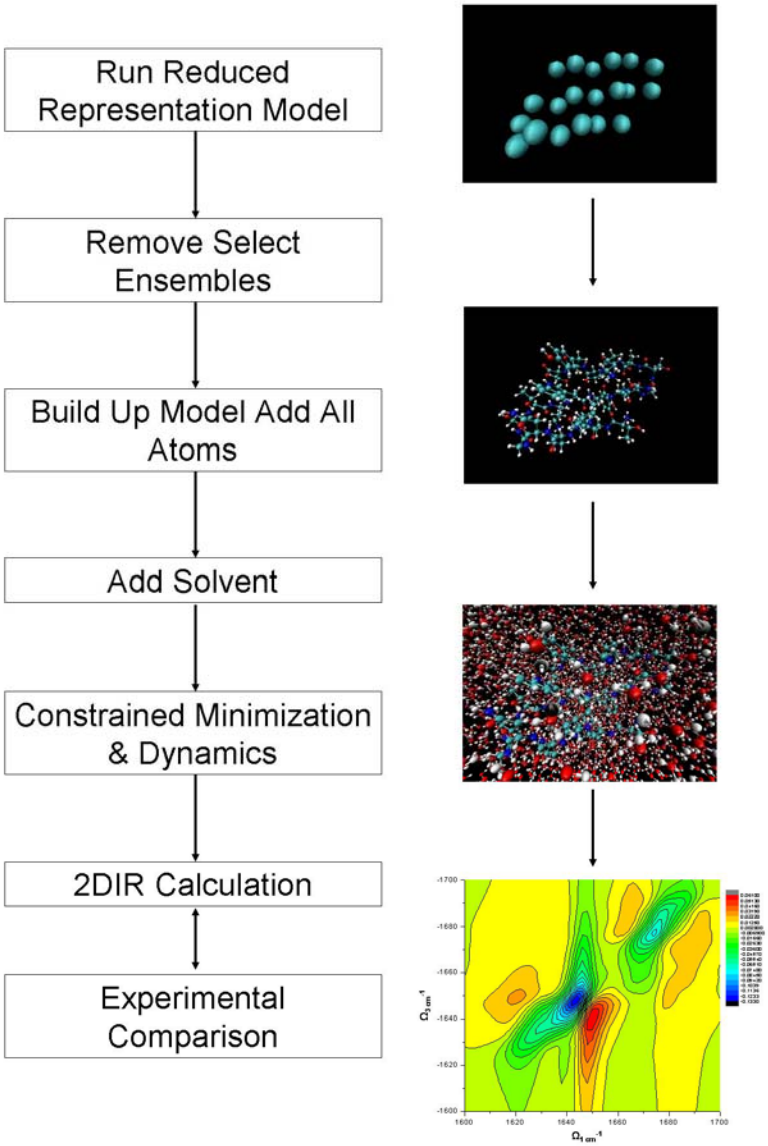


Figure 7-1: 2DIR simulation protocols for adapting reduced representation models.
The stepwise practical procedure for simulation of 2DIR calculations is highlighted. Images on the right hand side, top to bottom, Cα model system, scaled up system (now all-atoms are includes), solvent added to all atom system, simulated 2DIR of full system.

7.3 Future Directions

In this dissertation the work and data presented on Beta3s folding mechanism and its 2DIR spectrum suggests several future projects. Specific to Beta3s specific isotopic labels have been suggested based upon spectral peaks that provide and indication of the conformational state. This data proposes that an experimental setup using these indicators could help track the folding process of Beta3s by 2DIR. Moreover, this work supports the implementation of computational techniques in conjunction with experimental 2DIR to study the folding mechanism of proteins. In addition to exploring the folding mechanism this work can be applied in combination with experiment to refine and validate current molecular dynamics force fields. Finally, a generalized method has been proposed in section 7.2 to allow for folding trajectories of reduced representation models to be simulated by 2DIR. The resultant calculated spectra of these folding pathways when compared with experimental 2DIR data will undoubtedly further our understanding of the protein folding problem.

8 References

1. Anfinsen CB: **Formation and Stabilization of Protein Structure.** *Biochemical Journal* 1972, **128:**737-&.
2. Anfinsen CB: **Principles That Govern Folding of Protein Chains.** *Science* 1973, **181:**223-230.
3. Dobson CM: **Protein folding and misfolding.** *Nature* 2003, **426:**884-890.
4. Eaton WA, Munoz V, Thompson PA, Henry ER, Hofrichter J: **Kinetics and dynamics of loops, alpha-helices, beta-hairpins, and fast-folding proteins.** *Accounts of Chemical Research* 1998, **31:**745-753.
5. Kang J, Lemaire HG, Unterbeck A, Salbaum JM, Masters CL, Grzeschik KH, Multhaup G, Beyreuther K, Muller-Hill B: **The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor.** *Nature* 1987, **325:**733-736.
6. Thomas PJ, Qu BH, Pedersen PL: **Defective protein folding as a basis of human disease.** *Trends Biochem Sci* 1995, **20:**456-459.
7. Dobson CM: **The structural basis of protein folding and its links with human disease.** *Philos Trans R Soc Lond B Biol Sci* 2001, **356:**133-145.
8. Horwich A: **Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions.** *J Clin Invest* 2002, **110:**1221-1232.
9. Bullock AN, Fersht AR: **Rescuing the function of mutant p53.** *Nat Rev Cancer* 2001, **1:**68-76.
10. Fersht AR: **Characterizing Transition-States in Protein-Folding - an Essential Step in the Puzzle.** *Current Opinion in Structural Biology* 1995, **5:**79-84.
11. Fersht AR, Daggett V: **Protein folding and unfolding at atomic resolution.** *Cell* 2002, **108:**573-582.
12. Gianni S, Guydosh NR, Khan F, Caldas TD, Mayor U, White GWN, DeMarco ML, Daggett V, Fersht AR: **Unifying features in protein-folding mechanisms.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100:**13286-13291.
13. Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SMV, Alonso DOV, Daggett V, Fersht AR: **The complete folding pathway of a protein from nanoseconds to microseconds.** *Nature* 2003, **421:**863-867.
14. Daggett V, Fersht AR: **Is there a unifying mechanism for protein folding?** *Trends in Biochemical Sciences* 2003, **28:**18-25.
15. Anfinsen CB, Haber E, Sela M, White FH: **Kinetics of Formation of Native Ribonuclease during Oxidation of Reduced Polypeptide Chain.** *Proceedings of the National Academy of Sciences of the United States of America* 1961, **47:**1309-&.
16. Levintha.C: **Are There Pathways for Protein Folding.** *Journal De Chimie Physique Et De Physico-Chimie Biologique* 1968, **65:**44-&.
17. Onuchic JN, LutheySchulten Z, Wolynes PG: **Theory of protein folding: The energy landscape perspective.** *Annual Review of Physical Chemistry* 1997, **48:**545-600.

18. Onuchic JN, Socci ND, LutheySchulten Z, Wolynes PG: **Protein folding funnels: The nature of the transition state ensemble.** *Folding & Design* 1996, **1:**441-450.
19. Baldwin RL: **Protein-Folding - Matching Speed and Stability.** *Nature* 1994, **369:**183-184.
20. Wolynes PG: **Solid-state physics - A fresh glass of frozen chaos.** *Nature* 1996, **382:**495-496.
21. Sali A, Shakhnovich E, Karplus M: **How Does a Protein Fold.** *Nature* 1994, **369:**248-251.
22. Dill KA, Chan HS: **From Levinthal to pathways to funnels.** *Nature Structural Biology* 1997, **4:**10-19.
23. Clementi C, Nymeyer H, Onuchic JN: **Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins.** *Journal of Molecular Biology* 2000, **298:**937-953.
24. Fersht AR, Daggett V: **Protein folding and unfolding at atomic resolution.** *Cell* 2002, **108:**573-582.
25. Morelli B: **Kinetic Experiment Using a Spring Powered, Stopped-Flow Apparatus.** *Journal of Chemical Education* 1976, **53:**119-122.
26. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M: **Simultaneous determination of protein structure and dynamics.** *Nature* 2005, **433:**128-132.
27. Park S, Kwak K, Fayer MD: **Ultrafast 2D-IR vibrational echo spectroscopy: a probe of molecular dynamics.** *Laser Physics Letters* 2007, **4:**704-718.
28. Zheng JR, Kwak K, Asbury J, Chen X, Piletic IR, Fayer MD: **Ultrafast dynamics of solute-solvent complexation observed at thermal equilibrium in real time.** *Science* 2005, **309:**1338-1343.
29. Kim YS, Hochstrasser RM: **Chemical exchange 2D IR of hydrogen-bond making and breaking.** *Proc Natl Acad Sci U S A* 2005, **102:**11185-11190.
30. Zheng J, Kwak K, Xie J, Fayer MD: **Ultrafast carbon-carbon single-bond rotational isomerization in room-temperature solution.** *Science* 2006, **313:**1951-1955.
31. Shim SH, Gupta R, Ling YL, Strasfeld DB, Raleigh DP, Zanni MT: **Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution.** *Proc Natl Acad Sci U S A* 2009, **106:**6614-6619.
32. Scheraga HA, Khalili M, Liwo A: **Protein-folding dynamics: Overview of molecular simulation techniques.** *Annual Review of Physical Chemistry* 2007, **58:**57-83.
33. Shea JE, Brooks CL: **From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding.** *Annual Review of Physical Chemistry* 2001, **52:**499-535.
34. Hummer G, Garcia AE, Garde S: **Helix nucleation kinetics from molecular simulations in explicit solvent.** *Proteins-Structure Function and Genetics* 2001, **42:**77-84.

35. Pillardy A, Czaplewski C, Liwo A, Lee J, Ripoll DR, Kazmierkiewicz R, Oldziej S, Wedemeyer WJ, Gibson KD, Arnautova YA, et al: **Recent improvements in prediction of protein structure by global optimization of a potential energy function.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98:**2329-2333.
36. Day R, Daggett V: **Sensitivity of the folding/unfolding transition state ensemble of chymotrypsin inhibitor 2 to changes in temperature and solvent.** *Protein Science* 2005, **14:**1242-1252.
37. Jemth P, Gianni S, Day R, Li B, Johnson CM, Daggett V, Fersht AR: **Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101:**6450-6455.
38. Li AJ, Daggett V: **Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations.** *Journal of Molecular Biology* 1996, **257:**412-429.
39. Ganim Z, Tokmakoff A: **Spectral signatures of heterogeneous protein ensembles revealed by MD simulations of 2DIR spectra.** *Biophysical Journal* 2006, **91:**2636-2646.
40. Chung HS, Khalil M, Tokmakoff A: **Protein denaturing studied with 2D IR and vibrational echo spectroscopy: Equilibrium and temperature-jump measurements.** *Biophysical Journal* 2004, **86:**526a-526a.
41. Arrondo JLR, Iloro I, Aguirre J, Goni FM: **A two-dimensional IR spectroscopic (2D-IR) simulation of protein conformational changes.** *Spectroscopy-an International Journal* 2004, **18:**49-58.
42. Arrondo JLR, Iloro L, Aguirre J, Goni FM: **A 2D-IR simulation of protein conformational changes.** *Biophysical Journal* 2003, **84:**5A-6A.
43. Dobson CM, Fersht AR: *Protein folding : a discussion.* Cambridge ; New York: Cambridge University Press; 1995.
44. Best RB, Hummer G: **Reaction coordinates and rates from transition paths.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**6732-6737.
45. Chavez LL, Onuchic JN, Clementi C: **Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates.** *Journal of the American Chemical Society* 2004, **126:**8426-8432.
46. Onuchic JN, Wolynes PG: **Theory of protein folding.** *Current Opinion in Structural Biology* 2004, **14:**70-75.
47. Anil B, Sato S, Cho JH, Raleigh DP: **Fine structure analysis of a protein folding transition state; distinguishing between hydrophobic stabilization and specific packing.** *Journal of Molecular Biology* 2005, **354:**693-705.
48. Cho JH, Raleigh DP: **Electrostatic interactions in the denatured state and in the transition state for protein folding: Effects of denatured state interactions on the analysis of transition state structure.** *Journal of Molecular Biology* 2006, **359:**1437-1446.
49. Cho JH, Sato S, Raleigh DP: **Thermodynamics and kinetics of non-native interactions in protein folding: A single point mutant significantly stabilizes**

the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *Journal of Molecular Biology* 2004, **338:**827-837.
50. Fulton KF, Main ERG, Daggett V, Jackson SE: **Mapping the interactions present in the transition state for unfolding/folding of FKBP12.** *Journal of Molecular Biology* 1999, **291:**445-461.
51. Garcia-Mira MM, Boehringer D, Schmid FX: **The folding transition state of the cold shock protein is strongly polarized.** *Journal of Molecular Biology* 2004, **339:**555-569.
52. Hamill SJ, Steward A, Clarke J: **The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology.** *Journal of Molecular Biology* 2000, **297:**165-178.
53. Horng JC, Raleigh DP: **Phi-Values beyond the ribosomally encoded amino acids: Kinetic and thermodynamic consequences of incorporating trifluoromethyl amino acids in a globular protein.** *Journal of the American Chemical Society* 2003, **125:**9286-9287.
54. Kim DE, Fisher C, Baker D: **A breakdown of symmetry in the folding transition state of protein L.** *Journal of Molecular Biology* 2000, **298:**971-984.
55. Kragelund BB, Osmark P, Neergaard TB, Schiodt J, Kristiansen K, Knudsen J, Poulsen FM: **The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP.** *Nature Structural Biology* 1999, **6:**594-601.
56. Martinez JC, Viguera AR, Berisio R, Wilmanns M, Mateo PL, Filimonov VV, Serrano L: **Thermodynamic analysis of alpha-spectrin SH3 and two of its circular permutants with different loop lengths: Discerning the reasons for rapid folding in proteins.** *Biochemistry* 1999, **38:**549-559.
57. Northey JGB, Maxwell KL, Davidson AR: **Protein folding kinetics beyond the Phi value: Using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state.** *Journal of Molecular Biology* 2002, **320:**389-402.
58. Petrovich M, Jonsson AL, Ferguson N, Daggett V, Fersht AR: **phi-Analysis at the experimental limits: Mechanism of beta-hairpin formation.** *Journal of Molecular Biology* 2006, **360:**865-881.
59. Raleigh DP, Plaxco KW: **The protein folding transition state: what are phi-values really telling us?** *Protein and Peptide Letters* 2005, **12:**117-122.
60. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D: **Experiment and theory highlight role of native state topology in SH3 folding.** *Nature Structural Biology* 1999, **6:**1016-1024.
61. Villegas V, Martinez JC, Aviles FX, Serrano L: **Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain.** *Journal of Molecular Biology* 1998, **283:**1027-1036.
62. Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM: **Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding.** *Nature Structural Biology* 1999, **6:**1005-1009.
63. Itzhaki LS, Otzen DE, Fersht AR: **The Structure of the Transition-State for Folding of Chymotrypsin Inhibitor-2 Analyzed by Protein Engineering**

Methods - Evidence for a Nucleation-Condensation Mechanism for Protein-Folding. *Journal of Molecular Biology* 1995, **254:**260-288.
64. Matouschek A, Kellis JT, Serrano L, Fersht AR: **Mapping the Transition-State and Pathway of Protein Folding by Protein Engineering.** *Nature* 1989, **340:**122-126.
65. Fersht AR, Matouschek A, Serrano L: **The Folding of an Enzyme .1. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding.** *Journal of Molecular Biology* 1992, **224:**771-782.
66. Serrano L, Matouschek A, Fersht AR: **The Folding of an Enzyme .3. Structure of the Transition-State for Unfolding of Barnase Analyzed by a Protein Engineering Procedure.** *Journal of Molecular Biology* 1992, **224:**805-818.
67. Sanchez IE, Kiefhaber T: **Origin of unusual Phi-values in protein folding: Evidence against specific nucleation sites.** *Journal of Molecular Biology* 2003, **334:**1077-1085.
68. Rao F, Settanni G, Guarnera E, Caflisch A: **Estimation of protein folding probability from equilibrium simulations.** *Journal of Chemical Physics* 2005, **122:**-.
69. Settanni G, Rao F, Caflisch A: **phi-Value analysis by molecular dynamics simulations of reversible folding.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**628-633.
70. Vendruscolo M, Paci E, Dobson CM, Karplus M: **Three key residues form a critical contact network in a protein folding transition state.** *Nature* 2001, **409:**641-645.
71. Paci E, Vendruscolo M, Dobson CM, Karplus M: **Determination of a transition state at atomic resolution from protein engineering data.** *Journal of Molecular Biology* 2002, **324:**151-163.
72. De Los Rios MA, Muralidhara BK, Wildes D, Sosnick TR, Marqusee S, Wittung-Stafshede P, Plaxco KW, Ruczinski I: **On the precision of experimentally determined protein folding rates and phi-values.** *Protein Science* 2006, **15:**553-563.
73. Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, Davidson AR, Chan HS: **Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105:**9999-10004.
74. Clementi C, Plotkin SS: **The effects of nonnative interactions on protein folding rates: Theory and simulation.** *Protein Science* 2004, **13:**1750-1766.
75. Krantz BA, Dothager RS, Sosnick TR: **Discerning the structure and energy of multiple transition states in protein folding using psi-analysis.** *Journal of Molecular Biology* 2004, **337:**463-475.
76. Alder BJ, Wainwright TE: **Phase Transition for a Hard Sphere System.** *Journal of Chemical Physics* 1957, **27:**1208-1209.
77. Karplus M, McCammon JA: **Molecular dynamics simulations of biomolecules.** *Nature Structural Biology* 2002, **9:**646-652.
78. Mccammon JA, Gelin BR, Karplus M: **Dynamics of Folded Proteins.** *Nature* 1977, **267:**585-590.

79. Frauenfelder H, Petsko GA, Tsernoglou D: **Temperature-Dependent X-Ray-Diffraction as a Probe of Protein Structural Dynamics.** *Nature* 1979, **280:**558-563.
80. Artymiuk PJ, Blake CCF, Oatley SJ: **X-Ray-Analysis of Protein Dynamics.** *Journal De Chimie Physique Et De Physico-Chimie Biologique* 1979, **76:**813-815.
81. Artymiuk PJ, Blake CCF, Grace DEP, Oatley SJ, Phillips DC, Sternberg MJE: **Crystallographic Studies of the Dynamic Properties of Lysozyme.** *Nature* 1979, **280:**563-568.
82. Brunger AT, Campbell RL, Clore GM, Gronenborn AM, Karplus M, Petsko GA, Teeter MM: **Solution of a Protein Crystal-Structure with a Model Obtained from Nmr Interproton Distance Restraints.** *Science* 1987, **235:**1049-1053.
83. Brunger AT, Kuriyan J, Karplus M: **Crystallographic R-Factor Refinement by Molecular-Dynamics.** *Science* 1987, **235:**458-460.
84. Nilsson L, Clore GM, Gronenborn AM, Brunger AT, Karplus M: **Structure Refinement of Oligonucleotides by Molecular-Dynamics with Nuclear Overhauser Effect Interproton Distance Restraints - Application to 5' D(C-G-T-a-C-G)2.** *Journal of Molecular Biology* 1986, **188:**455-475.
85. Verlet L: **Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules.** *Physical Review* 1967, **159:**98-&.
86. Weiner PK, Kollman PA: **Amber - Assisted Model-Building with Energy Refinement - a General Program for Modeling Molecules and Their Interactions.** *Journal of Computational Chemistry* 1981, **2:**287-303.
87. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M: **Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations.** *Journal of Computational Chemistry* 1983, **4:**187-217.
88. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF: **The GROMOS biomolecular simulation program package.** *Journal of Physical Chemistry A* 1999, **103:**3596-3607.
89. Wang JM, Cieplak P, Kollman PA: **How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?** *Journal of Computational Chemistry* 2000, **21:**1049-1074.
90. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong GM, Zhang W, Yang R, Cieplak P, Luo R, Lee T, et al: **A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.** *Journal of Computational Chemistry* 2003, **24:**1999-2012.
91. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al: **All-atom empirical potential for molecular modeling and dynamics studies of proteins.** *Journal of Physical Chemistry B* 1998, **102:**3586-3616.
92. Guvench O, MacKerell AD: **Automated conformational energy fitting for force-field development.** *Journal of Molecular Modeling* 2008, **14:**667-679.
93. Showalter SA, Bruschweiler-Li L, Johnson E, Zhang F, Bruschweiler R: **Quantitative lid dynamics of MDM2 reveals differential ligand binding**

modes of the p53-binding cleft. *Journal of the American Chemical Society* 2008, **130:**6472-6478.
94. Gundertofte K, Liljefors T, Norrby PO: **A comparison of conformational energies calculated by several molecular mechanics methods.** *Journal of Computational Chemistry* 1996, **17:**429-449.
95. Nicklaus MC: **Conformational energies calculated by the molecular mechanics program CHARMm.** *Journal of Computational Chemistry* 1997, **18:**1056-1060.
96. Price DJ, Brooks CL: **Modern protein force fields behave comparably in molecular dynamics simulations.** *Journal of Computational Chemistry* 2002, **23:**1045-1057.
97. Bunimovich LA, Sina*i IAGe: *Dynamical systems, ergodic theory, and applications.* 2nd, expanded and rev. edn. Berlin ; New York: Springer; 2000.
98. Billingsley P: *Ergodic theory and information.* Huntington, N.Y.: R. E. Krieger Pub. Co.; 1978.
99. Yang LJ, Shao Q, Gao YQ: **Comparison between integrated and parallel tempering methods in enhanced sampling simulations.** *Journal of Chemical Physics* 2009, **130:**-.
100. Huang XH, Bowman GR, Pande VS: **Convergence of folding free energy landscapes via application of enhanced sampling methods in a distributed computing environment.** *Journal of Chemical Physics* 2008, **128:**-.
101. Minary P, Martyna G, Tuckerman ME: **Enhanced sampling methods via nonlinear variable transformations: Model systems, model proteins and all-atom peptides.** *Abstracts of Papers of the American Chemical Society* 2004, **227:**U1030-U1030.
102. Andricioaei I, Dinner AR, Karplus M: **Self-guided enhanced sampling methods for thermodynamic averages.** *Journal of Chemical Physics* 2003, **118:**1074-1084.
103. Roitberg A, Elber R: **Modeling Side-Chains in Peptides and Proteins - Application of the Locally Enhanced Sampling and the Simulated Annealing Methods to Find Minimum Energy Conformations.** *Journal of Chemical Physics* 1991, **95:**9277-9287.
104. Feig M, Brooks CL: **Recent advances in the development and application of implicit solvent models in biomolecule simulations.** *Current Opinion in Structural Biology* 2004, **14:**217-224.
105. Im W, Feig M, Brooks CL: **An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins (vol 85, pg 2900, 2003).** *Biophysical Journal* 2004, **86:**3330-3330.
106. Feig M, Im W, Brooks CL: **Implicit solvation based on generalized Born theory in different dielectric environments.** *Journal of Chemical Physics* 2004, **120:**903-911.
107. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL: **Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures.** *Journal of Computational Chemistry* 2004, **25:**265-284.

108. Lazaridis T, Karplus M: **Effective energy function for proteins in solution.** *Proteins-Structure Function and Genetics* 1999, **35:**133-152.
109. Bashford D, Case DA: **Generalized born models of macromolecular solvation effects.** *Annual Review of Physical Chemistry* 2000, **51:**129-152.
110. Creighton TE: **Protein folding.** *Biochem J* 1990, **270:**1-16.
111. Dobson CM, Sali A, Karplus M: **Protein folding: A perspective from theory and experiment.** *Angewandte Chemie-International Edition* 1998, **37:**868-893.
112. Gruebele M: **Protein folding: the free energy surface.** *Current Opinion in Structural Biology* 2002, **12:**161-168.
113. Rhodes G: *Crystallography made crystal clear : a guide for users of macromolecular models.* 2nd. edn. San Diego: Academic Press; 2000.
114. Dobson CM, Hore PJ: **Kinetic studies of protein folding using NMR spectroscopy.** *Nature Structural Biology* 1998, **5:**504-507.
115. Balbach J, Forge V, Lau WS, vanNuland NAJ, Brew K, Dobson CM: **Protein folding monitored at individual residues during a two-dimensional NMR experiment.** *Science* 1996, **274:**1161-1163.
116. Schotte F, Lim MH, Jackson TA, Smirnov AV, Soman J, Olson JS, Phillips GN, Wulff M, Anfinrud PA: **Watching a protein as it functions with 150-ps time-resolved X-ray crystallography.** *Science* 2003, **300:**1944-1947.
117. Moffat K: **Time-resolved biochemical crystallography: A mechanistic perspective.** *Chemical Reviews* 2001, **101:**1569-1581.
118. Doniach S: **Changes in biomolecular conformation seen by small angle X-ray scattering.** *Chemical Reviews* 2001, **101:**1763-1778.
119. Pollack L, Tate MW, Darnton NC, Knight JB, Gruner SM, Eaton WA, Austin RH: **Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle x-ray scattering.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96:**10115-10117.
120. Uzawa T, Kimura T, Ishimori K, Morishima I, Matsui T, Ikeda-Saito M, Takahashi S, Akiyama S, Fujisawa T: **Time-resolved small-angle X-ray scattering investigation of the folding dynamics of heme oxygenase: Implication of the scaling relationship for the submillisecond intermediates of protein folding.** *Journal of Molecular Biology* 2006, **357:**997-1008.
121. Pfuhl M, Driscoll PC: **Protein nuclear magnetic resonance spectroscopy in the new millennium.** *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences* 2000, **358:**513-545.
122. Wüthrich K: *NMR of proteins and nucleic acids.* New York: Wiley; 1986.
123. Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J: **Fast kinetics and mechanisms in protein folding.** *Annual Review of Biophysics and Biomolecular Structure* 2000, **29:**327-359.
124. Gruebele M, Sabelko J, Ballew R, Ervin J: **Laser temperature jump induced protein refolding.** *Accounts of Chemical Research* 1998, **31:**699-707.
125. Bredenbeck J, Helbing J, Sieg A, Schrader T, Zinth W, Renner C, Behrendt R, Moroder L, Wachtveitl J, Hamm P: **Picosecond conformational transition and equilibration of a cyclic peptide.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100:**6452-6457.

126. Dyer RB, Gai F, Woodruff WH: **Infrared studies of fast events in protein folding.** *Accounts of Chemical Research* 1998, **31:**709-716.
127. Bredenbeck J, Helbing J, Behrendt R, Renner C, Moroder L, Wachtveitl J, Hamm P: **Transient 2D-IR spectroscopy: Snapshots of the nonequilibrium ensemble during the picosecond conformational transition of a small peptide.** *Journal of Physical Chemistry B* 2003, **107:**8654-8660.
128. Goldbeck RA, Thomas YG, Chen EF, Esquerra RM, Kliger DS: **Multiple pathways on a protein-folding energy landscape: Kinetic evidence.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96:**2782-2787.
129. Cho MH: **Coherent two-dimensional optical spectroscopy.** *Chemical Reviews* 2008, **108:**1331-1418.
130. Bredenbeck J, Helbing J, Kolano C, Hamm P: **Ultrafast 2D-IR Spectroscopy of transient species.** *Chemphyschem* 2007, **8:**1747-1756.
131. Hein M, Wegener AA, Engelhard M, Siebert F: **Time-resolved FTIR studies of sensory rhodopsin II (NpSRII) from Natronobacterium pharaonis: Implications for proton transport and receptor activation.** *Biophysical Journal* 2003, **84:**1208-1217.
132. Brudler R, Rammelsberg R, Woo TT, Getzoff ED, Gerwert K: **Structure of the I-1 early intermediate of photoactive yellow protein by FTIR spectroscopy.** *Nature Structural Biology* 2001, **8:**265-270.
133. Thompson HW: **The Study of Macromolecules by Infra-Red Spectroscopy.** *Journal of the Chemical Society* 1947**:**289-297.
134. Miyazawa T, Shimanouchi T, Mizushima SI: **Normal Vibrations of N-Methylacetamide.** *Journal of Chemical Physics* 1958, **29:**611-616.
135. Mukamel S: **Multidimensional femtosecond correlation spectroscopies of electronic and vibrational excitations.** *Annu Rev Phys Chem* 2000, **51:**691-729.
136. Tsuboi M, Onishi T, Nakagawa I, Shimanouchi T, Mizushima S: **Assignments of the Vibrational Frequencies of Glycine.** *Spectrochimica Acta* 1958, **12:**253-261.
137. Mantsch HH, Chapman D: *Infrared spectroscopy of biomolecules.* New York: Wiley-Liss; 1996.
138. Kwak KW, Park S, Fayer MD: **Dynamics around solutes and solute-solvent complexes in mixed solvents.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104:**14221-14226.
139. Krimm S, Bandekar J: **Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins.** *Advances in Protein Chemistry* 1986, **38:**181-364.
140. Mantsch HH C (Ed.). ***Infrared Spectroscopy of Biomolecules***. New York: Wiley-Liss; 1996.
141. Zanni MT, Hochstrasser RM: **Two-dimensional infrared spectroscopy: a promising new method for the time resolution of structures.** *Current Opinion in Structural Biology* 2001, **11:**516-522.
142. De Alba E, Santoro J, Rico M, Jimenez MA: **De novo design of a monomeric three-stranded antiparallel beta-sheet.** *Protein Science* 1999, **8:**854-865.

143. Khalil M, Demirdoven N, Tokmakoff A: **Coherent 2D IR spectroscopy: Molecular structure and dynamics in solution.** *Journal of Physical Chemistry A* 2003, **107:**5258-5279.
144. Woutersen S, Pfister R, Hamm P, Mu YG, Kosov DS, Stock G: **Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations.** *Journal of Chemical Physics* 2002, **117:**6833-6840.
145. Woutersen S, Hamm P: **Nonlinear two-dimensional vibrational spectroscopy of peptides.** *Journal of Physics-Condensed Matter* 2002, **14:**R1035-R1062.
146. Hamm P, Lim M, DeGrado WF, Hochstrasser RM: **The two-dimensional IR nonlinear spectroscopy of a cyclic penta-peptide in relation to its three-dimensional structure.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96:**2036-2041.
147. Hamm P, Lim MH, Hochstrasser RM: **Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy.** *Journal of Physical Chemistry B* 1998, **102:**6123-6138.
148. Asplund MC, Zanni MT, Hochstrasser RM: **Two-dimensional infrared spectroscopy of peptides by phase-controlled femtosecond vibrational photon echoes.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97:**8219-8224.
149. Demirdoven N, Cheatum CM, Chung HS, Khalil M, Knoester J, Tokmakoff A: **Two-dimensional infrared spectroscopy of antiparallel beta-sheet secondary structure.** *Journal of the American Chemical Society* 2004, **126:**7981-7990.
150. Cheatum CM, Tokmakoff A, Knoester J: **Signatures of beta-sheet secondary structures in linear and two-dimensional infrared spectroscopy.** *Journal of Chemical Physics* 2004, **120:**8201-8215.
151. Zhuang W, Abramavicius D, Hayashi T, Mukamel S: **Simulation protocols for coherent femtosecond vibrational spectra of peptides.** *Journal of Physical Chemistry B* 2006, **110:**3362-3374.
152. Piryatinski A, Tretiak S, Chernyak V, Mukamel S: **Simulations of two-dimensional femtosecond infrared photon echoes of glycine dipeptide.** *Journal of Raman Spectroscopy* 2000, **31:**125-135.
153. Chernyak V, Zhang WM, Mukamel S: **Multidimensional femtosecond spectroscopies of molecular aggregates and semiconductor nanostructures: The nonlinear exciton equations.** *Journal of Chemical Physics* 1998, **109:**9587-9601.
154. Mukamel S, Piryatinski A, Chernyak V: **Semiclassical simulations of multidimensional Raman echoes.** *Journal of Chemical Physics* 1999, **110:**1711-1725.
155. Piryatinski A, Chernyak V, Mukamel S: **Vibrational-exciton relaxation probed by three-pulse echoes in polypeptides.** *Chemical Physics* 2001, **266:**285-294.
156. Piryatinski A, Chernyak V, Mukamel S: **Two-dimensional correlation spectroscopies of localized vibrations.** *Chemical Physics* 2001, **266:**311-322.
157. Piryatinski A, Skinner JL: **Determining vibrational solvation-correlation functions from three-pulse infrared photon echoes.** *Journal of Physical Chemistry B* 2002, **106:**8055-8063.

158. Zhuang W, Hayashi T, Mukamel S: **Coherent Multidimensional Vibrational Spectroscopy of Biomolecules: Concepts, Simulations, and Challenges.** *Angewandte Chemie-International Edition* 2009, **48:**3750-3781.
159. Abramavicius D, Mukamel S: **Coherent third-order spectroscopic probes of molecular chirality.** *Journal of Chemical Physics* 2005, **122:**-.
160. Mukamel S: *Principles of nonlinear optical spectroscopy.* New York: Oxford University Press; 1995.
161. Schweigert IV, Mukamel S: **Double-quantum-coherence attosecond x-ray spectroscopy of spatially separated, spectrally overlapping core-electron transitions.** *Physical Review A* 2008, **78:**-.
162. Fersht A: *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding.* New York: W.H. Freeman; 1999.
163. Daggett V, Li A, Fersht AR: **Combined Molecular Dynamics and Φ-Value Analysis of Structure-Reactivity Relationships in the Transition State and Unfolding Pathway of Barnase: Structural Basis of Hammond and Anti-Hammond Effects.** In *Book Combined Molecular Dynamics and Φ-Value Analysis of Structure-Reactivity Relationships in the Transition State and Unfolding Pathway of Barnase: Structural Basis of Hammond and Anti-Hammond Effects* (Editor ed.^eds.), vol. 120. pp. 12740-12754. City; 1998:12740-12754.
164. Fersht AR, Matouschek A, Sancho J, Serrano L, Vuilleumier S: **Pathway of protein folding.** *Faraday discussions* 1992**:**183-193.
165. Daggett V: **Validation of protein-unfolding transition states identified in molecular dynamics simulations.** *Biochemical Society symposium* 2001**:**83-93.
166. Daggett V: **Molecular dynamics simulations of the protein unfolding/folding reaction.** *Accounts of chemical research* 2002, **35:**422-429.
167. Settanni G, Gsponer J, Caflisch A: **Formation of the folding nucleus of an SH3 domain investigated by loosely coupled molecular dynamics simulations.** *Biophysical journal* 2004, **86:**1691-1701.
168. Settanni G, Rao F, Caflisch A: **Phi-value analysis by molecular dynamics simulations of reversible folding.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**628-633.
169. Jemth P, Day R, Gianni S, Khan F, Allen M, Daggett V, Fersht AR: **The structure of the major transition state for folding of an FF domain from experiment and simulation.** *Journal of molecular biology* 2005, **350:**363-378.
170. Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR: **Structure of the transition state for folding of a protein derived from experiment and simulation.** *Journal of molecular biology* 1996, **257:**430-440.
171. Fersht AR: **Characterizing transition states in protein folding: an essential step in the puzzle.** *Curr Opin Struct Biol* 1995, **5:**79-84.
172. Kuhlman B, Yang HY, Boice JA, Fairman R, Raleigh DP: **An exceptionally stable helix from the ribosomal protein L9: implications for protein folding and stability.** *Journal of molecular biology* 1997, **270:**640-647.
173. Lillemoen J, Cameron CS, Hoffman DW: **The stability and dynamics of ribosomal protein L9: investigations of a molecular strut by amide proton**

exchange and circular dichroism. *Journal of molecular biology* 1997, **268:**482-493.
174. Hoffman DW, Davies C, Gerchman SE, Kycia JH, Porter SJ, White SW, Ramakrishnan V: **Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein.** *The EMBO journal* 1994, **13:**205-212.
175. Hoffman DW, Cameron CS, Davies C, White SW, Ramakrishnan V: **Ribosomal protein L9: a structure determination by the combined use of X-ray crystallography and NMR spectroscopy.** *Journal of molecular biology* 1996, **264:**1058-1071.
176. Horng JC, Moroz V, Raleigh DP: **Rapid cooperative two-state folding of a miniature alpha-beta protein and design of a thermostable variant.** *Journal of molecular biology* 2003, **326:**1261-1270.
177. Kuhlman B, Luisi DL, Evans PA, Raleigh DP: **Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9.** *Journal of molecular biology* 1998, **284:**1661-1670.
178. Kuhlman B, Boice JA, Fairman R, Raleigh DP: **Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding.** *Biochemistry* 1998, **37:**1025-1032.
179. Aprilakis KN, Taskent H, Raleigh DP: **Use of the novel fluorescent amino acid p-cyanophenylalanine offers a direct probe of hydrophobic core formation during the folding of the N-terminal domain of the ribosomal protein L9 and provides evidence for two-state folding.** *Biochemistry* 2007, **46:**12308-12313.
180. Cho JH, Sato, S., Kim, E.Y., Schindelin, H., Raleigh, D.P.: **Highly cooperative Interactions in the Denatured State of a Globular Protein.** In *Book Highly cooperative Interactions in the Denatured State of a Globular Protein* (Editor ed.^eds.). City; 2008.
181. Cho JH, Raleigh DP: **Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins.** *Journal of molecular biology* 2005, **353:**174-185.
182. Cho JH, Raleigh DP: **Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure.** *Journal of molecular biology* 2006, **359:**1437-1446.
183. Cho JH, Sato S, Raleigh DP: **Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state.** *Journal of molecular biology* 2004, **338:**827-837.
184. Anil B, Li Y, Cho JH, Raleigh DP: **The unfolded state of NTL9 is compact in the absence of denaturant.** *Biochemistry* 2006, **45:**10110-10116.
185. Horng JC, Moroz V, Rigotti DJ, Fairman R, Raleigh DP: **Characterization of large peptide fragments derived from the N-terminal domain of the ribosomal protein L9: definition of the minimum folding motif and characterization of local electrostatic interactions.** *Biochemistry* 2002, **41:**13360-13369.

186. Luisi DL, Kuhlman B, Sideras K, Evans PA, Raleigh DP: **Effects of varying the local propensity to form secondary structure on the stability and folding kinetics of a rapid folding mixed alpha/beta protein: characterization of a truncation mutant of the N-terminal domain of the ribosomal protein L9.** *Journal of molecular biology* 1999, **289:**167-174.
187. Evans SP, Bycroft M: **NMR structure of the N-terminal domain of Saccharomyces cerevisiae RNase HI reveals a fold with a strong resemblance to the N-terminal domain of ribosomal protein L9.** *Journal of molecular biology* 1999, **291:**661-669.
188. Luisi DL, Snow CD, Lin JJ, Hendsch ZS, Tidor B, Raleigh DP: **Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein l9.** *Biochemistry* 2003, **42:**7050-7060.
189. Luisi DL, Raleigh DP: **pH-dependent interactions and the stability and folding kinetics of the N-terminal domain of L9. Electrostatic interactions are only weakly formed in the transition state for folding.** *Journal of molecular biology* 2000, **299:**1091-1100.
190. Sato S, Luisi DL, Raleigh DP: **pH jump studies of the folding of the multidomain ribosomal protein L9: the structural organization of the N-terminal domain does not affect the anomalously slow folding of the C-terminal domain.** *Biochemistry* 2000, **39:**4955-4962.
191. Kuhlman B, Luisi DL, Young P, Raleigh DP: **pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions.** *Biochemistry* 1999, **38:**4896-4903.
192. Luisi DL, Wu WJ, Raleigh DP: **Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: evidence for stable native-like secondary structure in the unfolded state.** *Journal of molecular biology* 1999, **287:**395-407.
193. Sato S, Raleigh DP: **Kinetic isotope effects reveal the presence of significant secondary structure in the transition state for the folding of the N-terminal domain of L9.** *Journal of molecular biology* 2007, **370:**349-355.
194. Sato S, Raleigh DP: **pH-dependent stability and folding kinetics of a protein with an unusual alpha-beta topology: the C-terminal domain of the ribosomal protein L9.** *Journal of molecular biology* 2002, **318:**571-582.
195. Sato S, Xiang S, Raleigh DP: **On the relationship between protein stability and folding kinetics: a comparative study of the N-terminal domains of RNase HI, E. coli and Bacillus stearothermophilus L9.** *Journal of molecular biology* 2001, **312:**569-577.
196. Sato S, Kuhlman B, Wu WJ, Raleigh DP: **Folding of the multidomain ribosomal protein L9: the two domains fold independently with remarkably different rates.** *Biochemistry* 1999, **38:**5643-5650.
197. Kuhlman B, Raleigh DP: **Global analysis of the thermal and chemical denaturation of the N-terminal domain of the ribosomal protein L9 in H2O and D2O. Determination of the thermodynamic parameters, deltaH(o),**

deltaS(o), and deltaC(o)p and evaluation of solvent isotope effects. *Protein Sci* 1998, **7:**2405-2412.
198. Vugmeyster L, Kuhlman B, Raleigh DP: **Amide proton exchange measurements as a probe of the stability and dynamics of the N-terminal domain of the ribosomal protein L9: comparison with the intact protein.** *Protein Sci* 1998, **7:**1994-1997.
199. Gsponer J, Caflisch A: **Role of native topology investigated by multiple unfolding simulations of four SH3 domains.** *Journal of molecular biology* 2001, **309:**285-298.
200. Gsponer J, Caflisch A: **Molecular dynamics simulations of protein folding from the transition state.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99:**6719-6724.
201. Geierhaas CD, Paci E, Vendruscolo M, Clarke J: **Comparison of the transition states for folding of two Ig-like proteins from different superfamilies.** *Journal of molecular biology* 2004, **343:**1111-1123.
202. Interlandi G, Settanni G, Caflisch A: **Unfolding transition state and intermediates of the tumor suppressor p16INK4a investigated by molecular dynamics simulations.** *Proteins* 2006, **64:**178-192.
203. Myers JK, Pace CN, Scholtz JM: **Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding.** *Protein Sci* 1995, **4:**2138-2148.
204. Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich ES: **On the transition coordinate for protein folding.** *Journal of Chemical Physics* 1998, **108:**334-350.
205. Bryant Z, Pande VS, Rokhsar DS: **Mechanical unfolding of a beta-hairpin using molecular dynamics.** *Biophysical journal* 2000, **78:**584-589.
206. Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, Merz KM, Jr., Onufriev A, Simmerling C, Wang B, Woods RJ: **The Amber biomolecular simulation programs.** *Journal of computational chemistry* 2005, **26:**1668-1688.
207. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, et al: **A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.** *Journal of computational chemistry* 2003, **24:**1999-2012.
208. Onufriev A, Bashford D, Case DA: **Exploring protein native states and large-scale conformational changes with a modified generalized born model.** *Proteins* 2004, **55:**383-394.
209. Feig M, Karanicolas J, Brooks CL: **MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology.** *Journal of Molecular Graphics & Modelling* 2004, **22:**377-395.
210. Tanford C: **Protein denaturation. C. Theoretical models for the mechanism of denaturation.** *Advances in protein chemistry* 1970, **24:**1-95.
211. Weiser J, Weiser AA, Shenkin PS, Still WC: **Neighbor-list reduction: Optimization for computation of molecular van der Waals and solvent-accessible surface areas.** *Journal of computational chemistry* 1998, **19:**797-808.
212. Goldenberg DP, Creighton TE: **Energetics of protein structure and folding.** *Biopolymers* 1985, **24:**167-182.

213. Taskent H, Cho JH, Raleigh DP: **Temperature-dependent Hammond behavior in a protein-folding reaction: Analysis of transition-state movement and ground-state effects.** *Journal of Molecular Biology* 2008, **378:**699-706.
214. Kabsch W, Sander C: **Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.** *Biopolymers* 1983, **22:**2577-2637.
215. Geney R, Sun L, Pera P, Bernacki RJ, Xia S, Horwitz SB, Simmerling CL, Ojima I: **Use of the tubulin bound paclitaxel conformation for structure-based rational drug design.** *Chemistry & biology* 2005, **12:**339-348.
216. Cho JH, Raleigh DP: **Not Yet Published Phi-Values.** In *Book Not Yet Published Phi-Values* (Editor ed.^eds.). City; 2008.
217. Frauenfelder H, Sligar SG, Wolynes PG: **The energy landscapes and motions of proteins.** *Science* 1991, **254:**1598-1603.
218. Gnanakaran S, Garcia AE: **Validation of an all-atom protein force field: From dipeptides to larger peptides.** *Journal of Physical Chemistry B* 2003, **107:**12555-12557.
219. Chung HS, Ganim Z, Jones KC, Tokmakoff A: **Transient 2D IR spectroscopy of ubiquitin unfolding dynamics.** *Proc Natl Acad Sci U S A* 2007, **104:**14237-14242.
220. Torii H, Tasumi M: **Model calculations on the amide-I infrared bands of globular proteins.** *J Chem Phys* 1992, **96:**3379-3387.
221. Hahn S, Ham S, Cho M: **Simulation studies of amide IIR absorption and two-dimensional IR spectra of beta hairpins in liquid water.** *Journal of Physical Chemistry B* 2005, **109:**11789-11801.
222. Jansen TL, Knoester J: **Nonadiabatic effects in the two-dimensional infrared spectra of peptides: Application to alanine dipeptide.** *Journal of Physical Chemistry B* 2006, **110:**22910-22916.
223. Finkelstein IJ, Zheng J, Ishikawa H, Kim S, Kwak K, Fayer MD: **Probing dynamics of complex molecular systems with ultrafast 2D IR vibrational echo spectroscopy.** *Physical Chemistry Chemical Physics* 2007, **9:**1533-1549.
224. Blanco F, Ramirez-Alvarado M, Serrano L: **Formation and stability of beta-hairpin structures in polypeptides.** *Current Opinion in Structural Biology* 1998, **8:**107-111.
225. Gellman SH: **Minimal model systems for beta sheet secondary structure in proteins.** *Current Opinion in Chemical Biology* 1998, **2:**717-725.
226. Cavalli A, Haberthur U, Paci E, Caflisch A: **Fast protein folding on downhill energy landscape.** *Protein science : a publication of the Protein Society* 2003, **12:**1801-1803.
227. Krivov SV, Muff S, Caflisch A, Karplus M: **One-dimensional barrier-preserving free-energy projections of a beta-sheet miniprotein: new insights into the folding process.** *J Phys Chem B* 2008, **112:**8701-8714.
228. Ferrara P, Caflisch A: **Folding simulations of a three-stranded antiparallel beta -sheet peptide.** *Proc Natl Acad Sci U S A* 2000, **97:**10780-10785.
229. Cavalli A, Ferrara P, Caflisch A: **Weak temperature dependence of the free energy surface and folding pathways of structured peptides.** *Proteins* 2002, **47:**305-314.

230. Muff S, Caflisch A: **Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a beta-sheet miniprotein.** *Proteins* 2008, **70:**1185-1195.
231. Carr JM, Wales DJ: **Folding pathways and rates for the three-stranded beta-sheet peptide Beta3s using discrete path sampling.** *J Phys Chem B* 2008, **112:**8760-8769.
232. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al: **CHARMM: the biomolecular simulation program.** *J Comput Chem* 2009, **30:**1545-1614.
233. Ferrara P, Apostolakis J, Caflisch A: **Evaluation of a fast implicit solvent model for molecular dynamics simulations.** *Proteins* 2002, **46:**24-33.
234. Kabsch W, Sander C: **Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features.** *Biopolymers* 1983, **22:**2577-2637.
235. Scheurer C, Piryatinski A, Mukamel S: **Signatures of beta-peptide unfolding in two-dimensional vibrational echo spectroscopy: a simulation study.** *J Am Chem Soc* 2001, **123:**3114-3124.
236. Ham S, Cho M: **Amide I modes in the N-methylacetamide dimer and glycine dipeptide analog: Diagonal force constants.** *J Chem Phys* 2003, **118:**6915-6922.
237. Zheng J, Kwak K, Fayer MD: **Ultrafast two-dimensional IR vibrational echo spectroscopy.** *Accounts of Chemical Research* 2007, **40:**75-83.
238. Zhuang W, Abramavicius D, Voronine DV, Mukamel S: **Simulation of two-dimensional infrared spectroscopy of amyloid fibrils.** *Proc Natl Acad Sci U S A* 2007, **104:**14233-14236.
239. Abramavicius D, Zhuang W, Mukamel S: **Peptide secondary structure determination by three-pulse coherent vibrational spectroscopies: A simulation study.** *Journal of Physical Chemistry B* 2004, **108:**18034-18045.
240. Lesar R, Kopelman R: **Vibrational Excitons, Resonant Energy-Transfer, and Local Structure in Liquid Benzene.** *Journal of Chemical Physics* 1977, **66:**5035-5041.
241. Zhang WM, Chernyak V, Mukamel S: **Multidimensional femtosecond correlation spectroscopies of electronic and vibrational excitons.** *J Chem Phys* 1999, **110:**5011-5028.
242. Bernstei.Er, Colson SD, Kopelman R, Robinson GW: **Electronic and Vibrational Exciton Structure in Crystalline Benzene.** *Journal of Chemical Physics* 1968, **48:**5596-&.
243. Lesar R, Kopelman R: **Vibrational Exciton Cluster States, Percolation and Pairwise Interactions - Benzene A2-Mu Band.** *Chemical Physics* 1978, **29:**289-297.
244. Torii H, Tasumi M: **Application of the three-dimensional doorway-state theory to analyses of the amide-I infrared bands of globular proteins.** *J Chem Phys* 1992, **97:**92-98.
245. Torii H, Tasumi M: **Ab initio molecular orbital study of the amide I vibrational interactions between the peptide groups in di- and tripeptides**

and considerations on the conformation of the extended helix. *Journal of Raman Spectroscopy* 1998, **29:**81-86.
246. Moran AM, Dreyer J, Mukamel S: **Ab initio simulation of the two-dimensional vibrational spectrum of dicarbonylacetylacetonato rhodium(I).** *Journal of Chemical Physics* 2003, **118:**1347-1355.
247. Zhuang W, Abramavicius D, Mukamel S: **Dissecting coherent vibrational spectra of small proteins into secondary structural elements by sensitivity analysis.** *Proc Natl Acad Sci U S A* 2005, **102:**7443-7448.
248. Cochran AG, Skelton NJ, Starovasnik MA: **Tryptophan zippers: Stable, monomeric beta-hairpins.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98:**5578-5583.
249. Ganim Z, Chung HS, Smith AW, Deflores LP, Jones KC, Tokmakoff A: **Amide I two-dimensional infrared spectroscopy of proteins.** *Acc Chem Res* 2008, **41:**432-441.
250. Zhuang W, Abramavicius D, Mukamel S: **Dissecting coherent vibrational spectra of small proteins into secondary structural elements by sensitivity analysis.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**7443-7448.
251. Fang C, Hochstrasser RM: **Two-Dimensional Infrared Spectra of the 13C:18O Isotopomers of Alanine Residues in an alpha -Helix.** *J Phys Chem B* 2005, **109:**18652-18663.
252. Chung HS, Ganim Z, Tokmakoff A: **2-D IR Spectroscopy of ubiquitin unfolding dynamics.** *Abstracts of Papers, 233rd ACS National Meeting, Chicago, IL, United States, March 25-29, 2007* 2007**:**PHYS-567.
253. Shim SH, Gupta R, Ling YL, Strasfeld DB, Raleigh DP, Zanni MT: **Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106:**6614-6619.