SIMULATING TEMPERATURE JUMPS FOR PROTEIN FOLDING STUDIES

By

SEONAH KIM

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2007

To my loving husband Jiho and my family

ACKNOWLEDGMENTS

At the completion of this work I take great pleasure in acknowledging the people who have supported me over the last couple of years. I gratefully thank and acknowledge my advisor, Prof. Adrian Roitberg, for the opportunity to work with him. His guidance, humor, understanding, and encouragement leaded me to overcome all the obstacles I had felt in studying. My personal goal would not have been realized without the unconditional support from Prof. Roitberg. Those words are not enough to thank my advisor Prof. Roitberg.

I would like to thank for kind advice and suggestions of my committee members, Prof. Alexander Angerhofer, and Prof. Gail Fanucci as well, especially Prof. Ronald Castellano who accepted to be my committee at the last moment. I am also greatly obligated to Prof. Steve Hagen to give helpful discussions and suggestions through all my research.

I am grateful to Prof. Carlos Simmerling, Prof. Jeff Krause and Prof. Valeria Kleiman for kindly providing invaluable advice within and beyond science, and to Prof. Thomas Cheatham for hosting while I was in Utah. I gratefully acknowledge Georgios Leonis' and Julio Palma's help with their kind, helpful advice, supporting in and out of science through all my Ph. D. study here in Gainesville. I also thank to all of my Quantum Theory Project (QTP) buddies (Josh, Andrew, Tom, Kelly, Joey, Lena and Martin) who share lots of party. I am so lucky to have all of my buddies here in qtp and enjoyed every moment. I will miss them all. It has been great pleasure to work with my group members (Dr. Gustavo Seabra, Ozlem Demir, Christina Crecca, Hui Xiong, Dan Sindhikara, Yilin MEng, Lena Dolghih, and Mehrnoosh Arrar).

Above all I thank my husband, Jiho, for his love, support, encouragement, endless corrections and personal sacrifice without which I would not have been able to accomplish this work. Finally, I am forever indebted to my parents and brother, Jihwan, for their love, understanding, prayer and encouragement when it was most required.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE:	Acetyl beginning group
CD:	Circular Dichroism
DSSP:	Definition of the Secondary Structure of Proteins
GB:	Generalized Born
HSQC:	Hetero-nuclear Single Quantum Coherence
IR:	Infrared
MD:	Molecular Dynamics
NME:	N-methylamine ending group
NMR:	Nuclear Magnetic Resonance
NOEs:	Nuclear Overhauser Effects
REMD:	Replica Exchange Molecular Dynamics
T-jump:	Temperature-jump

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

SIMULATING TEMPERATURE JUMPS FOR PROTEIN FOLDING STUDIES

By

Seonah Kim

December 2007

Chair: Adrian E. Roitberg
Major: Chemistry

Protein folding is described as a dynamic process of an ensemble of molecules reaching well-defined three dimensional structures to achieve biological activity from linear amino acids sequences. Many human diseases result from protein misfolding or aggregation. Enormous effort has been made both experimentally and theoretically for nearly 40 years to explain the basic principle and mechanism of protein folding and unfolding. Nonetheless, many of them are still unknown or incompletely understood, mainly due to the complexity of the systems and the fast folding time scale. Experimental and theoretical approaches are complementary with each other for the protein folding studies and hence, combination of the two is required to have better understanding.

One of the most popular experimental methods for the protein folding studies is laser-induced temperature-jump (T-jump), because it has nanosecond resolution. In the first project, the T-jump on the polyalanine peptides (Ala_{20}) was simulated as a proof-of-principle system to mimic the experimental measurements. Replica exchange molecular dynamics (REMD) were performed to obtain equilibrated ensembles as a proper conformational sampling, which was combined with multiplexed molecular dynamics to extract kinetic properties in line with experiments.

In the second project, the same methodology used in the first project was applied to real proteins. Effect of frictional coefficient in the solvent model was approximated using Langevin dynamics. Computationall results on the two related 14-residue peptides were chosen and compared with experimental results. A ratio of relaxation time of the two peptides was determined by calculated Circular Dichroism (CD) spectra by a factor of ~1.2, while the experimental results were ~1.1.

CHAPTER 1
INTRODUCTION

1.1 Prologue

Proteins and nucleic acid are a starting point of life science, as they have a role in all living processes. Protein folding studies are of great interest especially for many human diseases associated with protein misfolding, such as cystic fibrosis, Alzheimer's, Parkinson's disease and Mad cow disease.[1-5]

Proteins are built by various combinations of commonly twenty amino acids. All amino acids contain an amino group (NH_2), a carboxyl group (COOH), and a distinctive R group connected to a central carbon atom (C^{α}) (Figure 1-1).

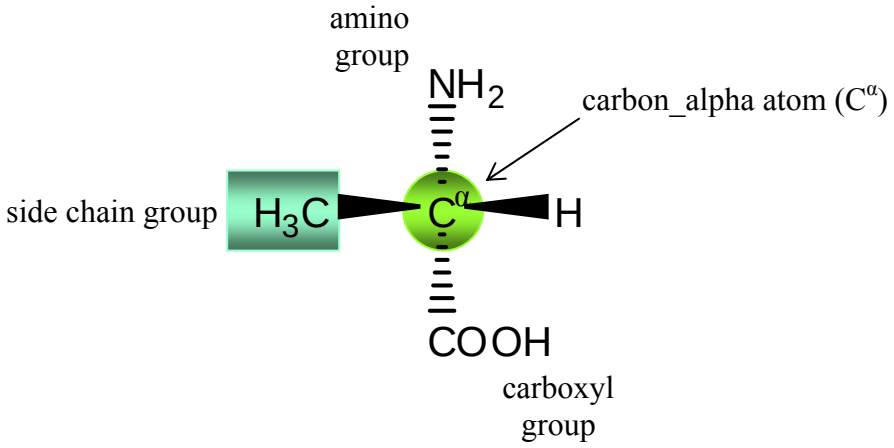


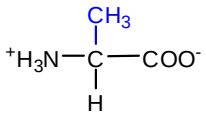
Figure 1-1. The general formula of an amino acid is showing a central carbon atom (C^{α}) is attached to an amino group (NH_2), a carbonyl group (COOH), a hydrogen atom, and a side chain (R).

The twenty amino acids found in proteins are shown in Figure 1-2. In general, four different groups are connected to the central C^{α} atom, making it a chiral center, except glycine, where two H atoms link to C^{α}. Chiral molecules can have optical isomers, L- and D- forms (Figure 1-3). Most of the amino acids in nature exist in L-form. Circular dichroism (CD) spectroscopy, discussed in section 1.4,[6] is a useful and critical tool to study chiral interactions.

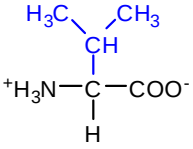
A protein is made of combination of amino acids joined via a peptide bond, where the

Nonpolar amino acids

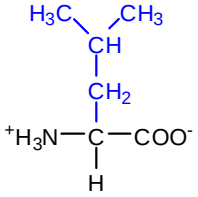
Ala, Alanine (A)



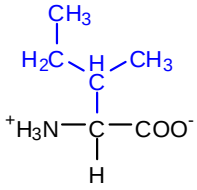
Val, Valine (V)



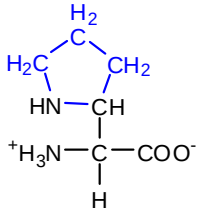
Leu, Leucine (L)



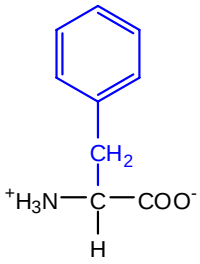
Ile, Isoleucine (I)



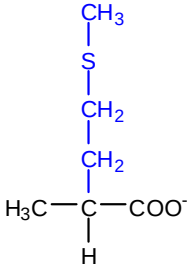
Pro, Proline (P)



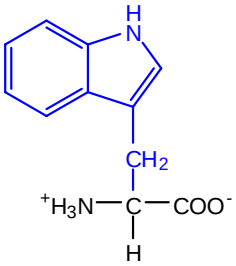
Phe, Phenylalanine (F)



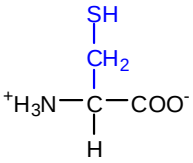
Met, Methionine (M)



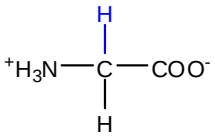
Trp, Tryptophan (W)



Cys, Cysteine (C)

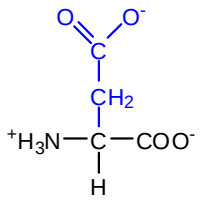


Gly, Glycine (G)

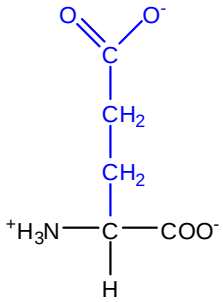


Charged polar amino acids

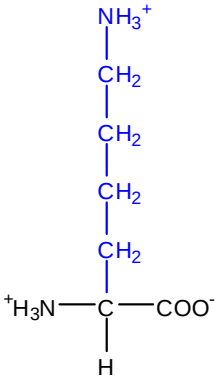
Asp, Aspartic acid (D)



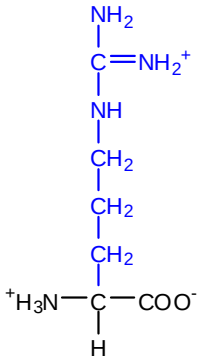
Glu, Glutamic acid (E)



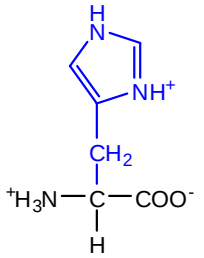
Lys, Lysine (K)



Arg, Arginine (R)

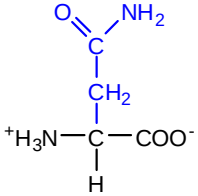


His, Histidine (H)

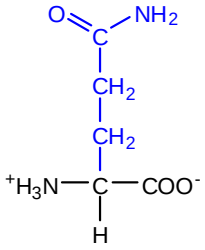


Uncharged polar amino acids

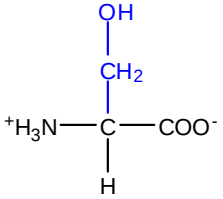
Asn, Asparagine (N)



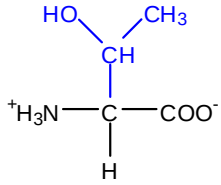
Gln, Glutamine (Q)



Ser, Serine (S)



Thr, Threonine (T)



Tyr, Tyrosine (Y)

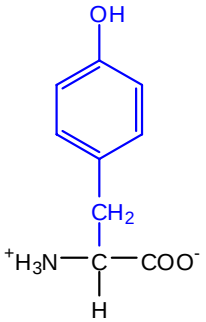
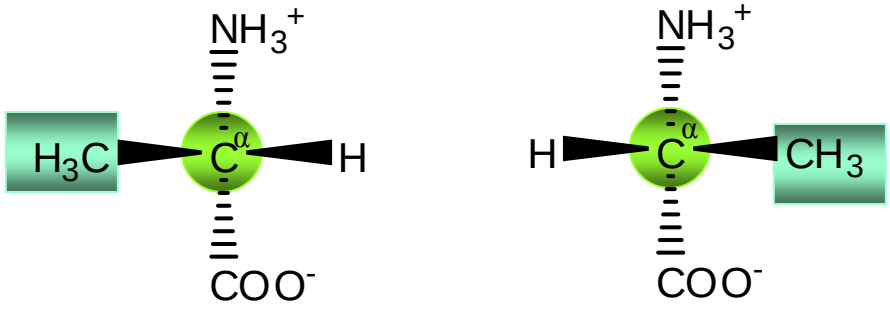


Figure 1-2. The twenty different amino acids found in proteins. Side chains are shown in blue.



L-alanine D-alanine

Figure 1-3. The optical isomers of alanine, L and D forms.

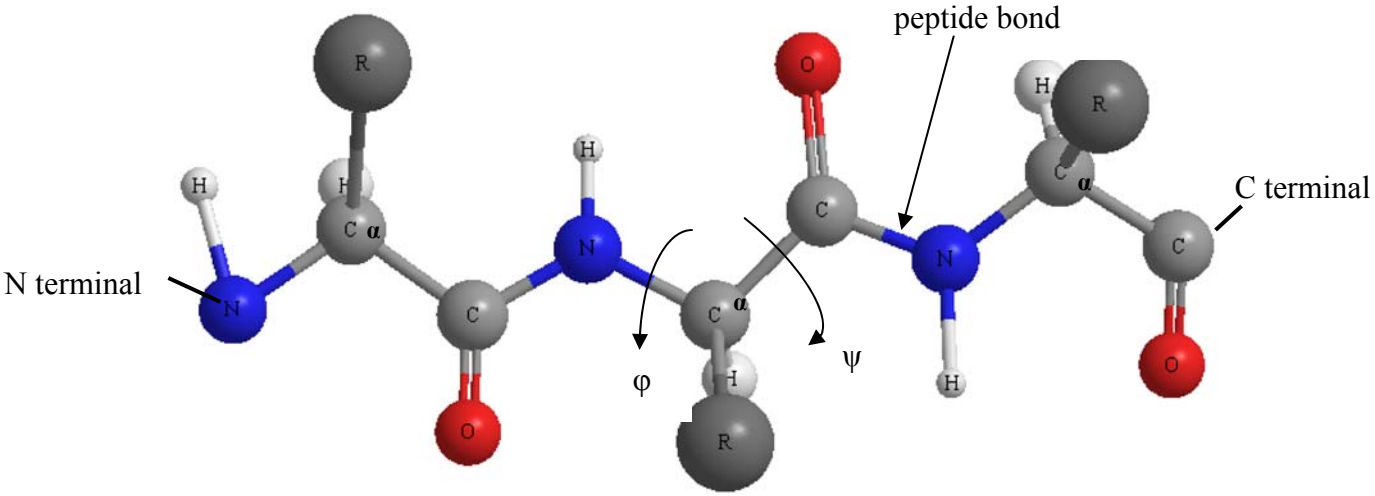


Figure 1-4. Part of the polypeptide chain shows to illustrate rigid peptide bond between C′ (from carbonyl group) and N (from amino group), two degrees of freedom, φ and ψ angels from rotations around N–C^{α} and C^{α}–C′ bonds, respectively.

carboxyl carbon atom (C-terminus) of one amino acid is bonded to the nitrogen atom (N-terminus) in the amino group of the next amino acid (Figure 1-4). The connection of multiple peptide bonds generates a backbone (or main chain) of a protein. The peptide group is generally rigid enough to remain planar and in *trans* configuration. The amino acid sequence which is linked by a peptide bond is called a primary structure.

Backbone rotations can occur around either the N–C^{α} or the C^{α}–C bonds; defined as phi (φ) and psi (ψ) angles, respectively. Different sets of values for phi and psi angles can denote different protein (or peptide) conformation. Thus, the backbone conformations can be described

by specifying these two angles. In Figure1-5, the pair of angles (φ and ψ) are plotted in a Ramachandran plot.[7] As shown in the figure, three main regions are allowed with respect to φ and ψ angles; one is the right-handed α-helix around φ=–57°, ψ=–47° (denoted α_R), another is the β sheet (parallel and antiparallel) around φ=–125°, ψ=+125° (denoted β) and the other is the left-handed α-helix around φ=+57°, ψ=+47° (denoted α_L). A left-handed polyproline II helix (PPII) which is often observed in proline-rich sequences and sterically forced conformation for polyproline, is also found in proteins around φ=–75°, ψ=+145° (denoted PPII) and is an important conformation in protein-protein interfaces.[8,9]

The three-dimensional form of local geometric arrangements within the peptide backbone is called secondary structure. For example, α-helices in proteins appear when the φ and ψ angles are approximately –57° and –47°, respectively and when hydrogen bond are formed between the carbonyl oxygen of the ith residue and the amido portion of the i + 4th residue.

The tertiary structure is the combination of secondary structural units joined by a loop (or turn). Furthermore, the quaternary structure can be defined when a protein involves more than one polypeptide chain. The diagram of hierarchy of protein structures Figure 1-6 shows this relationship.

1.2 Structural Biology

The subject of structural biology in proteins is the link between one-dimensional amino acid sequences to their three-dimensional structure. Structural biology is the study of molecular shape of biological macromolecules–proteins and nucleic acids– and their interactions. Studies of structural biology began with the exploration of biological materials by using early microscopy over 100 years ago. Thanks to development of new and powerful methods, such as X-ray crystallography and nuclear magnetic resonance (NMR), it became one of the most important

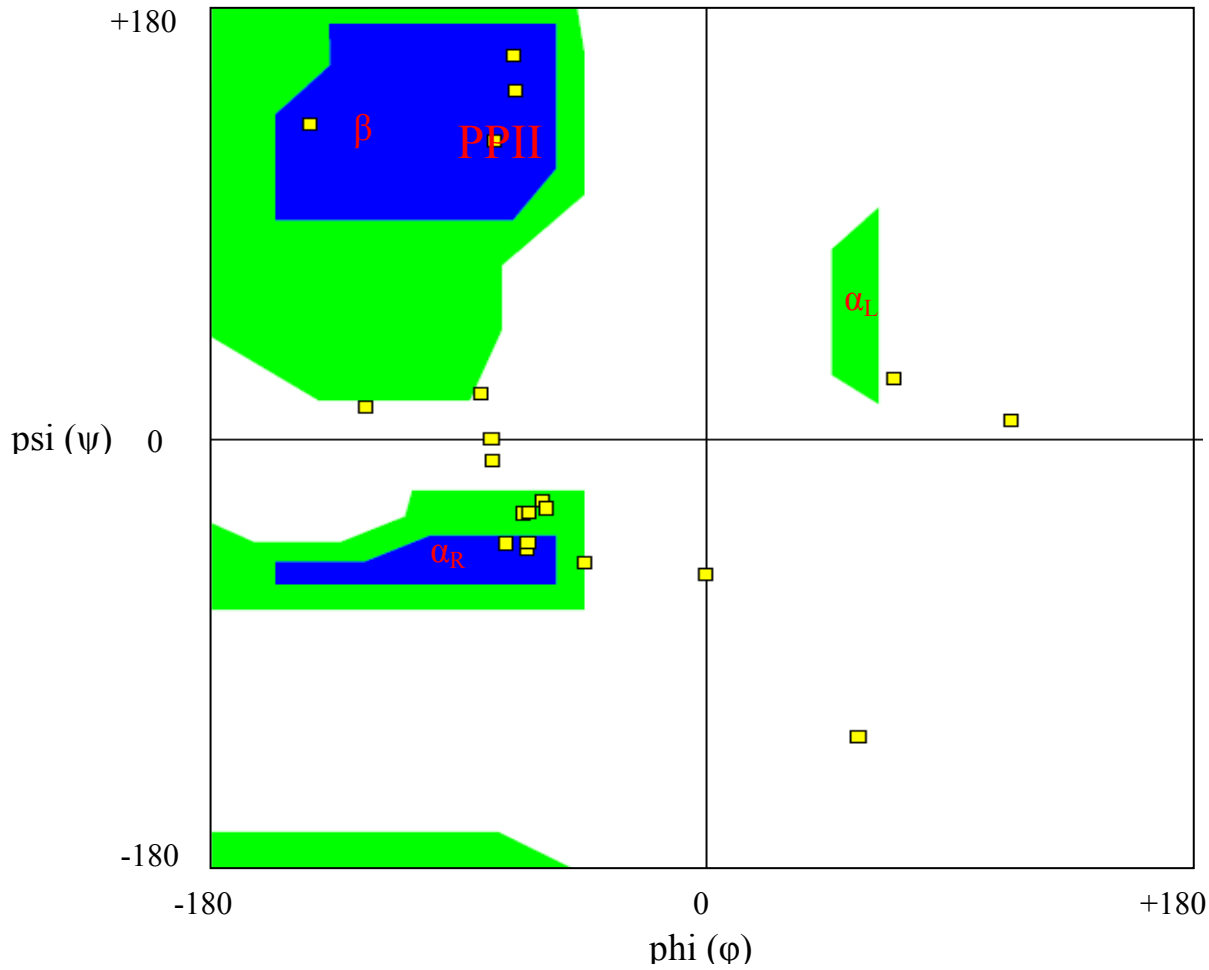


Figure 1-5. Ramachandran plot for the 20-residue Trp-cage protein (PDB ID: 1L2Y). The dots are created by each residue from the Trp-cage protein. The Trp-cage protein is one of smallest folding protein-like molecule. The structure is showed in Figure 1-6 C.

subjects in molecular biology.[10] There are two main experimental tools in structural biology; X-ray crystallography and NMR. X-ray crystallography is the first discovered and the most dominant tool, being widely used for structure determination of macromolecules. As of May 2007, 85% (37,101 out of total 43,633 structures) of released structures in Protein Data Bank were obtained based on X-ray crystallographic results. However, its application is not universal because it can be used only when high quality of crystal is available. Moreover, it provides averaged atomic positions, with atomic displacement parameters (or B-factors) used to infer the internal motions of proteins.[11]

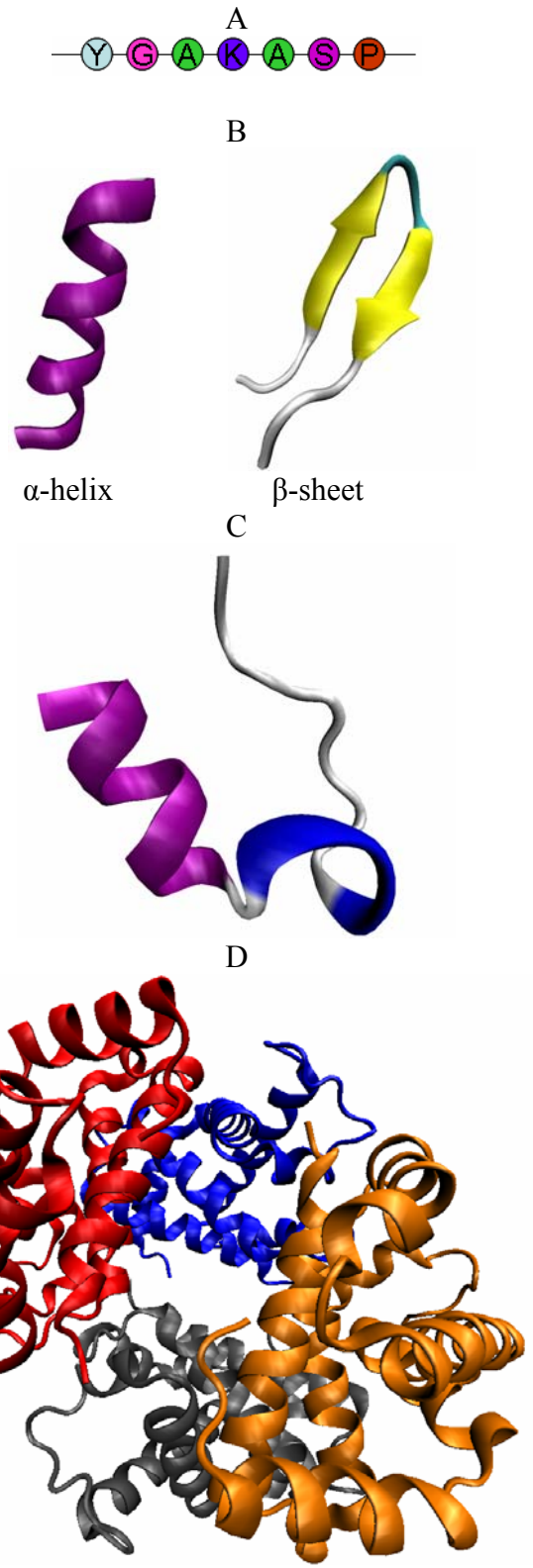


Figure 1-6. The diagram of protein structures. A) Primary structure, B) Secondary structure using an α-helix and β-sheet (PDB ID: MBH12), C) Tertiary structure (Trp-cage protein), and D) Quaternary structure showed using hemoglobin complex (1GZX)

The NMR technique has been also commonly used. It is, however, limited to smaller molecular systems, but sensitive enough to recognize mobile regions of macromolecules in aqueous solutions. In addition to effort on these technologies, theoretical work, especially computational studies, are routinely used to support and complement experimental data.

1.3 Protein Folding

The dynamical process where a protein forms its well-defined three dimensional structures to achieve biological activity is called "Protein Folding". For proteins that do fold, they usually do into one specific unique state in just a few seconds (or less) from any starting conformation. This state is defined as the "native state". Slight changes, such as pH, or temperature, can convert biologically active protein molecules in native state (folded) to biologically inactive denatured state (unfolded). Moreover, many human diseases result from protein misfolding or aggregation. Therefore, protein folding research is the one of the most important subjects in biology. Although enormous efforts were made from experimental and theoretical studies for nearly 40 years, the pathways and mechanisms of protein folding have not been yet fully understood due to the complexity of the systems and the fast time scale of folding.[12-14]

In the early 1960's, experimentally, Anfinsen *et al.* [15] studied the refolding of the denatured bovine pancreatic ribonuclease (RNase). RNase as non-functional protein, immediately returns to its native conformation (folding process) from a randomly coiled structure, upon removal of the denaturant (8M urea), which helps to restore its enzymatic activity. This behavior established the "thermodynamic hypothesis", which is directly related to the tertiary structure of a protein and more importantly, there is a thermodynamically the most stable minimum on the free energy profile, i.e., the free energy is being lowest in the native than in the unfolded state.[16] Thus, Anfinsen's observation opened a new era of protein folding research.

In 1968, Cyrus Levinthal showed that it is impossible that the protein folds into its native state by sampling all possible conformations.[17] For example, if a 150-residue protein molecule has only three stable conformations for each amino-acid (the allowed regions are α, β, and L from the Ramachandran plot), then $3^{150} \approx 10^{68}$ possible conformations exist. In order to obtain the native structure from a random search, it will take $\sim 10^{48}$ years, substantially longer than the age of the universe. This argument is popularly known as the "Levinthal's paradox". Therefore, Levinthal conjectured that proteins must fold to their native, stable conformation by using well-determined "folding pathways". These endeavors to find the correct pathways are continuing to these days with increasing success.

1.4 From the Experiment to the Simulation

The present study aims towards simulating experimentally observed processes, by computing thermodynamic and kinetic properties. Advances in computational power and speed have opened the way to investigations of new fields and new possibilities. Karplus and his coworkers first introduced molecular dynamics (MD) simulation of a biological macromolecule in 1977.[18] The MD simulation can generate the configurations of the system based on Newton's law of motion and hence, can provide ultimate details of individual atomic motions as a function of time. Hidden details of interest (including folding pathways) can be revealed by an MD simulation. Thus, the simulation can play an important role in interpreting experimental observations.

Experimental outcome is frequently compared with data calculated from MD simulations in order to validate the methodology and estimate systematic errors. The simulation can be carried out under conditions that are difficult or impossible to achieve in experiments, for example, under very high temperature or pressure. Therefore, combination of experimental and simulational information is of relevance in complementing and validating both approaches.

1.5 Temperature Jump Experiments

A major shortcoming of most experimental methods in protein folding is their limited time resolution while key events might occur much faster than that. Relaxation methods that can probe very fast time scales are of great interest in the studies of fast reactions.[19,20] A powerful relaxation method for the study of protein folding is temperature-jump (T-jump). The T-jump experiments were originally developed in the 1950s with the application of resistive heating. It was recently revisited and applied to protein folding dynamics with the use of modern laser heating.[21] Laser T-jump can reduce the dead time to nanosecond or picosecond scale and probe the earliest folding events. Laser T-jump is also advantageous with regard to the fast rise time coupled with a small amount of sample due to small heating volume. Fast T-jump experiments are, hence, adequate for studies of kinetics and later combined with theoretical results for a detailed description of biological systems.

A scheme of experimental T-jump relaxation kinetics is shown in Figure 1-7. The sample is initially in an equilibrium state at the initial temperature, $T_{initial}$. At t=0, a pulsed pump laser increases the temperature of the solution until the final temperature (T_{final}) is reached within a nanosecond or shorter timescale.[22] The final temperature should be sufficiently different from the initial temperature to perturb the equilibrium significantly. Temperature change (ΔT) is generally set up between 10°C and 20°C in an aqueous sample of protein. The unfolding process is initiated rapidly, and the subsequent relaxation kinetics is monitored and recorded at a time resolution of ~1ns [22,23] by using time-resolved infrared (IR) spectroscopy,[19,24] UV Circular Dichroism (CD),[20,25] or Trp-fluorescence,[4,26,27] until the system reaches a new equilibrium. The resulting kinetic spectra are then fitted to a proper relaxation curve; for example, single (for two-state mechanism) or double (for three-state mechanism) exponential model.

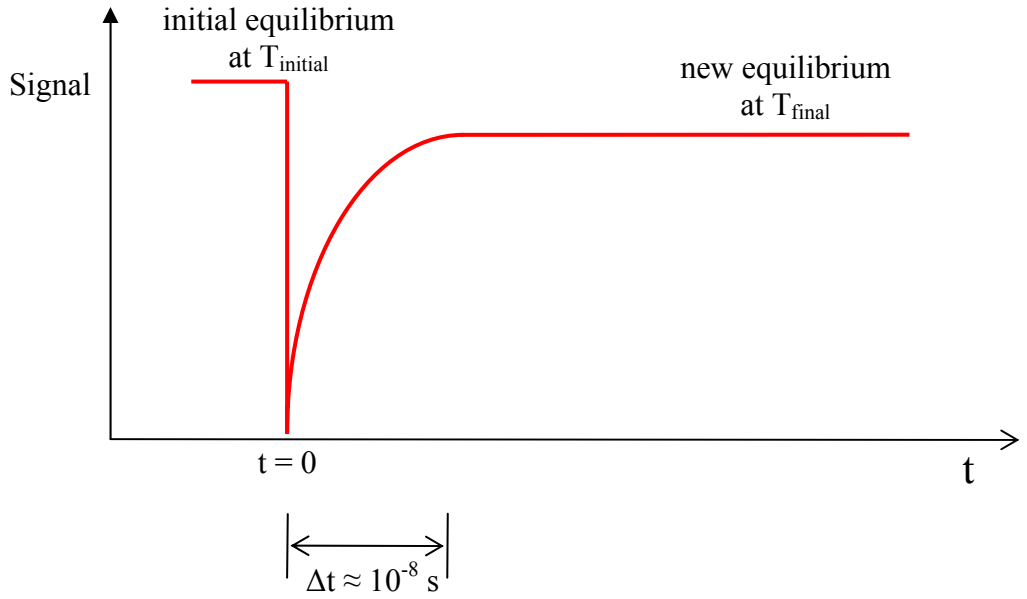


Figure 1-7. The general scheme of Temperature-jump (T-jump) relaxation kinetics.[28]

In a two-state folding mechanism, the process assumes only two states; the folded (F) and unfolded (U) ensembles with k_F and k_U, the folding and unfolding rates, respectively.

$$\text{Unfolded (U)} \underset{k_U}{\overset{k_F}{\rightleftharpoons}} \text{Folded (F)} \qquad (1\text{-}1)$$

The folding rate equations are derived as,

$$\frac{d[F]}{dt} = k_F[U] - k_U[F] \qquad (1\text{-}2)$$

$$\frac{d[U]}{dt} = k_U[F] - k_F[U] \qquad (1\text{-}3)$$

where $[U]$ and $[F]$ are the concentrations of unfolded and folded states, respectively. If we assume an initial concentration of 1, then this makes $[U] = 1 - [F]$. Thus, the time dependent populations can then be directly solved leading to:[29]

$$F(t) = C_1 e^{-\lambda t} + C_2 \qquad (1\text{-}4)$$

$$U(t) = 1 - F(t) \qquad (1\text{-}5)$$

The population relaxation is dominated by a single rate constant,

$$\lambda = k_U + k_F = k_F(1 + K^{-1}) = 1/\tau_{relax} = 1/\tau_U + 1/\tau_F \qquad (1\text{-}6)$$

In this equation, τ_F is the folding time (and τ_U is the unfolding time). The values of the two constants C_1 and C_2 are completely determined from the initial and final equilibrium concentrations and are then not adjustable parameters. The equilibrium constant K ($=\frac{k_U}{k_F}$) can be written as a ratio of the forward and backward rate constants and also determined by the equilibrium concentrations of the folded and unfolded states. Therefore, we could derive the folding (τ_F) and unfolding times (τ_U) by combining the relaxation results (Equation 1-6) and equilibrium constant.

1.6 Circular Dichroism

Circular Dichroism (CD) spectroscopy is a widely used technique for the study of secondary structures of polypeptides and proteins.[30-34] CD measures the differential absorption of choromophore –containing chiral molecules– between left circularly polarized light (LCPL) and right circularly polarized light (RCPL), CD=Abs(LPCL)–Abs(RPCL), which arises from structural asymmetry chirality. This, therefore, provides information about both conformation and its change. It also valuable for analysis of macromolecules, globular structures, and drug complexes in the field of biological, biochemical, chemical, and pharmaceutical sciences, because it can give more detailed data than most absorption or fluorescence spectroscopy techniques. [35,36]

The CD data are reported in units of absorbance (Abs(LPCL)–Abs(RPCL)) or ellipticity (θ). Molar ellipticity (θ) is usually utilized when CD involves molar concentration, and is defined as:

$$[\theta] = 100\theta / Cl \tag{1-7}$$

where C is the molar concentration (mol/L) and l is cell pathlength (cm) of the sample. The unit of molar ellipticity is typically reported as $deg \cdot cm^2 \cdot dmol^{-1}$ or $deg \cdot M^{-1} \cdot m^{-1}$.

CD data contain secondary structural information in terms of amide transitions in the backbone chain. For example, α-helical content consists of a positive band at 190nm and two negative bands at 208nm and 222nm,[37] whereas β-sheet contains two opposite signs at 215nm (minimum) and 198nm (maximum).[38] In particular, a strong negative band with molar ellipticity at 222nm is a key indication of helix formation of protein and peptides, because it is strongly affected by structural changes between folding and unfolding processes. Helical content can be monitored effectively in CD spectra and be used to understand protein folding and unfolding procedure.[39] Figure 1-8 shows the standard curves for CD spectra of poly-L-lysine in different secondary structure conformations (random coil, β-sheet and α-helix) from Campbell *et al.* [40]

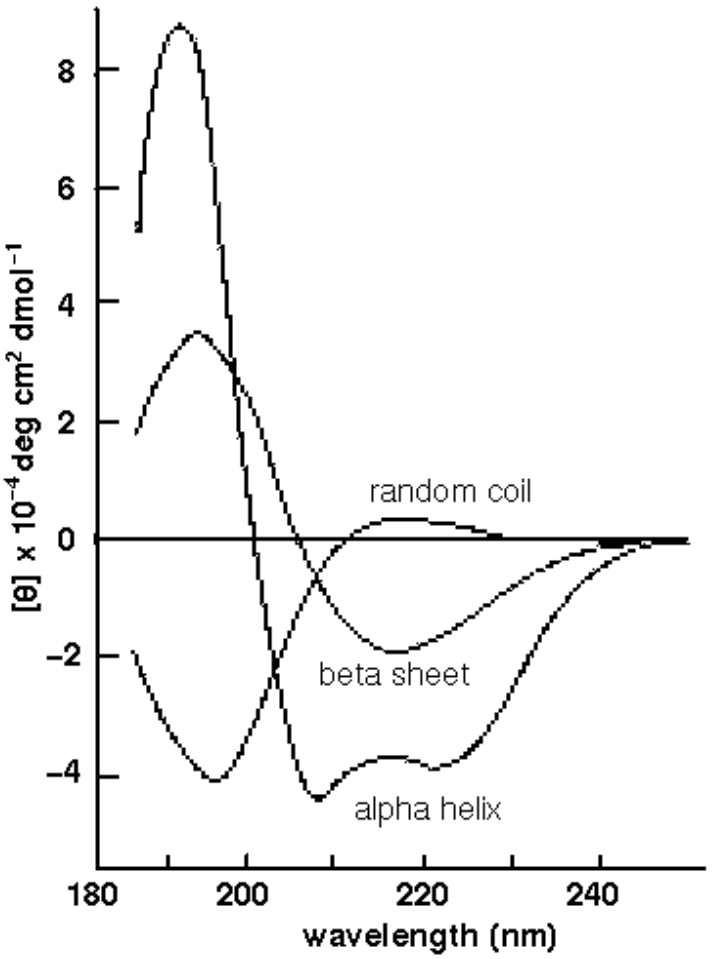


Figure 1-8. The standard curves for CD spectra of poly-L-lysine in different secondary structure conformations taken from Campbell & Dwek, 1984.[40]

1.7 NMR Spectroscopy

The spectroscopic measurements are extended to study more completely structural properties of proteins and peptides including the kinetic and thermodynamic results on protein folding.[41] Nuclear magnetic resonance (NMR) spectroscopy is also an important tool in these studies along with CD spectra. The most accessible quantities in NMR spectroscopy are chemical shifts,[42] nuclear Overhauser effects (NOEs), and scalar coupling constants.

Allerhand *et al.* showed significant differences between C^{α} chemical shifts in random coil and helical polypeptides.[43] These observations showed that the chemical shift might be applied to probe secondary structures of proteins. The chemical shifts of ^{1}H, α-^{13}C, and carbonyl-^{13}C are conformation-dependent, as shown by empirical [44,45] and *ab initio* studies.[46] Therefore, they can be used for determination of backbone conformations. Later, Dalgarno and his colleagues defined the 'secondary structure shift' (sometimes called 'conformational' or 'conformation-dependent' shift), $\Delta\delta_{ss}$, as:[47]

$$\Delta\delta_{ss} = \delta_{obs} - \delta_{r.coil} \tag{1-8}$$

where $\delta_{r.coil}$ refers to the standard chemical shift measured from random coil. The relative secondary structure shift ($\Delta\delta_{ss}$) could be correlated with the intensity of the helical CD signal (θ_{222}). For example, upfield shifts for ^{1}H, ^{15}N and downfield shifts for ^{13}C are observed for helix formation.[42]

Since one-dimensional (1D) NMR spectra are too complex to interpret due to severe overlapping signals, two-dimensional (2D) experiments are more popular in the studies of protein folding at individual residues. The hetero-nuclear single quantum coherence (HSQC) experiment is also frequently used in the field of protein NMR. The spectra are two-dimensional between ^{1}H and hetero nuclei (^{13}C or ^{15}N). They contain thermodynamic information such that the spectrum is well dispersed and all individual peaks are distinguishable when the protein is

folded. In this dissertation, we focused on conformational variations by direct use of chemical shift and HSQC experiments.[48]

1.8 Overview of Research Projects

The main goal of this work is to simulate the experimental T-jump setup for the study of protein folding. Computational methods used include replica exchange molecular dynamics (REMD), calculated CD spectra, structural cluster analysis, and computational NMR chemical shifts. Two projects have been performed: the first one is the T-jump simulation of $Alanine_{20}$, and the second project is the same type of study of two related 14-residue peptides.

1.8.1 First Project: Simulating Temperature Jumps for Protein Folding

My first project presents a new computational methodology aimed to calculate thermodynamic and kinetic properties of peptide folding, and designed to mimic the way experimental measurements of these properties are made. Particularly, I focus on T-jump simulations of folding rates, and show how a combination of REMD followed by multiplexed molecular dynamics starting from structures taken from the REMD runs can be used to extract properties in line with experiments. A model system, $Alanine_{20}$, was studied in this project as a proof of principle and description of the methodology.

1.8.2 Second Project: Folding Kinetics by Temperature-Jump Simulations of Two Related 14-residue Peptides

As follow-up of Project 1, REMD simulations of two closely related 14-residue peptides were performed to obtain equilibrated ensembles. Snapshots from this ensemble were used as initial structures for the T-jump simulations. These two 14-residue peptides are very similar but have different experimental folding rates by a factor of ~2.9.[49] They were selected by the experimental group to compare end-capping effects which can stabilize α-helix. The folding kinetics of the two 14-residue peptides is studied by using T-jump simulations, and their results

are analyzed using calculated CD spectra to obtain the folding and unfolding rate. For these two systems, the relaxation time and folding/unfolding rate constants are calculated and compared with experimental data.

CHAPTER 2
THEORY AND METHODS

In this chapter we will present some general discussion of theory and methods used in this dissertation. In particular, we will address issues of sampling, force fields, Langevin dynamics, generalized Born (GB) solvation model, and computation of circular dichroism (CD) and NMR chemical shifts.

2.1 Conformational Sampling

The MD simulations of biomolecules are still immature due to both inaccuracy of the force fields and inadequate conformational sampling associated with the number of degrees of freedom of system. The energy surface of biological systems is generally rough and rugged, such that it contains many local energy minima, which are isolated by high, insurmountable energy barriers.[50,51] MD simulations may often get trapped in a local minimum and never reach the global minimum. One way to overcome this sampling problem is to perform simulations in generalized ensembles, where the construction of the ensemble is weighted by a non-Boltzmann probability weight factor. Therefore, the resulting distribution guarantees a random walk in energy space, producing much better sampling in the conformational space. The results need to be properly re-weighted to give any thermodynamic quantity as a function of temperature.[52-54]

Many methods based on generalized ensemble algorithms have been introduced to overcome sampling problems of biological molecules; the multicanonical algorithm (MUCA,[55,56] also referred as entropic sampling[57] or adaptive umbrella sampling[58]), simulated tempering (ST),[59] 1/k-sampling,[60] Tsallis statistics [61] with simulated tempering,[62] replica-exchange method (REM),[63,64] and replica-exchange multicanonical algorithm (REMUCA).[65] The replica- exchange method (REM) (or parallel tempering [64,66]) is one of most widely used algorithms in a generalized-ensemble.

Replica Exchange Method (REM). In REM, the standard Boltzmann weight factor can be used. A number of non-interacting copies (replicas) can be simulated independently and simultaneously at different temperatures by the conventional MD or Monte-Carlo (MC) methods. Conformations are exchanged between different temperature replicas every few steps with a specified transition probability that is defined by the Metropolis criterion. This exchange process enforces random walks in temperature space, which in turn leads to random walks in potential energy space. Consequently, REM has been widely applied to protein and peptide folding research.[50,64,67,68] This dissertation makes use of the REM algorithm, modified to be combined with molecular dynamics, the so-called the replica exchange molecular dynamics (REMD).

In REM,[50] an artificial system composed of M non-interacting copies (or replicas) is considered at M different temperatures, T_m (m=1, 2,…, M). The state of this generalized ensemble is defined as $X = (x_1^{[i(1)]},...,x_M^{[i(M)]})$ with $x_m^{[i]} \equiv (p^{[i]}, q^{[i]})_m$ where $p^{[i]}$, $q^{[i]}$ represent momenta and coordinates, respectively, for replica *i* at temperature *m*. Since the replicas are non-interacting, the weight factor (*W*) for the state X is then given by the product of Boltzmann factors for each replica or temperature, as shown in Equation 2-1:

$$W_{REM}(X) = \exp\left\{-\sum_{i=1}^{M} \beta_{m(i)} H(p^{[i]}, q^{[i]})\right\} \tag{2-1}$$

where $\beta = \frac{1}{k_B T}$ (k_B is the Boltzmann constant) and the Hamiltonian $H(p^{[i]}, q^{[i]})$ is the sum of kinetic and potential energies.

Now, one can attempt to exchange temperatures between the *i*th and *j*th replicas at temperatures T_m and T_n, respectively. The new state of the system becomes:

$$X = (...,x_m^{[i]},...,x_n^{[j]},...) \quad \rightarrow \quad X' = (...,x_m^{[j]},...,x_n^{[i]},...) \tag{2-2}$$

The detailed balance condition needs to be applied to converge to an equilibrium ensemble:

$$W_{REM}(X)w(X \to X') = W_{REM}(X')w(X' \to X) \tag{2-3}$$

where $w(X \to X')$ is the transition probability from state X to X′ and $W_{REM}(X)$ is the weight factor of the state X. From Equation 2-1, 2-2, and 2-3, the exchange probability (P) is obtained:

$$\begin{aligned} P &= \frac{w(X \to X')}{w(X' \to X)} = \frac{W_{REM}(X')}{W_{REM}(X)} \\ &= \frac{e^{-\beta_j E_i} e^{-\beta_i E_j}}{e^{-\beta_i E_i} e^{-\beta_j E_j}} = e^{-\beta_j E_i - \beta_i E_j + \beta_i E_i + \beta_j E_j} = e^{(\beta_i - \beta_j)(E_i - E_j)} \\ &= e^{\Delta} \end{aligned} \tag{2-4}$$

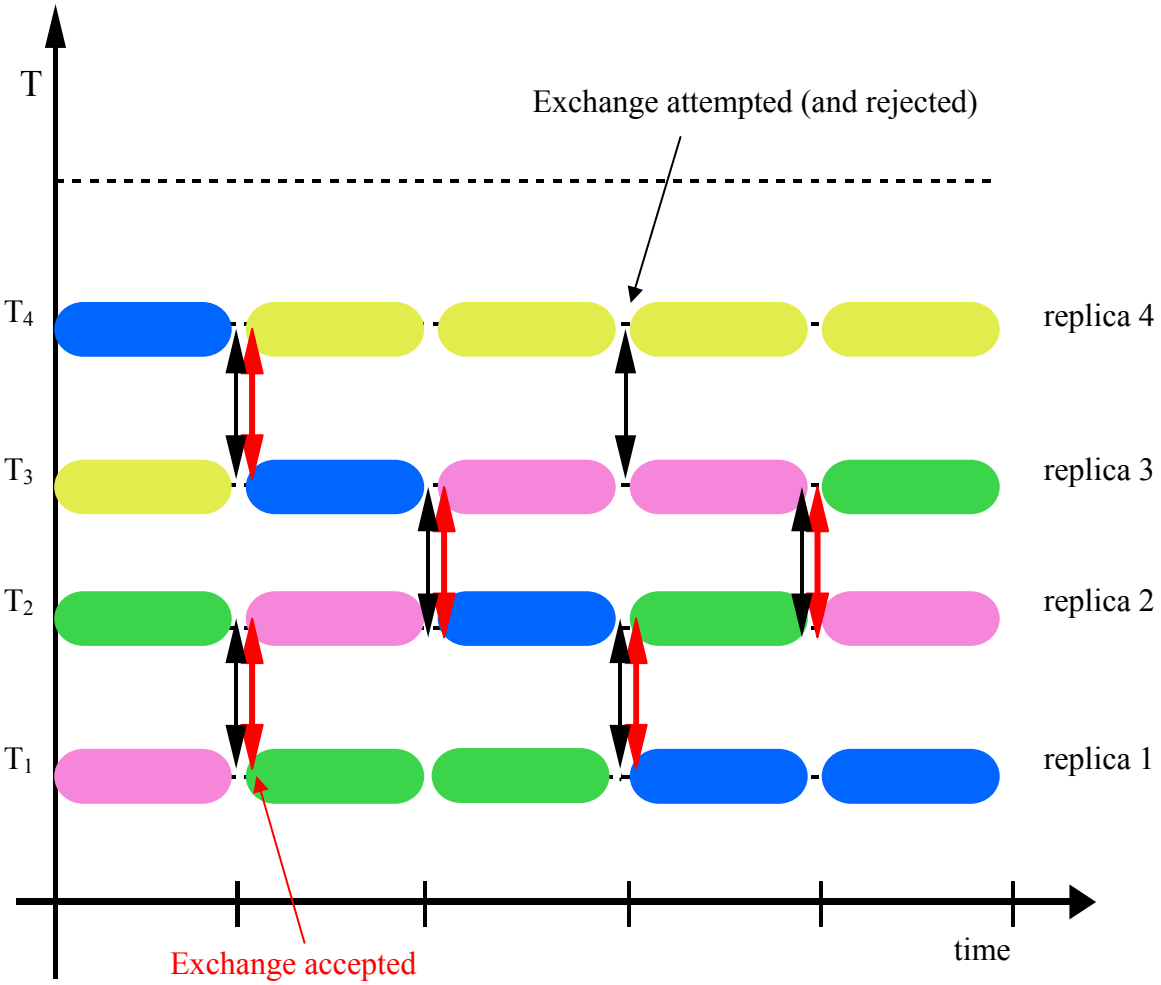


Figure 2-1. The sketch of Replica-exchange method (REM) simulation in amber molecular dynamics packages.

where $\Delta = (\beta_i - \beta_j)(E_i - E_j)$ and E is the potential energy of each replica. One can now obtain the acceptance probability of replica exchange (*P(accept)*) by using a Metropolis criterion (or Monte Carlo procedure):[69]

$$P(accept) = \min\left[1, e^{(\beta i - \beta j)(E_i - E_j)}\right] = \min\left[1, e^{\Delta}\right] \quad (2\text{-}5)$$

The general simulation of the REM performs the following steps:

1. Each replica is simulated-based on canonical ensemble-in parallel and independently for a certain number of MD steps.
2. Periodically, replicas with adjacent temperatures are swapped with acceptance probability, *P(accept)* from Equation 2-5.
3. Repeat the process.

In step 2, the exchanges are only allowed between adjacent replicas in temperature, because the acceptance ratio of the exchange decreases exponentially with increasing difference between the two temperatures.[63] Figure 2-1 shows a sketch of a REM simulation, describing the mechanism of replica exchange (or rejection) between different temperatures.

2.2 Force Field

The force field representing the collection of molecular interactions represents the behavior of all atoms and bonds with specific fitting parameters. Many different simulation packages have been developed over the years; for example, AMBER,[70,71] GROMACS,[72] OPLS,[73] and CHARMM.[74] Names of these packages generally imply the empirical force fields. In the framework of this dissertation, the AMBER force field, the most commonly used for biomolecular systems, was applied. Its potential energy function (*U(R)*) with a corresponding set of empirical parameters is shown below:[75]

$$
\begin{aligned}
U(R) = & \sum_{bonds} K_r (r - r_{eq})^2 && \textit{bonds} \\
& + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 && \textit{angles} \\
& + \sum_{dihedrals} \frac{V_n}{2} (1 + \cos[n\phi - \gamma]) && \textit{dihedrals} \\
& + \sum_{i<j}^{atoms} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} && \textit{van der Waals} \\
& + \sum_{i<j}^{atoms} \frac{q_i q_j}{\varepsilon R_{ij}} && \textit{electrostatic}
\end{aligned}
\qquad (2\text{-}6)
$$

where r_{eq} and θ_{eq} are equilibration structural parameters. K_r, K_θ, and V_n are force constants, n is multiplicity, and γ is the phase angle for the torsional angle parameters. In addition, A, B, and q are parameters related to the non-bonded potentials. Balance in parameterization can result in reasonably good compromise between accuracy and computational efficiency, thus, reproducing simulation results close to experimental ones.

For the non-bonded part, the van der Waals parameters are usually determined by thermodynamic properties of various pure liquids.[76,77] The electrostatic parameters are calibrated using a restrained electrostatic potential fit (RESP) model.[78,79] The parameters for the first three internal terms (bond, angle, and dihedral) of Equation 2-6, come from a combination of experimental data and high-level ab initio calculations.

Numerous MD simulations have been run for proteins and nucleic acids under these force fields and compared with experimental structures over the decades. Nevertheless, older AMBER type force fields present many deficiencies, such as the well-known over-stabilization of α-helices. In order to avoid this tendency and thus to get better backbone dihedral angles, Hornak *et al.* introduced new parameters, considering the energies of multiple conformations from high level ab initio calculations.[80] Since there are many approximations within force fields,

optimization of force fields remains to be improved along with advance in experimental technology.

2.3 Generalized Born (GB) Solvation Model

Simulations with an explicit treatment of solvent provide much improved accuracy, but is computationally expensive for larger molecules, for example, proteins or nucleic acids.[81-83] Alternatively, implicit solvation models have proven to be valuable tools for computational efficiency and also relative simplicity. In particular, the Generalized Born (GB) model is one of the most popular implicit solvation models. The solvent, such as water, in the GB model is treated as an infinite continuum medium with the corresponding dielectric properties.[81,84] Thus, the GB model calculates approximate values of the solute-solvent electrostatic free energies of solvation (G_{pol}) and gives rapid estimates of G_{pol} to save computation times in calculations.[81,85]

The electrostatic contribution of the solvation free energy in the reaction field, Φ_{reac}, is approximated by a system of simple ionic particles with radius α and charge q:[84]

$$\Delta G_{pol} = \frac{1}{2} q \Phi_{reac} = \frac{q^2}{2\alpha}\left(\frac{1}{\varepsilon} - 1\right) \tag{2-7}$$

where ε is the dielectric constant and this result is the well-known Born formula.[86] If the simple ion is expanded to a "molecule" consisting of spheres of radii ($\alpha_1, \alpha_2, \ldots, \alpha_N$) and charges ($q_1, q_2, \ldots, q_N$) with the separation distance r_{ij} between charges q_i and q_j, the polarization free energy (Generalized Born (GB) equation) has been approximated as follows:[81]

$$\Delta G_{pol} = \frac{1}{2}\left(\frac{1}{\varepsilon} - 1\right)\sum_{i}^{N}\sum_{j \neq i}^{N} \frac{q_i q_j}{f_{GB}} \tag{2-8}$$

where

$$f_{GB} = (r_{ij}^2 + \alpha_{ij}^2 e^{-D})^{0.5}, \quad \alpha_{ij} = (\alpha_i \alpha_j)^{0.5}, \quad D = r_{ij}^2 / (4\alpha_{ij}^2) \tag{2-9}$$

The f_{GB} is defined as a function of an "effective Born radius"(α_i) when the distance $r_{ij} \rightarrow 0$, while $f_{GB} \rightarrow r_{ij}$ as $r_{ij} \rightarrow \infty$.[87]

In the original model, the effective Born radius (α_i) was computed by a numerical integration procedure,81 but more recently "pairwise" approximations in which αi is estimated via a summation over atom pairs, has been proposed by several groups.[82,87,88] Therefore, effective Born radius (α_i) can be derived as:

$$\alpha_i^{-1} = \rho_i^{-1} - \sum_{j \neq i} g(r_i, r_j, \rho_i, \rho_j) \tag{2-10}$$

where ρ_i is an intrinsic radius for atom *i*, and g() is a positive function which depends on the positions and sizes of the atoms and also has scaling factor for an empirical correction.[83,88,89]

While the GB approximation is highly efficient for larger molecules, it is not so well balanced between protein-protein and protein-solvent interactions, compared to explicit solvent model. For example, over-stabilization of salt bridges has been frequently observed in GB model, causing a significant conformational difference from explicit solvent model.[90-93] In order to quantify the potential overstabilization of ion pairs for both models, Geney *et al.* performed Potential of Mean Force (PMF) method of salt bridge formation and found an excessive strength of salt bridges in GB.[94]

The lack of solvent friction in GB can accelerate conformational transition rates, resulting in faster conformational sampling and at the same time, correctly predicting the native conformations.[95-98] Langevin Dynamics is one of the commonly employed methods to overcome the frictional and high velocity collision problem.

2.4 Langevin Dynamics

Langevin dynamics complements Newton's second law to account for omitted (solvent) degrees of freedom. The Langevin equation includes a frictional term in the form of a stochastic differential equation and thus, it attempts to mimic the viscous aspect of a solvent. However, it is an incomplete implicit solvent model, since it ignores electrostatic or hydrophobic effect. Those

effects are included via the implicit solvation models, such as GB, described above. Furthermore, Langevin dynamics controls the temperature as a thermostat, thus approximating the canonical ensemble.

The Langevin equation for motion of a particle *i* can be written using a stochastic differential equation as:

$$m_i \frac{d^2 x_i(t)}{dt^2} = F_i\{x_i(t)\} - \gamma_i \frac{dx_i(t)}{dt} m_i + R_i(t) \quad (2\text{-}11)$$

where m_i and x_i are the mass and position of particle *i* in the simulation, respectively. F_i is an interaction force between a particle of interest and other particles and *R(t)* is the force on the particle due to random fluctuation by interaction with solvent molecules.[99] The collision frequency (γ) is derived from the friction coefficient (ξ) by $\gamma=\xi/m$ (m is the mass of the particle) and sometimes referred to as the friction coefficient in the literature. *R(t)* is a white thermal noise that obeys the fluctuation-dissipation theorem at temperature T,[100]

$$\left\langle R(t)R(t')\right\rangle = 2\gamma k_B T m \delta(t-t') \quad (2\text{-}12)$$

where k_B is the Boltzmann constant.

2.5 Computation of Circular Dichroism (CD)

The CD spectra generally provides a direct measure of "chirality" of the molecular structure, since the magnitude and sign of the CD spectrum depend on the geometrical variables and electronic structure of a molecule. The protein can be characterized as a collection of independent chromophores. An individual chromophore that is sensitive to secondary structural conformation and its interactions between the transitions on chromophores are the basis of calculations of CD. The matrix method is most commonly used to compute the CD spectra of proteins and peptides, where the excited states of each chromophore are subject to quantum chemical treatment, considering interactions between the chromophores based on classical physics.[101,102] The rotational strength, which gives the intensity of a CD band, can be

theoretically defined in terms of the imaginary part of the scalar product of the electric (μ) and magnetic (m) transition dipole moments of an electronic transition, using the Resenfeld equation and measures transitions of excited state.[103] For an electronic transition from ground (0) to excited (i) states (0 → i), the rotational strength can be calculated according to:[104]

$$R_{oi} = \mathrm{Im}\left(\langle \psi_0 \mid \mu_e \mid \psi_i \rangle\right) \bullet \left(\langle \psi_i \mid \mu_m \mid \psi_0 \rangle\right) \tag{2-13}$$

where Im represents imaginary part, ψ_0 and ψ_i are the ground and excited state wave function, respectively. μ_e and μ_m are the electronic and magnetic transition dipole moment, respectively.

Since an electronic excitation occurs only within a group, rather than between groups, a protein can be considered as a set of M non-interacting chromophoric groups in the matrix method. The excited-state wave function of the whole system (Ψ_T) is expressed as a linear combination of the basis functions (Φ_{ia}) for each chromophoric group with the ni excitation:

$$\Psi_T = \sum_i^M \sum_a^{n_i} c_{ia} \Phi_{ia} \tag{2-14}$$

Each basis function is a product of M monomer wave functions, such that:

$$\Phi_{ia} = \varphi_{10} \ldots \varphi_{ia} \ldots \varphi_{j0} \ldots \varphi_{M0} \tag{2-15}$$

where φ_{ia} corresponds to the wave function of chromophore *i* for ($0 \rightarrow a$) excitation. When the molecule is not symmetrical, the CD spectrum can be obtained from the sum of these rotational strengths derived as a nonzero value from each transition.

A Hamiltonian matrix of a protein is composed of the excitation energy of a single chromophore (forming the diagonal elements) and the interaction between different chromophoric groups (forming the off-diagonal elements). The off-diagonal elements can be simplified by charge distribution of electronic interactions.[105] Thus:

$$V_{ij,kl} = \sum_m \sum_n q_{ijm} q_{k\ln} / r_{ijm,k\ln} \tag{2-16}$$

where *m* and *n* correspond to the point charge of the transition *j* (on chromophore *i*) and transition *l* (on chromophore *k*), respectively, and *r* represents the distance between the point

charges. As an example, the Hamiltonian matrix for the amide electronic transitions between $n\pi^*$ (at 220nm) and $\pi\pi^*$ (at 193nm) is shown below:

$$H = \begin{pmatrix} E^1_{n\pi^*} & V^{11}_{n\pi^*\pi\pi^*} & V^{12}_{n\pi^*n\pi^*} & V^{12}_{n\pi^*\pi\pi^*} \\ V^{11}_{n\pi^*\pi\pi^*} & E^1_{\pi\pi^*} & V^{21}_{n\pi^*\pi\pi^*} & V^{12}_{\pi\pi^*\pi\pi^*} \\ V^{12}_{n\pi^*n\pi^*} & V^{21}_{n\pi^*\pi\pi^*} & E^2_{n\pi^*} & V^{22}_{n\pi^*\pi\pi^*} \\ V^{12}_{n\pi^*\pi\pi^*} & V^{12}_{\pi\pi^*\pi\pi^*} & V^{22}_{n\pi^*\pi\pi^*} & E^2_{\pi\pi^*} \end{pmatrix} \quad (2\text{-}17)$$

The diagonalization of the matrix H using a unitary transformation provides the eigenvalues and eigenvectors corresponding to all transitions of the protein. The eigenvalue gives information about the excitation energy and the eigenvector describes the mixing of localized transitions. The rotational strength (Equation 2-13) of each excited state can now be derived from the eigenvector and be used to calculate the CD. In our work we will use the programs design by Sreerama and Woody, where instead of performing a quantum chemical calculation they employed parameter sets consisting of a combination of experimental data and theoretical parameters.[105,106]

2.6 Computation of NMR Chemical Shifts

The NMR chemical shifts are affected by the environment. Calculating the shift is, therefore, important for interpretation of structural information on macromolecules. Empirical methods,[44,48,107-111] semi-empirical models [112,113] and ab initio quantum approaches [114-117] have been tried to calculate the chemical shift.

2.6.1 Calculation of Proton (^{1}H) Chemical Shift

An equation of the proton chemical shift is generally described in terms of various contributions, as below:[118]

$$\Delta\delta = \delta_{total} - \delta_{rc} = \delta_{tor} + \delta_{ring} + \delta_{HB} + \delta_e + \delta_{side} + \delta_{misc} \quad (2\text{-}18)$$

where δ_{rc} = the random coil chemical shift value of an amino acid residue, δ_{tor} = the backbone torsional contribution, δ_{ring} = the ring current contribution, δ_{HB} = the contribution arising from

hydrogen bond, δ_e = the electric field or local charge contribution, δ_{side} = the side chain torsional contribution, and δ_{misc} = other chemical shift contributions including solvent, temperature, motional averaging, and covalent bond geometry. The empirical model (Equation 2-18) was developed and parameterized to experimental shift data through literature analyses.[107,114,115,119,120] However, it simply represents rough and empirical knowledge of chemical shift propensity, rather than unique (or complete) and quantitative description of proton chemical shifts (^{1}H).[112]

Quantum chemical shift calculations were performed to improve the accuracy of the previous empirical models, considering ring current,[121] electrostatic effects, structural dependence of magnetic anisotropy, and close contact contributions.[115] As a result, a new empirical model developed via a combination of the empirical formula with the quantum calculation, was introduced with an improved prediction of proton shifts.[115]

2.6.2 Calculation of ^{15}N and ^{13}C Chemical Shifts

Xu *et al.* predicted ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$, and $^{13}C'$ (carbonyl C) chemical shift in proteins, using a mix of quantum chemistry and a database of experiments.[122] Figure 2-2 presents an outline of ^{15}N and ^{13}C shift prediction algorithm, where database of density-functional derived shifts in the program SHIFTS (version 4.1) was used. This database is also used for chemical shift calculations in this dissertation.[122] The SHIFTS program first takes a protein structure in Brookhaven (PDB) format and calculates the structural parameter of all the amino acids within a given protein, such as backbone conformation, side-chain orientation, and hydrogen bonding geometry.

The density-functional database was built based on the calculated chemical shift patterns of 1335 peptide sequences which are derived from 20 proteins. The calculated results identified

various significant potential contributions to the shift: the backbone φ and ψ torsion angles of the three consecutive residues (preceding (i - 1), self (i), and following residues (i + 1)), side-chain orientations of two consecutive residues and hydrogen bonding. Therefore, the total contribution is given by the sum of the individual ones:

$$\Delta(c) = \sum_{k} \Delta(k, c) \tag{2-19}$$

where k denotes one of the contributions, and c is either helix or sheet structure. The predicted chemical shift, $\delta_{pred}(c)$, is then derived as:

$$\delta^{pred}(c) = \delta_{REF}(c) + \Delta(c) \tag{2-20}$$

$\delta_{REF}(c)$ is a reference chemical shift for an amino acid, where c = α for helix and c = β for sheet. $\delta_{REF}(c)$ is ideally determined by DFT calculation using the standard structure parameters from the literature.[122] Finally, the process is followed by side-chain orientation refinement based on experimental shifts for an improved prediction.

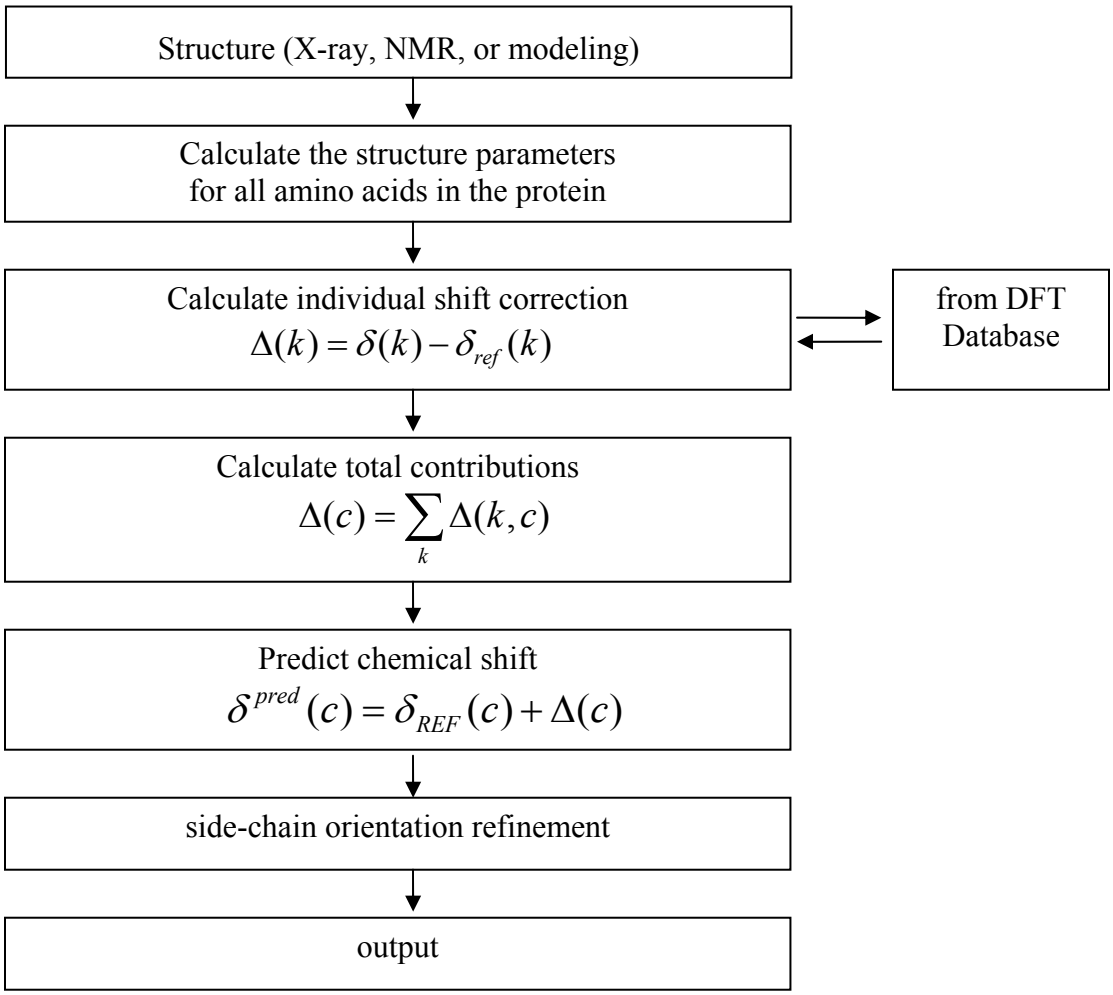


Figure 2-2. Outline of ^{15}N and ^{13}C chemical shift calculation in SHIFTS program.[122]

CHAPTER 3
SIMULATING TEMPERATURE JUMPS FOR PROTEIN FOLDING

3.1 Introduction

Understanding the structure, kinetics and thermodynamics of protein folding is one of the unsolved problems in biology. Many human diseases like Alzheimer's and mad cow disease [1-5] are directly associated with protein misfolding, unfolding and aggregation. When studying protein and peptide folding, there are three main questions of interest. First, given a primary sequence, is there a unique 3-D structure under physiological conditions and if so, what is that structure? Second, how does the peptide fold into its native structure? Third, how fast does it fold? [123-128] These questions are routinely answered in the laboratory using a mix of structural, thermodynamic and kinetic methods.[24,129] Both experimental and theoretical approaches to the protein folding problem have been used to address these questions.[25,28] It is basically impossible to probe all possible protein conformations experimentally because data, when available, is averaged over time and over many molecules (with the notable exception of single molecule experiments that still average over time).[24,130] Alternatively, molecular dynamics simulations can be used to provide a detailed description of the system.[131-134] In this chapter I focus on the speed of protein folding, but it is clear that underlying methodology, structure and thermodynamics are also available.

In recent years, the use of multiplexed molecular dynamics runs to study folding has become commonplace. The availability of very large numbers of processors for short times has permitted the existence of pioneering efforts such as Folding@Home.[4,135-137] This technique has been able to reproduce experimental folding rates for a number of systems.[4,138-140] In its most used, basic form, it starts from a very large number of initial conformations, and runs molecular dynamics for each of them for a pre-defined amount of time (a small number of nanoseconds).

The initial coordinates are usually chosen either fully extended or taken from a high temperature run. The procedure then monitors the evolution of each sample independently and, after the predefined amount of time has passed, simply counts the number of runs that produced a folded structure.[138,141,142] Under the assumption of a single exponential decay, one can extrapolate to long times and extract a rate constant for the process. In a single exponential decay assumption (over all time scales) with a time constant τ, and for M independent very short MD runs of time t, one expects around tM/τ runs to have folded.[137] For typical values of t=50ns, τ=10 μs and M=10,000 trajectories, one expects only 50 of the runs to have folded. These represent, by definition, the fastest component of the folding. The basic extrapolation relies on an assumption of a single exponential process over all time scales, even the very short ones. In other words, it is assumed that the fastest folders truly follow the same pathways as the overall ensemble.

However, some of the assumptions in the Folding@Home style methods become invalid under certain conditions. Several groups showed that an extrapolation of very short time decays to asymptotic exponential behavior might be unreliable.[143,144] Paci *et al.* [145] reported that the fastest folding events do not agree with a corresponding ensemble behavior obtained by the distributed computing for a three-stranded antiparallel β-sheet peptide. This is in some sense obvious: given a system with many free energy minima and a number of barriers (whose ensemble relaxation should be described in terms of a master equation description), one should not expect a single exponential decay to hold at ALL time scales.[146] This has been best expressed by Daggett and Fersht in ref.:[147] "It is our opinion that, at the molecular level, intermediates are always present. In other words, true two-state folding with only the denatured and native states occupying free energy minima is implausible."

Moreover, Pande and his colleagues[4] start from a fully extended conformation for their initial states. While they see evidence of very fast (tens of nanoseconds) relaxation to a collapsed, unstructured state in times substantially shorter than their simulation times, it is clear that this is a choice of unfolded ensemble that cannot be compared with typical kinetic measurements (such as temperature jumps) which usually measure relaxation times between two closely related equilibrium states.

Since in Folding@Home a folding event is only counted against a pre-determined coordinate set (the folded state), it is possible that a number of the simulations have ended in a different configuration. If this new state has lower (free) energy than the initial choice, then it should be properly called the folded state. The solution to this problem is to run every simulation long enough for a substantial percent of the ensemble to fold into the same state, which then is defined, a posteriori, as the folded state of the system.

Recently, Pande *et al.* and Levy *et. al*,[140,148-150] introduced a new method, namely, Markovian state models (MSMs), to predict protein folding rate constants. The MSMs can calculate both folding probability (P_{fold}) of all the configurations in a system and mean first passage time (MFPT) from the unfolded state to the folded state.[140]

In the present report, the folding kinetics of a polyalanine peptide is described and discussed based on MD simulation results. One of the most used experimental approaches to protein folding kinetics is laser-induced temperature-jump (T-jump) spectroscopy.[26,27,151] T-jump can raise the temperature extremely fast (nanosecond scale) and record the relaxation to equilibrium with a time resolution of ~1ns.[22,23] Since the usual distributed computing procedure (Folding@Home) is clearly not what is done experimentally, where, in the case of T-jump, relaxation of the ensemble from one equilibrium distribution to another one is realized, it is

important to ask if a protocol that closely resembles an experimental T-jump could be designed and explored. This chapter will show that this is indeed possible, focusing on a model system that should be considered as a proof of principle calculation.

In the computational T-jump which is first introduced in this chapter, the temperature will be increased suddenly, creating a system no longer at equilibrium and having to relax to a new state of equilibrium. In this method, the conformational space of a polyalanine peptide is pre-equilibrated at different temperatures using replica exchange molecular dynamics (REMD). In this way, we will have a-priori knowledge of what the ensemble looks like at the initial time, before the T-jump (equilibrium at T_{low}), and as time approaches infinity (equilibrium at T_{high}). Then we will monitor a number of trajectories started from structures taken from the T_{low} ensemble, run them at T_{high} for a pre-defined time, and then, at each time-slice, compute any property we desire. In a similar way with experiment, this data will be analyzed after watching the relaxation data, before deciding on what kinetic scheme will fit the data best. Additionally, we can then go into the ensemble, and ask detailed structural questions about rates and pathways.

3.2 Methods

3.2.1 Simulation Details

The initial structure of the alanine polypeptide, ACE–$(ALA)_{20}$–NME (ACE is acetyl beginning group and NME is N-methylamine ending group), was built in an extended conformation with the AMBER 8.0 simulation package[75] using recently published AMBER ff99SB force field,[76,152,153] which has been shown to provide improved agreement with experiment.[154,155] A cutoff of 16Å was used to compute long-range interactions, and the Hawkins, Cramer, Truhlar [88,89] pairwise generalized Born (GB) implicit solvent model, with parameters from Tsui and Case,[85] was then applied to mimic the effects of water solvation.[81,156] The system was initially subjected to 2,000 steps of minimization and the resulting

conformations were used as initial seeds for the REMD simulation.

The REMD method used the multisander implementation in the AMBER 9.0 molecular dynamics program. Sixteen replicas were simulated for 200ns each (total of 3.2μs) at exponentially spaced temperatures, from 153K to 542K (153, 166, 181, 197, 214, 233, 253, 276, 300, 326, 355, 386, 420, 457, 498, and 542K). These temperatures resulted in an average exchange rate of 15% between adjacent replicas. The SHAKE algorithm was used to constrain the lengths of all bonds involving hydrogen[157] and a 2-fs time step was used for every replica. Exchanges between replicas was attempted every 10ps, resulting in 20,000 attempted exchanges at each temperature. Conformations were recorded every 2ps from the simulation of each replica. The first 10ns of the simulation were discarded and the latter 190ns were saved for further calculations, giving a total of 19,000 configurations for analysis.

Conformations obtained from the REMD calculation were used as the initial structures for the T-jump experiment. 362 starting configurations were selected, equispaced in time, collected from the equilibrated simulations at 181K. Each member of this ensemble was then instantly and independently heated up to the T_{high} of 214K (the details of the T choices are presented later on) and MD was run for each member for 35ns to simulate the relaxation after the T-jump. All MD parameters are same as those used in REMD. Figure 3-1 shows a sketch of both the experimental T-jump scheme and computational T-jump setup to illustrate how these two methods are similar to each other.

In order to properly identify the structural elements of the ensemble, we performed a cluster analysis with moil-view [158] using backbone RMSD for residues 2 to 19 as a similarity criterion with average linkage.[159] The clusters were defined using a bottom-up approach with a similarity cutoff of 2Å, for C^{α}-RMSD. The representative structures from the REMD ensemble

at 153K were used to determine the composition of the folded ensemble.

3.2.2 Calculation of NMR Chemical Shifts

NMR Chemical shifts were computed in order to predict structural information. SHIFTS (version 4.1) program (by David Case group) was employed to estimate proton (amide proton, and H_α), as well as ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$, and $^{13}C'$ (from CO bonds) chemical shift of the polyalanine peptide.[112,121,122] SHIFTS program recognizes a protein structure in Brookhaven (PDB) format. It computes proton chemical shifts using empirical equations and ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$, and $^{13}C'$ chemical

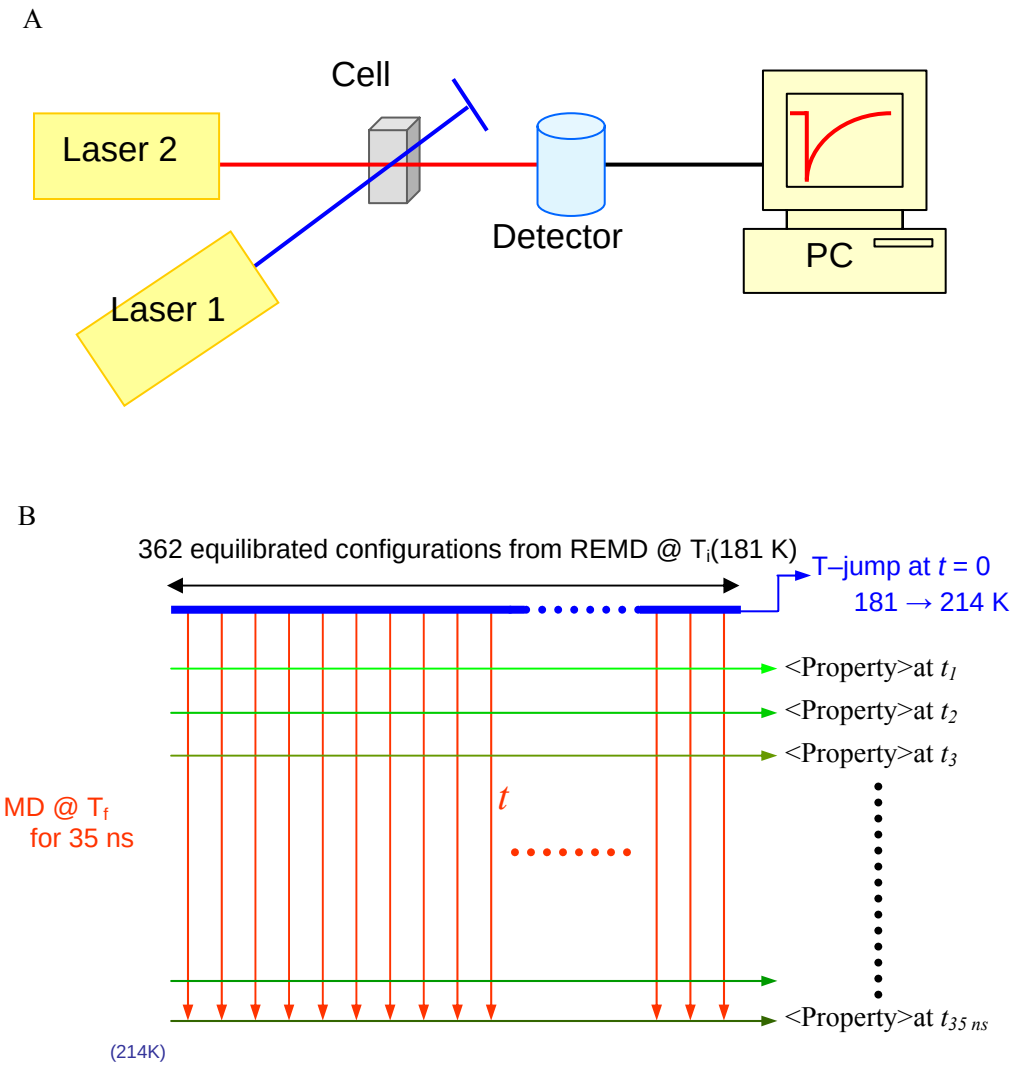


Figure 3-1. The diagram of the T-jump setup. A) Experimental T-jump scheme, B) Computational T-jump setup.

shifts from a database based on density functional calculations.

2D NMR spectra obtained using HSQC experiment and $^{13}C_{\alpha}/^{13}C_{\beta}$ crosspeaks relationship enable us to see structural distributions of folded and unfolded states.

3.2.3 Calculation of CD Spectra

Computations on CD spectra for estimating secondary structure and calculating folding properties were performed by Sreerama and Woody.[105,106] The matrix method [101] (in origin-independent formulation of rotational strength [102]) was employed with a transition parameter set consisting of a combination of experimental data and theoretical parameters: the authors used experimental data[160] for two amide ππ* transitions and parameters from intermediate neglect of differential overlap/spectroscopic (INDO/S) wavefunctions[161] for the nπ* transition. The rotational strength was computed in order to generate CD spectra through Gaussian band. The bandwidths assigned for the nπ* transition, and for two ππ* transitions were 10.5nm, 11.3nm, and 7.2nm, respectively.[39]

3.3 Results and Discussions

The conformations at the lowest temperature (153K) from REMD were sorted into clusters.[159] Clustering of the ensemble at 153K shows only two substantially populated clusters. The largest cluster contains 76% of the structures, whereas the next largest cluster is 9% populated, with no other cluster having a population higher than 6%. Figure 3-2 A shows the ensemble at 153K superimposed on the representative structure of the largest populated cluster. Figure 3-2 B shows the structure for that cluster, forming a well defined α-helix. This is then the reference structure for future analysis. For the ensembles at 181 and 214K, the C^{α}-RMSD (residues 2-19) are computed, versus the representative structure of most populated cluster, and obtain the histograms of Figure 3-3 A. There is a clear separation between the features in these plots, which allows us to define two different types of states, which are named folded (F) and

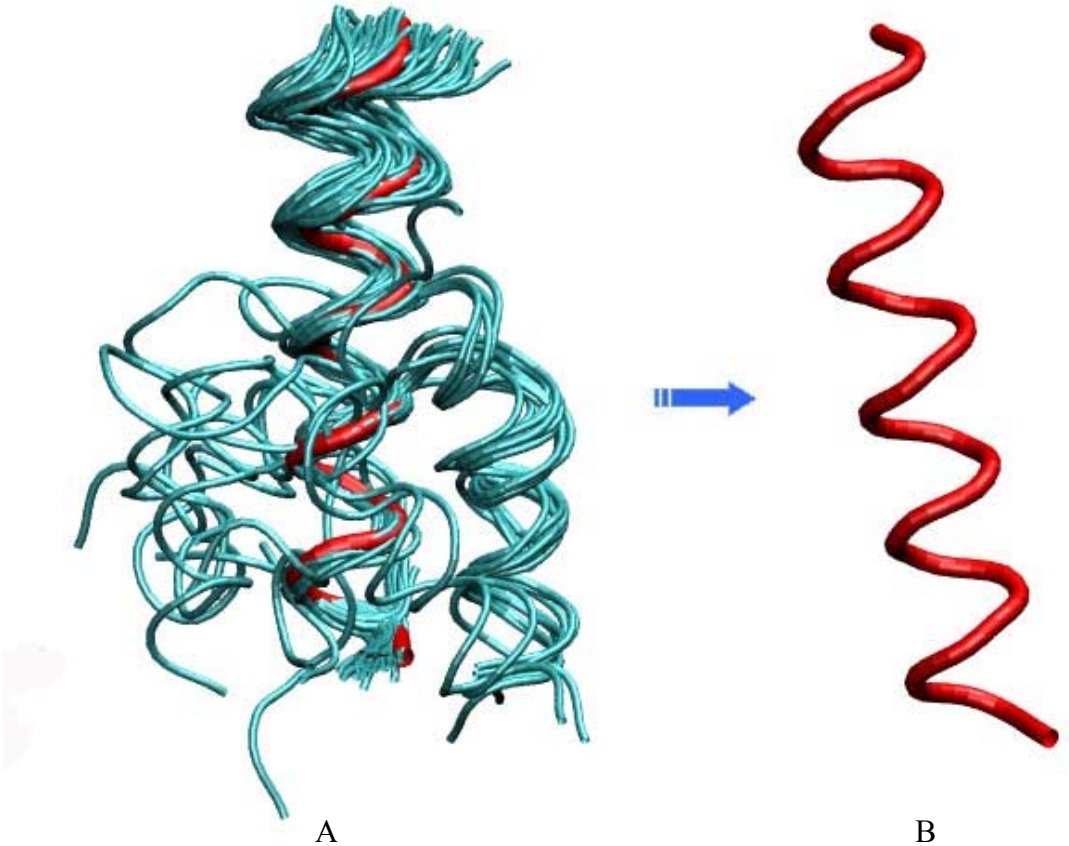
unfolded (U). Structures are defined as folded (F) (α-helical) if their C^{α}-RMSD is within 2.0Å of the reference structure, and labeled as unfolded (U) otherwise. Based on this clear structural separation, the process is treated as a two-state equation between the folded and unfolded ensembles.

$$\text{Unfolded (U)} \underset{k_U}{\overset{k_F}{\rightleftharpoons}} \text{Folded (F)} \qquad (3\text{-}1)$$

The folding rate equations for each species are written as,

$$\frac{d[F]}{dt} = k_F[U] - k_U[F] \qquad (3\text{-}2)$$

$$\frac{d[U]}{dt} = k_U[F] - k_F[U] \qquad (3\text{-}3)$$



A B

Figure 3-2. Cluster analysis of polyalanine. A) All conformations from REMD at 151K are superimposed on the reference structure of the largest cluster, B) α-helix reference structure from the cluster analysis.

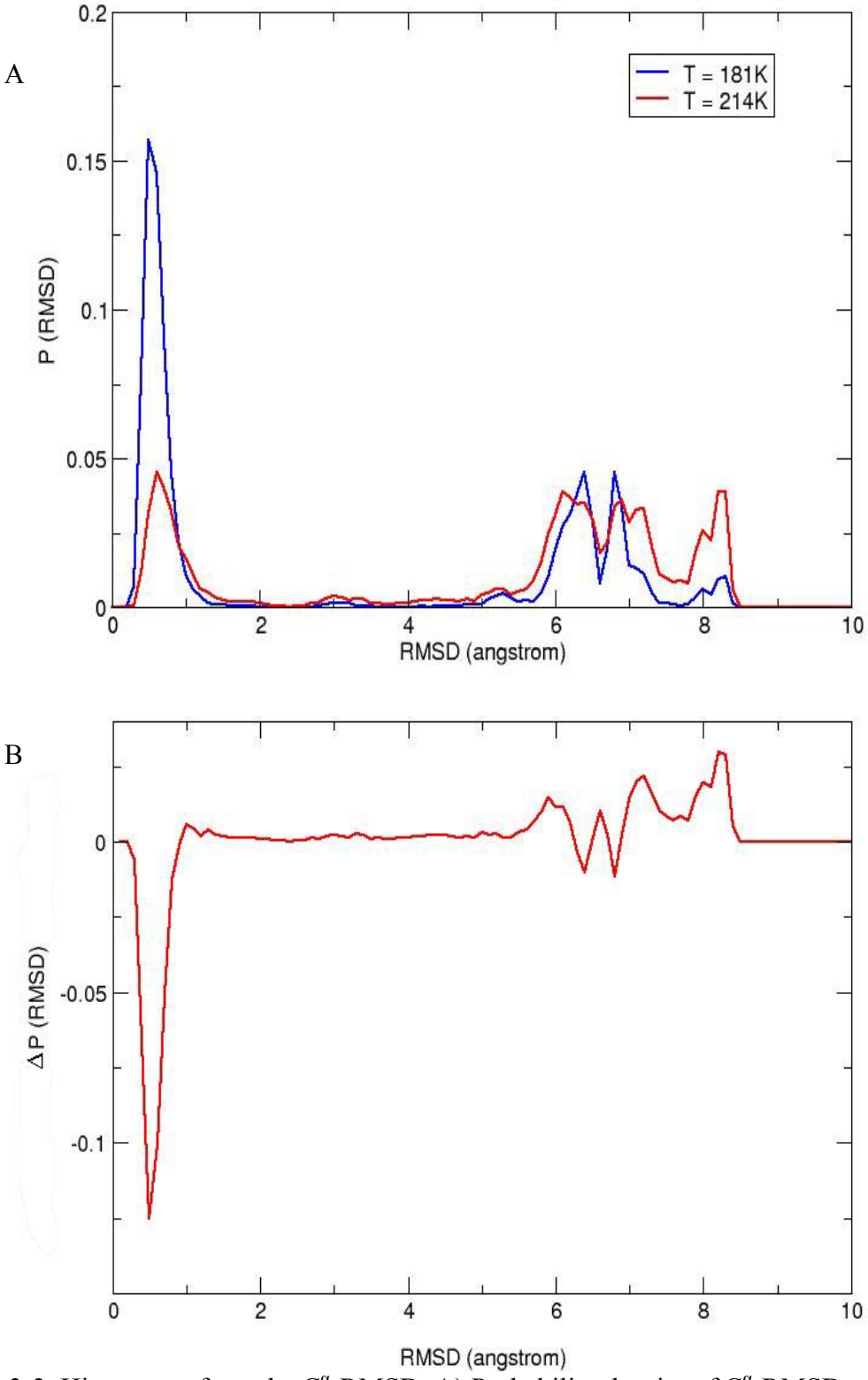


Figure 3-3. Histograms from the C^{α}-RMSD. A) Probability density of C^{α}-RMSD at 181K (blue line) and 214K (red line), B) Δ(probability at 214K - probability at 181K) vs. C^{α}-RMSD.

In the theoretical work concentrations were replaced by populations, which is equivalent of having a total initial concentration set to unity. This then makes (U) = 1 – (F). The time dependent populations can then be directly solved leading to:[29]

$$F(t) = C_1 e^{-\lambda t} + C_2 \quad (3\text{-}4)$$

$$U(t) = 1 - F(t) \quad (3\text{-}5)$$

The population relaxation is dominated by a single rate constant, $\lambda = k_U + k_F = 1/\tau_{relax} = 1/\tau_U + 1/\tau_F$. In this equation, τ_F is the folding time (and τ_U is the unfolding time). The values of the two constants C_1 and C_2 are completely determined from the initial and final concentrations and are then not adjustable parameters. By recalling that the equilibrium constant K can be written as a ratio of the forward and backward rate constants, the two can be then separately determined. The population of two F and U ensembles versus temperature is shown in Figure 3-4. They were computed from the C^{α}-RMSD histograms (referenced to the F structure) using a 2.0Å cutoff. At low temperature the system shows mostly

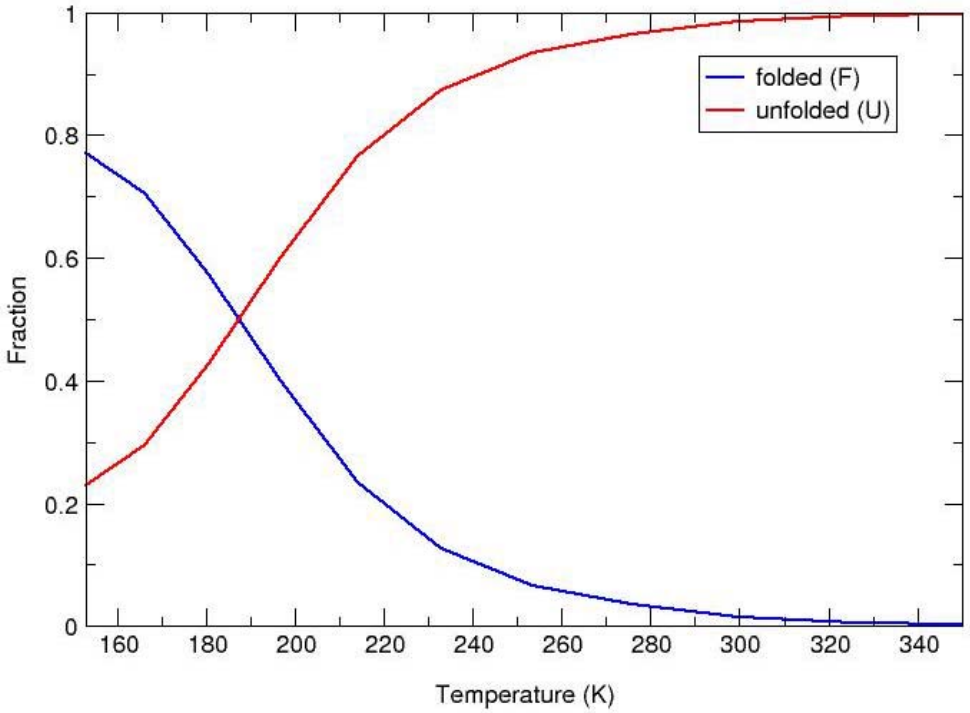


Figure 3-4. Fractions of α-helical folded state (F), and unfolded state (U) as a function of temperature using probability (C^{α}-RMSD).

α-helical folded states but the populations of unfolded states rise as temperature increases providing a reasonable melting curve. A clear melting temperature was found at ≈ 190K. The low melting temperature compared with regular experiments can be assigned directly to issues with the force field and solvation model.[1,94] It does not however subtract from the main point of this chapter which aims to present and test a new method. Based on the melting curve, T-jump simulations were performed with the temperature jump bracketing the melting temperature, from 181K to 214K. By having the complete description of the temperature dependent ensembles we can have complete knowledge of what both the initial t=0 (T=181K) and t=∞ (T=214K) ensembles look like. In Figure 3-3 A, we show the RMSD probability at both initial and final temperatures for the T-jump. In Figure 3-3 B, the probability difference (214K - 181K) validates choices of temperatures, by showing a significant change in (F) and (U). These figures completely determine the values of C_1 and C_2 in Equation 3-4.

3.3.1 Calculation of Chemical Shifts

Calculation of chemical shift was performed on the trajectories obtained from 10-200ns REMD simulations. All the chemical shifts ($^1H_\alpha$, $^1H_\beta$, ^1H (along the N–H bond), $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$ (along the carbonyl bond), ^{15}N) were computed using SHIFTS program (section 3.2.2). The spectra measured at 16 different temperatures were then averaged for each residue (total 1,900 structures). The averaged chemical shifts are plotted as a function of residue number (Figure 3-5). As shown in Figure 3-5, the highest temperature is placed at the bottom and all the averaged chemical shifts decrease with increasing temperature except $^{13}C_\beta$ and $^1H_\alpha$ chemical shifts. We can expect that our prediction of $^{13}C_\beta$ and $^1H_\alpha$ chemical shifts compared with experimental data (Table 3-1) shows prediction errors from side chain orientation.

Table 3-1 presents chemical shifts values obtained using REMD simulations at the lowest (153K) and the highest temperature (514K) for a residue 6 (A6) in polyalanine peptide. For

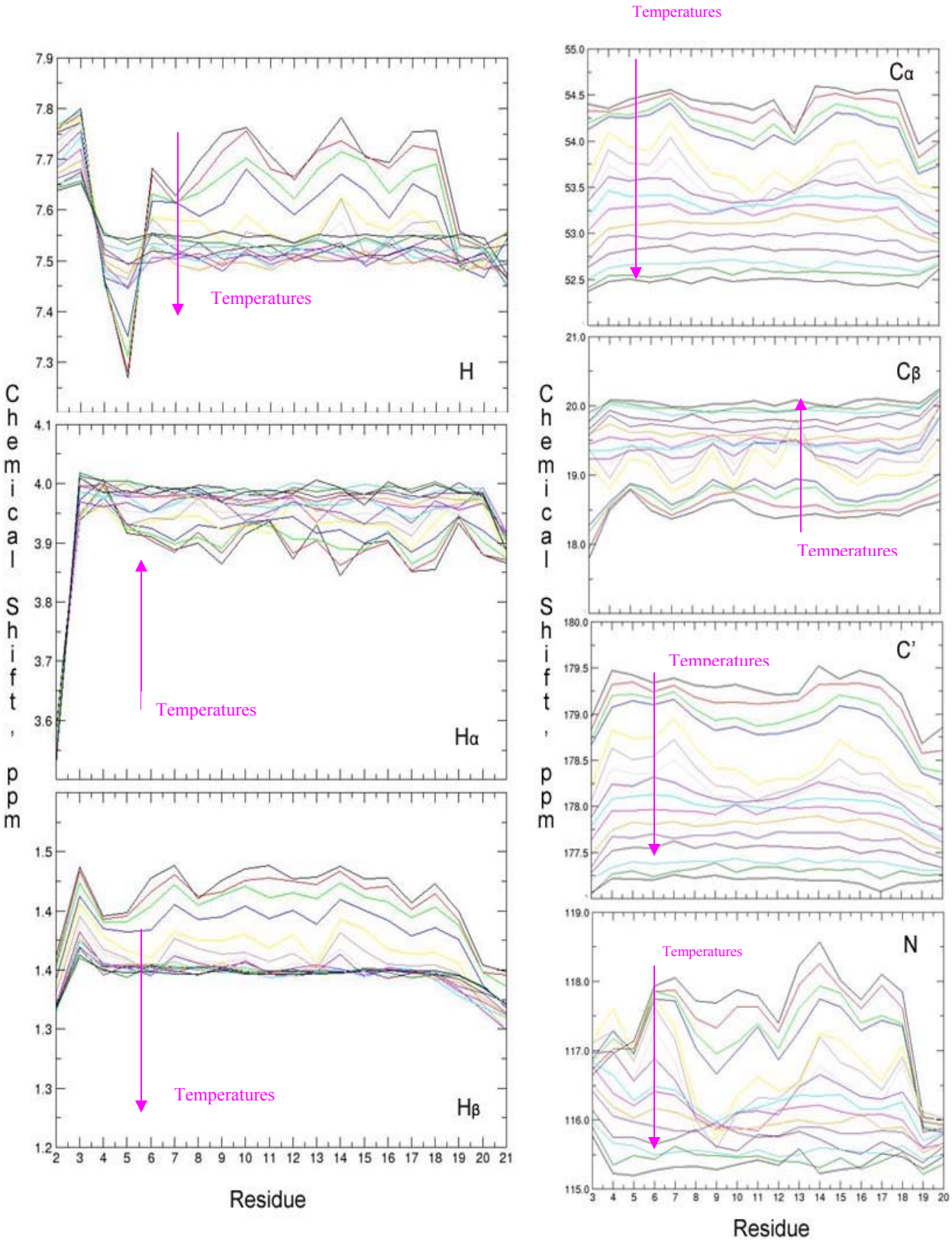


Figure 3-5. Averaged chemical shifts (proton, ^{13}C, and ^{15}N) for polyalanine peptide (ACE–$(ALA)_{20}$–NME) as a function of residue number, for all sixteen temperatures.

Table 3-1. Comparison of calculated chemical shifts for residue 6 (A6) in polyalanine peptide and experimental values for alanine residue from references [42,44,111]*

	Helix[a]	153K[b]	Random coil[c]	514K[d]	$\delta_f - \delta_{RC}$[e]	$\delta_l - \delta_h$[f]
$^{13}C_\alpha$	54.7	54.51	52.4	52.47	2.3	2.04
$^{13}C'$	179.6	179.34	177.6	177.20	2.0	2.14
$^{13}C_\beta$	19.74	18.48	19.26	20.06	0.48	-1.58
^{15}N		117.94	122.5	115.28		2.66
$\alpha-^1H$		3.91	4.33	3.98		-0.07
$N-^1H$		7.68	8.15	7.55		0.13

* All data are given in ppm.
[a, c] experimentally measured chemical shift values from references.
[b, d] computationally measured average chemical shift values from 10-200ns REMD simulations.
[e] chemical shift deviations, δ_f (helix) – δ_{RC} (random coil), from references.
[f] chemical shift deviations, δ_l (low temperature) – δ_h (high temperature), from 10-200ns REMD simulations.

comparison, it also provides the chemical shift values of both helix and random coil for a alanine residue, measured experimentally by Wishart *et al.* [42,111] and Spera *et al.*[44] As mentioned in the previous section (cluster analysis), the conformation at the lowest temperature (153K) is mainly well-defined α-helix (76% populated). The calculated chemical shifts (^{13}C) at 153K (Table 3-1) agree well with the helix chemical shifts values from the experiment. In contrast, the calculated values at 514K are comparable with the random coil values from the experiment. The calculation results make sense, disordered conformations of the peptide (random coil) are expected at the higher temperature and more structured conformations (folded or packed conformation) at the lower temperature. The results also indicate that with increasing temperature, there is an increase of the random coil population (unfolded state) accompanied by a decrease of α-helical structure (folded state (F)). Therefore, one can detect formation of the secondary structure by monitoring a change in the chemical shift (Δδ).

The calculated $^{13}C_\alpha$, and $^{13}C'$ chemical shift are in good agreement with experimental results. However, the calculated $^{13}C_\beta$ chemical shift differs from experimental one, due to errors associated with fitting of reference shifts in the database of the SHIFTS program.[112,121,122]

Wishart *et al.* commented that error in ^{13}C and ^{15}N chemical shift data is relatively larger than that in the proton chemical shift.[42]

The 2D spectrum, for example HSQC, can usually provide more complete structural details, compared to 1D one. Figure 3-6 shows ^{1}H–^{15}N HSQC spectrum of a polyalanine peptide at 181K and 300K. At higher temperature (300K, Figure 3-6 B), the spectrum is poorly resolved with slow fluctuation, due to the disordered state. This indicates unfolded states are dominated. On the contrary, the spectrum at lower temperature (181K, Figure 3-6 A) is well resolved and dispersed, as a result of folded states. Figure 3-6 included only the cross peaks providing the structural information. However, the kinetic analysis can be performed with additional intensity information.

The ^{13}C chemical shift for C_α and C_β has been often used to determine backbone conformation of proteins and peptides, as found in many of previous NMR studies, such as empirical [44,45] and *ab initio* [46] methods. This is because the C_α and C_β chemical shifts are mostly determined by the backbone φ and ψ torsional angle.[41] In this studies, $^{13}C_\alpha/^{13}C_\beta$ crosspeaks in the folded state are relatively sharp (2-4 ppm full width), narrow, and well resolved. The crosspeaks in the unfolded state are, however, broad, spreading toward upright. Figure 3-7 shows two-dimensional calculated $^{13}C_\alpha/^{13}C_\beta$ crosspeaks for polyalanine peptides (residue 5 (A5), 11 (A11), and 16 (A16)) obtained at 153, 181, 214, and 300K. The spectra measured at different temperatures were then averaged over the ensemble (total 1,900 structures). Three labeled residues are all similar in the general trend on temperature. The crosspeak at the lowest temperature (153K) is relatively ordered, sharp, and narrow, indicating significantly folded (or helical) conformation. As temperature increases, the crosspeak profile becomes broader toward upright, showing that conformational distribution is temperature dependent.

These results are in accord with the HSQC results that covered in Figure 3-6.

3.3.2 Calculation of Circular Dichroism

For each member of the ensemble and at each temperature in the REMD simulation, CD spectra were computed using the technique described in the methods section. The spectra for

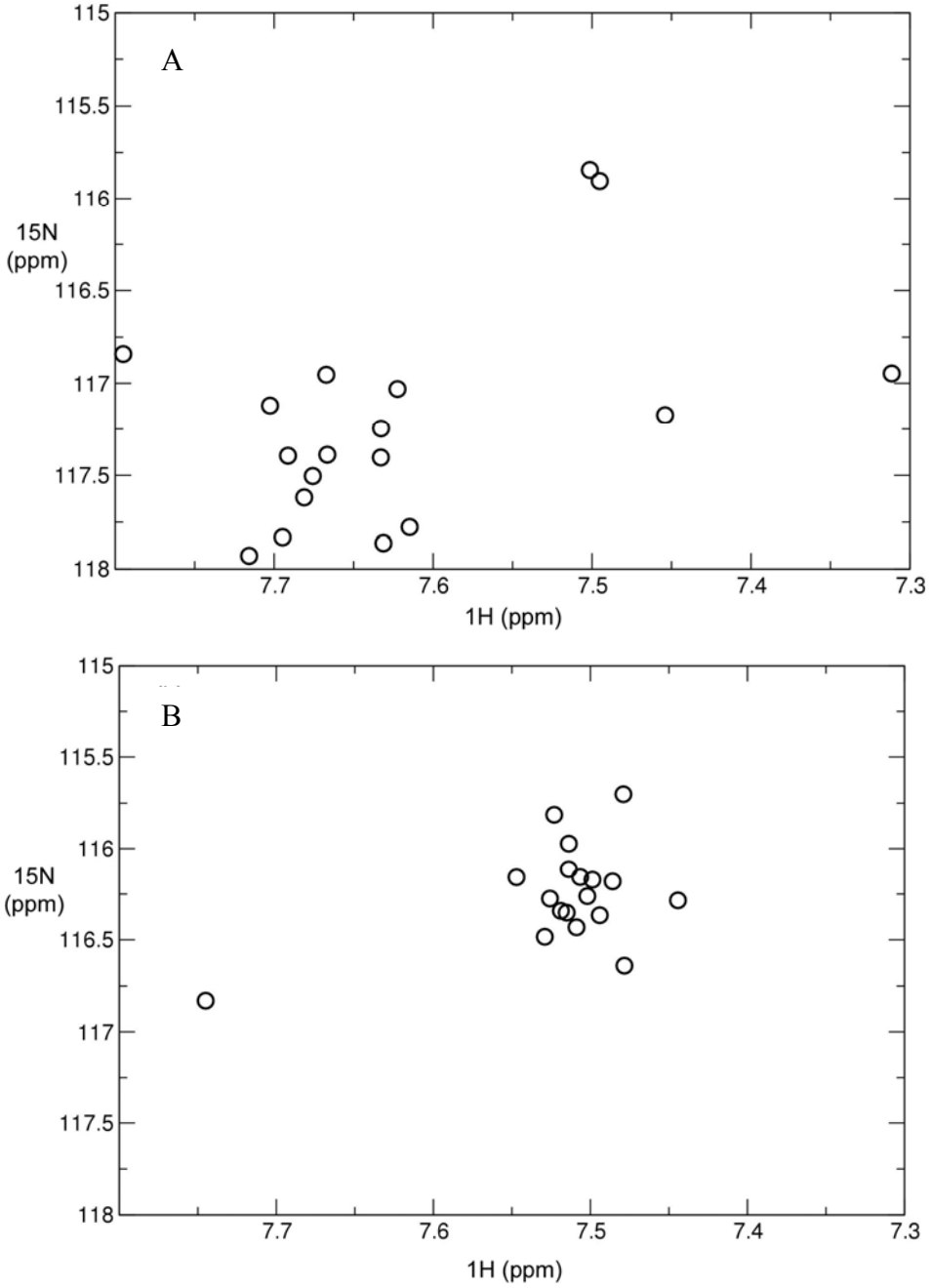


Figure 3-6. $^{1}H–^{15}N$ HSQC spectra of polyalanine peptide. A) At 181K and B) At 300K.

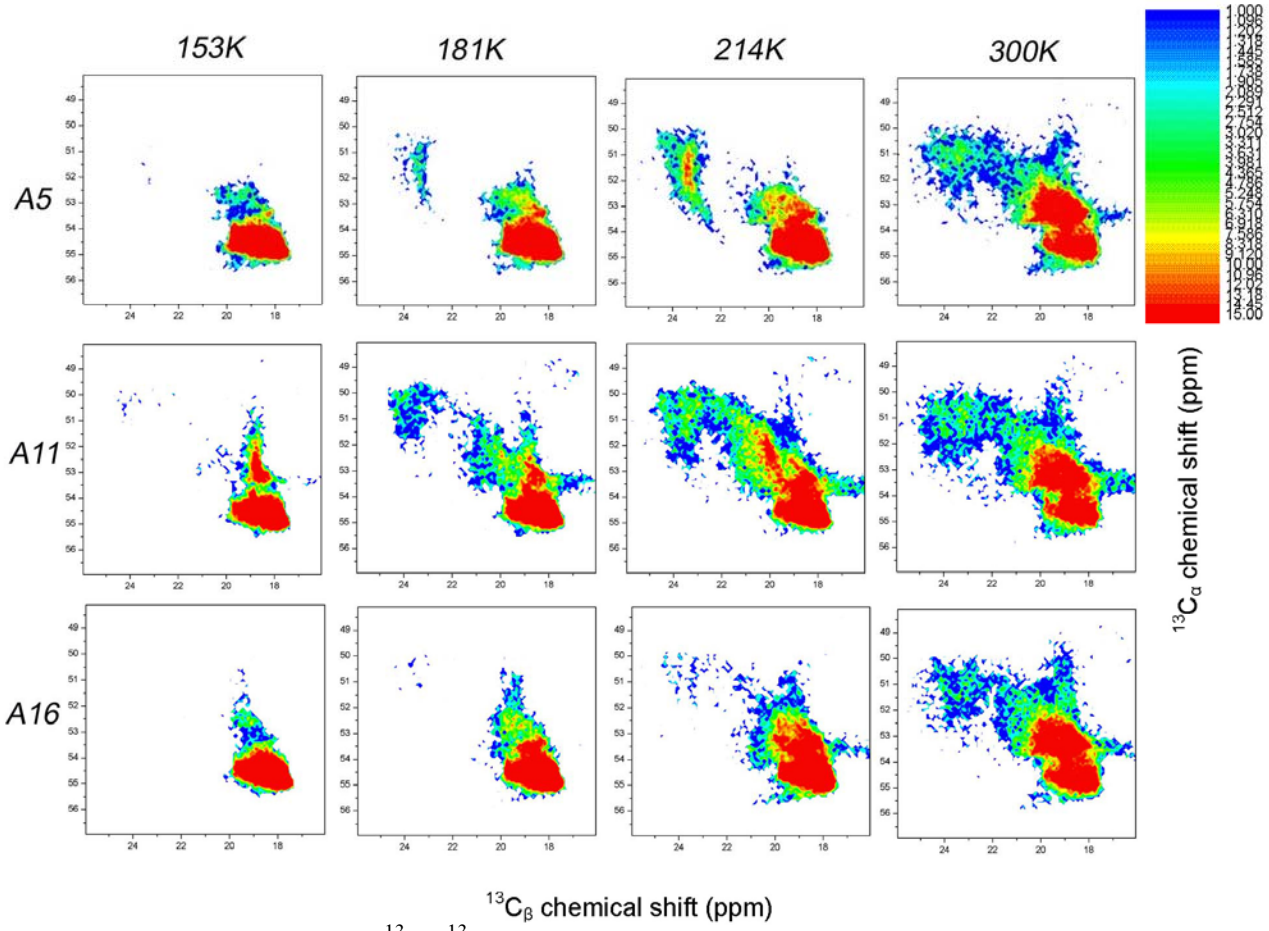


Figure 3-7. Two-dimensional $^{13}C_\alpha/^{13}C_\beta$ crosspeak shapes for polyalanine peptide, residue 5, 11 and 16 (A5, A11, and A16) from several temperatures. Color scheme are from the maximum signal intensity (red) to the minimum signal intensity (blue).

each temperature were then averaged over members of the ensemble (total 1,900 structures). The data is shown in Figure 3-8 A for the average spectra versus temperature. In agreement with the RMSD data, the lowest temperature spectrum resembles that of an α-helix, while as temperature increases, the system becomes disordered. The folded state has two minima in the CD spectrum, one at 208nm (π-π* amide transition) and another at 222nm (n-π* transition along the carbonyl bond). Experimental CD spectra for an α-helix shows a deeper minimum for 222nm than for 208nm.[162] The calculated local minimum at 208nm does not agree with the experimental results, indicating those transitions to be very sensitive. 100% α-helical poly-alanines, ACE–$(Ala)_{20}$–NME, were built using TINKER molecular dynamics program [163,164] and 200 steps of

minimizations were performed to investigate the difference. Minimum wavelengths are compared between two CD spectra (one is before minimization, i.e. pure 100% α-helix, and the other is after minimization) in Figure 3-9. The 100% α-helix has a minimum at 222nm while the structure after minimization has a minimum at 208nm.

This discrepancy is assigned to problems with the CD calculations parameters, and been in the process of investigating this. Figure 3-8 B presents the average value of CD at 222nm ($<CD_{222}>$) versus temperature. This is the type of signal one can follow when performing the T-jump experiment.

T-jump simulations were performed and the resulting time traces were fitted to calculate the folding and unfolding rates with initial and final temperatures determined by referring to the melting curve in Figure 3-4. 362 structures from the ensemble at 181K were used as starting points for molecular dynamic runs at 214K, hence simulating a T-jump. The temperature reaches the new value in a time scale of the order of the thermostat coupling ($0.1ps^{-1}$ time constant used) and should be thought as having a very short dead-time. These simulations were run for a total of 35ns each (total of 12.67μs). At each time slice, a CD spectrum was computed for each independent MD run, and the average performed. The folded and unfolded states were added and then $<CD_{222}>$ versus time is plotted for each state in Figure 3-10. The folded and unfolded states were defined in Figure 3-4 using C^{α}-RMSD probability. When fitting this kinetic trace, the power of this method can be clearly seen. This trace is bounded by the fact that we know, from the equilibrium runs, what the system looks like at t=0 and at t=∞. This leaves a single parameter to be determined via fitting, assuming a two-state system. The calculated relaxation data are well fitted using a single exponential equation, as shown in Figure 3-10, indicating that the suggested reversible two-state kinetic mechanism (Equation 3-1) during the cluster analysis is valid. From

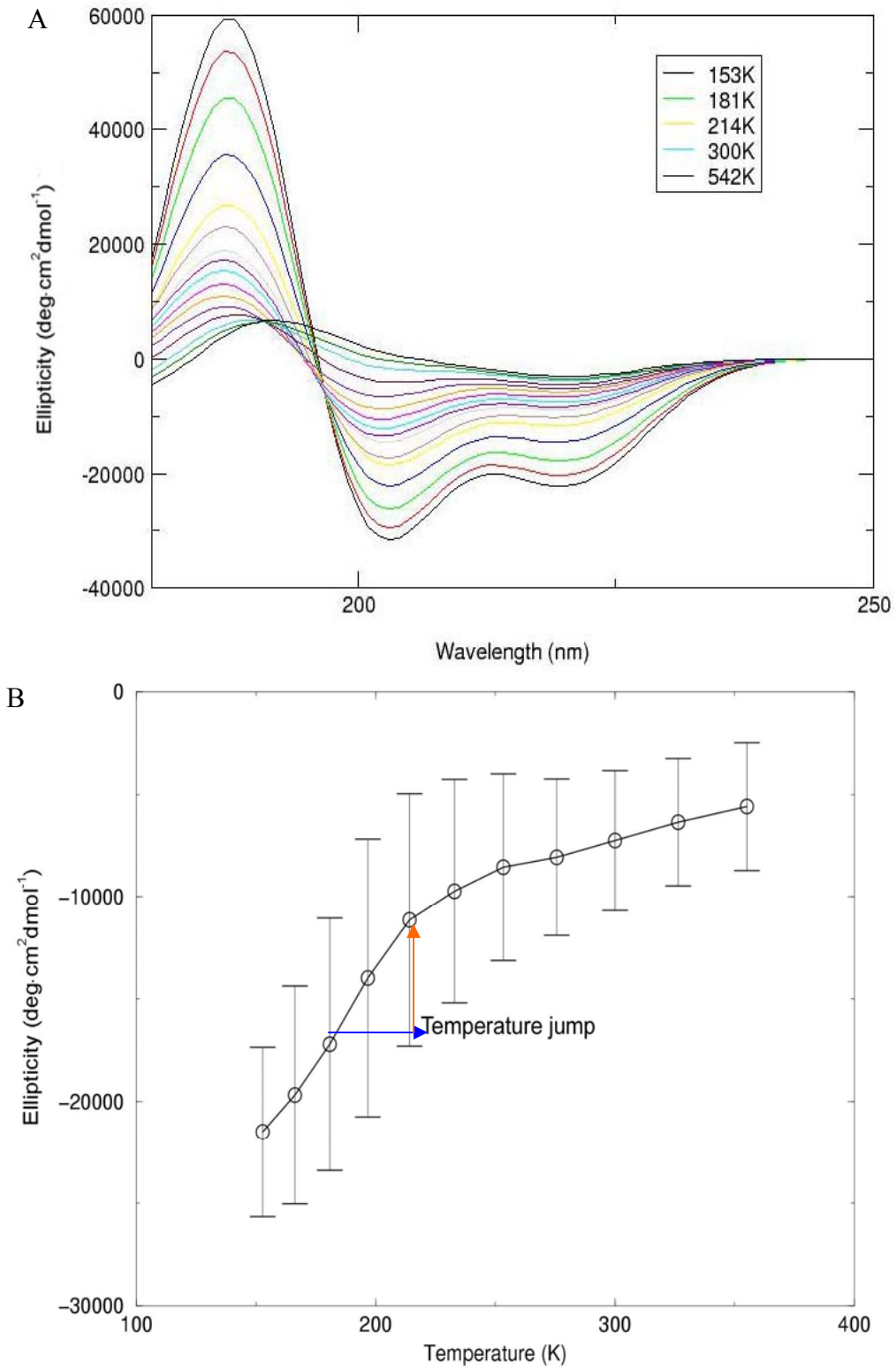


Figure 3-8. Results of Circular dichroism. A) Calculated circular dichroism spectra of ACE–$(ALA)_{20}$–NME from 153K to 542K, B) Average of CD_{222} ($<CD_{222}>$) of each simulated polyalanines. Error bars represent standard deviation.

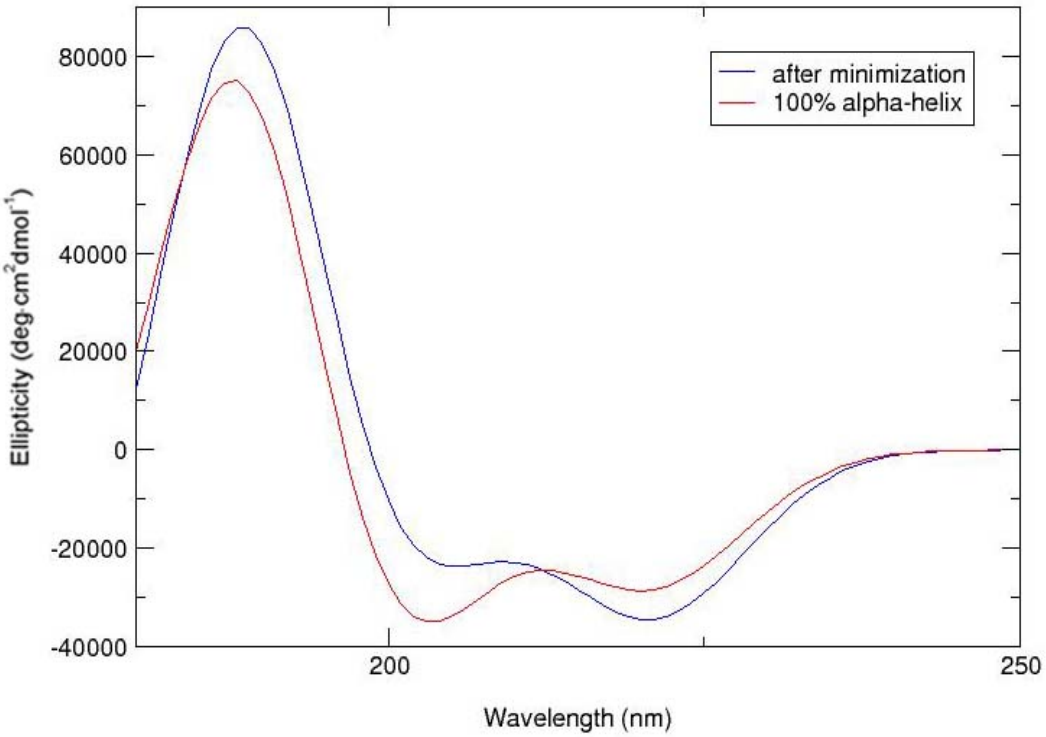


Figure 3-9. Two circular dichroism (CD) spectra, pure α-helix and α-helix after minimization, are compared to see different minimum wavelengths.

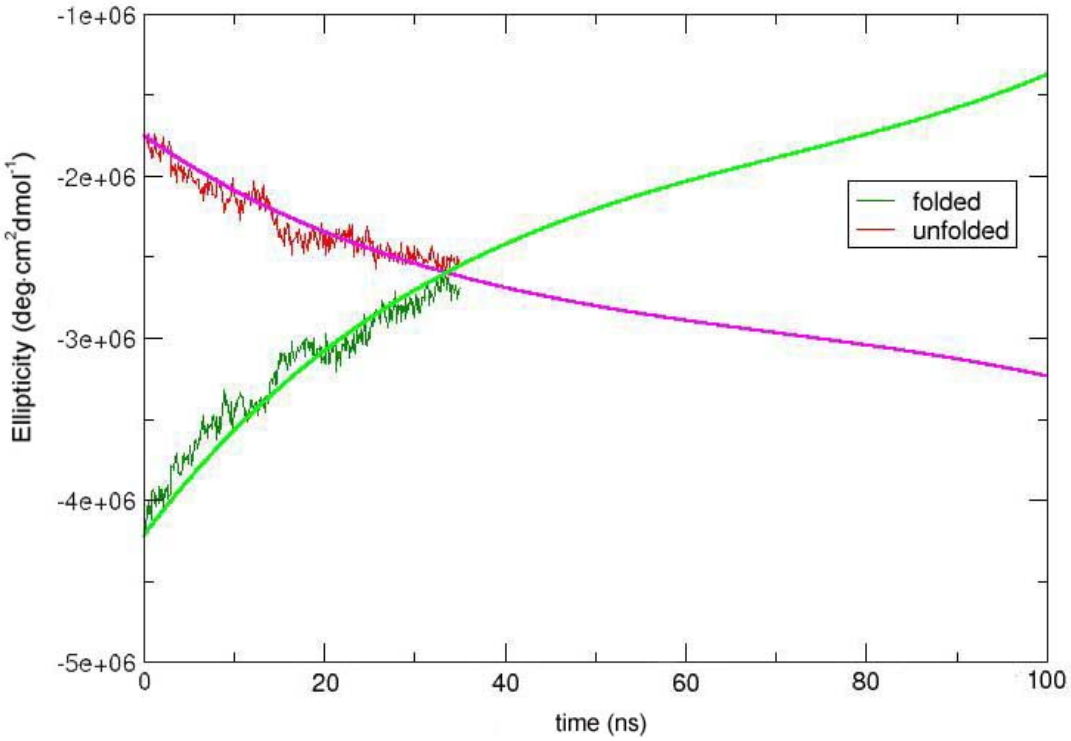


Figure 3-10. Average of CD_{222} ($<CD_{222}>$) of T-jump simulation data (red: folded (F), and green: unfolded (U) states) are fitted using a single exponential equation.

this fit the value of $\lambda = (k_U + k_F)$ in Equation 3-4 is determined to be $\frac{1}{36339}\ ps^{-1}$ and τ_{relax}=36.34ns. This relaxation time was used from CD to predict the time evolution of the F and U populations. Figure 3-11 has the simulated average populations versus time, and the populations predicted from Equation 3-4 also. The curves in Figure 3-11 are not fitted, but are instead simulated using the fit from Figure 3-10. The agreement between the simulated and raw population data is excellent. Using our knowledge of the limiting populations at the initial and final temperatures, equilibrium constants can be obtained. We can then separately determine the folding and unfolding rate constants at the final temperature.

The computed values are k_F=2.11 × 10-5ps^{-1}, and k_U=6.44 × 10-6ps^{-1}, corresponding to a folding time τ_F=47.5ns and an unfolding time τ_U=155ns. Williams *et al.* have simulated

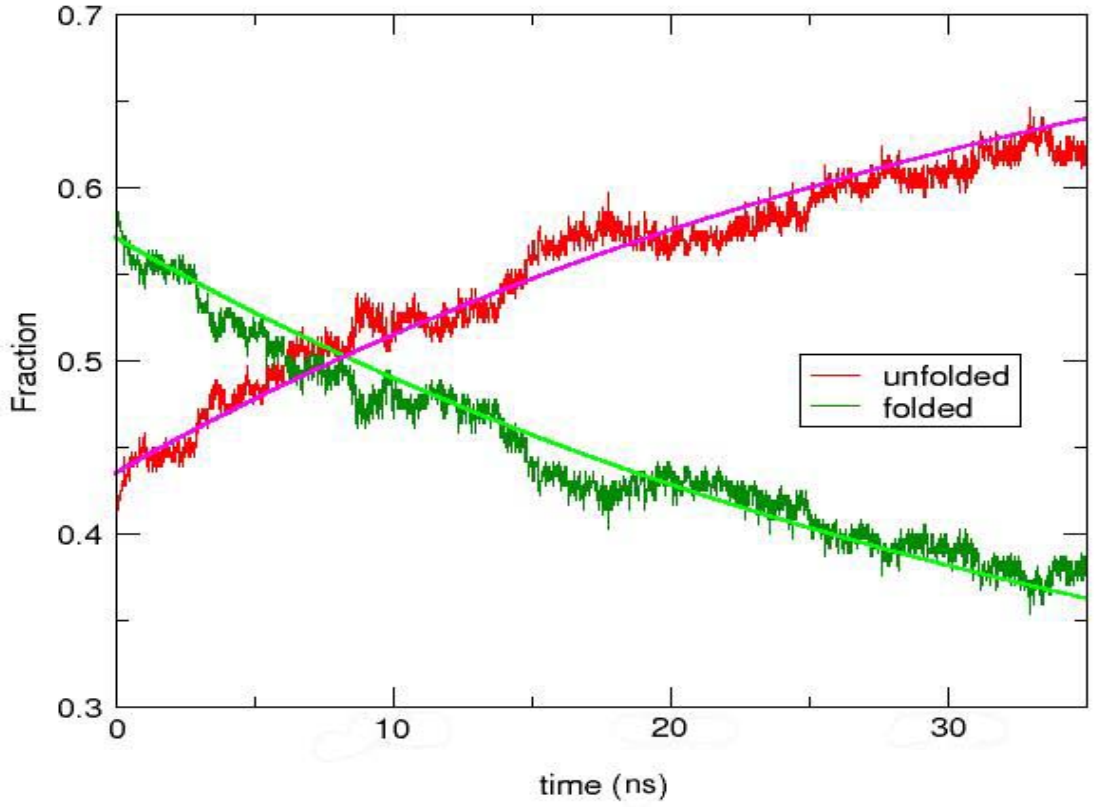


Figure 3-11. Folding (green) and unfolding (red) fractions from T-jump simulation data are calculated and fitted using same λ in Figure 3-10.

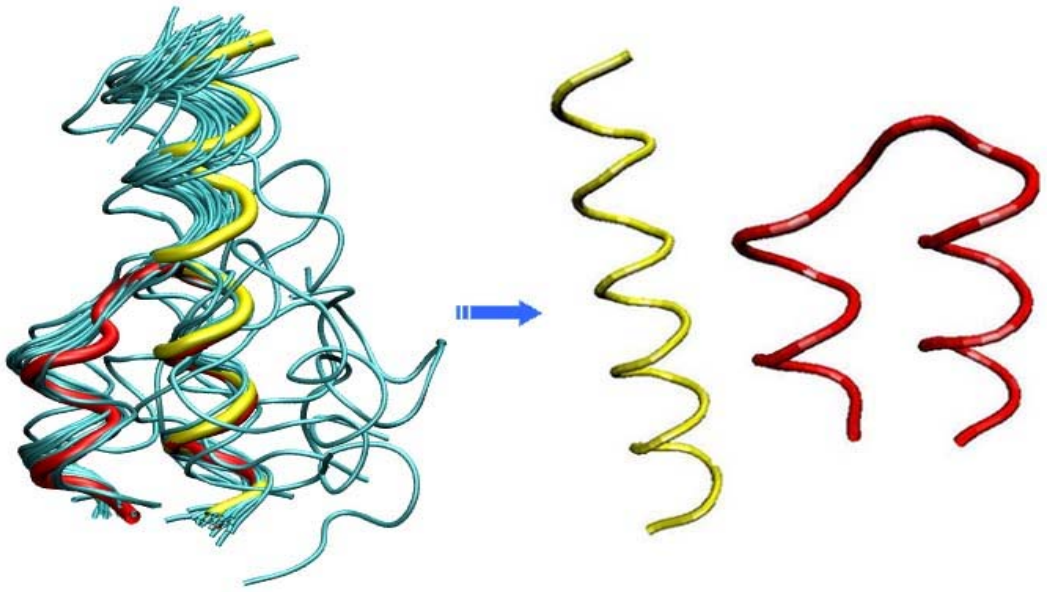
21-residue alanine-based peptide (Fs21 peptide, $–A_5–(A_3RA)_3A–$) and estimated folding time, 16 to 180ns using T-jump experiment from 9.3 to 27.4°C.[19]

In the present model we assume that the implicit solvent is a correct representation of structure and thermodynamics. However, it does not properly represent the friction of water solvent.[95] The time scales seen in this chapter are then much faster than they would be otherwise. The following chapter describes our continuing line of work on systems where experimental data is available and using Langevin dynamics (which incorporates friction).

3.3.3 Optimum Number of Unfolded States?

Two clusters were found in the cluster analysis earlier. The most populated cluster is α-helix (Figure 3-2 B) that occupies roughly 76% of the population. It is defined as a folded state in the two-state mechanism (α-helix and unfolded state). The second largest cluster is 9% populated coiled-coil α-helix (the structure is shown in Figure 3-12). If the coiled-coil α-helix is added as the second folded state, three-state kinetic mechanism (α-helix, coiled-coil α-helix, and unfolded state) can be suggested.

Three different states (α-helix, coiled-coil α-helix and unfolded states) are chosen for initial MD simulation run for 1μs at 214K in order to check how many different states exist. The simulation time (1μs) is chosen because it is estimated to be close to the folding time of small proteins (11-21 residues peptides) as the lower limit.[165] This simulation possibly shows all the intermediates involved in the protein folding process.[14] The C^{α}-RMSD (residues 2 to 19) was computed from the 1 μs trajectory with respect to α-helical, coiled-coil reference structures (Figure 3-13). These RMSD plot reflects the number of different states. By comparing α-helix reference (Figures 3-13 A, B, and C) and coiled-coil α-helix reference structure (Figures 3-13 D, E, and F), simulated structures are classified as follows: α-helix when C^{α}-RMSD1 < 2.0Å and C^{α}-RMSD2 > 5.0Å and coiled-coil α-helix when C^{α}-RMSD1 > 5.0Å and C^{α}-RMSD2



A B C

Figure 3-12. Reference structures from cluster analysis. A) All conformations from REMD at 151K are superimposed on the two reference structures, B) α-helix and C) Coiled-coil α-helix reference structures.

< 2.0Å, where C^{α}-RMSD1 was calculated from α-helix reference and C^{α}-RMSD2 from coiled-coil α-helix reference structure. Interestingly, several transitions, from coiled-coil α-helix to α-helix folded state based in the rmsd range (0-2, 2-3, 3-5, and above 5Å), are found in Figure 3-13 E. The fraction of four states was obtained by applying both C^{α}-RMSD1 and C^{α}-RMSD2, as shown in Figure 3-14. In order to show relationship between two folded states, C^{α}-RMSD1 as an x-axis is plotted against C^{α}-RMSD2 as an y-axis C^{α}-RMSD (Figure 3-15), where four different states are also found. Based on the results, the following kinetic mechanism can be suggested.

$$\text{α-helix} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{Unfolded } (U_1) \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \text{Unfolded } (U_2) \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \text{Coiled-coil } (CC) \quad (3\text{-}6)$$

In Equation 3-6, two different unfolded states are separated by two representative structures (U_1 and U_2). This implies that more representative structures can be defined without limit. Based on the present experiments, the number of unfolded states is counted, introducing the T-jump methodology. Optimal folding mechanism was not discussed in this chapter and should be dealt in the near future.

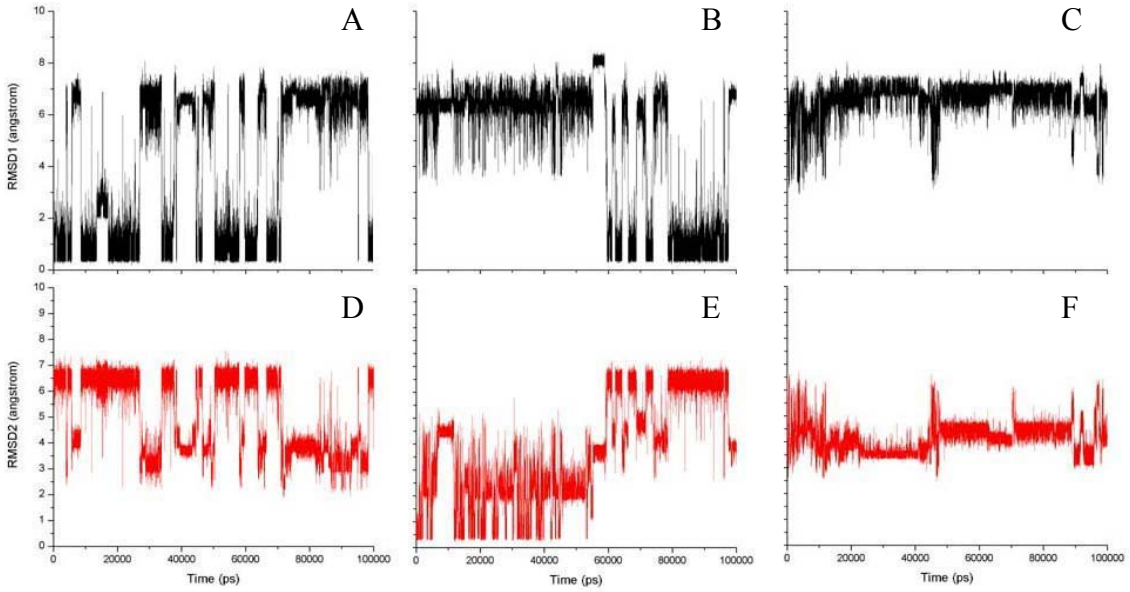


Figure 3-13. C^{α}-RMSD (residue 2-19) computed using different initial states from 1μs MD simulations. A) α-helix, B) Coiled-coil, and C) Unfolded initial states from C^{α}-RMSD1. D) α-helix, E) Coiled-coil, and F) Unfolded initial states from C^{α}-RMSD2.

3.4 Conclusion

This chapter presents the first attempt to using simulation conditions and predicting observables as close as possible to experiment. As far as we know, this is the first time a simulation of T-jump is used to calculate the folding rate constants.

By using a mix of very efficient sampling techniques (REMD) to properly populate ensembles and regular MD simulations to study the non-equilibrium relaxation of the populations, extracting kinetic and thermodynamic data are enabled. A set of folding/unfolding time (τ_F=47.5ns and τ_U =155ns) was calculated by data fitting of calculated CD spectra.

The chemical shifts were also calculated and it showed systematic change in conformational distributions depending on the temperatures. Finally, the number of folded and unfolded states was determined and discussed, based on 1μs long MD simulations.

In the future, implicit and explicit solvent models need to be compared and influence of friction remains to be determined. As an extension of the present work, the following chapter

describes results obtained on real proteins using Langevin dynamics which considers frictional effect.

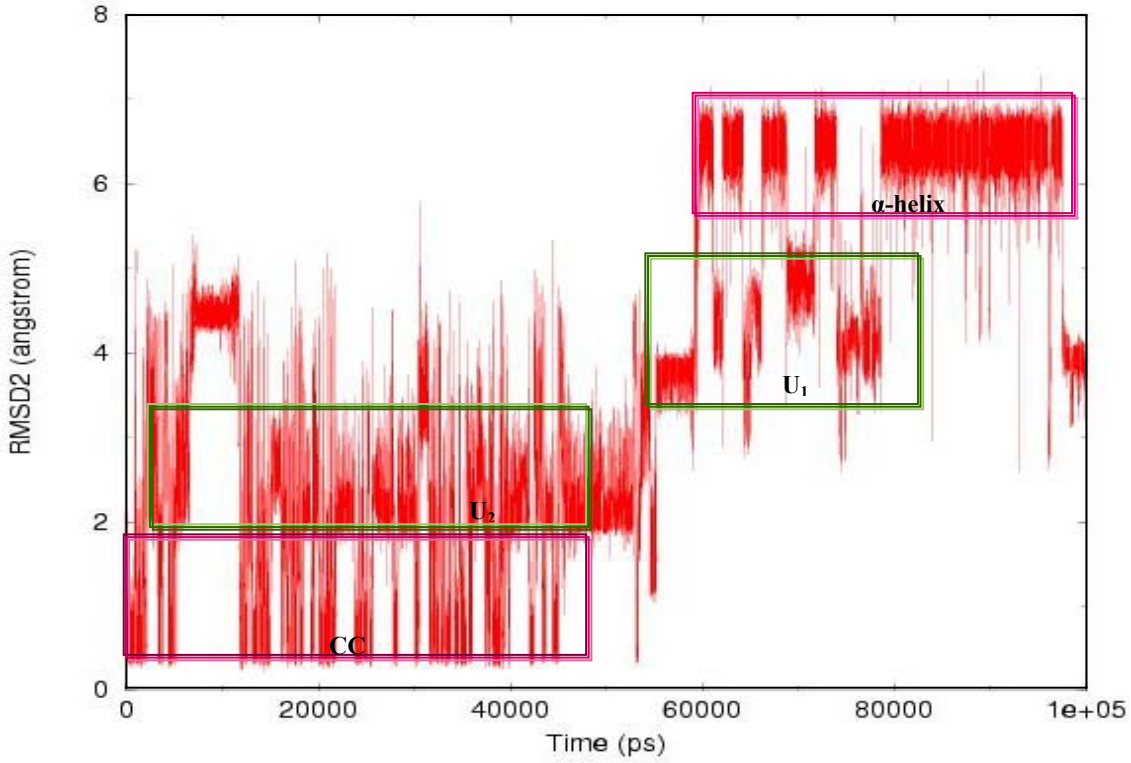


Figure 3-14. C^{α}-RMSD2 (with respect to coiled-coil α-helix) plots from 1μs MD simulation using coiled-coil α-helix initial state at 214K.

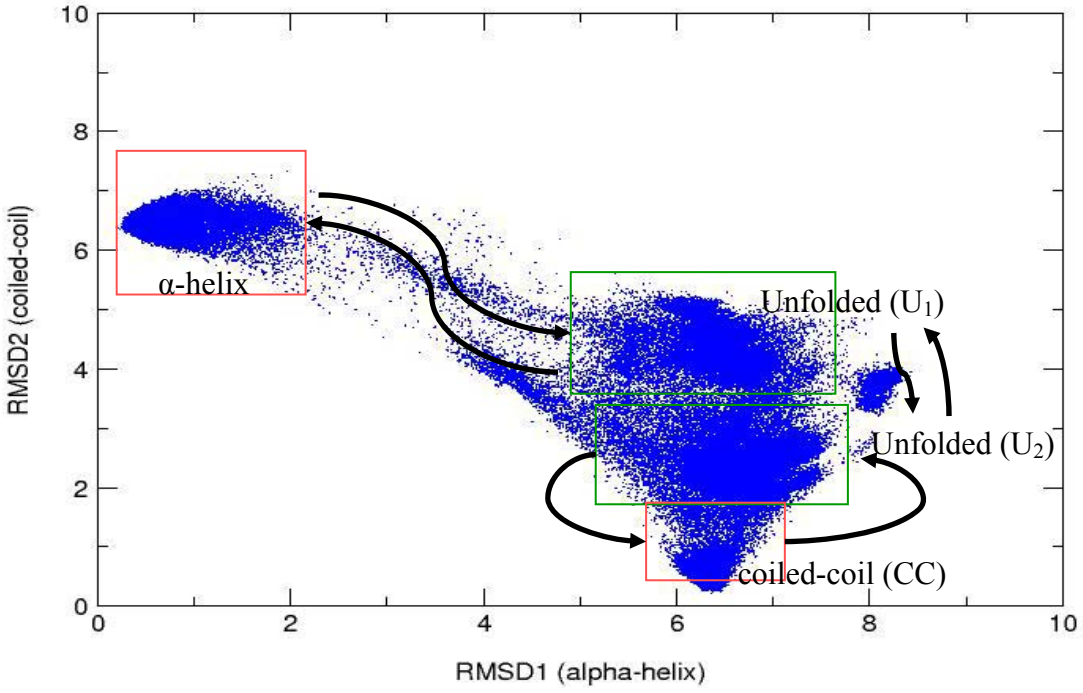


Figure 3-15. C^{α}-RMSD (residue 2-19) relation plot using C^{α}-RMSD1 (with respect to α-helix reference structure) as x-axis and C^{α}-RMSD2 (with respect to coiled-coil α-helix reference structure) as y-axis, respectively, at 214K.

CHAPTER 4
FOLDING KINETICS BY TEMPERATURE-JUMP SIMULATIONS OF TWO RELATED 14-RESIDUE PEPTIDES

4.1 Introduction

Protein folding is described as a process of an ensemble of molecules reaching their biologically active three dimensional structures from a linear chain of amino acids.[99] Both experimental and theoretical studies for over 40 years have converged to show and predict the basic principle and particular mechanism of folding and unfolding of proteins.[146] Their mechanism and kinetics are still unknown or incompletely understood and therefore, are being actively investigated in the molecular biology area. Particularly, simulations can describe the folding process microscopically and its atomic details which are unavailable in experiments.

However, conventional molecular dynamics is unable to yield complete conformational space sampling due to its high energy barriers and deep local minima in most of the systems. Thus, the development of efficient algorithms becomes a point at issue. One of the oldest methods is umbrella sampling technique that is combined with molecular dynamics simulations. [166] Brooks and coworkers [167] utilized this method extensively to understand thermodynamics and kinetics of folding for biomolecular (or biological) systems. It was very powerful for the investigation of the folding free energy landscape in several chosen reaction coordinates. However, it was limited to small proteins (or peptides),[168] and was not capable of determining the biasing potential.

Replica exchange molecular dynamics (REMD),[64] which adapts a random walk in potential energy space, was introduced to overcome those problems. This technique shows not only enhanced sampling but also increased speed of equilibration by treating the temperature as a control parameter.

Direct experimental measurement of folding rates becomes possible by using the temperature jump (T-jump) experiments in the nanosecond to microsecond time scale. The T-jump rapidly changes the position of the equilibrium between folded and unfolded states by the temperature change, and hence, reaction kinetics to the new equilibrium can be monitored [169]. Moreover, the laser-induced T-jump technique was extended to far-infrared absorption (5.88 - 6.67 μm), which can observe protein secondary structure.[29] Several groups applied this method to helix-coil transition [19,170] and folding transition of hairpin structure.[171,172]

Combination of computational and experimental methods in studying protein folding is advantageous.[25] For example, Folding@Home typically runs a very large number of initial conformations for a few nanoseconds, either fully extended or taken from a high temperature run. Under the assumption of a single exponential decay that is based on two-state folding kinetics, a folding time of 10 μs is calculated when 50 out of 10,000 trajectories have folded for a 50ns simulation. In order to verify the dynamics, the simulations were compared with T-jump spectroscopic results.[25,173]

However, several groups pointed out that two-state model might not be applicable under certain conditions. Daggett and Fersht [147] emphasize the problem in terms of intermediate states that always exist between a denatured and native state. Sabelko *et al.* found non-exponential folding kinetics of two proteins (yeast phosphoglycerate kinase (PGK) and a ubiquitin mutant) from a nanosecond T-jump.[173] F. Gai and coworkers also reported non-exponential behavior for helix-coil transition kinetics.[130] Therefore, the most important issue at this point is how we perform our simulation in the conditions close to real experiments without any kinetic assumption, while connecting theory and experiment.

In Chapter 3, we introduced Temperature-jump (T-jump) simulations of the polyalanine

peptides (Ala_{20}) as a proof-of-principle system.[174] Since alanine-20 has high hydrophobicity and insolubility in water,[175] no experimental data have been reported. In this chapter, we apply the same techniques (T-jump simulations) to model two related 14-residue peptides. Wang *et al.* [49] have previously studied the helix-coil kinetics of these two peptides using time-resolved infrared (IR) spectroscopy, coupled with laser-induced T-jump technique. They reported that the one is a general polyalanines derivative of Baldwin-type peptide,[176] and the other contains helix stabilizer and end-capping groups, showing difference in helicity and stability regarding the folding kinetics. In this chapter, we simulated T-jump and calculated folding and unfolding rates of the same peptides that Wang *et al.* [49] applied, using calculated circular dichroism (CD) spectra. The results were then compared with the experimental data. Most importantly, the present work was performed in a very similar way to the experiments, such that they were compared with each other, minimizing assumptions. Initially, REMD simulations were performed to obtain equilibrated ensembles to overcome energy barriers and accelerate the convergence. Snapshots from these ensembles were used as initial structures for T-jump simulations.

The Effect of Frictional Coefficient. To further validate our T-jump methodology, we focus on the influence of friction and random forces–introduced by the solvent–on the protein folding kinetics. Kramers [177] proposed in its simplest form that the reaction rate (k) in the high friction (γ) limit should be proportional to the inverse of the friction of the solvent, $k \propto 1/\gamma$. This inverse dependence of rates of protein folding on viscosity, η (or, equivalently, the friction, γ) has been frequently reported experimentally.[23,95,178,179]

The effect of frictional coefficient through Langevin dynamics was studied for the folding kinetics of two related 14-residue peptides. Different values of friction and random forces were applied and compared with experimental results from Wang *et al.*[49]

4.2 Methods

Alanine-based peptides were originally designed by Marqusee and Baldwin [176] for α-helix formation. Two related 14-residue peptides, Ac-YGAKAAAAKAAAAG-NH_2 (peptide 1), and Ac-YGSPEAAAKAAAA-r-NH_2 (peptide 2, where r represents D-Arg), were derived and tested by F. Gai group.[49] The initial structures for two related 14-residue peptides were constructed in fully extended conformations with the AMBER 9.0 molecular simulation package.[180] The AMBER ff99SB force field [76,152,153] was used for both peptides. Modified generalized Born (GB) implicit solvent model by A. Onufriev, D. Bashford and D. A. Case (GB^{OBC})[181] was applied. The system was initially subjected to 500 steps of minimization, and then equilibrated by using Langevin dynamics with collision frequency γ=1.0ps^{-1} for both peptides to account for frictional effects. The resulting conformations were used as initial seeds for the REMD simulation.

All REMD simulations reported here were carried out using the *multisander* implementation in the AMBER 9.0 simulation package. Sixteen replicas for peptide 1 were simulated for 500ns each (total of 8μs) at exponentially distributed temperatures, from 150K to 726K (150, 167, 185, 206, 228, 254, 282, 313, 348, 386, 429, 477, 529, 588, 653, and 726K). Eighteen replicas for peptide 2 were simulated for 500ns each (total of 9μs) from 150K to 834K (150, 166, 184, 203, 225, 248, 275, 304, 336, 372, 411, 455, 503, 557, 616, 681, 754, and 834K). The average exchange rate between adjacent replicas was 15%. The SHAKE algorithm was used to constrain all the bond lengths involving hydrogen atoms,[157] which allows an integration time step of 2fs for each replica. Replica exchanges were attempted at every 10ps, resulting in 50,000 attempted exchanges at each temperature. Conformations from the simulation of each replica were recorded every 2ps interval. The first 20ns of the simulation were discarded and the latter 480ns were saved for further calculation, giving a total of 48,000 configurations for both peptides.

To determine the secondary structure, the program designed by Kabsch and Sander, Definition of the Secondary Structure of Proteins (DSSP),[182] was used for both peptides. Percentage of α-helix (Helicity) was calculated for every residue and every replica, using this program. We also performed a cluster analysis with moil-view [158] based on backbone RMSD for residues 4-9 for peptide 1 and residues 5-10 for peptide 2 with a cutoff of 2Å.

Conformations obtained from the REMD calculation were used as the initial structures for the T-jump experiment. We selected 1,200 starting configurations collected from the equilibrated simulations for peptide 1 and peptide 2 at 282K and 304K, respectively. They are equispaced in time (200-500ns REMD). Each member of this ensemble was then instantly and independently heated up to the T_{high} (the details of the temperature choices are discussed later) and MD was run for 5ns to simulate the relaxation after the T-jump. For both of the peptides, all MD parameters are the same as those used in REMD.

Several collision frequencies (γ=1.0, 5.0, 10.0, and 20.0ps^{-1}) were used for both peptides to investigate the frictional effect on the rate of the protein folding.

In order to estimate secondary structure and folding properties, computations on CD spectra were performed, in the same manner used by Sreerama and Woody (chapter 3).[105] The matrix method [101] was employed with a transition parameter set consisting of a combination of experimental data and theoretical parameters: experimental data [160] for two amide $\pi\pi^*$ transitions and parameters from intermediate neglect of differential overlap/spectroscopic (INDO/S) wavefunctions [161] for the $n\pi^*$ transition. The rotational strength was computed to generate CD spectra through Gaussian band. The bandwidths assigned for the $n\pi^*$ transition, and for two $\pi\pi^*$ transitions were 10.5nm, 11.3nm, and 7.2nm, respectively.[39]

4.3 Results and Discussion

The two related 14-residue peptides in this study were selected to prove a new methodology (described in the chapter 3). Marqusee *et al.* [176] designed, synthesized, and tested alanine-based peptides as helix-forming peptides. In their work, the poly-Ala helix containing (i + 4) Glu^{-}…Lys^{+} salt bridges showed optimal behavior (≈80% helicity). D-Arg was also chosen because it is the most efficient α-helical C-capping residue.[183] According to Huang *et al.*, a tripeptide, Ser-Pro-Glu, was selected as helix-stabilizing N-terminal sequence that occurs most frequently at the N-terminus of helices in the WHATIF database of 1705 helices and also might stabilize the helix by electrostatic interactions.[130] Therefore, peptide 2 was built based on possible α-helix-stabilizing effects, while peptide 1 was considered as a regular poly-Ala peptide found in the experimental work.

We determined the secondary structure between two related 14-residue peptides to verify our new methodology on the study of properties of peptides. Figure 4-1 shows plots of the helicity of the secondary structure versus residue number to present the structural differences. The overall helicity is expected to be low, since these two peptides are only 14-residue long. However, significant differences are found between the two peptides in Figure 4-1. A high helicity region in peptide 2 is seen between residue 5 (Proline) and 10 (Lysine). In contrast, peptide 1 shows a low and broad distribution. The difference can be explained by a salt-bridge interaction of two charged residues, Glu^{-}(residue 6) and Lys^{+}(residue 10). When the two end terminals (N-terminal and C-terminal) in peptide 2 are compared, C-terminal end with a D-Arg residue shows 20-40% helicity, but N-terminal with an acetyl group shows almost zero helicity. Simulations results are comparable with experimental ones in a sense that peptide 2 is more helical than peptide 1.

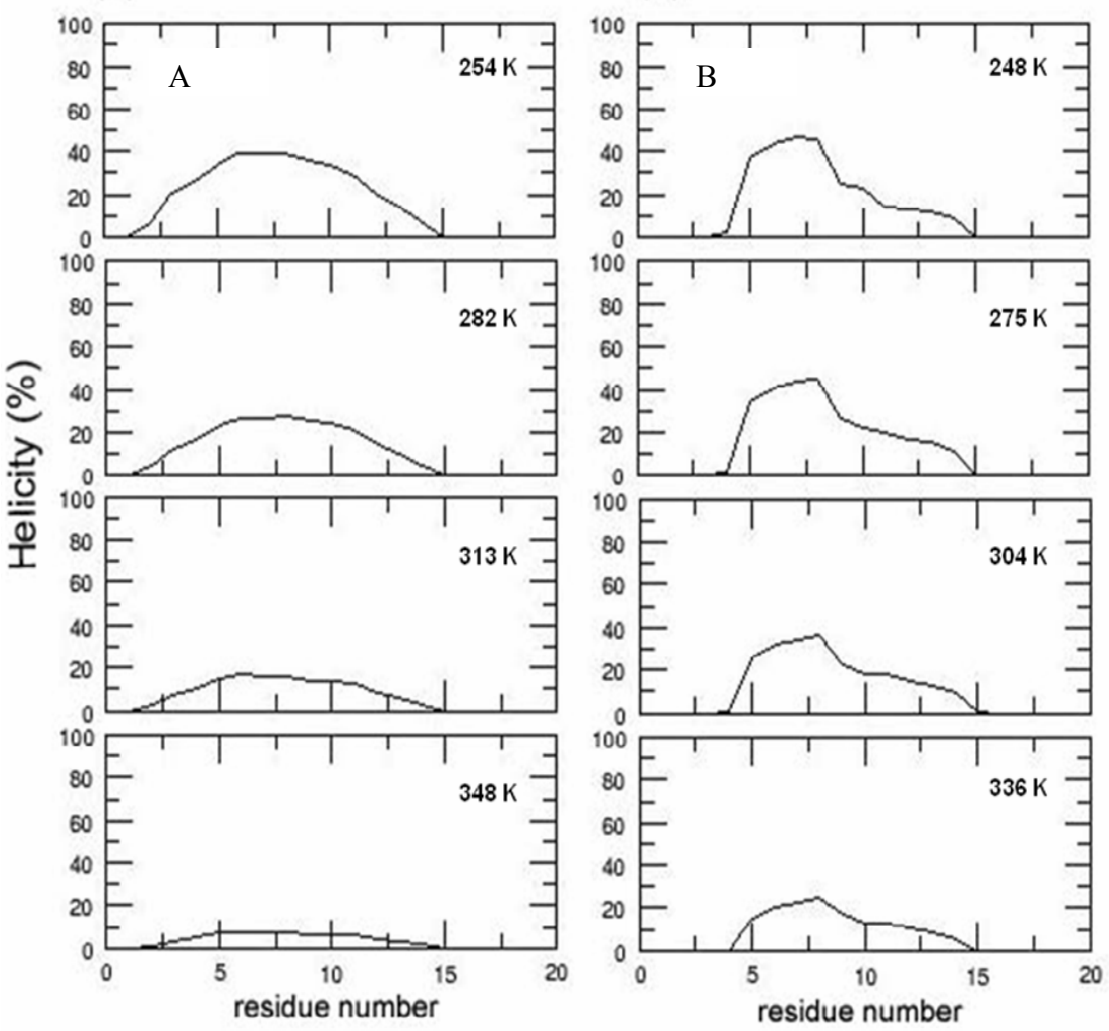


Figure 4-1. Calculated helicity. A) Peptide 1 versus residue number, B) Peptide 2 versus residue number based on DSSP method, respectively.

4.3.1 Cluster Analysis

The conformations at the lowest temperature (150K for both peptides) from REMD were sorted into clusters based on potential helical regions (residue 4 to 9 of peptide 1 and residue 5 to 10 of peptide 2) from DSSP method. The largest cluster in peptide 1 contains 71% of the structures and the next largest cluster is 13% populated. In peptide 2, the largest cluster is 78% and the next largest is 14% populated. No other clusters from both peptides have a population higher than 10%. Figure 4-2 A shows all conformations superimposed on the representative

structures of the largest populated clusters for both peptides (red and yellow for peptide 1 and 2, respectively) at 150K. Two representative structures from clusters are shown in Figure 4-2 B.

In Figure 4-3 the fractional sizes of each cluster in both peptides are plotted as a function of time to evaluate convergence of the REMD simulations, where the first 20ns of the simulations were discarded and the latter 480ns were used for calculation. Both peptides converge to their final populations after approximately 200ns. The biggest cluster of both peptides converges to a population of ~0.7-0.8 (70-80%). Therefore, all initial conformations were taken from 200-500ns REMD for future analysis.

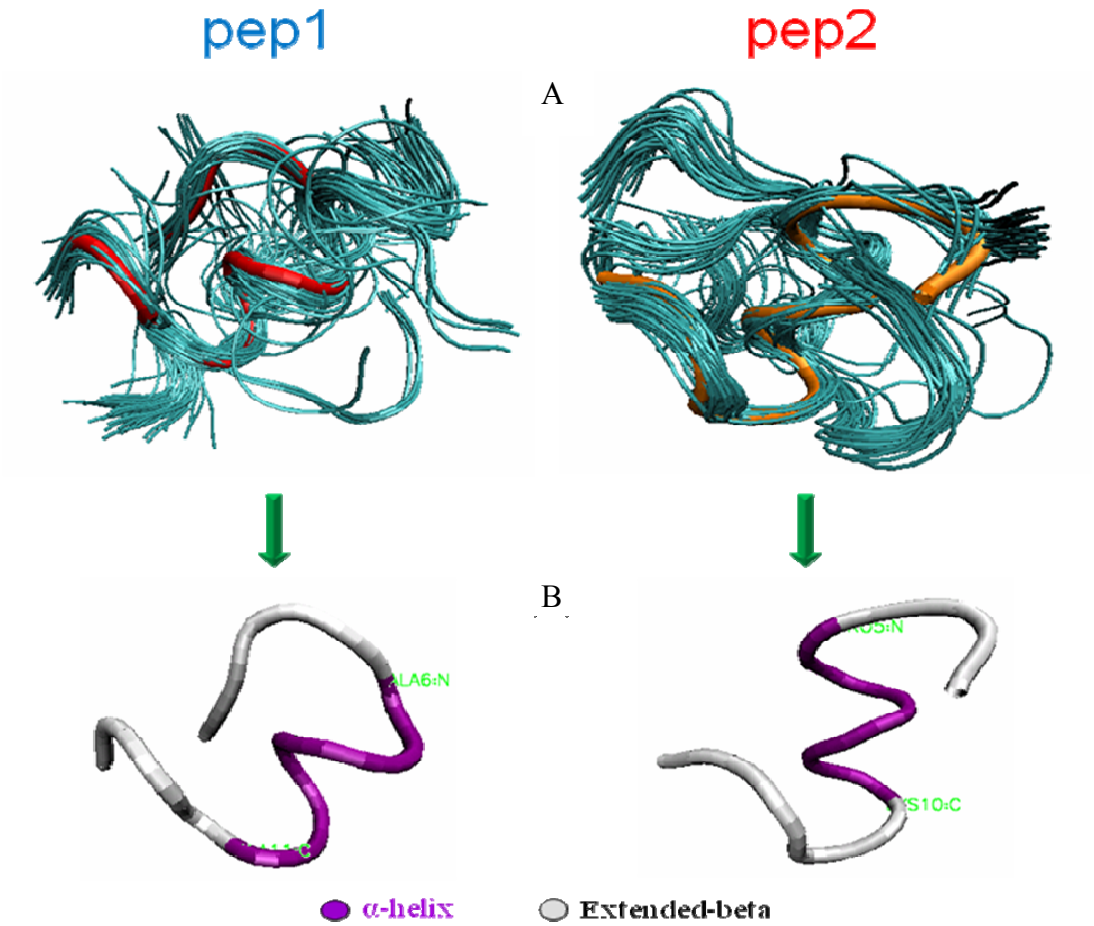


Figure 4-2. Reference structures from cluster analysis of peptide 1 and peptide 2. A) All conformations from REMD at 150K and are superimposed on the two reference structures, B) Two representative structures.

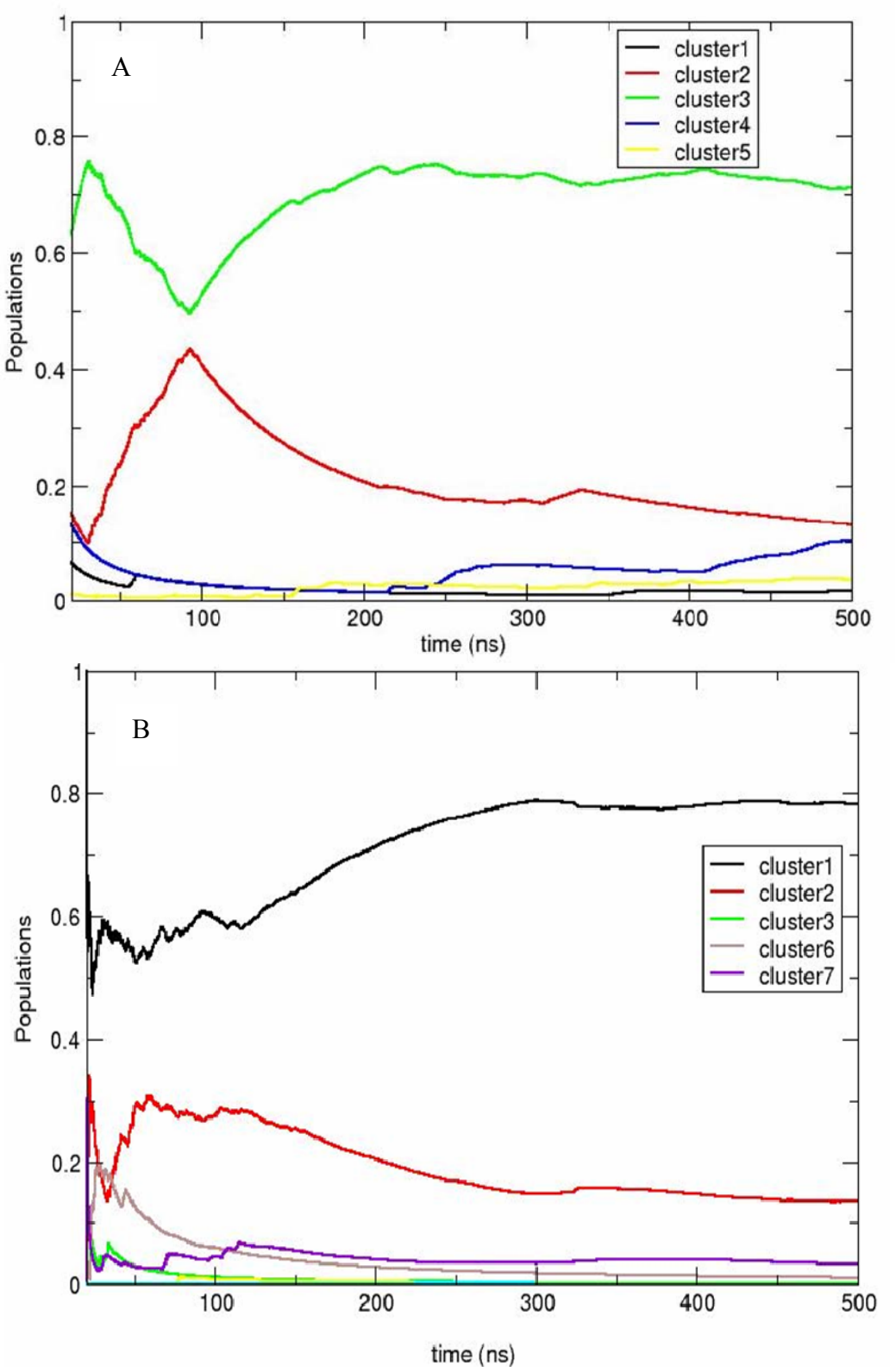


Figure 4-3. Populations of the representative clusters. A) Peptide 1 as a function of time from 20-500ns REMD, B) Peptide 2 as a function of time from 20-500ns REMD. At 200ns, simulations of both peptides are converged and stabilized.

4.3.2 Helicity

Theoretical fractional helicity (f_H) of a peptide can be calculated using the mean residue ellipticity at 222nm, $[\theta]_{222}$,[49,184,185]

$$f_H = \frac{[\theta]_{222}}{[\theta_\infty]_{222}(1-\frac{x}{n})} \quad (4\text{-}1)$$

where $[\theta_\infty]_{222}$ is the mean residue ellipticity of an ideal peptide with 100% helicity, n is the the length of the potential helical region, and x is an empirical correction. In order to assign those parameters, 14 polyalanines (ACE-$(ALA)_n$-NME, n=5-14, 16, 18, 20, 22) with 100% helix were built using HyperChem software [186] and subjected to 5,000 steps of initial minimization. The CD spectra ($[\theta]_{222}$) were computed and plotted ($[\theta]_{222}$ versus 1/n) from the Equation 4-1 (Figure 4-4). Values of x and $[\theta_\infty]_{222}$ are taken to be 2.38 and -31403 deg cm^2 $dmol^{-1}$, respectively, based on the plot and fit in Figure 4-4. These two values are used for further calculations. Figure 4-5 shows the fractional helicity of peptide 1 and peptide2 from 200 - 500ns REMD as a function of temperature (Equation 4-1).

Wang *et al.* [49] reported that the helicity of peptide 1 and peptide 2 at 11°C was approximately 11% and 29% (by experiment), respectively. Marqusee *et al.*[176] showed that the 16- or 18-residue Ala-based peptides contain approximately 25-50% helicity. Both theoretical and experimental results clearly indicate that the helicity of peptide 2 is larger than that of peptide 1.

4.3.3 Calculation of CD Spectra

For each member of the ensemble and at each temperature in the REMD simulation, CD spectra were computed as mentioned in the methods section 4.2. The spectra obtained for each peptide at a certain temperature were then averaged over members of the ensemble (49,000

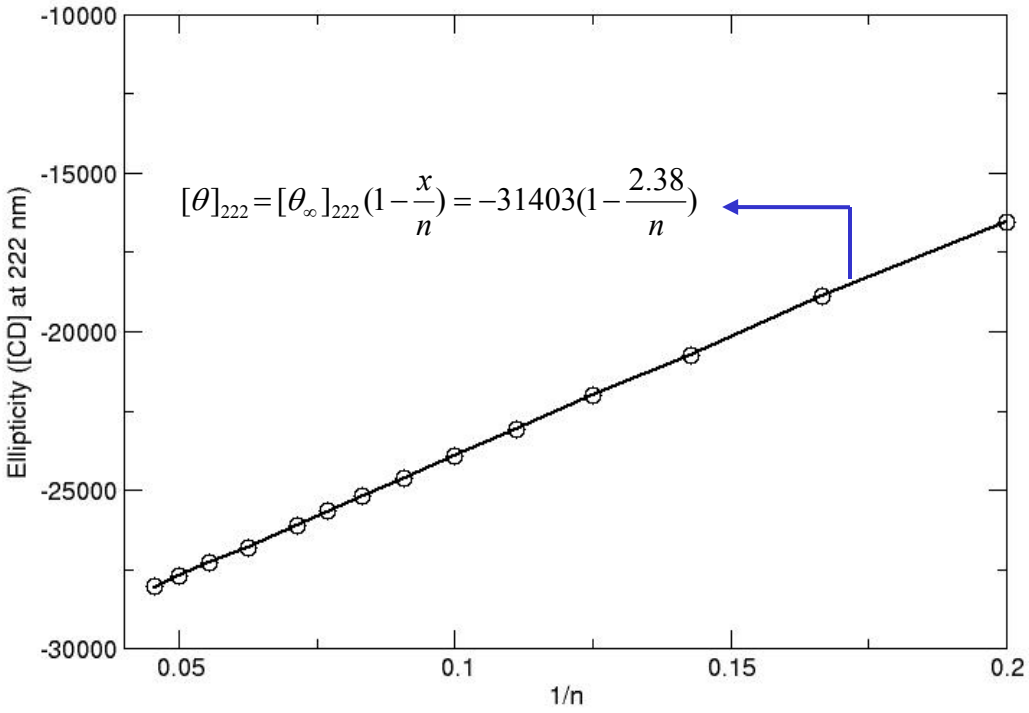


Figure 4-4. The resulting curve for the two parameter calculations ($[\theta_\infty]_{222}$ and x) from the plot $[\theta]_{222}$ versus 1/n from the Equation 4-1. The unit of molar ellipticity is deg·cm^2·dmol^{-1} of both peptides.

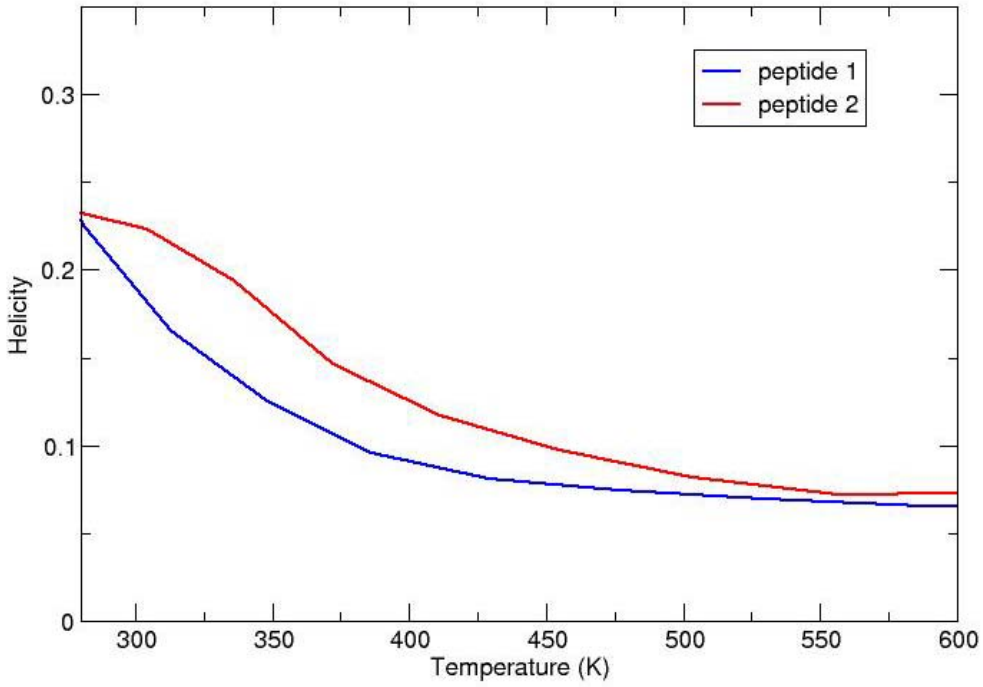


Figure 4-5. Fractional helicity (f_H) of peptide 1 (blue) and peptide 2 (red) as a function of temperature.

structures each). The resulting average spectra as a function of wavelength are shown in Figure 4-6. The CD spectra for an α-helix show two minima, one at 222nm (n-π* transition along the carbonyl bond) and the other at 208nm (π-π* amide transition).[162] In the plot, the molar ellipticities at 222nm (θ_{222}) are linearly related to the helical minimum.[184] Figure 4-7 shows average values of CD at 222nm ($<CD_{222}>$) for peptide 1 and peptide 2, as a function of temperature. The observation showing higher helical population in peptide 2 than peptide 1, are in a good agreement with the experimental results by Wang *et al.* [49], in the similar temperature region. This type of signal is useful for the T-jump experiment.

T-jump simulations were performed and the resulting time traces were fitted to calculate the folding and unfolding rates. This system appears to reach equilibrium after ~ 200ns of REMD from the cluster analysis in Figure 4-3. Therefore, 1,200 initial configurations were selected from 200 - 500ns REMD. The T-jump was then simulated from 313 to 348K for peptide 1 and 336 to 372K for peptide 2. Temperature change is achieved within a time scale of the order of the thermostat coupling (0.1ps time constant used), having a very short dead-time. The simulations were run for a total of 5ns each (total of 6 μs of each peptide). At each time slice, a CD spectrum was computed for each independent MD run, and then they are averaged. The resulting <CD222> are plotted against time and fitted (Figure 4-8). As we mentioned in the Chapter 3, the trace is bounded from the equilibrium runs, we know what the system looks like at t=0 and at t=∞. This leaves a single parameter (λ) to be determined via fitting from the Equation 3-4. The calculated relaxation data are fitted well using a single exponential function, indicating occurrence of reversible two-state folding. Details of the two-state mechanism are covered in Chapter 3.

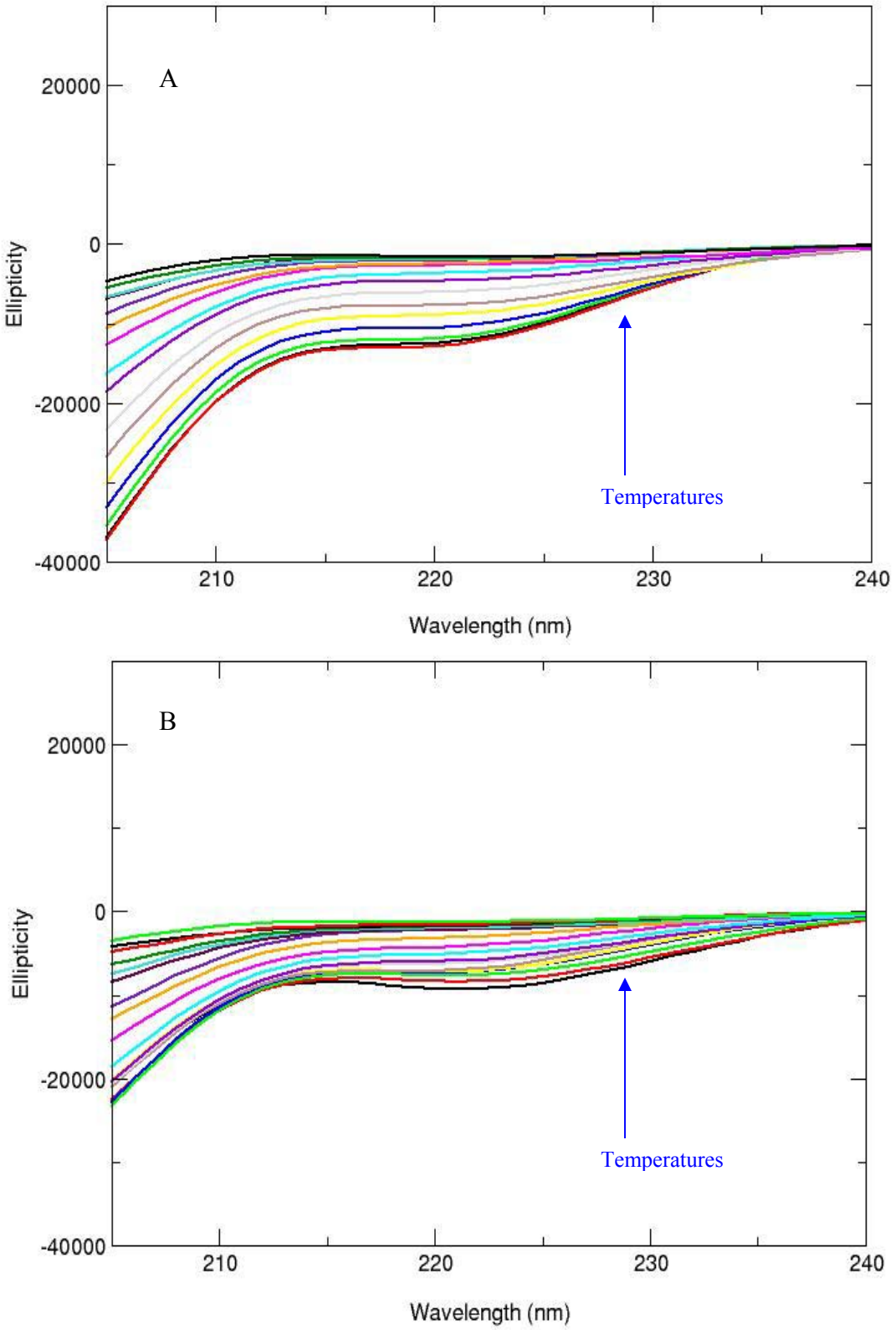


Figure 4-6. Calculated circular dichrioism (CD) spectra. A) Peptide 1 from 150-726K, B) Peptide 2 from 150-834K. The unit of molar ellipticity is deg·cm^2 ·dmol^{-1} of both peptides.

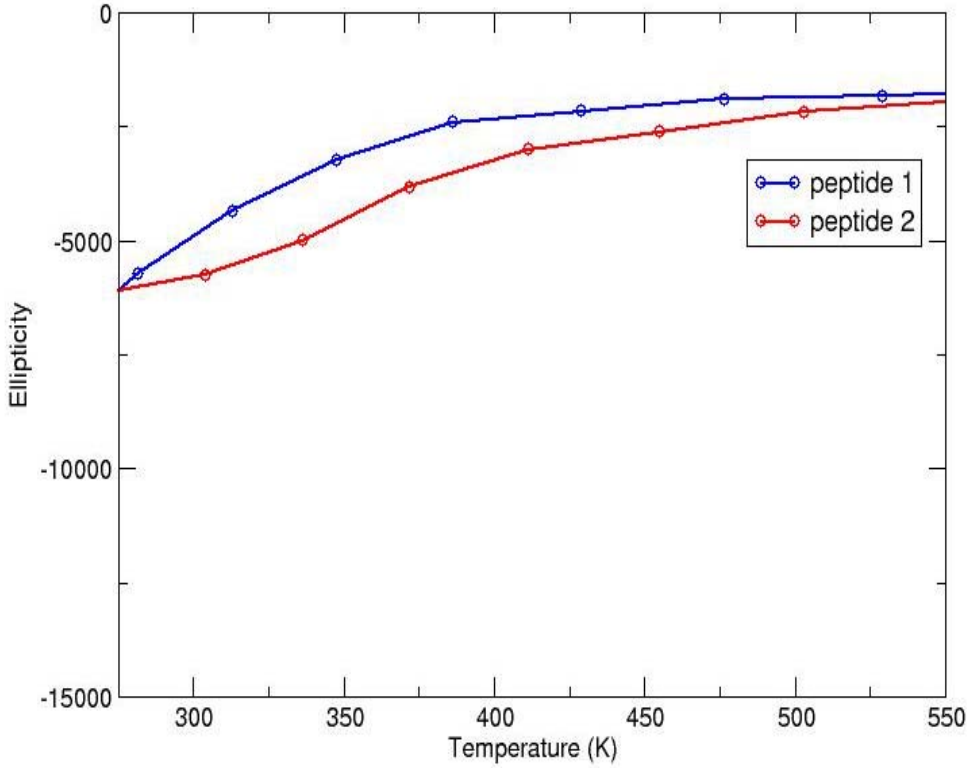


Figure 4-7. Average of mean residue ellipticities at 222ns (<CD_{222}>) of simulated peptide 1 (blue) and peptide 2 (red) are shown as a function of temperature. The unit of molar ellipticity is deg·cm^2 ·dmol^{-1} of both peptides.

In a two-state model, the relaxation time (τ_{relax}) is obtained from the sum of the folding rate (k_f) and unfolding rate (k_u),

$$\lambda = k_f + k_u = k_f\,(1 + K) = 1/\tau_{relax} = 1/\tau_f + 1/\tau_u \qquad (4\text{-}2)$$

In this equation, τ_f is the folding time, while τ_u is the unfolding time. The equilibrium constant K ($=\frac{k_u}{k_f}$) is calculated by fractional helicity (f_H) and is a ratio of the forward and backward rate constants. When only two states, folded (F) and unfolded (U) states, exists and fractional helicity of peptide 1 and peptide 2 are 12% at 348K and 15% at 372K, respectively, the K values of peptide 1 (K_1) and peptide 2 (K_2) are calculated with the final temperatures of T-jump simulation as the target temperatures, according to,

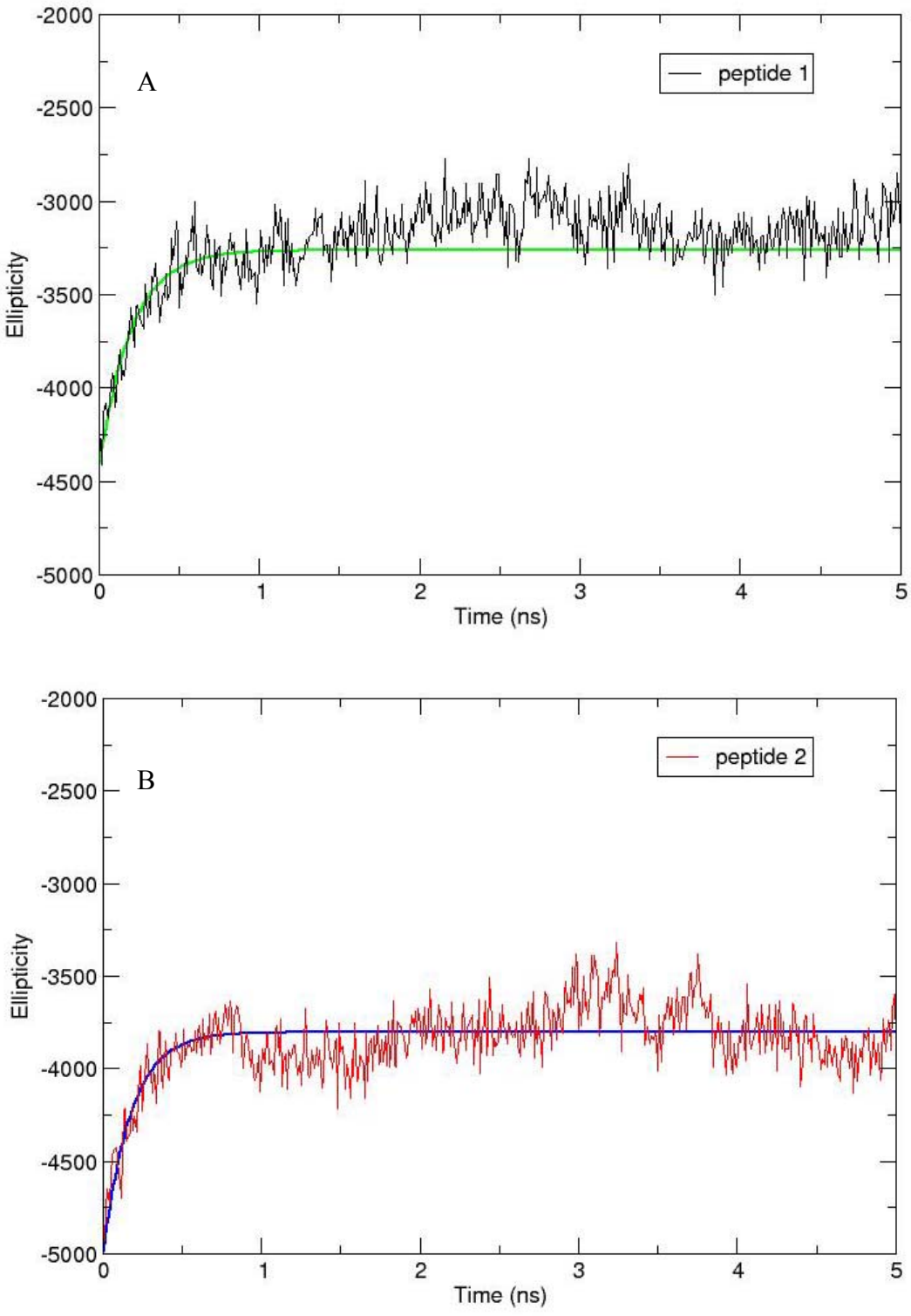


Figure 4-8. Average of CD_{222} ($\langle CD_{222} \rangle$) of temperature jump (T-jump) simulation data (collision frequency $\gamma = 1.0 ps^{-1}$). A) Peptide 1: black, B) Peptide 2: red with fitting curves, respectively. The unit of molar ellipticity is $deg \cdot cm^2 \cdot dmol^{-1}$ of both peptides.

$$K_1 = \frac{[U]}{[F]} = \frac{k_u}{k_f} = 7.3 \text{ and } K_2 = \frac{[U]}{[F]} = \frac{k_u}{k_f} = 5.7 \quad (4\text{-}3)$$

The relaxation time and folding time obtained from experimental and computational results are compared in the Table 4-1. The computed τ values are 0.20ns for peptide 1 and 0.17ns for peptide 2, which correspond to a folding time τ_f=1.65ns of peptide 1 and τ_f=1.16ns of peptide 2, respectively. Thus, the relaxation time of two peptides differ by a factor of ~1.2, which is the ratio $\left[(0.20ns)/(0.17ns)\right]$. Similarly,, difference in experimental results was by a factor of ~1.1, which is the ratio $\left[(222ns)/(204ns)\right]$.[49] This result validates our methodology for the kinetic studies. In addition, ratio of folding time between two peptides is ~1.4 from the computation and ~2.9 from the experiments.[49] The results show that peptide 2 with higher helicity folds faster than peptide 1, indicating correlation between overall helix stability and folding time.

Table 4-1. Relaxation and folding times of peptide 1 and peptide 2 from experimental and computational data.

	Relaxation and folding times (ns)			
	τ (exp.*)	τ (com.**)	τ_f (exp.*)	τ_f (com.**)
Peptide 1	222	0.20	~ 2000	1.65
Peptide 2	204	0.17	~ 700	1.16

* Experimental data obtained at 11°C (284K) by Wang *et al.* [49]
** Computational data obtained by calculated CD (<CD_{222}>) of T-jump simulations (peptide1 at 348K and peptide 2 at 372K) in Figure 4-7.

The time scales seen in theoretical data are much shorter than in experimental data. This mainly is due to the implicit solvent model used for the calculation, which accelerates the folding and unfolding times. The solvent effect on the folding time including different frictional coefficients will be discussed in the next section.

4.3.4 Effects of Frictional Coefficients

In order to generate data closer to the explicit solvent environment and also to show independence of friction versus folding time, we studied the effect of frictional coefficient. To investigate the frictional effect on the rate of the protein folding, we performed the T-jump

simulations (details in section 4.3.3) using Langevin dynamics with several collision frequencies (γ=1.0, 5.0, 10.0, and 20.0ps^{-1}), for both peptides. Figure 4-9 presents the simulation results in the plot of average values of CD at 222nm (<CD_{222}>) versus time. As we already mentioned in Chapter 3 and section 4.3.3, all traces for both peptides were fitted to only a single parameter (λ) with the same values of C_1 and C_2 from Equation 3-4 since we have complete knowledge of what the system looks like at t=0 and t=∞. The calculated relaxation data in Figure 4-9 are then fitted using a single exponential function, resulting in folding times (τ_f) from Equation 4-2. The folding times are plotted as a function of collision frequency (γ) in Figure 4-10 and the values of folding time for two peptides are compared in the Table 4-2.

Table 4-2. Folding times of the two peptides at different collision frequencies with Langevin dynamics.

	Folding times (τ_f, ns)	
Collision frequency, γ (ps^{-1})	Peptide 1	Peptide 2
1.0	1.65	1.16
5.0	4.45	4.11
10.0	8.94	5.94
20.0	16.7	11.2

The folding time with higher friction is much longer than lower friction for both peptides. The folding time is approximately linear to the collision frequency (Figure 4-10). However, it does not seem to be closely fitted to the Kramers' relation[1], $k_f^{-1}(=\tau_f) \propto \gamma$. The Kramers' model is not satisfactory, as its y-intercept is fixed to be zero.[23,177] Therefore, the results in Figure 4-10 were further analyzed using the following linear model (a modified simple Kramers' model, Equation 4-4) with *a* and *b* as variables which was suggested by Qiu *et al.*;[23]

$$\tau_f = a + b\gamma \tag{4-4}$$

[1] Kramer's equation in its simplest form could be described as, $k = (\omega_a \omega_b / 2\pi\gamma)\exp(-\Delta E / RT)$ where ω_a and ω_b are the curvature of the potential energy surface at the bottom and top of the barrier, respectively, E is the size of the energy barrier to conformational change, R is the gas constant, T is the temperature, and γ is the friction of the solvent.

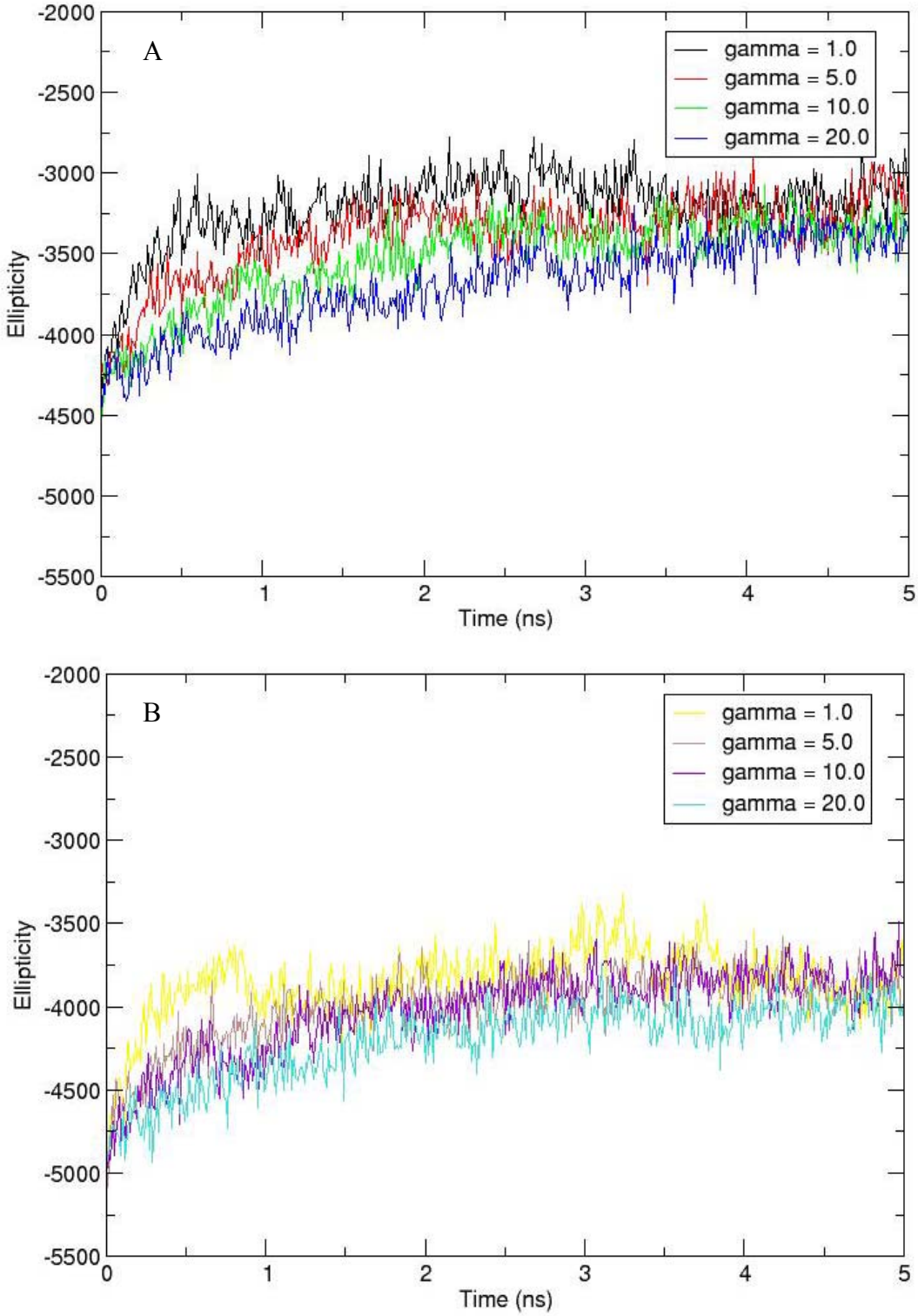


Figure 4 -9. Comparisons of different collision frequencies, γ=1.0, 5.0, 10.0, and 20.0ps^{-1}. A) Peptide1, B) Peptide 2 using average of CD_{222} (<CD_{222}>) of T-jump simulation data, respectively. The unit of molar ellipticity is deg·cm^2 ·dmol^{-1} of both peptides.

The plots were well fitted with $R^2 = 0.999$ for peptide 1 and $R^2 = 0.99$ for peptide 2 (values of *a* and *b* are given in Figure 4-11). This equation describes that the Kramers' relationship $\tau_f \propto \gamma$ is preserved except the positive value of y-intercept (*a*) whereas $a = 0$ in simple Kramers' model.

Ansari *et al.* [187] also suggested the relaxation rate equation using a modified Kramers' model according to;[2]

$$k^{-1} \propto \sigma + \eta_s \tag{4-5}$$

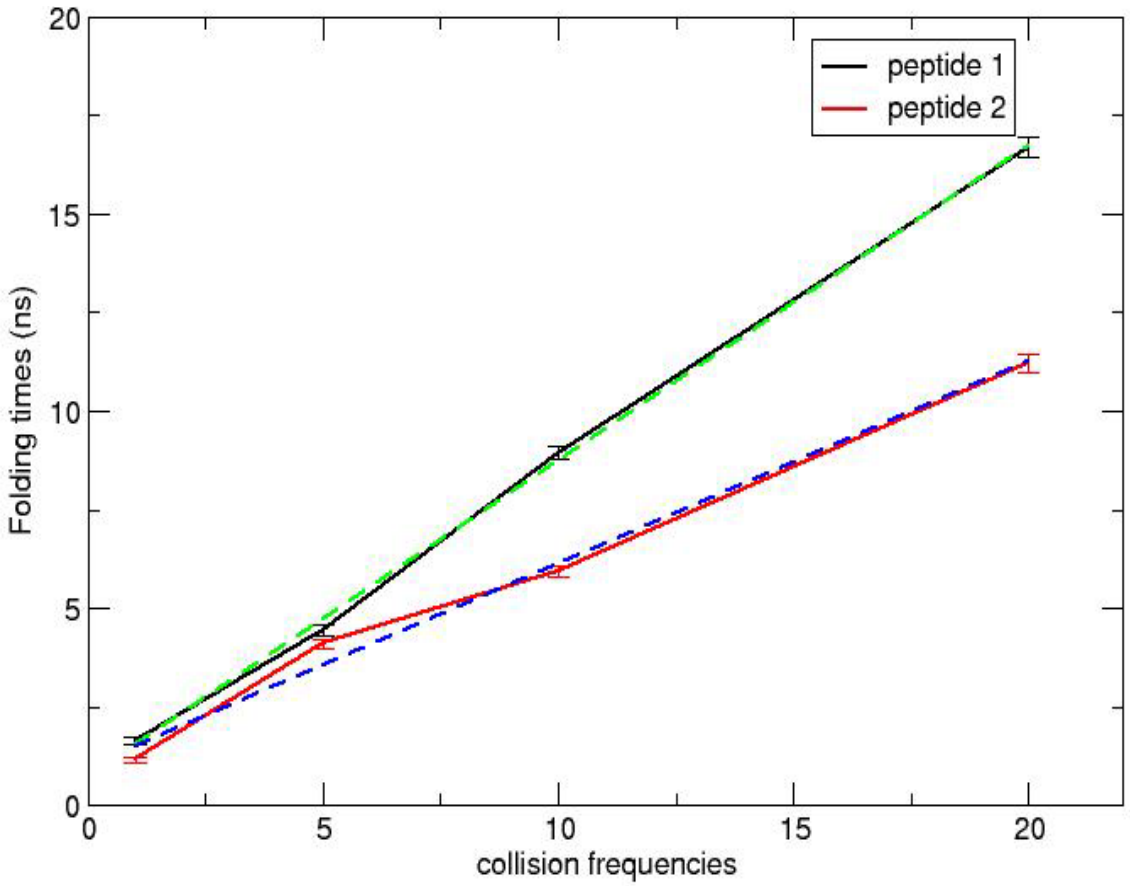


Figure 4-10. Comparisons of the folding times at different collision frequencies of two peptides with error bars and linear fits (dotted lines). The folding times and associate errors were calculated from fitting curves of average of CD_{222} (<CD_{222}>) from Figure 4-9.

[2] The rate equation can be written as $k = \frac{C}{\sigma + \eta_s}\exp(-E_0 / RT)$, where R is the gas constant, T is the temperature, E_0 is the average height of the potential energy barrier, C is adjustable parameter, σ is the protein friction, and η_s is the solvent friction.

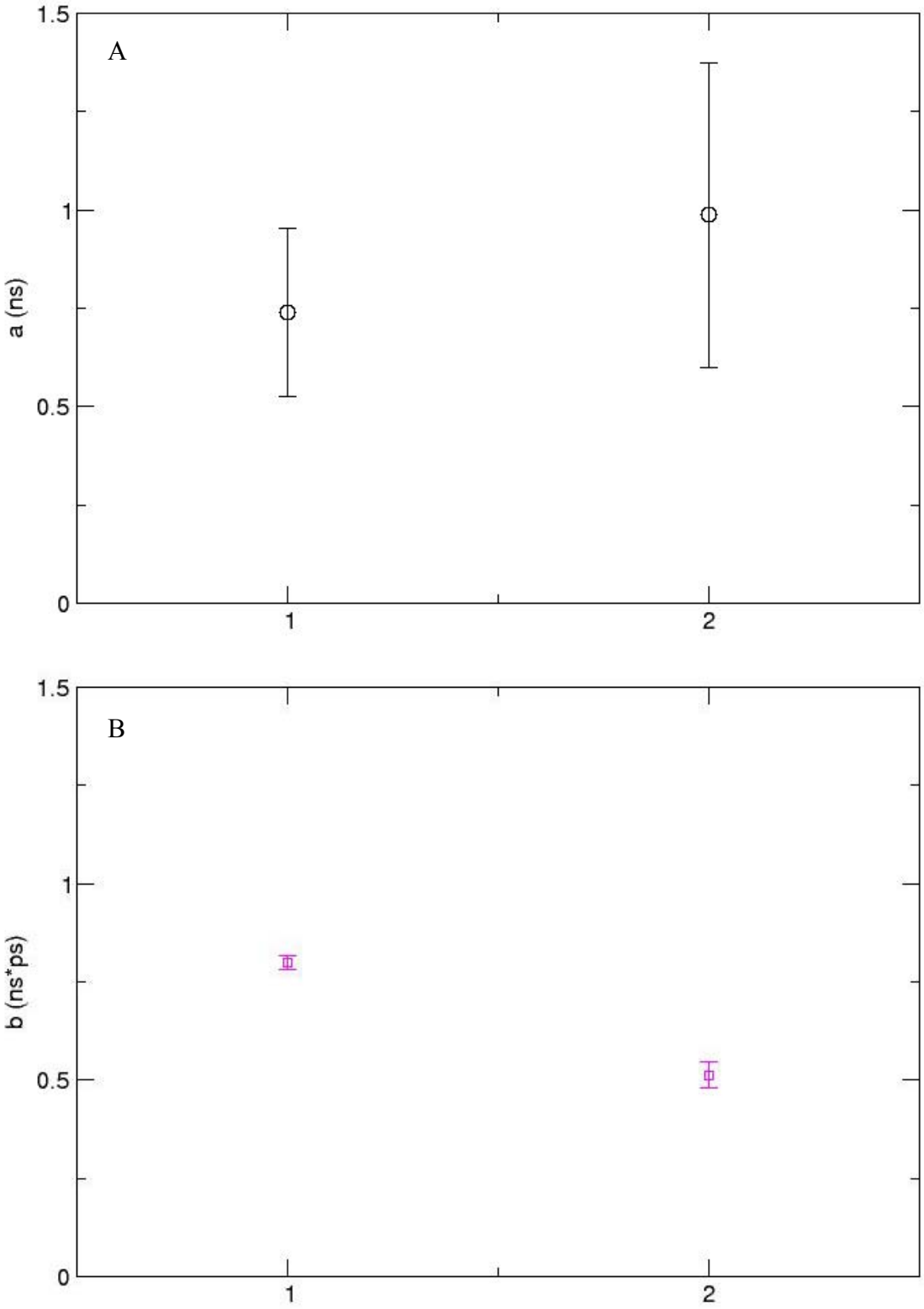


Figure 4-11. Friction dependence of peptide 1 (marked as **1**) and peptide 2 (marked as **2**) by T-jump simulations. A) Y-intercept, B) Slope of folding time (τ) obtained from linear fits in Figure 4-10.

where σ is the protein friction (or internal friction) and η_s is the solvent friction. Ansari *et al.* [187] mentioned that the folding rate can be determined as the sum of the solvent friction (η_s) and the protein friction (or internal friction, σ) from Equation 4-5. Related with Ansari model,[187] we could suggest that y-intercept (a) is correlated with internal friction.

The simple Kramers' model ($\tau_f \propto \gamma$) is well associated with the analysis of the protein folding kinetics when folding is relatively slow ($1/k_f$ ~ms).[23] However, our model shows better agreement with Equation 4-4, instead of simple Kramer model, and thus internal friction can influence the very fast folding reactions since the values of a and b in Figure 4-11 show fastest (nanoseconds) timescales.

Zagrovic and Pande [95] simulated TrpCage (TC5b) using the distributed computing technique and analyzed the dependence of solvent viscosity. According to their results, if the protein initially collapses into a random conformation and this continues until the protein folds (unfolded → random → folded), the first step is mainly controlled by the solvent friction and the second step is controlled only by the internal friction. Thus, internal friction would play a major role for the folding rate when the second step becomes the rate-limiting step.

The time scale of present folding (Figure 4-10) is much shorter than that from the experiments (microsecond scale).[23] This is due to the difference in the composition of the solvent between the experiment and simulation. The simulated results, however, show similar ratio of folding times between two peptides to the experimental results.

4.4 Conclusion

We studied folding kinetics of two related 14-residue peptides by using T-jump simulations that was recently introduced as a new computational methodology. Helicity and folding kinetics of two alanine-based peptides were investigated and compared with

experimental data. Very efficient sampling techniques, such as REMD, are used to populate equilibrium ensembles. Multiplexed MD simulations were then run to obtain kinetic information, particularly the non-equilibrium relaxation of the populations. We found that peptide 2 having more helicity, folds faster than peptide 1. The ratio of relaxation time of two peptides differed by a factor of ~1.2, while the corresponding experimental results were ~1.1. Therefore, our new methodology seems in good agreement with experimental data.

The effect of friction on the protein folding was also studied using Langevin dynamics. We performed data fitting, using the modified Kramers' linear model, on simulated results from different frictional coefficients. The observed nanosecond time scale of folding for both peptides indicates that internal friction can influence the very fast folding reactions. The composition of solvent made a significant effect on the folding kinetics, such that the nanosecond time scale in the simulation was obtained for the microsecond time scale in the experiments.

CHAPTER 5
CONCLUSIONS

Simulations opened the way to investigations of new fields and possibilities since Karplus *et al.* first introduced MD simulation of biomolecular systems.[18] MD simulations can provide ultimate details of individual atomic motions including pathways while experimental results usually show only averaged structure. However, approximations of molecular interactions are one of the well known drawbacks of MD simulations. Therefore, combining experimental and computational results address in complementing and overcoming limitations of both approaches.

Moreover, understanding the structure, kinetics and thermodynamics of protein folding remains one of the unsolved problems for both computational and experimental biophysists. Laser-induced T-jump is one of the most popular experimental methods in protein folding research since the timescale of T-jump extends from nanoseconds to milliseconds, which is an appropriate range for studies of folding kinetics.

In the first project, we introduced the computational T-jump: the temperature increases suddenly, creating the relaxation of the ensemble from one equilibrium distribution (at low temperature) to another one (at high temperature) by using proper conformational sampling method, which is REMD. Our method is designed and explored closely resembling to an experimental T-jump while the usual distributed computing procedure uses multiplexed MD runs to study folding. The alanine polypeptide (ACE–$(ALA)_{20}$–NME) was used as our model system in this project as a proof of principle and description of the methodology. A set of folding/unfolding time (τ_F=47.5ns and τ_U=155ns) was calculated from fitting the results of CD spectra and the changes of conformational distributions were shown through computation of NMR chemical shifts.

The second project extended the method to real proteins to get consistent results using Langevin dynamics which includes frictional effects and random forces. In this project two related 14-residue peptides were chosen and compared with the experimental data. The ratio of relaxation times of the two peptides is by a factor of ~1.2, while the corresponding experimental result was ~1.1.

LIST OF REFERENCES

1. Zhou, R. Trp-cage: folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci.* **100**, 13280-5 (2003).

2. Zhou, Y. & Karplus, M. Interpreting the folding kinetics of helical proteins. *Nature* **401**, 400-3 (1999).

3. Brooks, C.L., 3rd, Gruebele, M., Onuchic, J.N. & Wolynes, P.G. Chemical physics of protein folding. *Proc. Natl. Acad. Sci.* **95**, 11037-8 (1998).

4. Snow, C.D., Nguyen, H., Pande, V.S. & Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**, 102-6 (2002).

5. Daggett, V. Protein folding simulation. *Chem. Rev.* **106**, 1898-1916 (2006).

6. Huang, X., Nakanishi, K. & Berova, N. Porphyrins and metalloporphyrins: versatile circular dichroic reporter groups for structural studies. *Chirality* **12**, 237-255 (2000).

7. Ramachandron, G.N. & Sasisekharan, V. Conformation of polypeptides and proteins. *Adv. Prot. Chem.* **23**, 283-437 (1968).

8. Kay, B.K., Williamson, M.P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB Journal* **14**, 231-241 (2000).

9. Bielska, A.A. & Zondlo, N.J. Hyperphosphorylation of tau induces local polyproline II helix. *Biochemistry* **45**, 5527-5537 (2006).

10. Campbell, I.D. The march of structural biology. *Nat. Rev. Mol. Cell Biol.* **3**, 377-381 (2002).

11. Karplus, M. Molecular dynamics simulations of biomolecules. *Acc. Chem. Res.* **35**, 321-323 (2002).

12. Dill, K.A. & Chan, H.S. From levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* **4**, 10-19 (1997).

13. Dobson, C.M. & Ptitsyn, O.B. Folding and binding from theory to therapy. *Curr. Opin. Struct. Biol.* **7**, 1-2 (1997).

14. Duan, Y. & Kollman, P.A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**, 740-744 (1998).

15. Anfinsen, C.B., Haber, E., Sela, M. & White, F.H. The Kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Nat. Acad. Sci.* **47**, 1309-1314 (1961).

16. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).

17. Levinthal, C. Are there pathways for protein folding? . *J. Chem. Phys.* **65**, 44-45 (1968).

18. McCammon, J.A., Gelin, B.R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590 (1977).

19. Williams, S. et al. Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **35**, 691-697 (1996).

20. Nolting, B., Golbik, R. & Fersht, A.R. Submillisecond events in protein folding. *Proc. Nat. Acad. Sci.* **92**, 10668-10672 (1995).

21. Turner, D.H., Flynn, G.W., Lundberg, S.K., Faller, L.D. & Sutin, N. Dimerization of proflavin by the laser raman temperature-jump method. *Nature* **239**, 215-217 (1972).

22. Ervin, J., Sabelko, J. & Gruebele, M. Submicrosecond real-time fluorescence sampling: application to protein folding. *J. Photochem. Photobiol., B* **54**, 1-15 (2000).

23. Qiu, L. & Hagen, S.J. Internal friction in the ultrafast folding of the tryptophan cage. *Chem. Phys.* **312**, 327-333 (2005).

24. Callender, R. & Dyer, R.B. Probing protein dynamics using temperature jump relaxation spectroscopy. *Curr. Opin. Struct. Biol.* **12**, 628-633 (2002).

25. Snow, C.D. et al. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc. Nat. Acad. Sci.* **101**, 4077-4082 (2004).

26. Ballew, R.M., Sabelko, J. & Gruebele, M. Observation of distinct nanosecond and microsecond protein folding events. *Nat. Struct. Mol. Biol.* **3**, 923-926 (1996).

27. Gruebele, M., Sabelko, J., Ballew, R. & Ervin, J. Laser temperature jump induced protein refolding. *Acc. Chem. Res.* **31**, 699-707 (1998).

28. Mayor, U., Johnson, C.M., Daggett, V. & Fersht, A.R. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Nat. Acad. Sci.* **97**, 13518-13522 (2000).

29. Nolting, B. *Protein foliding kinetics - biophysical methods*, (Springer, 1999).

30. Woody, R.W. Optical rotatory properties of biopolymers. *J. Polym. Sci.: Macromol. Rev.* **12**, 181-320 (1977).

31. Woody, R.W. *The peptides (Hruby, V. J., Ed.)*, 15-114 (1985).

32. Yang, J.T., Wu, C.-S.C., Martinez, H.M. & Timasheff, C.H.W.H.a.S.N. Calculation of protein conformation from circular dichroism. *Methods in Enzymology* **130**, 208-269 (1986).

33. Johnson, W.C. Secondary structure of proteins through circular dichroism spectroscopy. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 145-166 (1988).

34. Johnson, W.C. Protein secondary structure and circular dichroism: A practical guide. *Proteins Struct. Funct. Genet.* **7**, 205-214 (1990).

35. Nakanishi, K., Berova, N. & Woody, R.W. *Circular dichroism principles and application*, (1994).

36. Fasman, G.D. Circular dichroism and the conformational analysis of biomolecules. *Plenum Press, New York* (1996).

37. Gratzer, W.B., Holzwarth, G.M. & Doty, P. Polarization of the ultraviolet absorption bands in alpha -helical polypeptides. *Proc. Nat. Acad. Sci.* **47**, 1785-1791 (1961).

38. Manning, M.C., Illangasekare, M. & W. Woody, R. Circular dichroism studies of distorted [alpha]-helices, twisted [beta]-sheets, and [beta]-turns. *Biophys. Chem.* **31**, 77-86 (1988).

39. Chin, D.-H., Woody, R.W., Rohl, C.A. & Baldwin, R.L. Circular dichroism spectra of short, fixed-nucleus alanine helices. *Proc. Nat. Acad. Sci.* **99**, 15416-15421 (2002).

40. Campbell, I.D. & Dwek, R.A. in *Biological Spectroscopy (Biophysical Techniques Series)* (Benjamin-Cummings Publishing Company, 1984).

41. Havlin, R.H. & Tycko, R. Probing site-specific conformational distributions in protein folding with solid-state NMR. *Proc. Nat. Acad. Sci.* **102**, 3284-3289 (2005).

42. Wishart, D.S., Sykes, B.D., Thomas, L.J. & Oppenheimer, N.J. Chemical shifts as a tool for structure determination. *Methods in Enzymology* **239**, 363-392 (1994).

43. Allerhand, A. & Oldfield, E. Determination of rotational mobilities of backbone and side-chain carbons of poly(r-benzyl L-glutamate) in the helical and random-coil states from measurements of carbon-13 relaxation times and nuclear Overhauser enhancements. Vol. 12 3428-3433 (1973).

44. Spera, S. & Bax, A. Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**, 5490-5492 (1991).

45. Wishart, D.S. & Sykes, B.D. The 13C chemical-shift Index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR* **4**, 171-180 (1994).

46. Oldfield, E. Chemical shifts in amino acids, peptides, and proteins: from quantum chemistry to drug design. *Annu. Rev. Phys. Chem.* **53**, 349-378 (2002).

47. Dalgarno, D.C., Levine, B.A. & Williams, R.J.P. Structural information from NMR secondary chemical shifts of peptide α C−H protons in proteins. *Biosci. Rep.* **3**, 443-452 (1983).

48. Szilagyi, L. Chemical shifts in proteins come of age. *Prog. Nucl. Magn. Reson. Spectrosc.* **27**, 325-430 (1995).

49. Wang, T., Du, D. & Gai, F. Helix-coil kinetics of two 14-residue peptides. *Chem. Phys. Lett.* **370**, 842-848 (2003).

50. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141-151 (1999).

51. Cheng, X., Cui, G., Hornak, V. & Simmerling, C. Modified replica exchange simulation methods for local structure refinement. *J. Phys. Chem. B* **109**, 8220-8230 (2005).

52. Ferrenberg, A.M. & Swendsen, R.H. New monte carlo technique for studying phase transitions. *Phys. Rev. Lett.* **61**, 2635 (1988).

53. Ferrenberg, A.M. & Swendsen, R.H. New monte carlo technique for studying phase transitions. *Phys. Rev. Lett.* **63**, 1658 (1989).

54. Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H. & Kollman, P.A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011-1021 (1992).

55. Berg, B.A. & Neuhaus, T. Multicanonical algorithms for first order phase transitions. *Phys. Lett. B* **267**, 249-253 (1991).

56. Berg, B.A. & Neuhaus, T. Multicanonical ensemble: a new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* **68**, 9 (1992).

57. Lee, J. New monte carlo algorithm: entropic sampling. *Phys. Rev. Lett.* **71**, 211 (1993).

58. Mezei, M. Adaptive umbrella sampling: self-consistent determination of the non-boltzmann bias. *J. Comput. Phys.* **68**, 237-248 (1987).

59. Lyubartsev, A.P., Martsinovski, A.A., Shevkunov, S.V. & Vorontsov-Velyaminov, P.N. New approach to monte carlo calculation of the free energy: method of expanded ensembles. *J. Chem. Phys.* **96**, 1776-1783 (1992).

60. Hesselbo, B. & Stinchcombe, R.B. Monte carlo simulation and global optimization without parameters. *Phys. Rev. Lett.* **74**, 2151 (1995).

61. Tsallis, C. Possible generalization of boltzmann-gibbs statistics. *J. Statist. Phys.* **52**, 479-487 (1988).

62. Munakata, T. & Mitsuoka, S. Langevin dynamics for generalized thermodynamics. *J. Phys. Soc. Jpn.* **69**, 92-96 (2000).

63. Mitsutake, A., Sugita, Y. & Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Peptide Science* **60**, 96-123 (2001).

64. Hansmann, U.H.E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140-150 (1997).

65. Sugita, Y. & Okamoto, Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.* **329**, 261-270 (2000).

66. Okamoto, Y. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graphics Modell.* **22**, 425-439 (2004).

67. Cheng, X., Cui, G., Hornak, V. & Simmerling, C. Modified Replica Exchange Simulation Methods for Local Structure Refinement. *J. Phys. Chem. B* **109**, 8220-8230 (2005).

68. Sugita, Y., Kitao, A. & Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **113**, 6042-6051 (2000).

69. Pitera, J.W. & Swope, W. Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. *Proc. Natl. Acad. Sci.* **100**, 7587-92 (2003).

70. Weiner, S. J. et al. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **7**, 230-252 (1986).

71. Weiner, S.J. et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765-784 (1984).

72. Berendsen, H.J.C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43-56 (1995).

73. Jorgensen, W.L., Maxwell, D.S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225-11236 (1996).

74. Brooks, B.R. et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187-217 (1983).

75. Case, D.A. et al. AMBER 8. (University of California, San Francisco, 2004).

76. Cornell, W.D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197 (1995).

77. MacKerell, A.D., Wiorkiewicz-Kuczera, J. & Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **117**, 11946-11975 (1995).

78. Bayly, C.I., Cieplak, P., Cornell, W. & Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269-10280 (1993).

79. Cornell, W.D., Cieplak, P., Bayly, C.I. & Kollmann, P.A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **115**, 9620-9631 (1993).

80. Hornak, V. et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **65**, 712-725 (2006).

81. Still, W.C., Tempczyk, A., Hawley, R.C. & Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127-6129 (1990).

82. Qiu, D., Shenkin, P.S., Hollinger, F.P. & Still, W.C. The GB/SA continuum model for solvation. a fast analytical method for the calculation of approximate born radii. *J. Phys. Chem. A* **101**, 3005-3014 (1997).

83. Tsui, V. & Case, D.A. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **122**, 2489-2498 (2000).

84. Bashford, D. & Case, D.A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129-152 (2000).

85. Tsui, V. & Case, D.A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers* **56**, 275-291 (2000).

86. Born, M. *Z. Phys.* **1**, 45-48 (1920).

87. Srinivasan, J., Trevathan, M.W., Beroza, P. & Case, D.A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **101**, 426-434 (1999).

88. Hawkins, G.D., Cramer, C.J. & Truhlar, D.G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **246**, 122-129 (1995).

89. Hawkins, G.D., Cramer, C.J. & Truhlar, D.G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **100**, 19824-19839 (1996).

90. Simmerling, C., Strockbine, B. & Roitberg, A.E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **124**, 11258-11259 (2002).

91. Ruhong, Z. Free energy landscape of protein folding in water: Explicit vs. implicit solvent. *Proteins Struct. Funct. Genet.* **53**, 148-161 (2003).

92. Zhou, R. & Berne, B.J. Can a continuum solvent model reproduce the free energy landscape of a beta -hairpin folding in water? *Proc. Nat. Acad. Sci.* **99**, 12777-12782 (2002).

93. Felts, A. K. et al. Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins: Struct., Funct., Bioinf.* **56**, 310-321 (2004).

94. Geney, R., Layten, M., Gomperts, R., Hornak, V. & Simmerling, C. Investigation of salt bridge stability in a generalized born solvent model. *J. Chem. Theory Comput.* **2**, 115-127 (2006).

95. Zagrovic, B. & Pande, V.S. Solvent viscosity dependence of the folding rate of a small protein: Distributed computing study. *J. Comput. Chem.* **24**, 1432-1436 (2003).

96. Karplus, M. Aspects of protein reaction dynamics: deviations from simple behavior. *J. Phys. Chem. B* **104**, 11-27 (2000).

97. Takano, M., Yamato, T., Higo, J., Suyama, A. & Nagayama, K. Molecular dynamics of a 15-residue poly(L-alanine) in water: helix formation and energetics. *J. Am. Chem. Soc.* **121**, 605-612 (1999).

98. Haran, G., Haas, E. & Rapaport, D.C. Molecular dynamics simulations of simple peptide models: solvent effects and comparison with experiment. *J. Phys. Chem.* **98**, 10294-10302 (1994).

99. Leach, A.R. *Molecular modelling - principles and applications*, (Prentice Hall, 2001).

100. Sancho, J.M., Romero, A.H., Lacasta, A.M. & Lindenberg, K. Langevin dynamics of A + A reactions in one dimension. *J. Phys.: Condens. Matter* **19**, 065108 (2007).

101. Bayley, P.M., Nielsen, E.B. & Schellman, J.A. Rotatory properties of molecules containing two peptide groups: theory. *J. Phys. Chem.* **73**, 228-243 (1969).

102. Goux, W.J. & Hooker, T.M. Chiroptical properties of proteins. I. Near-ultraviolet circular dichroism of ribonuclease S. *J. Am. Chem. Soc.* **102**, 7080-7087 (1980).

103. Rosenfeld, L. Quantenmechanische theorie der natürlichen optischen aktivität von flüssigkeiten und gasen. *Zeitschrift für Physik A Hadrons and Nuclei* **52**, 161-174 (1929).

104. Hirst, J.D., Bhattacharjee, S. & Onufriev, A.V. Theoretical studies of time-resolved spectroscopy of protein folding. *Faraday Disscuss.* **122**, 253-267 (2003).

105. Sreerama, N. & Woody, R.W. Computation and analysis of protein circular dichroism spectra. *Methods Enzymol.* **383**, 318-351 (2004).

106. Woody, R.W. & Sreerama, N. Comment on ``improving protein circular dichroism calculations in the far-ultraviolet through reparametrizing the amide chromophore" *J. Chem. Phys.* **111**, 2844-2845 (1999).

107. Iwadate, M., Asakura, T. & Williamson, M.P. Cα and Cβ carbon-13 chemical shifts in proteins from an empirical database. *J. Biomol. NMR* **13**, 199-211 (1999).

108. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289-302 (1999).

109. Gronwald, W., Boyko, R.F., Sönnichsen, F.D., Wishart, D.S. & Sykes, B.D. ORB, a homology-based program for the prediction of protein NMR chemical shifts. *J. Biomol. NMR* **10**, 165-179 (1997).

110. Wishart, D.S., Watson, M.S., Boyko, R.F. & Sykes, B.D. Automated 1H and 13C chemical shift prediction using the BioMagResBank. *J. Biomol. NMR* **10**, 329-336 (1997).

111. Wishart, D.S., Sykes, B.D. & Richards, F.M. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**, 311-333 (1991).

112. Osapay, K. & Case, D.A. A new analysis of proton chemical shifts in proteins. *J. Am. Chem. Soc.* **113**, 9436-9444 (1991).

113. Williamson, M.P. & Asakura, T. Empirical comparisons of models for chemical-shift calculation in proteins. *J. Magn. Reson., Ser B* **101**, 63-71 (1993).

114. de Dios, A.C., Pearson, J.G. & Oldfield, E. Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. *Science* **260**, 1491-1496 (1993).

115. Sitkoff, D. & Case, D.A. Density functional calculations of proton chemical shifts in model peptides. *J. Am. Chem. Soc.* **119**, 12262-12273 (1997).

116. de Dios, A.C. Ab initio calculations of the NMR chemical shift. *Prog. Nucl. Magn. Reson. Spectrosc.* **29**, 229-278 (1996).

117. Ando, I., Kameda, T., Asakawa, N., Kuroki, S. & Kurosu, H. Structure of peptides and polypeptides in the solid state as elucidated by NMR chemical shift. *J. Mol. Struct.* **441**, 213-230 (1998).

118. Wishart, D.S., Case, D.A. & Thomas L. James, V.D.U.S. Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.* **338**, 3-34 (2002).

119. Wishart, D.S. & Nip, A.M. Protein chemical shift analysis: a practical guide. *Biochem. Cell Biol.* **76**, 153-163 (1998).

120. Le, H. & Oldfield, E. Ab Initio studies of amide-15N chemical shifts in dipeptides: applications to protein NMR spectroscopy. *J. Phys. Chem.* **100**, 16423-16428 (1996).

121. Case, D.A. Calibration of ring-current effects in proteins and nucleic acids. *J. Biomol. NMR* **6**, 341-346 (1995).

122. Xu, X.-P. & Case, D. Automated prediction of 15N, 13Cα, 13Cβ and 13C′ chemical shifts in proteins using a density functional database. *J. Biomol. NMR* **21**, 321-333 (2001).

123. Shakhnovich, E.I. Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* **7**, 29-40 (1997).

124. Gruebele, M. The fast protein folding problem. *Annu. Rev. Phys. Chem.* **50**, 485-516 (1999).

125. Nolting, B. & Andert, K. Mechanism of protein folding. *Proteins Struct. Funct. Genet.* **41**, 288-298 (2000).

126. Oliveberg, M. Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis. *Curr. Opin. Struct. Biol.* **11**, 94-100 (2001).

127. Ghosh, A., Elber, R. & Scheraga, H.A. An atomically detailed study of the folding pathways of protein A with the stochastic difference equation. *Proc. Nat. Acad. Sci.* **99**, 10394-10398 (2002).

128. Nolting, B., Julich, D., Vonau, W. & Andert, K. Evolutionary computer programming of protein folding and structure predictions. *J. Theor. Biol.* **229**, 13-18 (2004).

129. d'Avignon, A.D., Bretthorst, G.L., Holtzer, M.E. & Holtzer, A. Thermodynamics and kinetics of a folded-folded' transition at valine-9 of a GCN4-like leucine zipper. *Biophys. J.* **76**, 2752-2759 (1999).

130. Huang, C.-Y. et al. Helix formation via conformation diffusion search. *Proc. Nat. Acad. Sci.* **99**, 2788-2793 (2002).

131. Gee, P.J. & van Gunsteren, W.F. Terminal-group effects on the folding behavior of selected beta-peptides. *Proteins: Struct., Funct., Bioinf.* **63**, 136-143 (2006).

132. Zagrovic, B. & van Gunsteren, W.F. Comparing atomistic simulation data with the NMR experiment: How much can NOEs actually tell us? *Proteins: Struct., Funct., Bioinf.* **63**, 210-218 (2006).

133. Cheng, Y. et al. Continuum simulations of acetylcholine diffusion with reaction-determined boundaries in neuromuscular junction models. *Biophys. Chem.* **127**, 129-139 (2007).

134. Babakhani, A., Gorfe, A.A., Gullingsrud, J., Kim, J.E. & McCammon, J.A. Peptide insertion, positioning, and stabilization in a membrane: Insight from an all-atom molecular dynamics simulation. *Biopolymers* **85**, 490-497 (2007).

135. Zagrovic, B., Sorin, E.J. & Pande, V. [beta]-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **313**, 151-169 (2001).

136. Shirts, M. & Pande, V.S. Computing: screen savers of the world unite! *Science* **290**, 1903-1904 (2000).

137. Pande, V.S. et al. Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* **68**, 91-109 (2003).

138. Zagrovic, B., Snow, C.D., Shirts, M.R. & Pande, V.S. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* **323**, 927-937 (2002).

139. Rhee, Y.M., Sorin, E.J., Jayachandran, G., Lindahl, E. & Pande, V.S. Simulations of the role of water in the protein-folding mechanism. *Proc. Nat. Acad. Sci.* **101**, 6456-6461 (2004).

140. Singhal, N., Snow, C.D. & Pande, V.S. Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* **121**, 415-425 (2004).

141. Zagrovic, B., Snow, C.D., Khaliq, S., Shirts, M.R. & Pande, V.S. Native-like mean structure in the unfolded ensemble of small proteins. *J. Mol. Biol.* **323**, 153-164 (2002).

142. Larson, S.M., Snow, C.D., Shirts, M.R. & Pande, V.S. Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology, To appear in Computational Genomics. (2002).

143. Fersht, A.R. On the simulation of protein folding by short time scale molecular dynamics and distributed computing. *Proc. Nat. Acad. Sci.* **99**, 14122-14125 (2002).

144. Eaton, W.A. et al. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327-359 (2000).

145. Paci, E., Cavalli, A., Vendruscolo, M. & Caflisch, A. Analysis of the distributed computing approach applied to the folding of a small {beta} peptide. *Proc. Nat. Acad. Sci.* **100**, 8217-8222 (2003).

146. Fersht, A.R. & Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **108**, 573-582 (2002).

147. Daggett, V. & Fersht, A.R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18-25 (2003).

148. Jayachandran, G., Vishal, V. & Pande, V.S. Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece. *J. Chem. Phys.* **124**, 164902-12 (2006).

149. Andrec, M., Felts, A.K., Gallicchio, E. & Levy, R.M. Chemical theory and computation special feature: protein folding pathways from replica exchange simulations and a kinetic network model. *Proc. Nat. Acad. Sci.* **102**, 6801-6806 (2005).

150. Andrec, M., Levy, R.M. & Talaga, D.S. Direct determination of kinetic rates from single-molecule photon arrival trajectories using hidden markov models. *J. Phys. Chem. A* **107**, 7454-7464 (2003).

151. Dimitriadis, G. et al. Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump. *Proc. Natl. Acad. Sci.* **101**, 3809-14 (2004).

152. Wang, J., Cieplak, P. & Kollman, P.A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049-1074 (2000).

153. Hornak, V. et al. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **65**, 712-725 (2006).

154. Showalter, S.A. & Bruschweiler, R. Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field. *J. Chem. Theory Comput.* **3**, 961-975 (2007).

155. Showalter, S.A. & Bruschweiler, R. Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. *J. Am. Chem. Soc.* **129**, 4158-4159 (2007).

156. Ghosh, A., Rapp, C.S. & Friesner, R.A. Generalized born model based on a surface integral formulation. *J. Phys. Chem. B* **102**, 10983-10990 (1998).

157. Ryckaert, J.P., Ciccotti, G. & Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327-341 (1977).

158. Simmerling, C.L., Elber, R. & Zhang, R. *MOIL-View - a program for visualization of structure and dynamics of biomolecules and STO - a program for computing stochastic paths. In modeling of biomolecular structures and mechanisms*, 241 - 265 (Kluwer Academic Publishers: Netherlands, 1995).

159. Okur, A. et al. Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *J. Chem. Theory Comput.* **2**, 420-433 (2006).

160. Clark, L.B. Polarization assignments in the vacuum UV spectra of the primary amide, carboxyl, and peptide groups. *J. Am. Chem. Soc.* **117**, 7974-7986 (1995).

161. Ridley, J. & Michael, Z. An intermediate neglect of differential overlap technique for spectroscopy: Pyrrole and the azines. *Theor. Chem. Acc.* **32**, 111-134 (1973).

162. Toumadje, A., Alcorn, S.W. & Curtis Johnson, J.W. Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures. *Anal. Chem.* **200**, 321-331 (1992).

163. Ponder, J.W. & Richards, F.M. An efficient newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **8**, 1016-1024 (1987).

164. Ren, P. & Ponder, J.W. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* **23**, 1497-1506 (2002).

165. Hagen, S.J., Hofrichter, J., Szabo, A. & Eaton, W.A. Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. *Proc. Nat. Acad. Sci.* **93**, 11615-11617 (1996).

166. Torrie, G.M. & Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187-199 (1977).

167. Shea, J.-E. & Brooks Iii, C.L. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52**, 499-535 (2001).

168. Bursulaya, B.D. & Brooks, C.L. Comparative study of the folding free energy landscape of a three-stranded beta-sheet protein with explicit and implicit solvent models. *J. Phys. Chem. B* **104**, 12378-12383 (2000).

169. Gilmanshin, R., Williams, S., Callender, R.H., Woodruff, W.H. & Dyer, R.B. Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc. Natl. Acad. Sci.* **94**, 3709-13 (1997).

170. Thompson, P.A., Eaton, W.A. & Hofrichter, J. Laser temperature jump study of the helix<->coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* **36**, 9200-9210 (1997).

171. Munoz, V., Henry, E.R., Hofrichter, J. & Eaton, W.A. A statistical mechanical model for beta -hairpin kinetics. *Proc. Nat. Acad. Sci.* **95**, 5872-5879 (1998).

172. Munoz, V., Thompson, P.A., Hofrichter, J. & Eaton, W.A. Folding dynamics and mechanism of [beta]-hairpin formation. *Nature* **390**, 196-199 (1997).

173. Sabelko, J., Ervin, J. & Gruebele, M. Observation of strange kinetics in protein folding. *Proc. Nat. Acad. Sci.* **96**, 6031-6036 (1999).

174. Kim, S. & Roitberg, A. Simulating temperature jumps for protein folding. *J. Phys. Chem.*, submitted (2007).

175. Heitmann, B., Job, G.E., Kennedy, R.J., Walker, S.M. & Kemp, D.S. Water-solubilized, cap-stabilized, helical polyalanines: calibration standards for NMR and CD analyses. *J. Am. Chem. Soc.* **127**, 1690-704 (2005).

176. Marqusee, S. & Baldwin, R.L. Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. *Proc. Nat. Acad. Sci.* **84**, 8898-8902 (1987).

177. Kramers, H.A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284-304 (1940).

178. Klimov, D.K. & Thirumalai, D. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* **79**, 317 (1997).

179. Jagielska, A. & Scheraga, H.A. Influence of temperature, friction, and random forces on folding of the B-domain of staphylococcal protein A: All-atom molecular dynamics in implicit solvent. *J. Comput. Chem.* **28**, 1068-1082 (2007).

180. Case, D.A. et al. AMBER 9. (University of California, San Francisco, 2006).

181. Onufriev, A., Bashford, D. & Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **55**, 383-394 (2004).

182. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).

183. Schneider, J.P. & DeGrado, W.F. The design of efficient alpha-helical C-capping auxiliaries. *J. Am. Chem. Soc.* **120**, 2764-2767 (1998).

184. Chen, Y.-H., Yang, J.T. & Chau, K.H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350-3359 (1974).

185. Gans, P.J., Lyu, P.C., Manning, M.C., Woody, R.W. & Kallenbach, N.R. The helix-coil transition in heterogeneous peptides with specific side-chain interactions: theory and comparison with CD spectral data. *Biopolymers* **31**, 1605-1614 (1991).

186. http://www.hyperchem.com.

187. Ansari, A., Jones, C.M., Henry, E.R., Hofrichter, J. & Eaton, W.A. The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **256**, 1796-1798 (1992).

BIOGRAPHICAL SKETCH

Seonah Kim, was born in Daegu, South Korea. She went to college at Yeoungnam University, majoring in industrial chemistry and graduated with a bachelor's degree. In August 1999, she started graduate school at the University of Houston, in the Computer Science Department and earned her master's degree. She moved to Gainesville, Florida, in July 2003. There she entered the University of Florida's Ph.D. program in chemistry, specializing in computational chemistry.