

UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

TESTING RECENTLY-DEVELOPED MOLECULAR DYNAMICS PROTOCOLS
FOR STRUCTURE PREDICTION OF POLYPEPTIDES AND PRION PROTEINS

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By
TIMOTHY HOWARD CLICK
Norman, Oklahoma
2007

UMI Number: 3304578

Copyright 2007 by
Click, Timothy Howard

All rights reserved.

UMI[®]

UMI Microform 3304578

Copyright 2008 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

TESTING RECENTLY-DEVELOPED MOLECULAR DYNAMICS PROTOCOLS
FOR STRUCTURE PREDICTION OF POLYPEPTIDES AND PRION PROTEINS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY

Ralph A. Wheeler

Paul F. Cook

Ann H. West

Wai Tak Yip

David P. Nagle

Acknowledgements

As I sit writing my dissertation, I ponder the life that I have lived. I think about all the events that have led me to this point. I wonder how I made it this far, and I honestly am amazed at all the people that have aided me in this process. I realize that I cannot list everyone who has offered their support, their suggestions, or simply their friendship, but I am grateful to each person who has played a role in shaping me and dealing with me.

I thank my parents who have seen me through this life. They taught me that I can succeed at whatever I pursue. They instilled values and ethic in me. They showed me that I was truly loved, and no matter where I went in life, they would support me. They encouraged me in my pursuit of knowledge, and when I needed an ear, they were willing to listen. They have given me more than I can ever repay, and I appreciate all that they have done for me.

To my wife, Megan, thank you. From the initial days of courtship through our engagement and wedding, you saw me in the good times and the bad. You encouraged me to complete this chapter in my life. You also gave me distractions that were needed. Your willingness to support me has given me the extra energy to finish. I continue to be blessed by you daily.

I offer my gratitude to Professors Ken Brown and Cynthia Pfeifer-Hill at Northwestern Oklahoma State University. They encouraged me to keep moving forward even when I did not know which direction to take. They drove me to succeed. They challenged me to limits that I did not know existed. Prof. Brown showed me that seeking the answer means knowing the right question, and Prof. Pfeifer-Hill inspired me to study

amyloid diseases through a simple *Nova* video. They have been there through my time beyond Northwestern, and I appreciate their continued friendship.

I owe Prof. Ralph A. Wheeler who was willing to accept me into his group though I lacked any research experience. He embraced my idea to use molecular dynamics to study amyloid diseases. He helped me as I applied for predoctoral fellowships. He gave me an opportunity to flourish. He has supported my endeavors, and when my research was not going well for four years, he never lost sight that I would eventually succeed at my goal.

My graduate committee consisted of Profs. Cook, West, Yip, Nagle, and formerly Ballard. They allowed me to find my way. I stumbled around figuring out what I was supposed to do, and they saw me through it all. They offered advice, and they were willing to write letters of recommendation when asked. Even when I thought that I may fail the General Exam, they saw something more in me than I perceived allowing me to pass. I appreciate their willingness to be on my committee.

I am thankful for my present and former colleagues in the Wheeler research group, particularly Adam Hixson, Scott Boesch, Zunnan Huang, Jermont Chen, Zhanyong Guo, Kurt Brorsen, and Caron Provatopolous. I consider it an honor and a privilege to have worked with you.

I am grateful to the students and faculty within the Biochemistry Division. Through your friendships and conversations, I have learned more about biochemistry. We may not fully comprehend each other's research, but you have allowed me to learn about the experimental research that is available. Through the seminars, you gave me the opportunity to share my research to an audience who does not understand the finer points

of computational biochemistry, but I learned to hone my presentation skills such that you could have a glimpse into my world.

I am appreciative to my friends outside of the department. You may not have fully comprehended what I was doing and what I went through, but you were willing to stand by my side. You allowed me to unwind and to enjoy life outside of research. You encouraged me and accepted my quirks. You granted me the opportunity to seem normal and to have a life. Thank you.

I offer my gratitude to the support staff. The office staff worked with me and guided me when I had no idea what needed to be done. The other staff offered a friendly smile when I would seek assistance or simply want to converse. My time at the University has been enriched by your kindness.

Finally, I thank the sources who provided me with funding, not only financially, but through computer time. The National Institutes of Health provided me with a predoctoral fellowship for three years (5 F31 GM067560). The Department of Education provided me with a Graduate Assistance in Areas of National Need (GAANN) fellowship for my final semester. The National Science Foundation and the Oklahoma Supercomputing Center for Education and Research (OSCER) allowed me access to their supercomputers for research.

Table of Contents

Acknowledgements.....	iv
Table of Contents.....	vii
List of Tables	ix
List of Figures.....	x
List of Abbreviations	xi
Abstract.....	xi
1. Introduction and background: molecular dynamics simulations and more specialized techniques for locating global energy minima.....	1
1.1. Introduction.....	1
1.2. Simulation of biomolecules with molecular dynamics.....	4
1.3. Potential energy (PE) surface.....	4
1.4. MD simulations of polypeptides.....	7
1.4.1. Components of a MD simulation.....	7
1.4.2. Types of MD simulations.....	13
1.4.3. Disrupted Velocity (DIVE) and Divergent Path (DIP) search protocols..	17
1.5. References.....	21
2. Finding the most stable conformation for C-peptide of ribonuclease A.....	29
2.1. Introduction.....	29
2.2. Algorithms and simulations	32
2.2.1. Algorithms	32
2.2.2. Simulation details.....	33
2.2.3. Analysis.....	36
2.3. Simulation results and discussion.....	39
2.3.1. Divergent path (DIP) simulations.....	39
2.3.2. Disrupted velocity (DIVE) simulations	46
2.4. Conclusions.....	50
2.5. References.....	52
3. Exploring low-potential energy structures of a tryptophan zipper	56
3.1. Introduction.....	56
3.2. Algorithms and simulations	58
3.2.1. Theory.....	58
3.2.2. Data collection	60
3.2.3. Data analysis.....	62
3.3. Simulation results and discussion.....	64
3.3.1. Divergent path (DIP) simulations.....	64
3.3.2. Disrupted velocity (DIVE) simulations	70
3.4. Conclusions.....	74
3.5. References.....	77
4. Structure of the B1 domain(41–56) of protein G as a β hairpin	82
4.1. Introduction.....	82
4.2. Simulation protocol.....	85
4.2.1. Protocol description and simulation setup	85
4.2.2. Data analysis.....	89
4.3. Simulation results and discussion.....	92

4.3.1.	Divergent path (DIP) results and discussion.....	92
4.3.2.	Disrupted velocity (DIVE) results and discussion.....	100
4.4.	Conclusions.....	104
4.5.	References.....	106
5.	Exploration of the potential energy surface of $\beta\beta\alpha 5$	111
5.1.	Introduction.....	111
5.2.	Simulation details.....	113
5.3.	Simulation results and discussion.....	119
5.3.1.	Divergent path (DIP) simulations.....	119
5.3.2.	Disrupted velocity (DIVE) simulations.....	129
5.4.	Conclusions.....	134
5.5.	References.....	136
6.	Prediction of a protein's unknown conformation.....	138
6.1.	Introduction.....	138
6.2.	Simulation procedure and analysis.....	139
6.3.	Simulation results and discussion.....	143
6.3.1.	Divergent path (DIP) results and discussion.....	143
6.3.2.	Disrupted velocity (DIVE) results and discussion.....	151
6.4.	Conclusions.....	154
6.5.	References.....	156
7.	Conclusions.....	160
7.1.	Closing thoughts.....	160
7.2.	References.....	162

List of Tables

Table 2.1 List of coupling constants for C-peptide of RNase A.....	38
Table 4.1 List of coupling constants for B1 domain(41–56) of protein G	91
Table 6.1 Results of cluster analysis	148

List of Figures

Figure 1.1 Disconnectivity graph	6
Figure 1. Molecular dynamics cycle	9
Figure 1.3 Conformation selection	17
Figure 1.4 Typical DIP simulation	18
Figure 2.1 Overlay of 32 NMR models of the C-peptide of RNase A	29
Figure 2.2 Geometric orientation of aromatic rings	37
Figure 2.3 Average conformations of C-peptide from DIP simulations.....	41
Figure 2.4 Disconnectivity graph	45
Figure 2.5 Potential energy graph of DIP simulations for C-peptide	46
Figure 2.6 C-peptide conformations from DIVE simulations	47
Figure 3.1 Trpzip2.....	56
Figure 3.2 Average conformation of trpzip2 from DIP simulations.....	66
Figure 3.3 Potential energy graph of DIP simulations for trpzip2.....	69
Figure 3.4 Disconnectivity graph	70
Figure 3.5 Trpzip2 conformations from DIVE simulations	71
Figure 4.1 B1 domain of Streptococcal protein G	82
Figure 4.2 Geometric orientation of aromatic rings	90
Figure 4.3 B1 domain(41–56) of protein G average conformations from DIP simulations.	94
Figure 4.4 Potential energy graph of DIP simulations for B1 domain(41–56) of protein G.	98
Figure 4.5 Disconnectivity graph	99
Figure 4.6 B1 domain(41–56) of protein G conformations from DIVE simulations.....	101
Figure 5.1 $\beta\beta\alpha 5$	112
Figure 5.2 Initial conformations for $\beta\beta\alpha 5$ simulations.....	114
Figure 5.3 Selection of conformations from DIVE simulations.....	117
Figure 5.4 Average $\beta\beta\alpha 5$ conformations from DIP simulations.....	120
Figure 5.5 Average $\beta\beta\alpha 5$ conformations (cont'd)	122
Figure 5.6 Overlay of initial two conformations from DIVE simulations.....	125
Figure 5.7 Potential energy graph of DIP simulations for $\beta\beta\alpha 5$	128
Figure 5.8 Disconnectivity graph	128
Figure 5.9 $\beta\beta\alpha 5$ conformations from DIVE simulations.....	130
Figure 6.1 C-domain of Ure2p as a dimer	138
Figure 6.2 Average Ure2p(1–64) conformations from DIP simulations	144
Figure 6.3 Potential energy graph of DIP simulations for Ure2p(1–64)	150
Figure 6.4 Ure2p(1–64) conformations from DIVE simulations	153
Figure 7.1 Disconnectivity graph	161

List of Abbreviations

AMBER	Assisted model building with energy refinement
DIVE	disrupted velocity
DIP	divergent path
EXACT	ensembles extracted from atomic coordinate transformations
FE	free energy
GB/SA	Generalized Born/surface area
LES	locally-enhanced sampling
MD	molecular dynamics
MMTK	Molecular Modeling ToolKit
MMTSB	Multiscale Modeling Tools for Structural Biology
NMR	nuclear magnetic resonance
PE, V	potential energy
REMD	replica exchange molecular dynamics
RNase	ribonuclease
r.m.s.d.	root mean square deviation
STRIDE	STRuctural IDentification program
trpzip	tryptophan zipper

Abstract

Two MD protocols were recently developed that allow a polypeptide to search the PE surface in search of the global PE minimum, which should correspond to the experimental structure. α -helical secondary structures have previously been tested. We tested an additional α helix (C-peptide of ribonuclease A), and we extended the tests with two β -hairpin secondary structures (tryptophan zipper 2 and the B1 domain(41–56) of protein G). For the C-peptide of ribonuclease A, the α helix was the dominate secondary structure, but a β hairpin was found, which to our knowledge had not previously been reported. For the tryptophan zipper 2 and the B1 domain(41–56) of protein G, the β hairpin was reproduced but alternative conformations were also found. After these test cases, we simulated a small protein ($\beta\beta\alpha 5$) that contained both secondary structural motifs and an overall tertiary structure. The secondary structures were reproduced, but the tertiary structure was not maintained. Finally, we attempted to predict possible conformations for a 64-residue protein, Ure2p, which is implicated in amyloid diseases of yeast. We found that Ure2p(1–64) was dominated by helical conformations. The DIVE and DIP protocols will need to be tested further with different polypeptides and proteins and using more recent force fields. Ure2p(1–64) should be simulated from additional secondary structures such as a β sheet or a combination of α helices and β hairpins.

1. Introduction and background: molecular dynamics simulations and more specialized techniques for locating global energy minima

1.1. Introduction

Molecular dynamics (MD) simulations have been used for several years now to characterize and monitor the motions of biomolecules such as proteins, DNA, and carbohydrates (1-3). As the years have progressed, both computer power and speed have increased considerably. This has allowed simulations to run longer, as evidenced by a number of long simulations such as the monumental 1- μ s simulation of the villin headpiece (4). The passage of time has also given rise to various methods for sampling a huge variety of conformations, including, but not limited to, locally-enhanced sampling (5), ensembles extracted from atomic coordinate transformations (EXACT approximation) (6-8), simulated annealing (9), and replica exchange (10-13).

The various methods mentioned all came about because of the need to locate the global potential or free energy minimum. MD was found to have an inherent problem — the inability to overcome energy barriers within a limited simulation time. Ideally, a protein would sample all conformations during a simulation, but the simulation would require almost an infinite amount of time compared to the actual time that it requires to locate the native conformation (14, 15). MD simulations can become trapped within a particular local PE well, or minimum, which limits the number of conformations sampled during the simulation. According to Anfinsen (16), the native structure represents the conformation within the global free energy minimum. The specialized methods listed previously were developed to allow the protein to overcome energy barriers and sample

more conformations within the limited simulation time in search of the native conformation.

Most recently, replica exchange molecular dynamics (*10-13*) (REMD) has become a popular choice for simulations because of its versatility. REMD allows a system to sample multiple conformations within a limited amount of simulation time and offers the possibility of mapping the free energy surface (FE surface) of the system. REMD begins with an initial conformation and creates multiple copies, or replicas. Each replica begins at a different initial temperature, and as the coordinate histories progress, replica temperatures are swapped on the basis of the Boltzmann criterion. This, in turn, allows the system to travel along the energy surface and become equilibrated before kinetic energy is inserted to encourage the system to overcome energy barriers. When the simulation ends, several conformations have been located, and upon analysis, the conformation with the lowest relative free energy can be determined.

Despite its benefits, REMD has its limitations in sampling an optimal number of conformations within the limited simulation time. Moreover, REMD primarily assists movement over energy barriers instead of circumventing the barriers. REMD, like most MD techniques, may be limited in its capabilities and therefore, has not been used much for protein structure prediction. Recently, two molecular dynamics protocols were developed within our research group (*17, 18*), which assist both energy barrier crossing and energy barrier circumvention. The first MD protocol, disrupted velocity (DIVE) search protocol, perturbs both the energy and the atomic velocity direction at a constant energy. The perturbations allow the system to traverse the potential energy surface (PE surface) in search of the global potential energy minimum near 0 K. The second

protocol, divergent path (DIP) search protocol, permits the system to travel along the PE surface at a constant temperature, e.g., physiological temperatures. The system begins with an initial conformation traveling in several different directions simultaneously yet independently, which permits the overall system to search the PE surface for the global minimum with less chance of becoming trapped within a single local minimum.

When the two MD protocols are used together, a broader picture of the PE surface can be painted. The system is given an opportunity to sample multiple conformations seeking the global potential energy minimum near 0 K. The multiple conformations can then be run using DIP at a desired constant temperature. The system, in turn, is allowed to locate the local energy minima at the desired temperature. Ideally, one or more conformations should converge upon the experimental structure and should have the lowest relative energy, according to Anfinsen (*16*). Our group has been rigorously testing the two protocols (*17, 18*).

Our efforts to test the new MD protocols are presented in chapters 1–5. In chapter 1, we begin with a well-known α -helical polypeptide, C-peptide (residues 1–13) of bovine pancreatic ribonuclease A as a continuation of testing the DIVE and DIP protocols with an α -helical conformation. We then simulate two different β hairpins, which are more difficult to reproduce than α helices, in chapters 3 and 4. We then attempt to reproduce the secondary and tertiary structures of a de novo designed protein whose tertiary structure containing a β hairpin and an α helix, a $\beta\beta\alpha$ motif, in chapter 5. Finally, we offer possible conformations for the N-domain (residues 1–64) of a yeast prion protein in chapter 6.

General reasons for simulating biomolecules with MD techniques are presented in section 1.2. A description of the PE surface is offered in section 1.3. As a general background, MD, various MD techniques in use, and finally the DIVE and DIP protocols are described in more detail in section 1.4.

1.2. Simulation of biomolecules with molecular dynamics

Proteins are an essential part of organisms; they aid in cellular structure and function. Proteins are folded into secondary and tertiary structures depending upon the function of the protein. The folding of a protein within a cell is rather complex and mimicking the folding process experimentally is not simple (19). If a protein “misfolds” or folds into a conformation that is non-native, functionality is lost, and in some cases, aggregation and fibril formation can occur (19-23).

MD simulations complement experimental techniques. Because proteins fold at a rapid rate, some experimental techniques are unable to observe the rapid transition from an unfolded state into the native state, but MD offers a way to examine the folding process (1, 24-27). MD simulations allow for observation of the protein-solvent interactions that cannot be easily seen by experimental techniques, and through the concept of microreversibility, the unfolding and limited folding events of a protein can be witnessed.

1.3. Potential energy (PE) surface

Potential energy (PE, for example eq. (1.1)) V is function of relative atomic positions. In eq. (1.1), q_1 and q_2 are the charges of the two atoms, ϵ_0 is the dielectric constant of the medium between the two atoms ($\epsilon_0 = 1.0$ for gas, $\epsilon_0 = 78.5$ for water), and r is the distance between the two atoms. With the addition of PE and kinetic energy E_k ,

the energy of motion, the internal energy U can be calculated (eq. (1.2)). When temperature T and entropy S are also considered, the free energy A of a system can be determined (eq. (1.3)). In this case, A corresponds to Helmholtz free energy because the system will be at a constant volume; if Gibbs free energy G is used (eq. (1.5)), the system will be under constant pressure (eq. (1.4)).

$$V = \frac{q_1 q_2}{4\pi\epsilon_0 r} \quad (1.1)$$

$$U = V + E_k \quad (1.2)$$

$$A = U - TS \quad (1.3)$$

$$H = U + PV \quad (1.4)$$

$$G = H - TS \quad (1.5)$$

MD simulations account for movement along the PE surface. Typically, free energy is considered when observing the conformations found within a simulation because the structure representing the global free energy minimum should correspond to the experimental structure. PE, however, is easier to calculate, and the global free energy minimum corresponds either to a low PE minima or to the global PE minimum (28).

Wales, et al. (29-35) have studied the PE surface for various systems and then related their findings to the free energy (FE) surface. They have noted that the FE surface may correspond to a smooth funnel compared to the rough PE surface. In essence, as the PE decreases over its rough surface, the FE surface generally descends into its global minimum. The Wales group has also categorized typical PE surfaces on the basis minima and barriers (33-35). This is evident in their depiction of disconnectivity graphs (Figure 1.1).

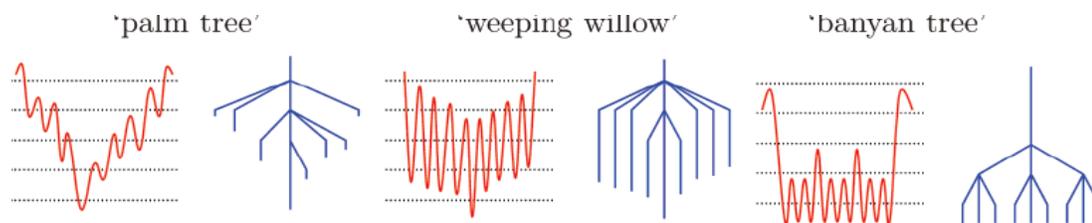


Figure 1.1 Three pairs of PE landscapes (left) and their corresponding disconnectivity graphs (right). The graphs are drawn as PE (vertical axis, arbitrary units) relative to some unspecified coordinates. The endpoints of the disconnectivity graphs represent PE minima, and the points where the branches are joined correspond to a common PE “superbasin”. For the PE landscape, the wells represent the minima of a system surrounded by PE barriers. Reprinted by permission from (33). Copyright 2006 American Chemical Society

The ‘palm tree’ graph is represented by a potential energy landscape with relatively low potential energy barriers and a well-defined global minimum. As the system moves along the PE surface from a relatively high energy state, several transition states may be encountered before the global potential energy minimum is found. According to the Wales group, proteins tend to fold along this particular pathway. The ‘weeping willow’ graph is indicative of a potential energy landscape still with a well-defined global minimum but with higher energy barriers between minima. A system can become trapped within a local energy minimum and lack the kinetic energy to overcome the high energy barriers. Unlike its counterparts, the ‘banyan tree’ graph is found when several minima with comparable energies exist but no distinct global energy minimum exists. Several energy barriers of varying heights are present within the ‘banyan tree’ model.

Kinetic energy traps are another concern encountered during simulations (28, 31-33, 35). Kinetic energy traps occur when the system locates a PE well which is unable to leave during the simulation time. The minimum may, in essence, be a local PE minimum; if the temperature is increased, the system is frequently able to cross the PE barrier in search of the global PE minimum.

1.4. MD simulations of polypeptides

1.4.1. Components of a MD simulation

Molecular dynamics (MD) is a computational method allowing one to observe the motions of molecules based upon Newtonian motion (36). Unlike its quantum mechanical counterpart, classical mechanics allows large systems such as biomolecules to be simulated within short time periods. This allows for the observation of protein folding in solution. MD is an iterative, multi-step process, which involves the calculation of coordinates, atomic velocities, and acceleration from the classical equations of motion (Figure 1.2). Several components are required for a MD simulation to function properly, and a few of the components are described in the following paragraphs.

One of the first components to consider is the description of atomic interactions, which equates to the potential energy of the system — the force field. The atomic interactions are typically described as the generic force field defined by eq. (1.6) (37). Bond stretching, angle bending, and angle twisting are defined by the first three terms in the equation. These terms describe interactions between atoms separated by one, two, and three covalent bonds, respectively. For interactions between atoms separated by more than three bonds, Lennard-Jones and electrostatic (or Coulombic), interactions between two atoms are described by the fourth term. The attraction (r^6) and repulsion (r^{12}) of two atoms by van der Waals interactions are modeled by Lennard-Jones interactions. The electrostatic forces between two atoms via atom-centered partial charges are represented by Coulombic interactions.

Within eq. (1.6), the bond lengths and bond angles are represented by a single harmonic oscillator. b_0 represents the bond length at equilibrium, and the equilibrated

bond angle is symbolized by Θ_0 . b and Θ are the bond length and bond angle at any given time during the simulation. K_b and K_Θ are the harmonic oscillator constants for bond stretching and angle bending, respectively. The angle twisting term is denoted by a truncated Fourier series where K_φ , n , φ , and δ stand for magnitude, periodicity, instantaneous dihedral angle, and phase of the torsion, respectively. For the non-bonded term, r is the distance between two atoms; A and C are constants related to the Lennard-Jones well depth and radii of the two atoms, respectively. The first term relates to the repulsion potential between the two atoms, whereas the second term defines the attraction potential between the two atoms. q_1 and q_2 are the atomic partial charges and ϵ is the environmental dielectric constant ($\epsilon = 1.0$ in gas and $\epsilon = 78.5$ in implicit water solvent).

$$\begin{aligned}
 V = & \frac{1}{2} \sum_{\substack{\text{bond} \\ \text{lengths}}} K_b (b - b_0)^2 + \frac{1}{2} \sum_{\substack{\text{bond} \\ \text{angles}}} K_\Theta (\Theta - \Theta_0)^2 \\
 & + \frac{1}{2} \sum_{\substack{\text{dihedral} \\ \text{angles}}} K_\varphi [1 + \cos(n\varphi - \delta)] + \sum_{\substack{\text{nonbonded} \\ \text{pairs}}} \left(\frac{A}{r^{12}} - \frac{C}{r^6} + \frac{q_1 q_2}{\epsilon r} \right)
 \end{aligned} \tag{1.6}$$

Atomic positions either are arbitrarily assigned on the basis of distances between covalent bonds or are obtained by preassigned coordinates (e.g., a Protein Data Bank file). Typically, velocity vectors are randomly assigned through a Gaussian distribution (39), and the magnitudes are scaled to a constant temperature. As Figure 1.2 illustrates, the force upon the system and the acceleration can then be calculated. The force upon the system is determined by the potential energy from the force field and the positions of all atoms included in the system (eq. (1.9)).

$$\mathbf{f}_i(t) = - \frac{\delta V(\mathbf{r}_1(t), \mathbf{r}_2(t), \dots, \mathbf{r}_n(t))}{\delta \mathbf{r}_i(t)} \tag{1.7}$$

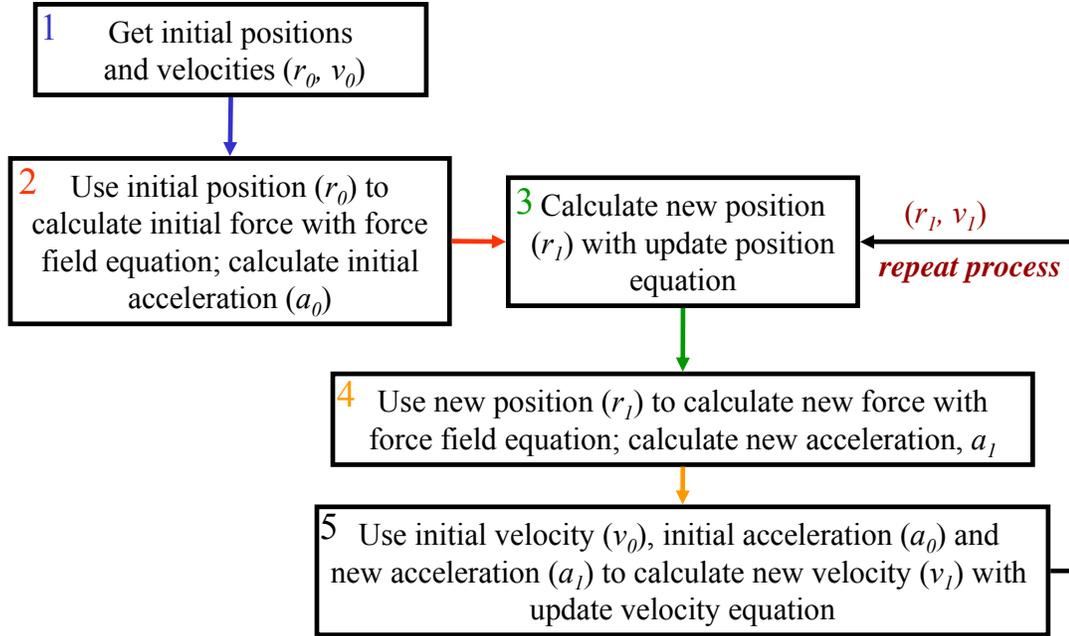


Figure 1.2 Molecular dynamics cycle (courtesy of Jermont Chen) (38)

\mathbf{f}_i is the net force on atom i from all atoms at positions $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i$. From this, atomic acceleration can be evaluated via $\mathbf{F} = m\mathbf{a}$ where \mathbf{a} is defined as

$$\mathbf{a}_i(t) = \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \frac{\mathbf{f}_i(t)}{m_i} \quad (1.8)$$

\mathbf{a}_i is the acceleration of atom i , which has mass m_i and force \mathbf{f}_i . Eqs. (1.7) and (1.8) are then iterated during the simulation until the predetermined number of steps is completed.

The Amber force fields are utilized within our simulations and define specific values of the constants in eq. (1.6). Four different Amber force field definitions are in popular use today: Amber94 (37), Amber96 (40), Amber99 (41), and Amber03 (42). The first three differ only by torsion angle definitions. Subtle nuances have arisen, which affect the folding of proteins. Amber03 (42) has both different definitions of the torsion angles and different partial charges. Extended/ β -hairpin conformations are favored by Amber96 (40), whereas Amber94 (37) and Amber99 (41) are biased towards helical conformations (43-49). To counter these biases, variations of Amber99 (41) have been

developed to achieve a better balance between α helices and β sheets and thus allow proteins to fold to conformations resembling experimental data (45, 46, 50-52).

Simulations can be run in either a gas phase or within a solvent. Biomolecules are typically simulated in water. Two types of water solvent models are typically used — explicit and implicit. As the name implies, explicit water solvent has individual water molecules surrounding the solute molecule(s). Explicit solvent has been represented in several forms including, but not limited to, TIP3P (53), TIP4P (54), TIP5P (55), and SPC/E (56). AMBER (57-59) typically uses one of these models. Explicit solvent allows for a more accurate representation of molecular motion within water. However, explicit water requires more computational time because of the increased number of interatomic distances calculated during each step of an MD simulation compared with gas phase.

Implicit solvent (60-64), also called continuum models of solvent, offers a solution to the computational strain by removing the necessity of tracking actual atomic waters. Solvent is represented by a mathematical equation (eq. (1.9)) modeling a dielectric continuum. The system is simulated in a medium similar to gas phase with a dielectric constant added to represent a homogenous solvent. Therefore, the system is provided an approximation of solvation effects, is allowed to move freely in space, and is able to search for different conformations without the limitation of explicit solvent intervening between different parts of a solute such as a protein.

In eq. (1.9), G_{pol} defines the solvation free energy of the solute-solvent electrostatic polarization term, and ϵ_p is the dielectric value inside of the protein, and ϵ_w is the solvent dielectric constant. r_{ij} determines the separation distance of particles i and j , q_i and q_j represent the atomic charges, and α_i and α_j correspond to the effective Born radii

(65). f_{gb} symbolizes a complex function of r_{ij} , α_i and α_j . The effective Born radius determines the charge distance between the solute and the continuum dielectric boundary. Typically, a simulation defines the following parameters: $\epsilon_w = 78.5$, $\epsilon_p = 1.0$, a surface tension of $0.005 \text{ kcal/mol-Å}^2$, and $r_{ij} = 0.9 \text{ Å}$.

$$G_{pol} = -\frac{1}{2} \left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w} \right) \sum_{ij} \frac{q_i q_j}{f_{gb}}$$

$$f_{gb} = \left[r_{ij}^2 + \alpha_i \alpha_j \exp \left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j} \right) \right]^{\frac{1}{2}} \quad (1.9)$$

Implicit solvent does have its disadvantages. Implicit solvent tends to overstabilize salt bridges and does not always mimic the free energy surface of explicit solvent (66). One reason for the overstabilization of salt bridges stems from the exposure of side chains, which allows for electrostatic interactions between amino acid side chains that may otherwise be hidden within a solvation shell (67). According to Geney and coworkers (68), another reason for non-native salt bridge formation is that the intrinsic radius of hydrogen bonded to charged nitrogen is larger than necessary. Bias towards helices probably stems from the adjustment in hydrogen bonds present (69). Mortenson and Wales (28) have found that different implicit solvents can affect the PE surface; the surface can either exhibit the “banyan tree” model or the “palm tree” model depending upon the definition of the dielectric constant.

As Figure 1.2 depicts, the atomic velocities for the system must be calculated based upon eqs. (1.6), (1.7), and (1.8). The method selected for our simulations is known as the velocity-Verlet method (70). This method allows the force, atomic velocity, and atomic acceleration to be calculated at the same time step, Δt , as defined by the two equations.

$$\dot{r}(t + \Delta t) = \dot{r}(t) + \dot{v}(t)\Delta t + \frac{1}{2}\dot{a}(t)\Delta t^2 \quad (1.10)$$

$$\dot{v}(t + \Delta t) = \dot{v}(t) + \frac{1}{2}[\dot{a}(t) + \dot{a}(t + \Delta t)]\Delta t \quad (1.11)$$

The accuracy of the forces, atomic velocity, and atomic acceleration depend upon Δt , which is called the time step. The time step must be approximately an order of magnitude smaller than the fastest bond stretching vibration. If the time step is too large, accuracy is lost and the energy of the system can increase rapidly. Within biomolecules, the fastest vibrations occur between atoms covalently bonded to hydrogen. Bond length constraints (SHAKE (71, 72)) are used to limit the motion of these vibrations allowing for a larger time step to be used.

$$\dot{r}_{ij}^2 - d_{ij}^2 = 0 \quad (1.12)$$

Eq. (1.12) defines the ideal way in which SHAKE should respond. As two atoms move during the simulation, the bond constraint counters the vibrational movement to prevent overstretching. This, however, is ideal; in reality, a tolerance level, ϵ , is set giving the stretching an upper limit (eq. (1.13)).

$$s_{ij} = (\dot{r}_{ij}^2 - d_{ij}^2)/d_{ij} < \epsilon \quad (1.13)$$

Up to this point, our description of MD simulations has assumed that a molecule will be simulated in an ensemble with a constant number of atoms, constant volume, and constant total energy (constant N, V, E). For some simulations, a microcanonical simulation is acceptable, but to mimic reality, temperature must remain constant, in a constant (N, V, T) simulation. The constant N,V,T simulation can be achieved by coupling the system of interest to a larger system, or thermal bath. The bath allows heat to be transferred into and out of the system.

Several methods exist to keep the temperature constant during a simulation. The oldest, and perhaps simplest, method rescales the atomic velocities to maintain the desired temperature (73). This method, invented by Berendsen (74) provides a weak-coupling algorithm by utilizing an atomic velocity-rescaling constant for all atoms, but the method does not give the correct ensemble properties.. Imaginary random collisions offers another way to randomize atomic velocities (75, 76); however, the collisions create discontinuous changes in atomic velocities, which means that a PE surface cannot be easily mapped.

A more common algorithm for temperature control is the Nosé-Hoover chain method (75-77). This method creates a chain of thermostats attached to the system. In other words, the system is coupled to a thermal bath A , and thermal bath A is coupled to thermal bath B , etc. The chained thermostats control the fluctuations of the kinetic energy for the entire system, thermostats included. The Nosé-Hoover chain method (75-77) has been shown to represent a constant temperature simulation well compared to the previous methods (78).

1.4.2. Types of MD simulations

A conventional MD simulation traverses the potential energy (PE) surface in search of the global free energy minimum. The global free energy minimum is assumed to correspond to the native state of the simulated system. One drawback to conventional MD simulations is the entrapment of the simulation into PE minima. Any minimum may or may not be the global PE minimum, and if the minimum is not the global PE minimum, the system has encountered a local PE minimum. When a PE minimum is located, the simulation generally does not continue long enough to leave the minimum

energy geometry. Ideally, a simulation that runs for an infinite amount of time will cover the entire PE surface, entering and leaving local PE minima throughout the course of the simulation, and eventually locating the global PE minimum.

To overcome entrapment in local minima, several techniques have been developed. Most of these techniques utilize creative measures to overcome PE barriers. By “jumping over” barriers, a simulation can move to the next PE minimum in search of the global PE minimum. Four techniques are described in this section: simulated annealing (9) and mean-field methods, including locally-enhanced sampling (5), extracted atomic coordinate transformation (EXACT) approximation (6-8), and replica exchange (10-13). Each technique has its value but also its drawbacks.

Simulated annealing (9) allows a system to overcome PE barriers by perturbing the temperature, i.e., the kinetic energy of the system. The temperature is increased via input of kinetic energy to a designated temperature (e.g., 1000 K), and the system is allowed to “walk” across the PE surface. The temperature is slowly cooled via kinetic energy removal to permit the system to enter gradually into a PE minimum.

A single heating and cooling cycle may lead to entrapment of the system in a local PE minimum (79). The heating and cooling cycle must be repeated multiple times to ensure the discovery of the global PE minimum. Ideally, the global PE minimum should be located with its corresponding conformation after several cycles have been completed. The global PE minimum has the highest probability of being found by using multiple simulated annealing cycles. However, simulated annealing does not guarantee that the global PE minimum will be found because of limitations in simulation time.

Locally-enhanced sampling (5) (LES) offers another alternative to conventional MD simulations. LES takes a system and separates it into two parts: a subsystem of primary interest and a bath containing the less interesting part of the system. Several copies of the more interesting subsystem, such as an amino acid side chain, are simulated. The individual copies travel along the PE surface independent of the other copies, and each copy interacts with the bath. The bath, on the other hand, feels the average interaction of all copies. This allows for a favorable movement of the subsystem of interest to cover a broader area of the PE surface than in a conventional MD simulation while decreasing PE barriers. The reduction in PE barriers via LES was shown to enhance with conformational changes of proteins and peptides (80).

However, LES has some drawbacks (39, 81-83). One major drawback to LES is the appearance of pseudo-minima (39). LES alters the PE surface by decreasing PE barriers, but it also creates local PE minima which do not correspond to real local PE minima. A system, therefore, may find artifacts of the PE surface that have no physical meaning. The global PE minimum is, however, retained in the LES simulation and the global minimum found by using LES is still physically relevant.

Hixson et al. (6-8) derived a rigorous method, the ensembles extracted from atomic coordinate transformation (EXACT) approximation, that can remove for the pseudo-minima by adding an additional mathematical term absent from LES. By accounting for the additional mathematical term, the problems prevalent in LES (39, 80-83) are rigorously addressed by this method. The additional term can also be adjusted such that the EXACT approximation will mimic either LES or conventional MD. The

method has been shown to give the correct conformation and increase sampling on small atomic systems (6, 8) and on melatonin (7).

The replica exchange method (10-13, 84-87) (REM) is another recent MD method. A simulation is begun with multiple copies (replicas) of a system at different initial temperatures (e.g., 1000 K, 600 K, 300 K, 100 K, etc.). Each replica is simulated independently from the other replicas; yet, all simulations are run simultaneously. As the simulation progresses, replicas adjacent in temperature (1000 K and 600 K, for example) are exchanged at a preset time based upon the Boltzmann criterion. This exchange between replicas occurs multiple times throughout the simulation period, which allows for an increased number of minima to be located on the PE surface.

REM locates multiple conformations and can be used to calculate thermodynamic properties such as free energy. However, REM, like other methods, has its limitations. The major limitation that REM encounters is the PE surface area which it traverses. Systems have kinetic energy both input and removed during the simulation period. This exchange of energy allows the system to cross PE barriers but leaves the option of circumventing the barriers to random chance. In other words, several conformations may be found, but the PE range in which the conformations are located may be small.

1.4.3. Disrupted Velocity (DIVE) and Divergent Path (DIP) search protocols

Disrupted velocity and divergent path search protocols were developed by Huang et al. (17, 18) as alternative MD protocols to search for the global PE minimum by mapping the PE surface. DIVE allows the system to travel across the rugged PE surface by crossing and by circumventing PE barriers. The system is, in the process, allowed to cool to temperatures near 0 K, which allows for the location of potential energy minima (Figure 1.3). By having several initial conformations, a broad map of the PE surface can be formed and statistically, the global PE minimum and its corresponding conformation are more likely located.

DIP simulations, on the other hand, do not seek to cross PE barriers in the same fashion as DIVE simulations. Systems are simulated at a constant temperature but have multiple coordinate histories, or trajectories, simultaneously running. The individual trajectories evolve differently by the assignment of a different direction to identical atoms in each trajectory. Thus, a system can cover more of the PE surface at a constant temperature than a single MD simulation can (Figure 1.4). Where a typical MD

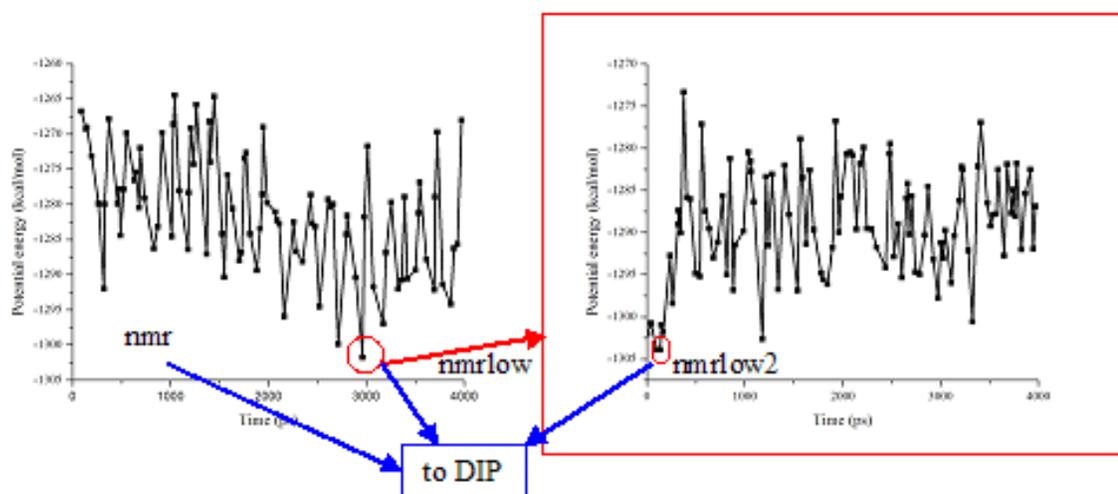


Figure 1.3 How conformations are selected during DIVE simulations for further simulation either with the DIVE or the DIP protocol. Red represents DIVE simulations while the blue indicates a DIP simulation.

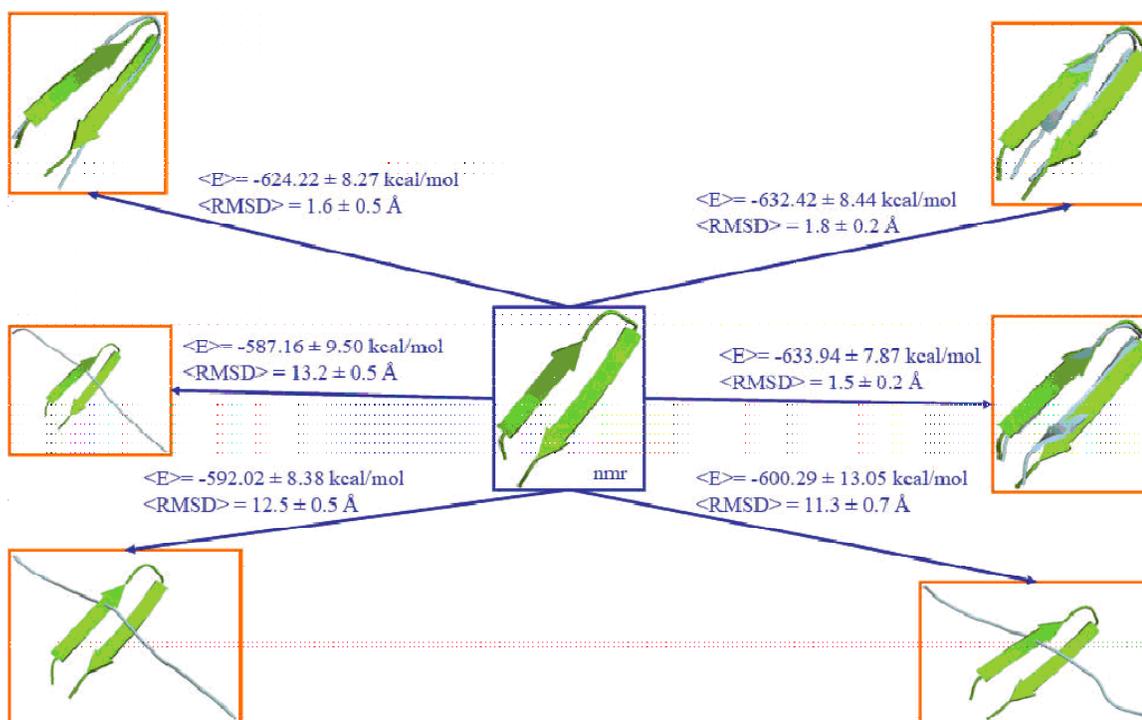


Figure 1.4 Typical DIP simulation from an initial conformation (green) in the center with average conformations (blue) on the left and right. The average conformation is overlaid with the initial conformation. Average potential energies and backbone atom rmsd along with standard deviations are shown relative to NMR model.

simulation may become trapped within a local energy minimum, DIP gives the system an opportunity to circumvent energy barriers by evolving in different directions.

Simulations begin with a predefined set of coordinates from, for example, an experimental structure or a fully extended structure. The coordinates are copied n times in DIP simulations, atoms of the individual copies are each assigned random atomic velocity directions. (DIVE simulations also randomly assign different atomic velocity directions and in addition assign different velocity magnitudes, causing the individual copies to begin with different kinetic energies and therefore different temperatures.) On the other hand, DIP simulations maintain a constant temperature. Thus DIVE simulations are run using the microcanonical ensemble (N, V, E); DIP simulations utilize the canonical ensemble (N, V, T).

During DIVE simulations, the different trajectories are allowed to evolve for a defined time before the atomic velocities are reassigned according to $\mathbf{p}' \equiv \sigma^{1/2} \mathbf{p}$ where \mathbf{p} and \mathbf{p}' are the momenta of the particles before and after atomic velocity reassignment, respectively. σ is the scaling parameter that determines the magnitude of the simulation temperature after atomic velocity reassignment. The atomic velocity magnitudes are therefore rescaled and the directions are changed when the atomic velocities are reassigned.

The scaling parameter, σ , can be set to a value either above or below 1. $\sigma > 1$ indicates heating, and $0 < \sigma < 1$ denotes cooling. On the other hand, the scaling parameter may be calculated from the scaling temperature T by $\sigma = |T - \Delta T|/T$. The scaling temperature is defined as the difference between the temperatures before and after atomic velocity rescaling (ΔT). Alternatively, σ can be defined using a target temperature T_{target} after atomic velocity rescaling with $\sigma = T_{\text{target}}/T$.

A threshold temperature is defined as well. During a simulation, kinetic energy is added when the temperature falls below the threshold, and kinetic energy is removed after a period when the temperature rises above the threshold. This cycle allows the system to cross energy barriers, and because the threshold temperature is typically set near 0 K, PE minima can be mapped within approximately 10 kcal/mol. Heating of the system typically happens once to approximately 500–1000 K during one heating and cooling cycle, whereas cooling of the system occurs multiple times by smaller magnitudes.

DIP simulations, as mentioned earlier, maintain a constant temperature while the individual trajectories are initially assigned different atomic velocity directions based

upon a Gaussian distribution. The trajectories then evolve into unique conformations while traversing the PE surface. One simulation may potentially contain n unique conformations because the individual copies locate different potential energy minima. Huang et al. (88) found that polyalanine can actually locate a β hairpin, which is higher in PE than the expected α helix. This result indicates that DIP simulations can locate conformations that may not typically be found with other MD algorithms; yet, the simulations can still locate conformations in good agreement with experimental data.

Chapters 1–5 present work on various polypeptides used to test the DIVE and DIP protocols. Chapter 1 gives details on a 13-residue α -helical structure, C-peptide of ribonuclease A (89). Chapters 3 and 4 offer two studies using β hairpins: tryptophan zipper (90) and the B1 domain(41-56) of Streptococcal protein G (91). The final test involves a structure containing both common secondary structures of proteins—an α helix and a β hairpin. This polypeptide, $\beta\beta\alpha$ (92-94), offers the challenge of folding both motifs correctly and reproducing their relative orientation correctly. Finally, we attempt to predict the secondary and tertiary structures of the structurally-undetermined N-terminus of Ure2p, which is implicated in prion formation of Ure2p into [URE3] (95-118)

1.5. References

1. Daggett, V. (2001) Molecular dynamics simulations of protein unfolding/folding, *Methods Mol Biol* 168, 215-247.
2. Daggett, V., and Fersht, A. (2003) The present view of the mechanism of protein folding, *Nat Rev Mol Cell Bio* 4, 497-502.
3. Norberg, J., and Nilsson, L. (2003) Advances in biomolecular simulations: methodology and recent applications, *Q Rev Biophys* 36, 257-306.
4. Duan, Y., and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science* 282, 740-744.
5. Elber, R., and Karplus, M. (1990) Enhanced Sampling in Molecular Dynamics: Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon Monoxide Diffusion through Myoglobin, *J Am Chem Soc* 112, 9161-9175.
6. Hixson, C. A., and Wheeler, R. A. (2001) Rigorous classical-mechanical derivation of a multiple-copy algorithm for sampling statistical mechanical ensembles, *Phys Rev E* 6402, -.
7. Hixson, C. A., Chen, J., Huang, Z. N., and Wheeler, R. A. (2004) New perspectives on multiple-copy, mean-field molecular dynamics methods, *J Mol Graph Model* 22, 349-357.
8. Hixson, C. A., and Wheeler, R. A. (2004) Practical multiple-copy methods for sampling classical statistical mechanical ensembles, *Chem Phys Lett* 386, 330-335.
9. Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983) Optimization by Simulated Annealing, *Science* 220, 671-680.
10. Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding, *Chem Phys Lett* 314, 141-151.
11. Sugita, Y., and Okamoto, Y. (2000) Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape, *Chem Phys Lett* 329, 261-270.
12. Sugita, Y., Kitao, A., and Okamoto, Y. (2000) Multidimensional replica-exchange method for free-energy calculations, *J Chem Phys* 113, 6042-6051.
13. Mitsutake, A., Sugita, Y., and Okamoto, Y. (2001) Generalized-ensemble algorithms for molecular simulations of biopolymers, *Biopolymers* 60, 96-123.
14. Levinthal, C. (1968) Are There Pathways for Protein Folding, *J Chim Phys Pcb* 65, 44-&.
15. Levinthal, C. (1969) How to fold gracefully, in *Mossbauer Spectroscopy in Biological Systems, Proceedings of a Meeting Held at Allerton House, Monticelli, Illinois* (Debrunner, P., Tsibris, J. C. M., and Munck, E., Eds.), pp 22-24, University of Illinois Press, Urbana, IL.
16. Anfinsen, C. B. (1973) Principles That Govern Folding of Protein Chains, *Science* 181, 223-230.
17. Huang, Z. N., and Wheeler, R. A. Replica Microcanonical Search Methods for Mapping Potential Energy Landscapes and Conformations of miniproteins, *In preparation*.

18. Huang, Z. (2005) Design of new molecular dynamics global minimum search protocols for mapping energy landscapes and conformations of folded polypeptides and mini-proteins, in *Chemistry and Biochemistry*, p 282, The University of Oklahoma, Norman.
19. Clark, P. L. (2004) Protein folding in the cell: reshaping the folding funnel, *Trends Biochem Sci* 29, 527-534.
20. Prusiner, S. B. (1998) Prions, *P Natl Acad Sci USA* 95, 13363-13383.
21. Lansbury, P. T., Costa, P. R., Griffiths, J. M., Simon, E. J., Auger, M., Halverson, K. J., Kocisko, D. A., Hendsch, Z. S., Ashburn, T. T., Spencer, R. G. S., Tidor, B., and Griffin, R. G. (1995) Structural Model for the Beta-Amyloid Fibril Based on Interstrand Alignment of an Antiparallel Sheet Comprising a C-Terminal Peptide, *Nat Struct Biol* 2, 990-998.
22. Lazo, N. D., and Downing, D. T. (1997) beta-helical fibrils from a model peptide, *Biochem Bioph Res Co* 235, 675-679.
23. Tjernberg, L. O., Callaway, D. J. E., Tjernberg, A., Hahne, S., Lilliehook, C., Terenius, L., Thyberg, J., and Nordstedt, C. (1999) A molecular model of Alzheimer amyloid beta-peptide fibril formation, *J Biol Chem* 274, 12619-12625.
24. Daggett, V. (2000) Long timescale simulations, *Curr Opin Struc Biol* 10, 160-164.
25. Daggett, V. (2002) Molecular dynamics simulations of the protein unfolding/folding reaction, *Accounts of Chemical Research* 35, 422-429.
26. Daggett, V., and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding?, *Trends Biochem Sci* 28, 18-25.
27. Beck, D. A. C., and Daggett, V. (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution, *Methods* 34, 112-120.
28. Mortenson, P. N., and Wales, D. J. (2001) Energy landscapes, global optimization and dynamics of the polyalanine Ac(ala)(8)NHMe, *J Chem Phys* 114, 6443-6454.
29. Doye, J. P. K., and Wales, D. J. (1996) On potential energy surfaces and relaxation to the global minimum, *J Chem Phys* 105, 8428-8445.
30. Wales, D. J., and Scheraga, H. A. (1999) Review: Chemistry - Global optimization of clusters, crystals, and biomolecules, *Science* 285, 1368-1372.
31. Carr, J. M., Trygubenko, S. A., and Wales, D. J. (2005) Finding pathways between distant local minima, *J Chem Phys* 122.
32. Evans, D. A., and Wales, D. J. (2004) Folding of the GB1 hairpin peptide from discrete path sampling, *J Chem Phys* 121, 1080-1090.
33. Wales, D. J., and Bogdan, T. V. (2006) Potential energy and free energy landscapes, *J Phys Chem B* 110, 20765-20776.
34. Komatsuzaki, T., Hoshino, K., Matsunaga, Y., Rylance, G. J., Johnston, R. L., and Wales, D. J. (2005) How many dimensions are required to approximate the potential energy landscape of a model protein?, *J Chem Phys* 122, -.
35. Wales, D. J. (2005) The energy landscape as a unifying theme in molecular science, *Philos T Roy Soc A* 363, 357-375.
36. Mitchell, M. (1991) Molecular Dynamics.
37. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A

- Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J Am Chem Soc* 117, 5179-5197.
38. Chen, J. (2003) Tests of a new molecular dynamics method for enhanced conformational sampling, in *Chemistry and Biochemistry*, p 58, University of Oklahoma, Norman.
 39. Stultz, C. M., and Karplus, M. (1998) On the potential surface of the locally enhanced sampling approximation, *J Chem Phys* 109, 8809-8815.
 40. Kollman, P., Dixon, R., Cornell, W., Fox, T., Chipot, C., and Pohorille, A. (1997) The development/application of a \"minimalist\" organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data, *Computer Simulation of Biomolecular Systems, Volume 3*, 83-96.
 41. Wang, J. M., Cieplak, P., and Kollman, P. A. (2000) 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.
 42. Duan, Y., and Kollman, P. A. (2001) Computational protein folding: From lattice to all-atom, *Ibm Syst J* 40, 297-309.
 43. Hu, H., Elstner, M., and Hermans, J. (2003) Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine \"dipeptides\" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution, *Proteins* 50, 451-463.
 44. Mackerell, A. D. (2004) Empirical force fields for biological macromolecules: Overview and issues, *J Comput Chem* 25, 1584-1604.
 45. Simmerling, C., Strockbine, B., and Roitberg, A. E. (2002) All-atom structure prediction and folding simulations of a stable protein, *J Am Chem Soc* 124, 11258-11259.
 46. Sorin, E. J., and Pande, V. S. (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations, *Biophys J* 88, 2472-2493.
 47. Wang, T., and Wade, R. C. (2006) Force field effects on a beta-sheet protein domain structure in thermal unfolding simulations, *J Chem Theory Comput* 2, 140-148.
 48. Yoda, T., Sugita, Y., and Okamoto, Y. (2004) Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations, *Chem Phys* 307, 269-283.
 49. Yoda, T., Sugita, Y., and Okamoto, Y. (2004) Comparisons of force fields for proteins by generalized-ensemble simulations, *Chem Phys Lett* 386, 460-467.
 50. Garcia, A. E., and Sanbonmatsu, K. Y. (2002) α -Helical stabilization by side chain shielding of backbone hydrogen bonds, *P Natl Acad Sci USA* 99, 2782-2787.
 51. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins-Structure Function and Bioinformatics* 65, 712-725.

52. Okur, A., Strockbine, B., Hornak, V., and Simmerling, C. (2003) Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins, *J Comput Chem* 24, 21-31.
53. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water, *J Chem Phys* 79, 926-935.
54. Jorgensen, W. L., and Madura, J. D. (1985) Temperature and Size Dependence for Monte-Carlo Simulations of Tip4p Water, *Mol Phys* 56, 1381-1392.
55. Mahoney, M. W., and Jorgensen, W. L. (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, *J Chem Phys* 112, 8910-8922.
56. Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987) The Missing Term in Effective Pair Potentials, *J Phys Chem-US* 91, 6269-6271.
57. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham III, T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (2002) AMBER 7, University of California, San Francisco, San Francisco.
58. Case, D. A., Darden, T. A., Cheatham, I., T.E. , Simmerling, C. L., Wang, J., Duke, R. R. E., Merz, L., K.M. , Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) AMBER 8, 8 ed., University of California, San Francisco, San Francisco.
59. Case, D. A., Darden, T. A., Cheatham, I., T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Pearlman, D. A., Crowley, M., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., and Kollman, P. A. (2006) AMBER 9, 9 ed., University of California, San Francisco, San Francisco.
60. Eisenberg, D., and McLachlan, A. D. (1986) Solvation Energy in Protein Folding and Binding, *Nature* 319, 199-203.
61. Kang, Y. K., Nemethy, G., and Scheraga, H. A. (1987) Free-Energies of Hydration of Solute Molecules .3. Application of the Hydration Shell-Model to Charged Organic-Molecules, *J Phys Chem-US* 91, 4118-4120.
62. Kang, Y. K., Nemethy, G., and Scheraga, H. A. (1987) Free-Energies of Hydration of Solute Molecules .1. Improvement of the Hydration Shell-Model by Exact Computations of Overlapping Volumes, *J Phys Chem-US* 91, 4105-4109.
63. Kang, Y. K., Nemethy, G., and Scheraga, H. A. (1987) Free-Energies of Hydration of Solute Molecules .2. Application of the Hydration Shell-Model to Nonionic Organic-Molecules, *J Phys Chem-US* 91, 4109-4117.
64. Ooi, T., Oobatake, M., Nemethy, G., and Scheraga, H. A. (1987) Accessible Surface-Areas as a Measure of the Thermodynamic Parameters of Hydration of Peptides, *P Natl Acad Sci USA* 84, 3086-3090.

65. Born, M., and von Karman, T. (1912) On fluctuations in spatial grids., *Phys Z* 13, 297-309.
66. Zhou, R. H., and Berne, B. J. (2002) Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water?, *P Natl Acad Sci USA* 99, 12777-12782.
67. Dixit, S. B., Bhasin, R., Rajasekaran, E., and Jayaram, B. (1997) Solvation thermodynamics of amino acids - Assessment of the electrostatic contribution and force-field dependence, *J Chem Soc Faraday T* 93, 1105-1113.
68. Geney, R., Layten, M., Gomperts, R., Hornak, V., and Simmerling, C. (2006) Investigation of salt bridge stability in a generalized born solvent model, *J Chem Theory Comput* 2, 115-127.
69. Nymeyer, H., and Garcia, A. E. (2003) Simulation of the folding equilibrium of alpha-helical peptides: A comparison of the generalized born approximation with explicit solvent, *P Natl Acad Sci USA* 100, 13934-13939.
70. Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A Computer-Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *J Chem Phys* 76, 637-649.
71. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes, *J Comput Phys* 23, 327-341.
72. Andersen, H. C. (1983) Rattle: a Velocity Version of the Shake Algorithm for Molecular Dynamics Calculations, *J Comput Phys* 52, 24-34.
73. Woodcock, L. V. (1971) Isothermal Molecular Dynamics Calculations for Liquid Salts, *Chem Phys Lett* 10, 257-&.
74. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular-Dynamics with Coupling to an External Bath, *J Chem Phys* 81, 3684-3690.
75. Andrea, T. A., Swope, W. C., and Andersen, H. C. (1983) The Role of Long Ranged Forces in Determining the Structure and Properties of Liquid Water, *J Chem Phys* 79, 4576-4584.
76. Hoover, W. G. (1985) Canonical Dynamics - Equilibrium Phase-Space Distributions, *Phys Rev A* 31, 1695-1697.
77. Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992) Nosé-Hoover Chains: the Canonical Ensemble via Continuous Dynamics, *J Chem Phys* 97, 2635-2643.
78. Tuckerman, M. E., and Martyna, G. J. (2000) Understanding modern molecular dynamics: Techniques and applications, *J Phys Chem B* 104, 159-178.
79. Huber, G. A., and McCammon, J. A. (1997) Weighted-ensemble simulated annealing: Faster optimization on hierarchical energy surfaces, *Phys Rev E* 55, 4822-4825.
80. Roitberg, A., and Elber, R. (1991) Modeling Side-Chains in Peptides and Proteins: Application of the Locally Enhanced Sampling and the Simulated Annealing Methods to Find Minimum Energy Conformations, *J Chem Phys* 95, 9277-9287.
81. Straub, J. E., and Karplus, M. (1991) Energy Equipartitioning in the Classical Time-Dependent Hartree Approximation, *J Chem Phys* 94, 6737-6739.

82. Ulitsky, A., and Elber, R. (1993) The Thermal-Equilibrium Aspects of the Time-Dependent Hartree and the Locally Enhanced Sampling Approximations - Formal Properties, a Correction, and Computational Examples for Rare-Gas Clusters, *J Chem Phys* 98, 3380-3388.
83. Zheng, W. M., and Zheng, Q. (1997) An analytical derivation of the locally enhanced sampling approximation, *J Chem Phys* 106, 1191-1194.
84. Sugita, Y., and Okamoto, Y. (2005) Molecular mechanism for stabilizing a short helical peptide studied by generalized-ensemble simulations with explicit solvent, *Biophys J* 88, 3180-3190.
85. Okamoto, Y., Fukugita, M., Nakazawa, T., and Kawai, H. (1991) Alpha-Helix Folding by Monte-Carlo Simulated Annealing in Isolated C-Peptide of Ribonuclease-A, *Protein Eng* 4, 639-647.
86. Okamoto, Y. (2004) Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations, *J Mol Graph Model* 22, 425-439.
87. Okamoto, Y. (1999) Tackling the multiple-minima problem in protein folding by Monte Carlo simulated annealing and generalized-ensemble algorithms, *Int. J. Mod. Phys. C* 10, 1571-1582.
88. Huang, Z. N., and Wheeler, R. A. Trajectory-randomized Replica Canonical Ensembles Simulation for the Folding and Unfolding studies of Polyalanine in GB/SA implicit Solvent Environment, *In preparation*.
89. Santoro, J., Gonzalez, C., Bruix, M., Neira, J. L., Nieto, J. L., Herranz, J., and Rico, M. (1993) High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy, *J Mol Biol* 229, 722-734.
90. Cochran, A. G., Skelton, N. J., and Starovasnik, M. A. (2001) Tryptophan zippers: Stable, monomeric beta-hairpins, *P Natl Acad Sci USA* 98, 5578-5583.
91. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G, *Science* 253, 657-661.
92. Struthers, M. D., Cheng, R. P., and Imperiali, B. (1996) Design of a monomeric 23-residue polypeptide with defined tertiary structure, *Science* 271, 342-345.
93. Struthers, M. D., Cheng, R. P., and Imperiali, B. (1996) Economy in protein design: Evolution of a metal-independent beta beta alpha motif based on the zinc finger domains, *J Am Chem Soc* 118, 3073-3081.
94. Struthers, M., Ottesen, J. J., and Imperiali, B. (1998) Design and NMR analyses of compact, independently folded BBA motifs, *Fold Des* 3, 95-103.
95. Wickner, R. B. (1994) [Ure3] as an Altered Ure2 Protein: Evidence for a Prion Analog in *Saccharomyces Cerevisiae*, *Science* 264, 566-569.
96. Masison, D. C., and Wickner, R. B. (1995) Prion-Inducing Domain of Yeast Ure2p and Protease Resistance of Ure2p Prion-Containing Cells, *Science* 270, 93-95.
97. Masison, D. C., Maddelein, M. L., and Wickner, R. B. (1997) The prion model for [URE3] of yeast: Spontaneous generation and requirements for propagation, *P Natl Acad Sci USA* 94, 12503-12508.

98. Fernandez-Bellot, E., Guillemet, E., Baudin-Baillieu, A., Gaumer, S., Komar, A. A., and Cullin, C. (1999) Characterization of the interaction domains of Ure2p, a prion-like protein of yeast, *Biochem J* 338, 403-407.
99. Maddelein, M. L., and Wickner, R. B. (1999) Two prion-inducing regions of Ure2p are nonoverlapping, *Mol Cell Biol* 19, 4516-4524.
100. Perrett, S., Freeman, S. J., Butler, P. J. G., and Fersht, A. R. (1999) Equilibrium folding properties of the yeast prion protein determinant Ure2, *J Mol Biol* 290, 331-345.
101. Taylor, K. L., Cheng, N. Q., Williams, R. W., Steven, A. C., and Wickner, R. B. (1999) Prion domain initiation of amyloid formation in vitro from native Ure2p, *Science* 283, 1339-1343.
102. Thual, C., Komar, A. A., Bousset, L., Fernandez-Bellot, E., Cullin, C., and Melki, R. (1999) Structural characterization of *Saccharomyces cerevisiae* prion-like protein Ure2., *Faseb J* 13, A1485-A1485.
103. Bousset, L., Thomson, N. H., Radford, S. E., and Melki, R. (2002) The yeast prion Ure2p retains its native alpha-helical conformation upon assembly into protein fibrils in vitro, *Embo J* 21, 2903-2911.
104. Edskes, H. K., and Wickner, R. B. (2002) Conservation of a portion of the *S-cerevisiae* Ure2p prion domain that interacts with the full-length protein, *P Natl Acad Sci USA* 99, 16384-16391.
105. Mornon, J. P., Prat, K., Dupuis, F., and Callebaut, I. (2002) Structural features of prions explored by sequence analysis I. Sequence data, *Cell Mol Life Sci* 59, 1366-1376.
106. Baudin-Baillieu, A., Fernandez-Bellot, E., Reine, F., Coissac, E., and Cullin, C. (2003) Conservation of the prion properties of Ure2p through evolution, *Mol Biol Cell* 14, 3449-3458.
107. Bousset, L., Briki, F., Doucet, J., and Melki, R. (2003) The native-like conformation of Ure2p in fibrils assembled under physiologically relevant conditions switches to an amyloid-like conformation upon heat-treatment of the fibrils, *J Struct Biol* 141, 132-142.
108. Zhu, L., Zhang, X. J., Wang, L. Y., Zhou, J. M., and Perrett, S. (2003) Relationship between stability of folding intermediates and amyloid formation for the yeast prion Ure2p: A quantitative analysis of the effects of pH and buffer system, *J Mol Biol* 328, 235-254.
109. Baxa, U., Ross, P. D., Wickner, R. B., and Steven, A. C. (2004) The N-terminal prion domain of Ure2p converts from an unfolded to a thermally resistant conformation upon filament formation, *J Mol Biol* 339, 259-264.
110. Bousset, L., Redeker, V., Decottignies, P., Dubois, S., Le Marechal, P., and Melki, R. (2004) Structural characterization of the fibrillar form of the yeast *Saccharomyces cerevisiae* prion Ure2p, *Biochemistry-Us* 43, 5022-5032.
111. Ripaud, L., Maillet, L., Immel-Torterotot, F., Durand, F., and Cullin, C. (2004) The [URE3] yeast prion results from protein aggregates that differ from amyloid filaments formed in vitro, *J Biol Chem* 279, 50962-50968.
112. Baxa, U., Cheng, N. Q., Winkler, D. C., Chiu, T. K., Davies, D. R., Sharma, D., Inouye, H., Kirschner, D. A., Wickner, R. B., and Steven, A. C. (2005) Filaments

- of the Ure2p prion protein have a cross-beta core structure, *J Struct Biol* 150, 170-179.
113. Fay, N., Redeker, V., Savistchenko, J., Dubois, S., Bousset, L., and Melki, R. (2005) Structure of the prion Ure2p in protein fibrils assembled in vitro, *J Biol Chem* 280, 37149-37158.
 114. Pierce, M. M., Baxa, U., Steven, A. C., Bax, A., and Wickner, R. B. (2005) Is the prion domain of soluble Ure2p unstructured?, *Biochemistry-Us* 44, 321-328.
 115. Talarek, N., Maillet, L., Cullin, C., and Aigle, M. (2005) The [URE3] prion is not conserved among *Saccharomyces* species, *Genetics* 171, 23-34.
 116. Bach, S., Tribouillard, D., Talarek, N., Desban, N., Gug, F., Galons, H., and Blondel, M. (2006) A yeast-based assay to isolate drugs active against mammalian prions, *Methods* 39, 72-77.
 117. Cecchini, M., Curcio, R., Pappalardo, M., Melki, R., and Caflisch, A. (2006) A molecular dynamics approach to the structural characterization of amyloid aggregation, *J Mol Biol* 357, 1306-1321.
 118. Edskes, H. K., Naglieri, B. M., and Wickner, R. B. (2006) Nitrogen source and the retrograde signalling pathway affect detection, not generation, of the [URE3] prion, *Yeast* 23, 833-840.

2. Finding the most stable conformation for C-peptide of ribonuclease A

2.1. Introduction

α helices are some of the most abundant secondary structures available to polypeptides and proteins. They are stable in aqueous solution. They have also been studied extensively using molecular dynamic (MD) simulations. The majority of studies involving α helices come about because of the abundance of this secondary structure, their stability in aqueous solution, and the bias of force fields towards helical conformations.

One such α -helical polypeptide is C-peptide of ribonuclease A (RNase A) (1) as shown in Figure 2.1. C-peptide consists of the first thirteen residues of the N-terminus of RNase A (KETAAAKFERQHM). The polypeptide has been studied extensively by experimental methods (2-16). The first two residues, Lys1 and Glu2, are considered a part of the flexible region; residues 3–13 form an α -helical conformation. This particular polypeptide's α -helical structure is stable in aqueous solution because of several factors. A salt bridge exists between Glu2 and Arg10 (2, 3) and between Glu9 and His12 (4, 5). π -Stacking occurs between Phe8 and His12 (11). The stability of C-peptide has also been confirmed through MD methods: simulated annealing (17), copy exchange (18, 19), and general multicanonical ensembles (20-24).

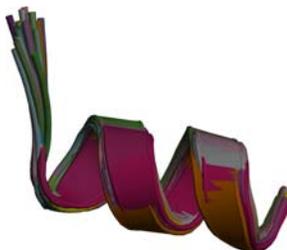


Figure 2.1 Overlay of all 32 NMR models of the C-peptide of RNase A

Conventional MD simulations run into one major problem — sampling on the potential energy (PE) surface. PE is simulated because it tends to dominate the free energy surface compared with the entropic contribution and because, unlike free energy, the PE is temperature-independent. The time that a typical polypeptide or protein requires to fold is currently inaccessible with today's technology. A polypeptide will therefore become trapped within a PE well—a particular conformation—for extended periods before it has an opportunity to leave its current state. High temperature simulations allow the polypeptide to sample more conformations within a shorter amount of time, but the experimental structure may not be located within the limited simulation time at higher temperatures. On the other hand, the experimental conformation, which in theory should correspond to the global free energy minimum, may be found as the global PE minimum, but the higher temperatures prevent the conformation from stabilizing because of increased kinetic energy.

Several methods have been developed over the years to overcome the conventional MD sampling problems. One of the earliest methods is simulated annealing (25) in which the temperature is increased and slowly decreased thus allowing for the protein to find the lowest possible PE well. Elber and Karplus developed locally enhanced sampling (LES) (26) to deal with the problem. LES lowers energy barriers by averaging forces. This, however, introduces pseudo-minima, which may not exist on the actual energy surface (27). More recently, the replica exchange method (28) has been used. In the replica exchange method, multiple copies of a system are run simultaneously but independently; and their coordinates are exchanged when the Boltzmann criterion is met.

Our group recently developed two protocols for geometry optimization (29, 30). The disrupted velocity (DIVE) search protocol perturbs both the kinetic energy (via the magnitudes of atomic velocity vectors) and velocity directions when predefined criteria are met. The system also experiences several cycles of velocity rescaling to accomplish heating and cooling. The perturbation allows a system to either overcome PE barriers (kinetic energy perturbation) or circumvent them (velocity direction perturbation). In the process, a system samples several conformations near 0 K and is able to locate several low PE conformations.

Independently or in conjunction with DIVE, one can use divergent path (DIP) search simulations. Through the DIP method, a system has multiple copies which are run simultaneously but independently, starting from the same initial geometry. They are run at a constant temperature, and each copy's trajectory develops separately because atoms in each of the copies are assigned different velocity directions. Thus, a system can cover a broader area of the PE surface in searching for low energy conformations.

Because of the recent development, we want to test our DIVE and DIP protocols. We selected C-peptide of RNase A because of its stability in aqueous solution. We want to determine whether C-peptide will form a stable α helix close to the structure of the NMR models and whether the α helix is the lowest PE conformation. In the process, we want to test the ability of DIVE and DIP to sample multiple conformations over a wide range of PEs.

2.2. Algorithms and simulations

2.2.1. Algorithms

Two protocols are implemented for our study. The disrupted velocity (DIVE) and the divergent path (DIP) protocols are discussed in previous literature (29, 30). As the name implies, the DIVE protocol perturbs the atomic velocity of the system permitting the conformation either to circumvent or to overcome PE barriers. The protocol allows a simulation to map a peptide's PE surface near 0 K. DIP, on the other hand, allows a conformation to traverse across the PE surface at a constant kinetic energy in search of PE minima. In the DIP protocol, multiple conventional MD simulations are run simultaneously by assigning atomic velocities of identical atoms in different simulations different directions, allowing more of the PE surface to be explored.

Each copy during a DIVE simulation undergoes a microcanonical simulation (constant number of atoms, volume, and energy) for a fixed time period before the atomic velocities are reassigned according to $\mathbf{p}' \equiv \sigma^{1/2} \mathbf{p}$ where \mathbf{p} and \mathbf{p}' are the momenta of the particles before and after atomic velocity reassignment, respectively. By convention, atomic velocity "reassignment" includes both "rescaling" of the magnitudes and changing the directions of atomic velocity vectors. σ is the scaling parameter that determines the magnitude of kinetic energy after atomic velocity reassignment.

The scaling parameter σ , where $\sigma > 1$ is used for adding kinetic energy and $0 < \sigma < 1$ is used for removing kinetic energy, may be initially set or *it may* be calculated during the simulation by $\sigma = |T - \Delta T|/T$. The scaling temperature (ΔT) is defined as the difference between the temperatures before and after atomic velocity rescaling.

Alternatively, σ can be defined using a target temperature T_{target} to be achieved after atomic velocity rescaling by defining $\sigma = T_{\text{target}}/T$.

A threshold temperature is also defined for the DIVE protocol. Above the threshold temperature, kinetic energy is removed by decreasing atomic velocities, and below the threshold temperature, kinetic energy is added by increasing atomic velocities. The threshold temperature is typically set near 0 K, so that the PE surface can be mapped at low temperatures. During one cycle, heating typically occurs once whereas cooling occurs multiple times.

For the divergent path (DIP) search simulations, the copies are simulated simultaneously but independently as in the DIVE simulations. The copies begin with the same initial coordinates and atomic velocity magnitudes, but unlike in DIVE, only the directions of the atomic velocity vectors are assigned from a Gaussian distribution (27). Each copy is allowed to traverse the PE surface in a canonical simulation (constant number of atoms, volume, and temperature) in which the temperatures are within a specified but limited range.

2.2.2. Simulation details

The first 13 residues from RNase A (PDB codes 2AAS (1) [32 NMR models] and 1KF5 (12) [x-ray crystal structure]), KETAAAKFERQHM, are used. Both termini are charged, and each acidic or basic amino acid has its side chains deprotonated or protonated, respectively in accordance with the NMR models solved at pH 4.0. Counterions are not used to counteract any of the charges. NMR model 1, the x-ray crystal structure with added hydrogens, and a fully extended conformation ($\phi = -180^\circ$ and $\psi = +180^\circ$) are used for our three initial simulations.

The PEs of all conformations are consequently minimized for 100 steps of steepest descent in a Generalized Born/surface area (GB/SA) implicit solvent model (31) using the Multiscale Modeling Tools for Structural Biology (MMTSB) (32) program. Upon the completion of minimization, the systems are then input into the Molecular Modeling Toolkit (MMTK) (33) package and are converted into files usable for our suite of programs. Through MMTK (33), the velocities are scaled to a temperature of 300.0 K.

Several parameters are included in the MD simulations. Simulations are run using a modified version of the Amber99 force field (34). The SHAKE (35, 36) algorithm is used to constrain distances between covalent bonds involving hydrogen, and the Nosé-Hoover Chain method (37) is used to maintain constant temperature. The velocity-Verlet method (38) is employed for integration of the equations of motion. All simulations are run in an a Generalized Born/surface area (GB/SA) implicit water solvent (31) as defined by

$$G_{pol} = -\frac{1}{2} \left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w} \right) \sum_{ij} \frac{q_i q_j}{f_{gb}}$$

$$f_{gb} = \left[r_{ij}^2 + \alpha_i \alpha_j \exp \left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j} \right) \right]^{\frac{1}{2}}$$

where G_{pol} is the solvation free energy of the solute-solvent electrostatic polarization term, ϵ_p is the dielectric value within of the protein, ϵ_w is the solvent dielectric constant, r_{ij} is the separation distance of particles i and j , q_i and q_j are the atomic charges, α_i and α_j are the corresponding effective Born radii (39), and f_{gb} is a complex function of r_{ij} , α_i and α_j . The effective Born radius determines the charge distance between the solute and the continuum dielectric boundary. $\epsilon_w = 78.5$, $\epsilon_p = 1.0$, a surface tension of 0.005 kcal/mol-

\AA^2 , and $r_{ij} = 0.9 \text{\AA}$. Periodic boundaries are not used; instead, electrostatic and Lennard-Jones cutoffs are set to 999\AA to represent an infinite cutoff.

In each simulation, six copies are simultaneously run for 4 million steps per copy with a 2 fs time step (4×10^6 steps/copy \times 2 fs/step \times 6 copies/simulation = 48 ns total simulation time), and data is output every 250 steps (0.5 ps). For DIVE simulations, each copy is initialized at a different temperature (10 K, 50 K, 100 K, 300 K, 600 K, and 1000 K), and the simulated conformations include the fully extended conformation, NMR model 1, the x-ray crystal structure, and the lowest PE conformations from the first three simulations. Conformations are designated by the derivation from the initial conformation. For instance, the second round of DIVE simulations obtained from the fully-extended conformation is assigned the name “ext”.

After 20,000 steps, each copy is perturbed by atomic velocity reassignment. The heating parameter is calculated from the target temperature T_{target} and the temperature T at the velocity- reassignment step ($\sigma = T_{\text{target}}/T$). The scaling parameter for kinetic energy removal is set to 0.25, and the heating and cooling threshold temperature is 10 K. Thus, when a system’s temperature rises above the threshold temperature, it is cooled to 0.25 of its temperature T . When a system’s temperature drops below the threshold temperature at the time of velocity reassignment, the system is heated to $T_{\text{target}} = 1000 \text{ K}$. Although the target temperature is 1000 K, the actual temperature achieved is 500 K because of rapid energy redistribution.

Ten conformations are also simulated using DIP. The simulated conformations include the fully extended conformation, NMR model 1, the x-ray crystal structure, the lowest PE conformations from each of the six DIVE simulations, and a β hairpin with

type IV turn located during a DIVE simulation. Each simulation is run at 300 ± 20 K. Identical atoms within each of the six copies of each simulation are begun at the same positions and are assigned the same atomic velocity magnitudes, but atomic velocities are assigned different directions, allowing each copy to search for different PE minima. All results reported within this chapter use the conformation corresponding to the lowest PE.

2.2.3. Analysis

The analysis module ptraj in the MD package AMBER 8 (40) is used to calculate the root mean squared deviations (rmsd's) relative to experimental NMR models for each conformation. Backbone atom rmsd's are calculated relative to NMR model 1 and to all of the NMR models (32 models total) (1). The backbone atom rmsd's are also calculated using only Tyr3–Met13, the residues comprising the α helix; they are calculated compared with both NMR model 1 and all NMR models.

Interatomic distances are calculated to compare with the NMR distance constraints. Two other distances (total 163 distance constraints) are also calculated from among those found in the NMR distance constraints: the distances between the positively charged nitrogen in Arg10 and the two carboxylic oxygens in the Glu2 side chain. After the distances are calculated, they are then compared with the distances listed in the NMR distance constraint file.

To determine whether salt bridges and π -stacking interactions are deemed possible, distances are calculated for each heavy atom pair involved in the putative interaction and if more than one pair of heavy atoms is considered, the lowest distance per step is used. Four polar side chain interactions are considered: Glu2 \cdots Arg10, Lys1 \cdots Glu2, Glu2 \cdots Arg10, Glu2 \cdots His12, Glu2 \cdots His12, Glu2 \cdots Lys7, and Lys7 \cdots Glu2.

Two heavy atoms in Arg, Glu, and His are included for distance calculations whereas a single nitrogen atom in Lys is included. Therefore, four distances are calculated between Glu and Arg and between Glu and His but only two distances are calculated between Lys and Glu. The charged polar side chains are considered in proper orientation if the distance between any two terminal heavy atoms of the side chains is less than 3.5 Å. The aromatic rings between Phe8 and His12 are implicated in π -stacking interactions; distance calculations are performed using the heavy atoms of the phenyl ring of Phe8 and the imidazole ring of His12. The geometric configuration between the two aromatic rings is conventionally defined as improper dihedral angles of $\pm 90^\circ$ between $C\delta_i-C\gamma_i\cdots C\gamma_j-C\delta_j$ of each aromatic ring and the overall distance between the heavy atoms of the two aromatic rings is below 6 Å (41) (Figure 2.2). Secondary structures are determined by the program STRIDE (42), which classifies secondary structures on the basis of backbone dihedral angles and H-bond interactions.

The $^3J_{\text{HN-H}\alpha}$ and $^3J_\phi$ coupling constants are calculated using the Karplus equations

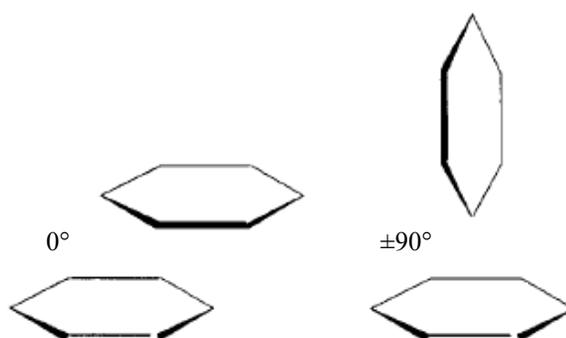


Figure 2.2 Geometric orientation of aromatic rings for possible π -stacking. The left figure shows aromatic rings parallel to each other and slightly out of phase (coplanar), whereas the right figure represents two aromatic rings perpendicular to one another for π -stacking (T-shaped). Figure from ref. (41).

(43, 44) (eq. (2.1) and (2.2)) for residues 3–13.

$${}^3J_{\text{HN-H}\alpha} = 6.7 \cos^2 \theta - 1.3 \cos \theta + 1.5 \quad (2.1)$$

$${}^3J_{\varphi} = {}^3J_{\text{CNC}\alpha\text{C}} = 1.8 \cos^2 \theta - 0.2 \cos \theta + 0.5 \quad (2.2)$$

where the angle θ is the dihedral angle for H–N–C α –H α or for C–N–C α –C. The ranges of ${}^3J_{\text{HN-H}\alpha}$ and ${}^3J_{\varphi}$ coupling constants are calculated from the respective dihedral angles of all NMR models. The J-coupling constants are calculated for all simulations and are compared with the minimum and maximum NMR model J-coupling constants (Table 2.1). For DIP simulations, an average J-coupling constant and its standard deviation are calculated after the conformations have reached equilibrium (4–8 ns).

For DIVE simulations, the atomic velocity and the kinetic energy perturbations occurs every 20,000 steps, and one conformation that has a temperature lower than 10 K is selected within a 20,000 step range because the conformation corresponds to the lowest PE within that range. For example, one conformation is selected within the range of 0–20,000 steps. A conformation is considered for further simulation and for further analysis when the conformation corresponds to the lowest PE from among all conformations selected from each segment of 20,000 steps.

Hydrogen bond (H-bond) calculations and cluster analysis are done for both individual conformations found in both DIVE and DIP simulations. H-bonds are

Table 2.1 Range for coupling constants. $J_{\text{HN-H}\alpha}$ from experiment and J_{φ} calculated

		J_{φ}										
		Thr3	Ala4	Ala5	Ala6	Lys7	Phe8	Glu9	Arg10	Gln11	His12	Met13
Min		0.49	0.81	0.68	0.69	0.63	0.97	0.91	0.73	0.51	0.78	0.49
Max		0.71	2.1	0.94	0.88	0.91	1.25	1.07	0.99	0.85	1.04	0.59

		$J_{\text{HN-H}\alpha}$										
		Glu2	Thr3	Ala4	Ala5	Ala6	Lys7	Phe8	Glu9	Arg10	Gln11	His12
Min		5.18	1.75	3.14	3.43	3.35	2.29	2.99	2.91	3.62	9.24	5.65
Max		8.9	4.9	4.6	4.58	4.88	3.26	3.66	4.3	6.35	9.5	7.57

conventionally defined by a range of $180.0^\circ \pm 60.0^\circ$ for the X–H \cdots X₁ angle where X and X₁ are polar heavy atoms, with a 3.5 Å distance between polar heavy atoms and with an appearance of more than 5% in the simulations. H-bonds for DIP simulations are calculated for the conformations corresponding to the lowest PE of each simulation.

Cluster analysis offers an alternative way to classify conformations on the basis of similar secondary structures; the analysis technique groups conformations into families without consideration of the PE. The technique also allows qualitative evaluation of the entropic contribution from a particular group of conformations. Cluster analysis is done both for the DIP and the DIVE simulations based on their rmsd's from the NMR structure and using the fixed-radius clustering algorithm using MMTSB (32). The algorithm uses an iterative process with an error tolerance of 0.5 and a least-squares fit for alignment of the conformations (45). To be classified within a cluster, a conformation must lie within a range of 3.0 Å compared with the cluster centroid, an average conformation within the cluster. The conformations are compared to each other by excluding the first two residues, Lys1 and Glu2, which are not within the α -helical region.

2.3. Simulation results and discussion

2.3.1. Divergent path (DIP) simulations

Ten conformations are simulated at 300 ± 20 K to determine whether the experimental conformation can be reproduced near physiological temperatures. The experimental conformation has an α helix extending from Tyr3–Met13 with the 1–4 backbone H-bonds. Three charged polar side chain pairs are implicated in stabilizing the α helix: E2R10, E2H12, and E9H12 (4, 6, 7, 9, 10, 21, 22, 46-48). π -stacking between

the aromatic rings of Phe8 and His12 are also suspected to stabilize the secondary structure (10, 11, 48, 49).

Four of the simulations have average backbone atom rmsd's below 2.0 Å relative to the experimental conformation: nmr (Figure 2.3b), nmr' (Figure 2.3e), nmr'' (Figure 2.3h), and xray (Figure 2.3c). Yet, the average PE of each simulation differs; nmr'' and nmr differ by 5.4 kcal/mol; the nmr' and xray simulations experience the least PE difference (0.3 kcal/mol). This is mostly likely because the flexible residues (Lys1 and Glu2) and the side chain configurations vary. The salt bridge between the side chains of Glu2 and Lys7 are common in all four simulations. Two of the interactions (E2R10 and E2H12), which are implicated in stabilizing the secondary structure, also fluctuate during the simulations; the close proximity between Glu2 and His12 is only found in the nmr' simulation whereas simulations starting from the xray structure do not exhibit interactions between Glu2 and Arg10. Glu9 and Arg10 are also found in close proximity for the simulations labeled nmr and xray, but this interaction is not documented in the literature. All four simulations reveal that the aromatic side chains of Phe8 and His12 are in a geometric configuration similar to experiment. The simulations exhibit between 40% and 60% of the NMR distance constraints in agreement with experimental data. 30–40% of the $^3J_{\text{HN-H}\alpha}$ coupling constants and 40–50% of the calculated $^3J_{\phi}$ coupling constants agree with the experimental data; in other words, 30–40% of the H–N–C α –H α dihedral angles and 40–50% of the C–N–C α –C dihedral angles correspond to the experimental structure within 13–15% error. Because the percent of coupling constants and distance constraints are unexpectedly low, one cannot solely depend upon these

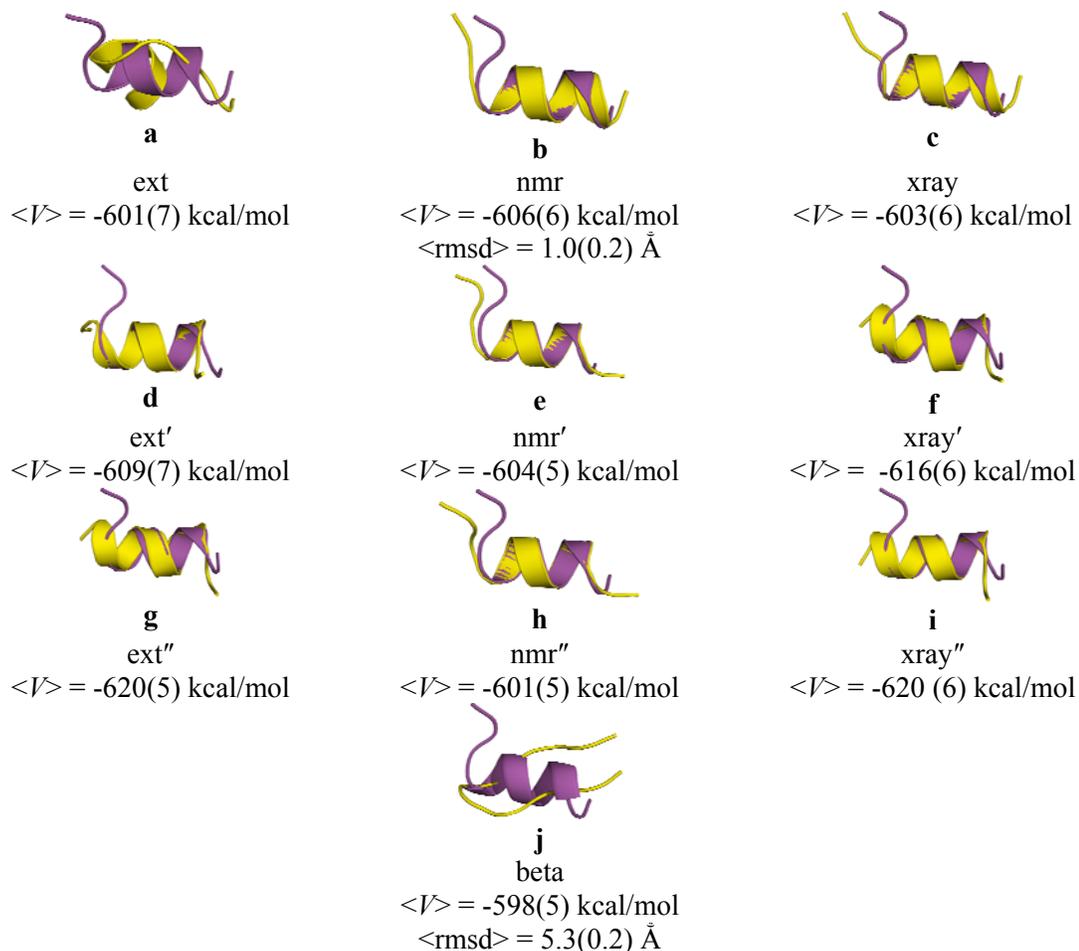


Figure 2.3 Representations of average equilibrated conformations (yellow) relative to NMR model 1 (magenta). Conformations are fitted to the NMR model about residues 3–13. PEs and backbone atom rmsd's (residues 3–13 relative to all NMR models) are listed with standard deviation in parentheses.

calculations as criteria for how well calculated structures agree with the NMR model structures.

The ext' (Figure 2.3d), ext'' (Figure 2.3g), xray' (Figure 2.3f), and xray'' (Figure 2.3i) simulations varied more considerably in their conformation than the previous simulations. For the ext' simulation, the average α helix extends from Thr3 to Gln11 whereas the ext'' and xray'' simulations are composed of α helices ranging from Glu2 to Arg10 and ranging from Glu2 to Gln11, respectively. The average α helix for the xray' simulation is similar to that found in the ext'' simulation, but backbone atom rmsd's for

xray' and ext'' relative to the experimental conformation vary because Gln11–Met13 vary in their orientation. The PE of ext', interestingly, is higher than that of the other three conformations (≈ 8 kcal/mol between ext' and xray') although its average backbone atom rmsd is lower relative to the NMR model. The side chains of Glu9 and Arg10 are consistently within proximity to form a salt bridge during the four simulations, but side chains of Lys1 and Glu9 appear similarly close to each other during 22% of the ext' simulation. The aromatic side chains of Phe8 and His12 achieve geometric configuration for π -stacking only for the xray' simulation, but the distance and the angle between the aromatic rings fluctuate considerably. Distances calculated for the four simulations are found to agree with 45–50% of the experimental distance constraints whereas the $^3J_{\text{HN-H}\alpha}$ coupling constants and the calculated $^3J_{\phi}$ coupling constants are respectively within 25–30% and 25–35% of the experimental values with 12–13% error. The percent agreement for the distance constraints and for the coupling constants appear lower for the conformations whose rmsd's from the NMR structure are lower. A slight shift in the dihedral angles can affect the calculations of the coupling constants, and if the atomic distances are not within the desired range although close, the distance is excluded from consideration. In other words, the distance constraints and coupling constants may be close to the expected values but are discounted because they are not within the acceptable range.

Two other conformations, which are not similar to the experimental conformation, are simulated. The ext simulation (Figure 2.3c) produces an average conformation composed of an α helix from Thr3 to Phe8 with Glu9–Met13 as a flexible region. We unexpectedly find a β hairpin (Figure 2.3j) in the DIVE simulation, and the secondary

structure, after equilibration, retains β strands extending from Glu2 to Ala4 and from Arg10 to His12 separated by a type I β turn. The ext conformation is 7 kcal/mol higher in energy than the α -helical ext' conformation but is 4 kcal/mol lower than the β hairpin. For the ext simulation, the side chains of Glu9 and Arg10 and of Glu2 and His12 are less than 3.5 Å apart, but because Glu9–Met13 is flexible, the aromatic side chains do not align properly. The β hairpin structure experiences possible interaction between the charged polar side chains of Glu2 and His12, but the aromatic side chains are not able to align because of backbone conformation. The distance constraints and the coupling constants are close those of calculated conformations described previously, but not surprisingly, the percent agreement between calculated coupling constants of the β hairpin and coupling constants measured the experimental structure is only $45 \pm 2\%$. With only 163 distance constraints, the decent agreement is not too surprising.

Structural determination based solely upon PE and conformational analysis of individual structures may not give the full picture. Another way of analyzing conformations is through the use of cluster analysis. Cluster analysis permits the conformations to be grouped according to conformational similarities by rmsd comparisons and offers qualitative insight into the entropic component of free energy. The groups, or “clusters”, are classified on the basis of similar secondary structures by rmsd comparisons, and the average conformation of the cluster is called the centroid.

When the ten simulations are analyzed by cluster analysis, two clusters are composed of conformations with backbone atom rmsd's (for residues 3–13) relative to NMR model 1 that are less than 2.0 Å. The conformation that best represents the two centroids consists of α helices from Ala4 to Gln11. The representative conformation of

the largest cluster has a backbone atom rmsd (residues 3–13) of 2.5 Å relative to NMR model 1. Finally, the fourth largest cluster has a representative conformation consisting of an α helix from Thr3 to Lys7 with a backbone atom rmsd (residues 3–13) of 4.4 Å relative to NMR model 1. From cluster analysis, we surmise that the experimental conformation can be reproduced and maintained near physiological temperatures. Because clusters whose conformations are similar to the experimental conformation contain the most representatives, we can qualitatively ascertain that they should make a high entropic contribution to the free energy, and we can also qualitatively conclude that the non-native conformations, although appearing to have favorable entropy contributions to the free energy, are not as favored as the experimental conformation.

The DIP protocol, when used in conjunction with the DIVE protocol, allows the simulated conformations to sample the PE surface near physiological temperatures. Some of the noncovalent interactions observed experimentally are present during the DIP simulations although others are not. The increase in kinetic energy from the initial conformations contributes to the alteration of some noncovalent interactions. The β hairpin might be expected to undergo a conformational change and collapse into the native α helix, but the β hairpin is stable despite the increase in kinetic energy.

Although the global PE minimum located from the DIP simulations displays an α helix extending from Glu2 to Gln11, one major feature stands out—all of the average equilibrated PEs of the equilibrated structures are within a relatively small range of 23 ± 7 kcal/mol (Figure 2.5). This may indicate that the helical conformations are located within a single PE minimum containing numerous substates. The minimum, in fact, may contain substantial energy barriers separating the various substates and preventing their

interconversion. The height of the barriers has not been determined, but this hypothesis certainly seems feasible on the basis of the data presented.

In other words, the simulations are experiencing a kinetic energy trap (50-54), which occurs when a non-native conformation that is unable to cross the PE barrier. The DIP protocol finds several local minima on the PE surface, which are close in PE but do not converge into a single low-energy conformation. The lack of conformational convergence suggests that the minima are surrounded by high PE barriers. Based upon the three possible disconnectivity graphs (Figure 2.4), the ‘weeping willow’ model probably represents our present findings best. If the ‘palm tree’ model was relevant, the conformations should converge over time into a single conformation with similar PE, or if the ‘banyan tree’ model was valid, all conformations should have similar PE but not converge to a single conformation. Standard deviations of the simulations suggest that the average PEs should overlap, but the standard errors of the mean (≈ 0.1 kcal/mol) of the simulations, which quantifies the accuracy of the average PE, suggest that the PEs are fairly accurate.

The “weeping willow” model can be explained from a biological standpoint. Helices and hairpins differ within their overall makeup (e.g., H-bond formation and

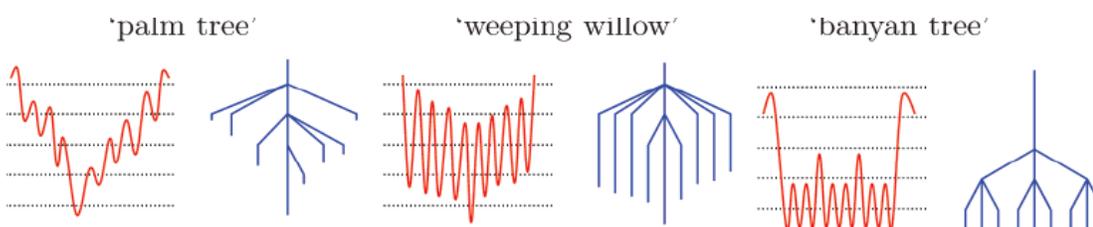


Figure 2.4 Three pairs of PE landscapes (left) and their corresponding disconnectivity graphs (right). The graphs are drawn as PE (vertical axis) relative to arbitrary coordinates (horizontal axis). The endpoints of the disconnectivity graphs represent PE minima, and the points where the branches are joined correspond to a common PE “superbasin”. For the PE landscape, the wells represent the minima of a system surrounded by PE barriers. Reprinted with permission from ref. (54). Copyright 2006 American Chemical Society

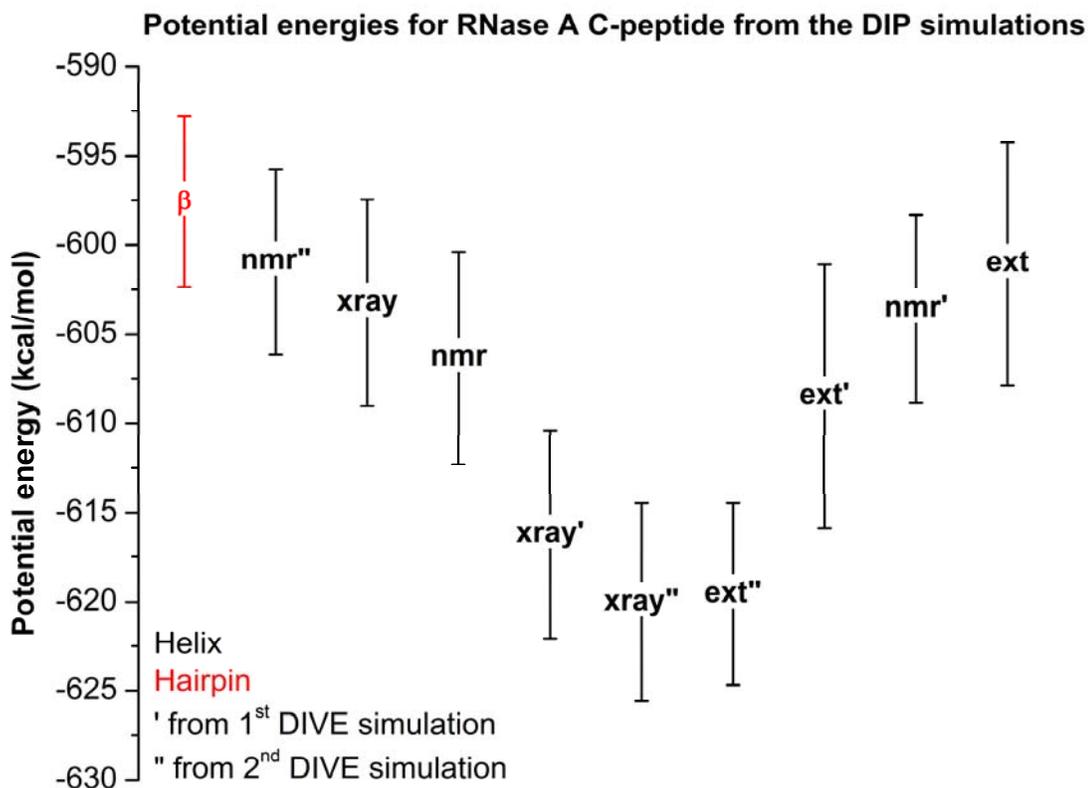


Figure 2.5 Average potential energies with standard deviation bars.

backbone dihedral angles). To transition from one conformation to another (e.g., β hairpin \rightarrow α helix) requires noncovalent bond breaking (salt bridges and H-bonds) and noncovalent bond re-formation. The difference between two PE minima may seem small, but the amount of energy required to break and re-form bonds translates into high PE barriers and prevents transitions between conformations. Alternatively, the experimental conformation may lie within a wide-basin local minimum, whereas the global PE minimum resides in a deep, narrow well.

2.3.2. Disrupted velocity (DIVE) simulations

Several conformations are sampled in the six simulations using the disrupted velocity protocol. Any conformation found below 10 K and corresponding to the lowest

PE within a 20,000-step range was considered for further simulation and for further analysis. For the DIVE simulations of C-peptide, 531 conformations meet this criteria and the PE for the conformations vary by 385 kcal/mol. A conformation from each of the six simulations is selected because the conformation meets the aforementioned criteria and corresponds to the lowest PE within the simulation.

In the following simulations, five charged polar side chain pairs (E2R10, E9R10, E2K7, E2H12, and E9H12) are capable of forming salt bridges. Two of the side chain pairs (E2R10 and E9H12) are suspected to stabilize the α helix (2-5). The aromatic rings of Phe8 and His12 are documented to be in a geometric configuration that supports π -stacking (11), but none of the calculated conformations below exhibit this configuration primarily because the angle between the aromatic rings is more than $\pm 90^\circ$ (neither coplanar nor T-shaped). Most of the conformations have 45–60% calculated distances within 2–5% of the experimentally-calculated distance constraints. Relative to the experimental data, 35–55% of the $^3J_{\text{HN-H}\alpha}$ coupling constants and 35–60% of the $^3J_\phi$

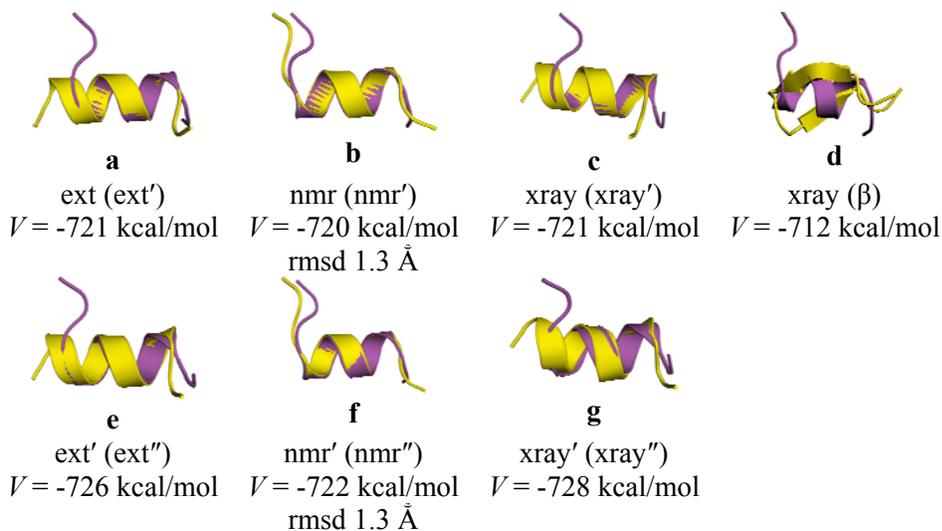


Figure 2.6 Overlays of low PE conformations (yellow) relative to NMR model 1 (magenta) of C-peptide and fit to residues 3–13. Conformations are listed by the simulation in which they are derived and by a unique identifier, in parenthesis.

coupling constants are maintained within 10–15% of experimental values.

Of the structures calculated using the DIVE protocol, nmr' (Figure 2.6b) and nmr'' (Figure 2.6f) are most similar to the experimental conformation. They both are composed of an α helix extending from Ala4 to Gln11 whereas Ala3–Met13 comprise the experimentally observed α helix. Between the nmr' and nmr'', the backbone atom rmsd's of the α helical region differs by 0.1 Å, but the nmr' conformation is 2 kcal/mol higher in PE than the nmr'' conformation. Both conformations similarly exhibit two charged polar side chain pairs (E2R10 and E2K7).

The ext' (Figure 2.6a), ext'' (Figure 2.6e) xray' (Figure 2.6c), and xray'' (Figure 2.6g) conformations have α -helical secondary structures that are shorter than that observed in the experimental conformation. An α helix extending from Glu2 to Arg10 comprise both the ext'' and xray' conformations whereas the ext' and xray'' conformation have an α helix extending from Glu2 to Gln11. The xray'' conformation has the lowest calculated PE of the six conformations. The ext' conformation is 7 kcal/mol higher than the xray'' conformation, and the ext'' and the xray' conformations differ by 5 kcal/mol. Unlike the other conformations found with the DIVE protocol, $^3J_\phi$ coupling constants for ext'' agree with only 9% of the experimental $^3J_\phi$ coupling constants, which suggests that the conformation is not similar to the experimental conformation. However, xray' has a similar secondary structure and maintains 46% of the $^3J_\phi$ coupling constants. The C–N–C α –C dihedral angles differ slightly between the two conformations (0.3 Å), which may account for the variation in calculated coupling constant.

An unusual conformation, a β hairpin with a type IV turn (Figure 2.6d), is found among the conformations calculated by using the DIVE protocol. The β hairpin originates

in the simulation from the x-ray crystal structure. Residues 2–6 and 9–13 comprise the β strands whereas residues 7 and 8 form the type IV β turn. The charged polar side chains of both Glu2 and Glu9 are in close proximity to His12. The two glutamines are positioned close to different imidazole nitrogens. Furthermore, the N- and C-termini of the peptide are less than 3.5 Å apart. The β hairpin retains a similar percentage of the distance constraints and of the $^3J_\phi$ coupling constant compared to the other calculated structures. The non-native conformation should have a lower percentage of the distance constraints and the coupling constants that agree with experiment, but the similarities to other calculated conformations suggest that relying solely upon the percentage of coupling constants or distance constraints may give a misleading picture.

For the DIVE simulations, 456 conformations are classified into conformational families by cluster analysis. The largest cluster (103 conformations) has a representative conformation consisting of an α helix extending from Ala4 to Gln11. The second largest cluster (56 conformations) has a representative conformation composed of an α helix from Thr3 to Gln11. Three families of conformations contain 21 conformations, so the three conformations may share a qualitatively similar entropic contribution. Two of the clusters contain α helices of different lengths whereas one cluster has no detectable secondary structure.

Overall, the DIVE protocol is able to reproduce the experimental conformation and samples several additional conformations. Six clusters show some helical conformation. Some of the conformations apparently have a more flexible C-terminus than the ordered structure of the experimental conformation. Surprisingly, a β hairpin is observed in several simulated structures. The secondary structure is quite unusual for this

polypeptide but not improbable. To our knowledge, a β hairpin has not been reported for this peptide, but it is not an improbable structure.

The DIVE simulations offer the opportunity to map the PE surface. We find that the "xray" conformation resides in the lowest PE minimum. Near 0 K, the entropy does not contribute much compared with the PE and therefore, we suggest that the "xray" conformation resides in the global free energy minimum at low temperatures. We suspect that the other conformations may eventually converge to this conformation, but at low temperatures, the PE barriers may be relatively high and prevent the transition between conformations.

2.4. Conclusions

We have tested both the DIVE and DIP protocols for their abilities to reproduce the observed α -helical structure of the C-peptide of RNase A near 0 K (DIVE simulations) and near 300 K (DIP simulations). Our simulations show that the experimental conformation can be reproduced within 1.0 Å backbone atom rmsd of the NMR model by using DIVE simulations of an initial α helix structure and within 0.9 Å rmsd of the NMR model by using DIP simulations, starting from an α helix structure. Cluster analysis of the DIP simulations shows that the two most populous clusters each have an rmsd less than 2.0 Å from the NMR structure and together contain 40% of the structures analyzed. Similarly, cluster analysis of the DIVE simulations shows that the two largest clusters display α -helical conformations and contain 35% of the structures analyzed. Thus, even though the structure representing the global potential energy (PE) minimum found by using each of the two methods is an α helix extending from Glu2 to Gln11 rather than an α helix extending from Thr3 to Met13, the experimental α -helical

conformation is only 14–18 kcal/mol higher in energy than the global minimum and α -helical conformations appear most frequently among the PE minima located.

We also find a β hairpin conformation that is a local PE minimum only 22 kcal/mol above the global PE minimum. We propose that C-peptide of RNase A can fold into a β hairpin but is more likely to fold into the experimental conformation under the conditions of the experiment. On the basis of the cluster analysis, we further propose that the experimental conformation lies within a wide-basin local PE minimum surrounded by high PE barriers, which prevent simulated structures from refolding into the conformation residing in the global PE minimum. We have shown that the DIVE protocol is able to locate many different conformations within PE minima near 0 K using different starting structures. DIVE allows a simulation to map the PE surface of a system near 0 K, whereas DIP used in conjunction with DIVE, can determine whether the low PE conformations are truly the conformations maintained at higher temperatures.

Finally, our simulations have shown that the experimental conformation can be reproduced well. We also report that a β hairpin conformation has been found. The non-native β hairpin appears stable during our simulations. We suspect that C-peptide of RNase A can fold into a β hairpin but the native α -helical conformation is more stable. We propose that the experimental conformation lies within a wide-basin local PE minimum surrounded by high PE barriers, which prevent it from refolding into the conformation residing in the nearby global PE minimum.

2.5. References

1. Santoro, J., Gonzalez, C., Bruix, M., Neira, J. L., Nieto, J. L., Herranz, J., and Rico, M. (1993) High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy, *J Mol Biol* 229, 722-734.
2. Brown, J. E., and Klee, W. A. (1969) Conformational Studies of a Series of Overlapping Peptides from Ribonuclease and Their Relationship to Protein Structure, *Biochemistry-U.S.* 8, 2876-2879.
3. Brown, J. E., and Klee, W. A. (1971) Helix-Coil Transition of Isolated Amino Terminus of Ribonuclease, *Biochemistry-U.S.* 10, 470-476.
4. Bierzynski, A., Kim, P. S., and Baldwin, R. L. (1982) A Salt Bridge Stabilizes the Helix Formed by Isolated C-Peptide of Rnase A, *P Natl Acad Sci-Biol* 79, 2470-2474.
5. Rico, M., Nieto, J. L., Santoro, J., Bermejo, F. J., Herranz, J., and Gallego, E. (1983) Low-Temperature H-1-Nmr Evidence of the Folding of Isolated Ribonuclease S-Peptide, *Febs Lett* 162, 314-319.
6. Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., and Baldwin, R. L. (1985) Nature of the Charged-Group Effect on the Stability of the C-Peptide Helix, *P Natl Acad Sci USA* 82, 2349-2353.
7. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., and Baldwin, R. L. (1986) The Role of a Glu 2- ... Arg 10+ Salt Bridge in Stabilizing the Isolated C-Peptide Helix, *Biophys J* 49, A290-A290.
8. Fairman, R., Shoemaker, K. R., York, E. J., Stewart, J. M., and Baldwin, R. L. (1989) Further-Studies of the Helix Dipole Model: Effects of a Free Alpha-Nh3+ or Alpha-Coo- Group on Helix Stability, *Proteins* 5, 1-7.
9. Osterhout, J. J., Baldwin, R. L., York, E. J., Stewart, J. M., Dyson, H. J., and Wright, P. E. (1989) H-1-Nmr Studies of the Solution Conformations of an Analog of the C-Peptide of Ribonuclease-A, *Biochemistry-U.S.* 28, 7059-7064.
10. Strehlow, K. G., and Baldwin, R. L. (1989) Effect of the Substitution Ala->Gly at Each of 5 Residue Positions in the C-Peptide Helix, *Biochemistry-U.S.* 28, 2130-2133.
11. Shoemaker, K. R., Fairman, R., Schultz, D. A., Robertson, A. D., York, E. J., Stewart, J. M., and Baldwin, R. L. (1990) Side-Chain Interactions in the C-Peptide Helix: Phe 8 - His 12+, *Biopolymers* 29, 1-11.
12. Tilton, R. F., Dewan, J. C., and Petsko, G. A. (1992) Effects of Temperature on Protein-Structure and Dynamics - X-Ray Crystallographic Studies of the Protein Ribonuclease-a at 9 Different Temperatures from 98-K to 320-K, *Biochemistry-U.S.* 31, 2469-2481.
13. Maliwal, B. P., Lakowicz, J. R., Kupryszewski, G., and Rekowski, P. (1993) Fluorescence Study of Conformational Flexibility of Rnase S-Peptide - Distance-Distribution, End-to-End Diffusion, and Anisotropy Decays, *Biochemistry-U.S.* 32, 12337-12345.
14. Moye-Sherman, D., Jin, S., Ham, I., Lim, D. Y., Scholtz, J. M., and Burgess, K. (1998) Conformational preferences of RNase a C-peptide derivatives containing a highly constrained analogue of phenylalanine, *J Am Chem Soc* 120, 9435-9443.

15. Kumaran, S., and Roy, R. P. (1999) Helix-enhancing propensity of fluoro and alkyl alcohols: influence of pH, temperature and cosolvent concentration on the helical conformation of peptides, *J Pept Res* 53, 284-293.
16. Berisio, R., Sica, F., Lamzin, V. S., Wilson, K. S., Zagari, A., and Mazzarella, L. (2002) Atomic resolution structures of ribonuclease A at six pH values, *Acta Crystallogr D* 58, 441-450.
17. Kinoshita, M., Okamoto, Y., and Hirata, F. (1999) Analysis on conformational stability of C-peptide of ribonuclease A in water using the reference interaction site model theory and Monte Carlo simulated annealing, *J Chem Phys* 110, 4090-4100.
18. Felts, A. K., Harano, Y., Gallicchio, E., and Levy, R. M. (2004) Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model, *Proteins-Structure Function and Bioinformatics* 56, 310-321.
19. Ohkubo, Y. Z., and Brooks, C. L. (2003) Exploring Flory's isolated-pair hypothesis: Statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A, *P Natl Acad Sci USA* 100, 13916-13921.
20. Hansmann, U. H. E. (1997) Parallel tempering algorithm for conformational studies of biological molecules, *Chem Phys Lett* 281, 140-150.
21. Hansmann, U. H. E., and Okamoto, Y. (1998) Tertiary structure prediction of C-peptide of ribonuclease a by multicanonical algorithm, *J Phys Chem B* 102, 653-656.
22. Hansmann, U. H. E., and Okamoto, Y. (1999) Effects of side-chain charges on alpha-helix stability in C-peptide of ribonuclease A studied by multicanonical algorithm, *J Phys Chem B* 103, 1595-1604.
23. Okamoto, Y. (1999) Tackling the multiple-minima problem in protein folding by Monte Carlo simulated annealing and generalized-ensemble algorithms, *Int. J. Mod. Phys. C* 10, 1571-1582.
24. Sugita, Y., and Okamoto, Y. (2005) Molecular mechanism for stabilizing a short helical peptide studied by generalized-ensemble simulations with explicit solvent, *Biophys J* 88, 3180-3190.
25. Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983) Optimization by Simulated Annealing, *Science* 220, 671-680.
26. Elber, R., and Karplus, M. (1990) Enhanced Sampling in Molecular Dynamics: Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon Monoxide Diffusion through Myoglobin, *J Am Chem Soc* 112, 9161-9175.
27. Stultz, C. M., and Karplus, M. (1998) On the potential surface of the locally enhanced sampling approximation, *J Chem Phys* 109, 8809-8815.
28. Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding, *Chem Phys Lett* 314, 141-151.
29. Huang, Z. N., and Wheeler, R. A. Replica Microcanonical Search Methods for Mapping Potential Energy Landscapes and Conformations of miniproteins, *In preparation*.
30. Huang, Z. (2005) Design of new molecular dynamics global minimum search protocols for mapping energy landscapes and conformations of folded

- polypeptides and mini-proteins, in *Chemistry and Biochemistry*, p 282, The University of Oklahoma, Norman.
31. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular Simulations, *Biopolymers* 56, 275-291.
 32. Feig, M., Karanicolas, J., and Brooks, C. L. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology, *J Mol Graph Model* 22, 377-395.
 33. Hinsen, K. (2000) The molecular modeling toolkit: A new approach to molecular simulations, *J Comput Chem* 21, 79-85.
 34. Okur, A., Strockbine, B., Hornak, V., and Simmerling, C. (2003) Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins, *J Comput Chem* 24, 21-31.
 35. Andersen, H. C. (1983) Rattle: a Velocity Version of the Shake Algorithm for Molecular Dynamics Calculations, *J Comput Phys* 52, 24-34.
 36. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes, *J Comput Phys* 23, 327-341.
 37. Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992) Nosé-Hoover Chains: the Canonical Ensemble via Continuous Dynamics, *J Chem Phys* 97, 2635-2643.
 38. Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A Computer-Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *J Chem Phys* 76, 637-649.
 39. Born, M., and von Karman, T. (1912) On fluctuations in spatial grids., *Phys Z* 13, 297-309.
 40. Case, D. A., Darden, T. A., Cheatham, I., T.E. , Simmerling, C. L., Wang, J., Duke, R. R. E., Merz, L., K.M. , Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) AMBER 8, 8 ed., University of California, San Francisco, San Francisco.
 41. McGaughey, G. B., Gagne, M., and Rappe, A. K. (1998) pi-stacking interactions - Alive and well in proteins, *J Biol Chem* 273, 15458-15463.
 42. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins* 23, 566-579.
 43. Kao, L. F., and Barfield, M. (1985) Conformational Dependencies of Vicinal C-13(O)-N-C-Alpha-C-13 and C-13(O)-N-C-Alpha-H-1 Coupling Constants in Compounds Which Model the Peptide Backbone, *J Am Chem Soc* 107, 2323-2330.
 44. Ludvigsen, S., Andersen, K. V., and Poulsen, F. M. (1991) Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of phi-angles, *J Mol Biol* 217, 731-736.
 45. Lloyd, S. P. (1982) Least-Squares Quantization in Pcm, *Ieee T Inform Theory* 28, 129-137.

46. Okamoto, Y., Fukugita, M., Nakazawa, T., and Kawai, H. (1991) Alpha-Helix Folding by Monte-Carlo Simulated Annealing in Isolated C-Peptide of Ribonuclease-A, *Protein Eng* 4, 639-647.
47. Rico, M., Gallego, E., Santoro, J., Bermejo, F. J., Nieto, J. L., and Herranz, J. (1984) On the Fundamental Role of the Glu 2- ... Arg 10+ Salt Bridge in the Folding of Isolated Ribonuclease A S-Peptide, *Biochem Bioph Res Co* 123, 757-763.
48. Neira, J. L., and Rico, M. (1997) Folding studies on ribonuclease A, a model protein, *Fold Des* 2, R1-R11.
49. Rashin, A. A. (1992) Continuum Electrostatics of the C-Peptide: Anatomy of the Problem, *Proteins* 13, 120-131.
50. Mortenson, P. N., and Wales, D. J. (2001) Energy landscapes, global optimization and dynamics of the polyalanine Ac(ala)(8)NHMe, *J Chem Phys* 114, 6443-6454.
51. Evans, D. A., and Wales, D. J. (2004) Folding of the GB1 hairpin peptide from discrete path sampling, *J Chem Phys* 121, 1080-1090.
52. Carr, J. M., Trygubenko, S. A., and Wales, D. J. (2005) Finding pathways between distant local minima, *J Chem Phys* 122.
53. Wales, D. J. (2005) The energy landscape as a unifying theme in molecular science, *Philos T Roy Soc A* 363, 357-375.
54. Wales, D. J., and Bogdan, T. V. (2006) Potential energy and free energy landscapes, *J Phys Chem B* 110, 20765-20776.

3. Exploring low-potential energy structures of a tryptophan zipper

3.1. Introduction

β hairpins are a common tertiary structural motif of proteins. Compared to α helices though, they tend to have more hydrophobic side chains exposed to the solution. In aqueous solutions, this can be detrimental to conformational stability; exposed hydrophobic cores can lead to aggregation as in amyloid diseases (1). Several force fields tend to favor helical conformations over the native β hairpin (2-4). With the advent of modern force fields, β hairpins have been simulated successfully (1, 4-24).

One family of polypeptides, tryptophan zippers (25) (trpzips), are stabilized by π -stacking of four alternating tryptophans. Structures of six different trpzips with different residues and different lengths were determined by NMR spectroscopy; three are 12-mers and three are modeled after the B1 domain(41-56) of Streptococcal protein G. Trpzip2 (Figure 3.1) is a 12-residue polypeptide that is the most stable of the 12-residue trpzips. Residues 6 and 7 form a type I' β turn, and Trp2, Trp4, Trp9, and Trp11 are implicated in π -stacking characteristic of tryptophan zippers. Trpzip2 has been simulated by canonical simulations, replica exchange, and generalized ensembles (2, 6, 15, 18, 26-39).

Tryptophan zippers have been studied extensively both experimentally (25, 26, 40-42) and computationally (6, 15, 18, 27-36, 38, 42) and results suggest that the

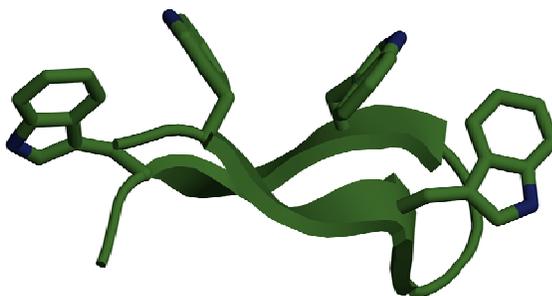


Figure 3.1 Trpzip2 NMR model 1 in cartoon with tryptophan side chains displayed.

tryptophan zipper family is composed of β hairpins. Reportedly, the major reason for the stability of tryptophan zippers is, as the name implies, the alternating tryptophans and their putative π -stacking interaction. The turn region and hydrogen bonds also play an important role in the stabilization of the conformation (15, 26, 36, 41, 42). Meanwhile, the termini are more flexible.

Our group recently developed two molecular dynamics (MD) protocols: the disrupted velocity (DIVE) search protocol and the divergent path (DIP) search simulations (39, 43). Both simulation protocols allow the system to traverse the potential energy (PE) surface more quickly than conventional MD simulations. The DIVE protocol perturbs both the magnitudes (kinetic energy) and directions of atomic velocities allowing the system both to overcome and to circumvent PE barriers. This, in turn, allows the system to locate conformations in PE minima near 0 K. The DIP protocol allows a system to traverse the PE surface in multiple directions simultaneously at a constant kinetic energy. Therefore, a system can locate PE minima by circumventing PE barriers. When the two protocols are used in conjunction with each other, several conformations may be found giving the system an opportunity to locate the lowest PE conformation near physiological temperatures.

The protocols have been tested on a 13-residue model of alanine (39, 43), Trp-cage (39), the 18-residue peptide F (39), and C-peptide of ribonuclease A (chapter 1). Each of these previous simulations involved an α helix, but we wanted to expand our protocol validation to β hairpins. We decided upon a simple case study by simulating tryptophan zipper 2 (trpzip2) (25). The goal is to determine whether the new protocols

can locate the lowest potential energy conformations of trpzip2 and whether the lowest potential energy conformation corresponds to the native conformations.

3.2. Algorithms and simulations

3.2.1. Theory

Proteins fold into a native conformation by locating the global free energy minimum. The free energy surface, however, is not easy to represent using molecular dynamics (MD) because of the temperature dependence and because of the challenge of calculating entropic contributions. MD instead simulates systems like protein folding on the temperature-independent potential energy surface. Potential energy is typically the major contributor to the free energy and can therefore give a close approximation to the free energy, especially at temperatures near 0 K. Our protocols focus on mapping the PE surface because of the aforementioned reasons.

Simulations using both the DIVE and the DIP protocols begin with an initial conformation simulated simultaneously with multiple, independent copies. Each copy is randomly assigned an initial velocity. The disrupted velocity (DIVE) search protocol is a microcanonical simulation (constant number of atoms, volume, and energy), but the divergent path (DIP) search protocol is a canonical ensemble simulation (constant number of atoms, volume, and temperature).

The atoms within each copy are assigned different initial atomic velocity magnitudes in the DIVE protocol. The different magnitudes allow the simulations to begin with several different initial kinetic energies, thus different initial temperatures. The simulations evolve for an assigned period before the atomic velocities are reassigned according to $\mathbf{p}' \equiv \sigma^{1/2} \mathbf{p}$ where \mathbf{p} and \mathbf{p}' are the momenta of the particles before and after

atomic velocity reassignment, respectively. σ is a scaling parameter that determines the magnitude of the simulation's kinetic energy after atomic velocity reassignment. When $\sigma > 1$, kinetic energy is added by increasing atomic velocities, but when $0 < \sigma < 1$, kinetic energy is removed by decreasing atomic velocities. When the atomic velocities are reassigned, both the magnitudes of the velocity vectors are rescaled and their directions are changed.

The scaling parameter σ may be chosen before starting the simulation or it may be calculated during the simulation using one of two different algorithms. σ may be calculated during the simulation by $\sigma = |T - \Delta T|/T$, where ΔT is defined as the difference in temperature before and after atomic velocity rescaling. The parameter may alternatively be defined by using a target temperature T_{target} to be achieved after atomic velocity rescaling with $\sigma = T_{\text{target}}/T$.

The DIVE protocol also uses a predefined threshold temperature. When simulation temperatures fall below the threshold temperature, kinetic energy is added to increase the temperature. If the simulation temperatures rise above the threshold temperature, especially after velocity rescaling, kinetic energy is removed thus lowering the simulation temperature. During one typical cycle of a DIVE simulation, heating occurs once whereas cooling occurs multiple times. Heating is used to surmount energy barriers, whereas cooling allows the system to sample potential energy minima near 0 K.

For DIP simulations, the atoms are assigned initial atomic velocities randomly according to a Gaussian distribution (44). Within different copies, the magnitude of an atom's velocity is the same but the directions are different, which allows the simulations

to maintain a constant temperature, while each copy samples different energies on the PE surface (39).

3.2.2. Data collection

For the simulations, the sequence of trpzip2 (pdb code: 1LE1, sequence: SWTWENGKWTWK-NH2) is selected. Three starting conformations are used for the simulations: fully-extended, β -hairpin, and α -helical conformations. As represented by the NMR models, Glu5 is deprotonated and Lys8 and Lys12 are protonated. For the β hairpin, model 1 from the 20 NMR models is selected because it best represents the conformation according to the PDB file header. To create the fully extended and the α -helical conformations, we used AMBER 8's (45) xleap feature. The appropriate ϕ and ψ dihedral angles are selected to form the fully extended conformation ($\phi = -180^\circ$ and $\psi = +180^\circ$) and the α -helical conformation ($\phi = -60^\circ$ and $\psi = -40^\circ$). From these three initial simulations, the lowest potential energy conformations are located using both the DIVE and DIP protocols. The lowest potential energy conformations from the second round of DIVE simulations are further simulated using the DIP protocol. In total, six DIVE simulations and nine DIP simulations are run. Conformations are designated by their derivation from the initial conformation. For instance, the second round of DIVE simulations obtained from the fully-extended conformation is assigned the name "ext".

The PEs of all conformations are minimized for 100 steps using steepest descent. Minimization is done in implicit solvent (46) using the Multiscale Modeling Tools for Structural Biology (MMTSB) (47) program. After minimization, the Molecular Modeling Toolkit (MMTK) (48) is used to convert the coordinate files into files suitable

for our suite of programs. The atomic velocities are scaled to an initial temperature of 300 K (used in DIP simulations only).

Other parameters are included during the simulations. Covalent bond distances to hydrogen are constrained with the SHAKE (49, 50) algorithm. Temperature is maintained by the Nosé-Hoover Chain method (51), and the equations of motion are integrated by the velocity-Verlet method (52). All simulations are run with a Generalized Born/surface area (GB/SA) implicit water solvent (46) with an external dielectric constant of 78.5, an internal dielectric constant of 1.0, surface tension at 0.005 kcal/mol-Å², and an offset of 0.9 Å. Lennard-Jones and electrostatic cutoffs are set to 999 Å to represent an infinite cutoff. A modified version of Amber99 (2) is used for minimization and data collection. Six copies are simultaneously simulated using the DIVE and DIP protocols, and simulations are run for 4 million steps/copy with a 2 fs time step (4×10^6 steps/copy \times 2 fs/step \times 6 copies/simulation = 48 ns total simulation time). Data is output every 250 steps (0.5 ps).

For DIVE simulations, each copy is initialized with atomic velocities scaled to a different temperature (10 K, 50 K, 100 K, 300 K, 600 K, and 1000 K). Velocity reassignment for each copy occurs after 20,000 steps. The scaling parameter for heating is calculated from $\delta = T_{target}/T$, and the scaling parameter for removal of kinetic energy is set to 0.25. The threshold temperature is 10 K. So simulations are heated to a $T_{target} = 1000$ K when the system temperature drops below the threshold temperature, but when the system temperature rises above the threshold temperature at the time of velocity reassignment, it is cooled to 0.25 of its temperature T .

For DIP simulations, all simulations are run at a temperature of 300 ± 20 K. As mentioned above, identical atoms within each of the six copies of each simulation are assigned the same atomic velocity magnitudes to maintain the desired temperature, but atomic velocities are assigned different directions. Thus, each copy is allowed to search for different PE minima.

3.2.3. Data analysis

The suite of MD programs, AMBER 8 (45), is used to calculate backbone atom root-mean-square deviations (rmsd's), all-atom rmsd's, tryptophan-only (indole ring) rmsd's, distances and angles involving atoms in hydrogen bonds (N–H···O), two charged polar side chain distances (Glu5···Lys8 and Glu5···Lys12), distances and improper dihedral angles between the indole rings, H–N–C α –H α and ϕ dihedral angles (for $^3J_{\text{HN-C}\alpha}$ and $^3J_{\phi}$ coupling constants calculations). Secondary structures are determined by the program STRIDE (53), which classifies secondary structures on the basis of backbone dihedral angles and H-bond interactions. The backbone atom rmsd's are calculated relative to NMR model 1 and all other NMR models. The $^3J_{\text{HN-C}\alpha}$ and $^3J_{\phi}$ coupling constants are calculated using the Karplus equation (54, 55) (eqs. (3.1) and (3.2)). The range of $^3J_{\text{HN-H}\alpha}$ coupling constants are found from experimental values, but the range of $^3J_{\phi}$ coupling constants are calculated from the C–N–C α –C dihedral angles of all 20 NMR models.

$$^3J_{\text{HN-H}\alpha} = 6.7 \cos^2 \theta - 1.3 \cos \theta + 1.5 \quad (3.1)$$

$$^3J_{\phi} = ^3J_{\text{CNC}\alpha\text{C}} = 1.8 \cos^2 \theta - 0.2 \cos \theta + 0.5 \quad (3.2)$$

where the angle θ is the dihedral angle for H–N–C α –H α or for C–N–C α –C. The ranges of $^3J_{\text{HN-H}\alpha}$ and $^3J_{\phi}$ coupling constants are calculated from the respective dihedral angles of

all NMR models. According to convention, hydrogen bonds (H-bonds) are defined by N–H \cdots O angles in a range of $180.0^\circ \pm 60.0^\circ$, N \cdots O distances less than 3.5 Å, and an appearance during more than 5% of the simulation. Each low energy conformation is analyzed to locate H-bonds. The aromatic ring distances are determined between the heavy atoms of the indole rings of the tryptophan side chains, and the improper dihedral angles are calculated for C $\delta_{\text{Trp}i}$ –C $\varepsilon_{\text{Trp}i}$ \cdots C $\varepsilon_{\text{Trp}j}$ –C $\delta_{\text{Trp}j}$ between tryptophans *i* and *j* (56). By convention, the proper geometric configuration for π -stacking between the Trp-pairs occurs when the distances between heavy atoms are below 6 Å with a dihedral angle within a range of $0^\circ \pm 90^\circ$.

Cluster analysis is convenient way to analyze a large number of conformations by grouping or “clustering” those with similar conformational features. Each cluster is defined by the comparison of rmsd’s between conformations, to categorize the conformations within a group. The average conformation of the cluster is called the centroid. The clustering procedure using MMTSB (47) consists of an iterative process with an error tolerance of 0.5 and a least-square fit rmsd comparison of the conformations (57) . Clusters are defined as having a radius of 3.0 Å. The analysis is done for simulations carried out using both the DIVE and DIP protocols. Cluster analysis for the DIVE protocol utilized 477 conformations. For the DIP protocol, conformations are obtained from all ten simulations. The conformations used are obtained after the equilibration of the polypeptide (as many as 8000 conformations per simulation). Every fourth conformation is extracted for the actual cluster analysis of DIP simulations.

3.3. *Simulation results and discussion*

By using the disrupted velocity (DIVE) and divergent path (DIP) protocols, the native conformation of trpzip2 is located. Other low-PE, non-native conformations are also encountered. The non-native conformations are lower in potential energy (PE) than the native conformation when simulated near 0 K; yet, the native conformations has a lower average PE near physiological temperatures. More detailed analysis of the simulations allows some insight into the stability of the non-native conformations near physiological temperatures as well as at lower temperatures.

The remainder of this section describes the various conformations encountered during the simulations. The two conformations closely resembling the native β hairpin shall be described first. These two conformations are equilibrated with the DIP protocol. Alternative conformations that are equilibrated by the DIP protocol shall then be mentioned. Following the aforementioned descriptions, the conformations from the DIVE protocol shall be described. The conformational descriptions shall follow the same order of the corresponding equilibrated conformations. Finally, a discussion of the significance of the simulations and cluster analysis results shall be provided.

3.3.1. Divergent path (DIP) simulations

Two β hairpins are simulated using the DIP protocol. NMR model 1 and a β hairpin from the DIVE simulations are equilibrated to determine whether the secondary structure will undergo a conformational change. Both β hairpin simulations (β , Figure 3.2b and β^* , Figure 3.2j) maintain their secondary structure with backbone atom rmsd's below 1.0 Å relative to the experimental NMR model 1. Unlike the native conformation (β strands: Trp2–Glu5 and Lys8–Trp11), the average equilibrated conformation of β

consists of β strands ranging from Trp2 to Trp4 and from Trp9 to Trp11 with a type IV β turn in between. β^* , on the other hand, is composed of β strands similar to the native conformation, but the β turn shifts from a type I' (backbone dihedral angles of $60^\circ/30^\circ$ and $90^\circ/0^\circ$ with a 20° range for residues $i+1$ and $i+2$, respectively) to a type IV turn. The all heavy-atom and indole ring rmsd's are below 2.0 \AA relative to the native conformation. The two charged polar side chain pairs (Glu5 \cdots Lys8 and Glu5 \cdots Lys12) maintain proper distance during more than 30% of the simulation (with $\angle\text{O}\cdots\text{H}-\text{N} < 120^\circ$). The Trp2–Trp11 and Trp4–Trp9 indole rings are in close proximity for proper π -stacking (ring distance $< 6 \text{ \AA}$, ring angle $< \pm 90^\circ$), but similar to the native conformation, the indole rings between Trp4 and Trp11 are not aligned. On average, more than 70% of the $^3J_{\text{HN}-\text{H}\alpha}$ coupling constants and more than 60% of the $^3J_\phi$ coupling constants agree with experimental data.

Near 300 K, the stability of the β hairpin is attributed to interactions between charged side chain pairs and aromatic ring pairs. Both β hairpins have the Glu5 \cdots Lys8 and Glu5 \cdots Lys12 side chains in close proximity implying potential salt bridge formation. These structural features were not reported in other published simulation studies using explicit solvent and may be due to over-stabilization of salt bridges by the implicit solvent model (58, 59). Nonetheless, the tryptophans are aligned within 2.0 \AA for both β -hairpin conformations as observed by Cochran, et al. (25)

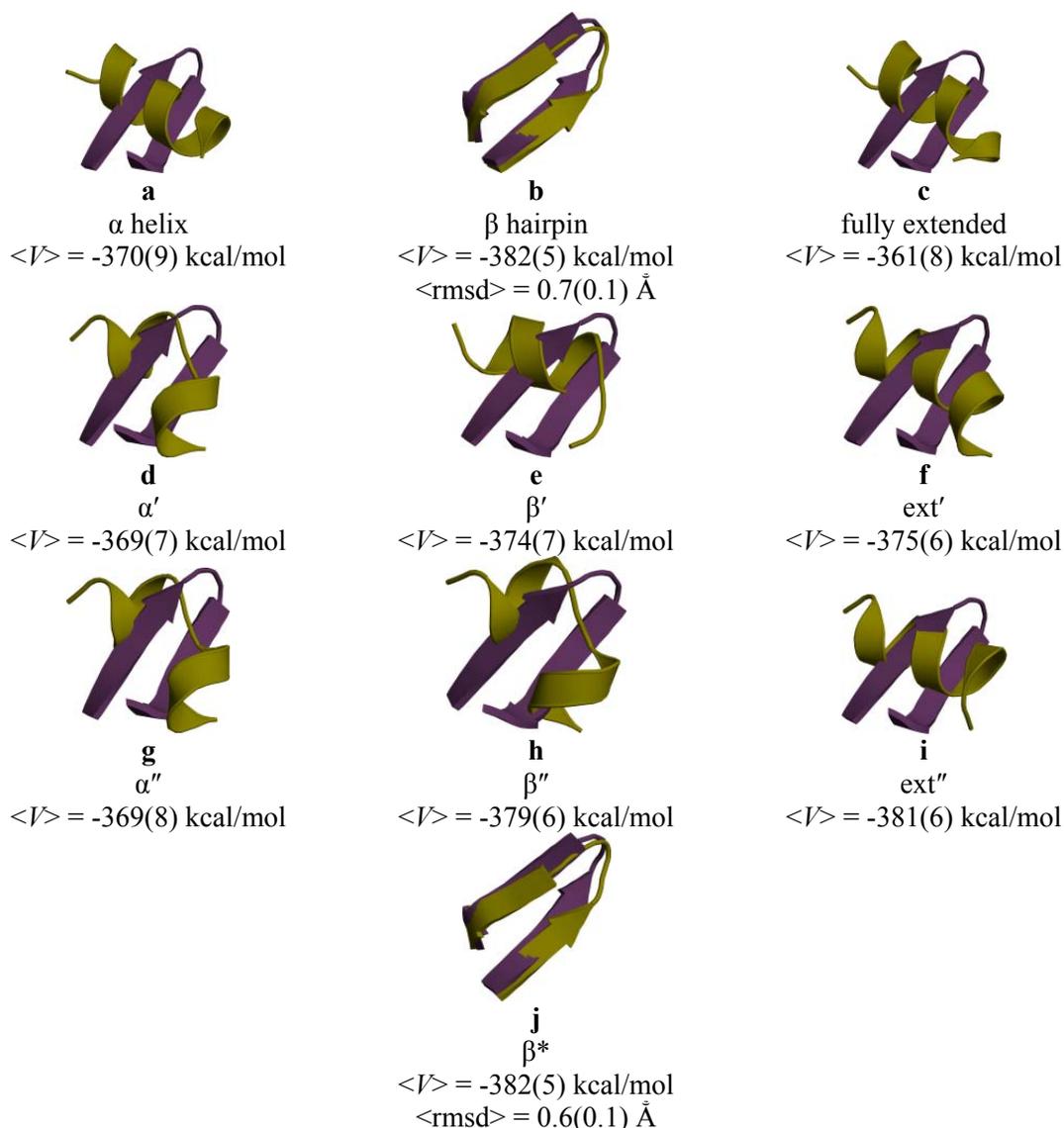


Figure 3.2 Average conformation from DIP simulations (yellow) overlaid with NMR model 1 (magenta). Conformations listed by the initialized conformation. Backbone atom rmsd's relative to all NMR models.

The other conformations that are presented do not fold into a β hairpin. Instead, we offer alternative conformations composed primarily of α helices. The α helix is predominant, but the conformation fluctuates between this and the 3_{10} or π helix. The distance between the Glu5 and Lys12 charged polar side chains is less than 3.5 Å during the β' simulation, whereas the Glu5 and Lys8 side chains are within close proximity during four simulations (α , β' , ext', and ext''). The native tryptophan pairs do not appear

during any of the following simulations; however, the indole rings of Trp4 and Trp11 are properly aligned during three simulations (α' , α'' , and β''). Unlike the β hairpin simulations, the following simulations exhibit 10–30% of the H–N–C α –H α dihedral angles and 45–55% of the C–N–C α –C dihedral angles relative to the coupling constants for the experimental data.

Eight simulations yield average conformations with an helical secondary structure of varying lengths. Simulations begun in both the α -helical (α , Figure 3.2a) and the fully-extended (ext, Figure 3.2c) conformations give low-energy conformations exhibiting similar α -helical conformations. The α helix simulation has the helix extending from Trp2 to Trp11, whereas the conformation labeled ext has the helix from Trp2 to Trp9. The ext' (Figure 3.2f) simulation on average is composed of an α helix from Trp2 to Trp11. When ext'' (Figure 3.2i) is simulated at 300 K, an α helix from Trp2 to Trp11 comprises the average conformation. α' (Figure 3.2d) and α'' (Figure 3.2g) are similar in secondary structure (0.2 Å backbone atom rmsd): an α helix from Trp2 to Asn6 and a 3_{10} helix from Lys8 to Thr10. β' (Figure 3.2e) is composed of an α helix from Trp2 to Lys8, whereas an α helix from Trp2 to Asn6 comprises β'' , (Figure 3.2h).

Because the potential energies for the various conformations are within such close proximity, we chose to use cluster analysis to judge the similarity of the conformations produced by each simulation. Cluster analysis allows the conformations to be grouped on the basis of conformation similarities and allows for a qualitative analysis of the entropic component of free energy. For the DIP protocol, 20,010 conformations are analyzed. Five clusters comprise more than 1000 conformations, and among these five clusters, the

β -hairpin motif is located in only one cluster, representing 4,004 conformations. The other five clusters are composed of helical conformations.

The β -hairpin cluster is the second largest cluster and allows for a plausible conformation. Basing our results on simulation data indicates that the β hairpin is a possible secondary structure, and cluster analysis implies that the β hairpin is a dominant conformation. The large number of conformations within the β -hairpin cluster qualitatively suggests that the entropy component of the free energy may favor β hairpins less than α helices solely on the basis of cluster analysis. One would expect the native conformation to be highly favored compared with α helices, but cluster analysis suggests differently. The stability of helical conformations indicates either that the experimentally-determined conformation is not, in fact, in the global free energy minimum, or that α helices may be favored entropically but the β hairpin may be energetically favored. However, when both cluster analysis and the simulations are taken into account, β hairpins may indeed be the global free energy minimum at 300 K.

One final point to note is the overlapping PE minima found during the DIP simulations (Figure 3.3). Yang, et al. (18) mentioned that the trpzip2 free energy surface is rough which agrees with our simulations near physiological temperatures. However, the overlapping potential energy minima also imply another interesting point. The modified Amber99 force field (2) does not seem to be as biased towards α helices as Amber94 or Amber99 (2, 3, 19, 60-66).

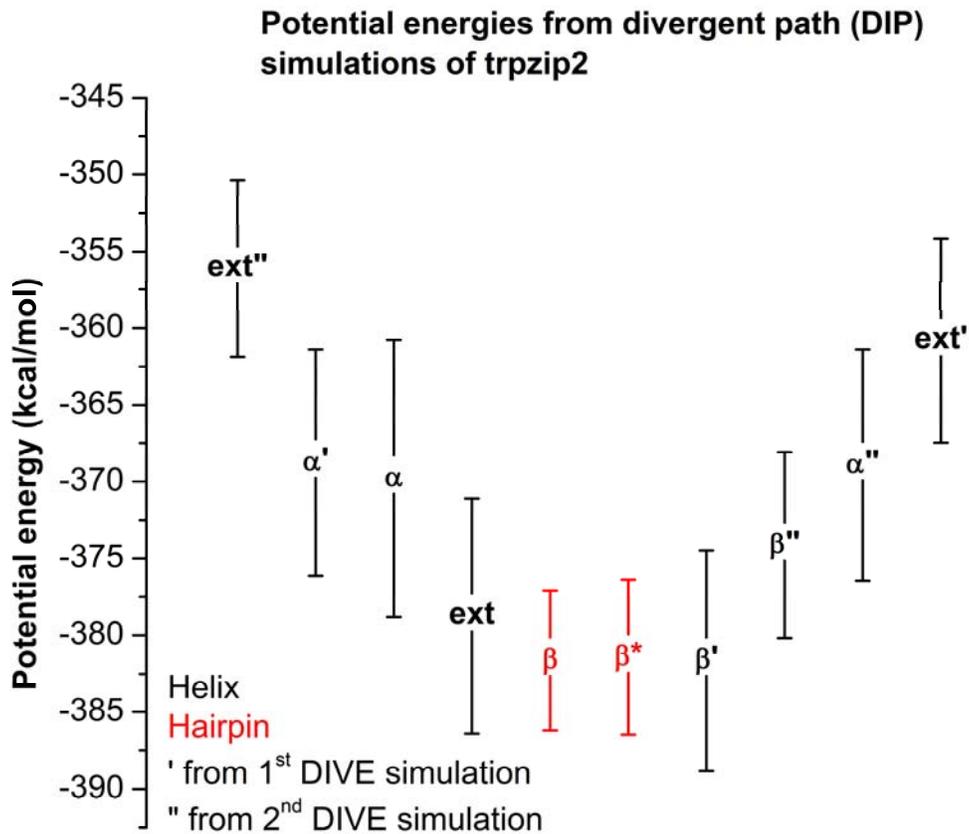


Figure 3.3 Average potential energies with standard deviation error bars for the ten simulations.

In the previous chapter, we noted that the potential energy landscape can be described by disconnectivity graphs. We suggest that a distinctive global PE exists for trpzip2, with PE minima separated by high-PE barriers. High PE barriers are inferred simply because the conformations fail to converge into a single conformation. According to Wales, et al. (67-69), this type of PE surface corresponds to the ‘weeping willow’ model (Figure 3.4) because the conformations fail to converge into a single conformation. The conformational energies appear to overlap on the basis of standard deviations, but according to the standard error (data not shown), the PEs are fairly accurate (± 0.1 kcal/mol). Within the accuracy of the simulations, each conformation appears to reside in a local PE minimum, and the β hairpins have converged into a single PE minimum—

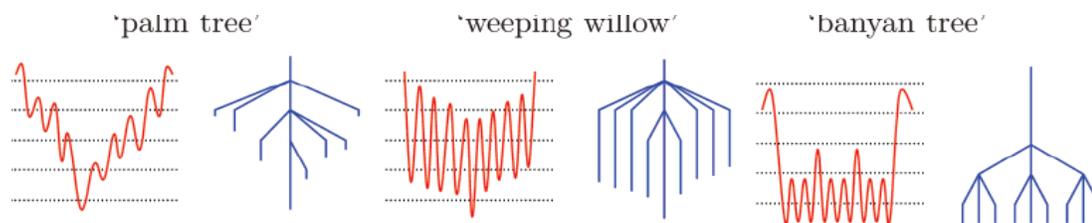


Figure 3.4 Three pairs of PE landscapes (left) and their corresponding disconnectivity graphs (right). The graphs are drawn as PE (vertical axis) relative to arbitrary coordinates (horizontal axis). The endpoints of the disconnectivity graphs represent PE minima, and the points where the branches are joined correspond to a common PE “superbasin”. For the PE landscape, the wells represent the minima of a system surrounded by PE barriers. Reprinted by permission from ref. (68). Copyright 2006 American Chemical Society

the global PE minimum. If the PE surface supported the ‘palm tree’ model, our conformations should have converged to a single PE minimum and its corresponding secondary structure.

According to our simulations, trpzip2 adopts the β hairpin structure as the global PE minimum near physiological temperature, but cluster analysis implies that α helices, supported by cluster analysis, may be equally valid conformations. Since α helices are reportedly not encountered during the actual folding process (6, 15, 18, 26-28, 30-35, 38, 41), the local PE minima for α helices may form a narrow basin compared with the native β hairpin, which may reside in a wide basin.

3.3.2. Disrupted velocity (DIVE) simulations

DIVE simulations located 486 potential energy minima, spanning a range of 43 kcal/mol ($V = -499$ kcal/mol to $V = -456$ kcal/mol). One conformation that corresponds to the experimental conformation is described below. The conformation closely resembling the experimental conformation is designated β^* because it originated from a simulation starting from the native β hairpin. Six other conformations corresponding to the lowest PE conformations are also described below.

Thirteen β -hairpins-like conformations are located during the DIVE simulations. The β hairpin (β^* , Figure 3.5d) corresponds to the lowest PE among the thirteen, but is located among the higher PE conformations compared with the other conformations listed. The β^* conformation compares well with the experimental structure, as both structures display β strands between Trp2–Glu5 and Lys8–Trp11 with a type IV turn between Glu5–Lys8. The backbone atom rmsd's are less than 1.0 Å between β^* and the experimental conformations, and the all heavy-atom rmsd is 1.9 Å compared with all NMR models. When the heavy atoms of the indole rings are compared with all NMR models, they are within 2.1 Å of the native conformation. Both experimentally-observed charged polar side chains (Glu5...Lys8 and Glu5...Lys12) are less than 3.5 Å apart. The aromatic side chains for Trp2 and Trp11 and for Trp4 and Trp9 are oriented for proper π -stacking (ring distance < 6.0 Å and ring angle < $\pm 90^\circ$). The Trp11 and Trp4 indole rings do not align because they are in a Y-formation as expressed within the native

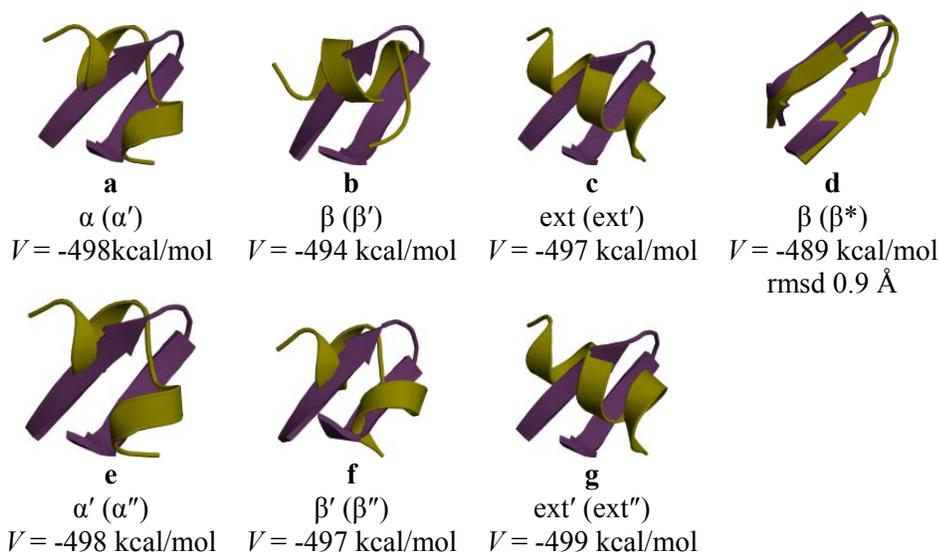


Figure 3.5 Conformations corresponding to the lowest PE (yellow) from DIVE simulations overlaid with NMR model 1 (magenta). The conformations are listed by the initialized conformation and with a unique identifier. The rmsd is backbone atom rmsd relative to all NMR models.

conformation. 80.0% of $^3J_{\text{HN-H}\alpha}$ coupling constants and 72.7% of $^3J_{\text{phi}}$ coupling constants agree with experimental data..

The alternative conformations are primarily composed of varying lengths of α helices. Ext'' (Figure 3.5c) corresponds to the lowest PE of the seven conformations and comprises an α helix spanning the amino acids Ser1–Thr10. The conformation ext' (Figure 3.5g) is similar in conformation to ext'' (backbone atom rmsd 0.1 Å). α' (Figure 3.5a) and α'' (Figure 3.5e) are similar to each other in secondary structure (backbone atom rmsd 0.1 Å). Their overall conformation comprises an α helix (Trp2–Asn6) and a 3_{10} helix (Lys8–Thr10). β' (Figure 3.5b) and β'' (Figure 3.5f) both consist of α helices from Trp2 to Lys8 and from Trp2 to Asn6, respectively.

The β hairpin is found in a higher PE minimum than the following alternative conformations that show enhanced stability near 0 K, compared to 300 K. Yet, a native conformation is located; however, we now present alternative conformations that may exist near 0 K. The side chains of Glu5 and Lys8 are at a distance of less than 3.5 Å for two conformations (ext' and ext''). None of the alternative conformations exhibit the side chains of Glu5 and Lys12 in close proximity. β'' is the only conformation that appears to have any tryptophan side chains properly oriented, but the orientation is between a non-native pair: Trp4 and Trp11. 10–20% of the coupling constants for the H–N–C α –H α dihedral angles, and 35–55% of the coupling constants for the C–N–C α –C dihedral angles agree with experimental data.

477 conformations are analyzed by cluster analysis. Nine clusters contained more than 10 conformations (40.9% of the 477 conformations). Cluster analysis finds that DIVE locates conformations resembling the native β hairpin but the clusters include few

conformations, so the native conformation appears qualitatively not to be entropically favored near 0 K. One cluster contained a representative conformation with a type IV β turn between Gln5 and Lys8 as exhibited by the equilibrated β hairpins but lacked the actual β strands. The two largest clusters contained more than 40 conformations and consist of an α helix and 3_{10} helix combined with an α helix. One cluster contained 26 conformations representing α helices from Trp2 to Lys8. In fact, the majority of the clusters contained helices, so cluster analysis qualitatively suggests that α helices are entropically favored over β hairpins near 0 K.

We are viewing snapshots of conformations and not the dynamic motion of the conformations. At low temperatures, the conformations are deep within their PE minima and are surrounded by high PE barriers. We sample several conformations near 0 K including β hairpins. Our conventional MD simulations do not sample multiple conformations near 0 K, but the DIVE protocol gives trpzip2 the opportunity to fold into alternative conformations.

According to Yang, et al. (18), the free energy surface of trpzip2 is rough with several local minima. Our DIVE simulations show that the PE surface has a similar shape, and near 0 K trpzip2 is stable in a helical conformation. Figure 3.5 presents the conformations corresponding to the lowest PE from among the six simulations, in addition to a β -hairpin conformation, which appears higher in PE compared with several helical conformations. One plausible explanation for the stability of α helices at low temperatures is the ease of hydrogen bonds formation compared with a hydrophobic collapse. According to Muñoz et al. (70), β -hairpin formation is not as favorable as α -helix formation because α helices compensate entropy loss by H-bond formation. In

contract, H-bond formation occurs after fixation of φ and ψ for β hairpins followed by hydrophobic side chain interactions for stabilization. The energy needed to overcome the van der Waals forces is not present, and therefore, more helical conformations are observed than β hairpins. The presumed entropy contribution and the low temperature prevent the helices from transitioning into β hairpins.

We suggest that the helical conformations near 0 K are metastable (18, 31, 38). As the kinetic energy increases, the β hairpin may reside within the global PE minimum. Other groups also find non-native conformations but at higher energies compared with the native state, which is considered to correspond to the global free energy minimum (18, 31, 38).

We have shown that, with the DIVE and DIP protocols, trpzip2 has the ability to sample unique conformations both near 0 K and at 300 K. With the DIVE protocol, we find that trpzip2 can fold into helical conformations, but when the conformations are subjected to higher kinetic energies, the β hairpin becomes lower in PE than the non-native conformations.

3.4. Conclusions

We have tested the ability of both the DIVE and the DIP protocols to reproduce the native β -hairpin structure of tryptophan zipper 2 near 0 K and around 300 K. Our simulations show that we can reproduce the native conformation within 0.7 Å backbone atom rmsd relative to the NMR models using the DIVE simulations and within 0.5 Å backbone atom rmsd relative to the NMR models using the DIP simulations. The β -hairpin structure represents the global potential energy (PE) minimum found by using the DIP method. The β -hairpin-like conformations are 0.2 kcal/mol lower than an α -helical

conformation. The β -hairpin conformations from both the DIVE and the DIP simulations were started from a native β -hairpin-like conformation. Cluster analysis of the simulations reveals that the β -hairpin conformations have a backbone atom rmsd less than 2.0 Å from the NMR structure and contains 16% of the structures analyzed from the DIP simulations and 2% for the DIVE simulations.

We also find several α -helical conformations that range from 0.2 kcal/mol to 25.5 kcal/mol above the global PE minimum. We propose that tryptophan zipper 2 can fold into an α helix but is more likely to fold into the experimentally observed β hairpin under the conditions of the experiment. On the basis of cluster analysis, we further propose that the native conformation lies within a narrow-basin global PE minimum surrounded by high PE barriers, which prevent simulated structures from refolding into the conformation residing in the global PE minimum.

We have shown that the disrupted velocity (DIVE) and the divergent path (DIP) protocols allow for a broad sampling of different conformations. The DIVE protocol permits trpzip2 to traverse the PE surface, map PE minima, and detect different PE minima lying within a small energy range. By using the DIP protocol together with DIVE, we are further able to search the PE surface near 300 K in search of the global PE minimum.

β hairpins present a challenge in molecular dynamics simulations because of the current force fields available, but we find that the two protocols can locate β hairpins from a primary sequence. For the DIVE protocol, the global PE minimum comprises a helical conformation, but the helical conformations are higher in average PE relative to the native conformation near 300 K. We therefore propose that the β hairpin corresponds

to the global PE minimum but that the helical conformations are viable alternative structures in local PE minima surrounded by high energy barriers.

3.5. References

1. Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riekel, C., Grothe, R., and Eisenberg, D. (2005) Structure of the cross-beta spine of amyloid-like fibrils, *Nature* 435, 773-778.
2. Okur, A., Strockbine, B., Hornak, V., and Simmerling, C. (2003) Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins, *J Comput Chem* 24, 21-31.
3. Yoda, T., Sugita, Y., and Okamoto, Y. (2004) Comparisons of force fields for proteins by generalized-ensemble simulations, *Chem Phys Lett* 386, 460-467.
4. Zhou, R. H. (2003) Free energy landscape of protein folding in water: Explicit vs. implicit solvent, *Proteins* 53, 148-161.
5. Ando, T., and Yamato, I. (2005) Free energy landscapes of two model peptides: alpha-helical and beta-hairpin peptides explored with Brownian dynamics simulation, *Mol Simulat* 31, 683-693.
6. Carr, J. M., Trygubenko, S. A., and Wales, D. J. (2005) Finding pathways between distant local minima, *J Chem Phys* 122.
7. De Alba, E., Rico, M., and Jimenez, M. A. (1997) Cross-strand side-chain interactions versus turn conformation in beta-hairpins, *Protein Sci* 6, 2548-2560.
8. Felts, A. K., Harano, Y., Gallicchio, E., and Levy, R. M. (2004) Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model, *Proteins-Structure Function and Bioinformatics* 56, 310-321.
9. Garcia, A. E., and Sanbonmatsu, K. Y. (2001) Exploring the energy landscape of a beta hairpin in explicit solvent, *Proteins* 42, 345-354.
10. Honda, S., Kobayashi, N., and Munekata, E. (2000) Thermodynamics of a beta-hairpin structure: Evidence for cooperative formation of folding nucleus, *J Mol Biol* 295, 269-278.
11. Malolepsza, E., Boniecki, M., Kolinski, A., and Piela, L. (2005) Theoretical model of prion propagation: A misfolded protein induces misfolding, *P Natl Acad Sci USA* 102, 7835-7840.
12. Monticelli, L., and Colombo, G. (2004) The influence of simulation conditions in molecular dynamics investigations of model beta-sheet peptides, *Theor Chem Acc* 112, 145-157.
13. Okamoto, Y., and Hansmann, U. H. E. (1995) Thermodynamics of Helix-Coil Transitions Studied by Multicanonical Algorithms, *J Phys Chem-US* 99, 11276-11287.
14. Okamoto, Y., Masuya, M., Nabeshima, M., and Nakazawa, T. (1999) beta-sheet formation in BPTI(16-36) by Monte Carlo simulated annealing, *Chem Phys Lett* 299, 17-24.
15. Snow, C. D., Qiu, L. L., Du, D. G., Gai, F., Hagen, S. J., and Pande, V. S. (2004) Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy, *P Natl Acad Sci USA* 101, 4077-4082.
16. Swope, W. C., Pitera, J. W., Suits, F., Pitman, M., Eleftheriou, M., Fitch, B. G., Germain, R. S., Rayshubski, A., Ward, T. J. C., Zhestkov, Y., and Zhou, R. (2004) Describing protein folding kinetics by molecular dynamics simulations. 2.

- Example applications to alanine dipeptide and beta-hairpin peptide, *J Phys Chem B* 108, 6582-6594.
17. Wei, G. H., Derreumaux, P., and Mousseau, N. (2003) Sampling the complex energy landscape of a simple beta-hairpin, *J Chem Phys* 119, 6403-6406.
 18. Yang, W. Y., Pitera, J. W., Swope, W. C., and Gruebele, M. (2004) Heterogeneous folding of the trpzip hairpin: Full atom simulation and experiment, *J Mol Biol* 336, 241-251.
 19. Yoda, T., Sugita, Y., and Okamoto, Y. (2004) Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations, *Chem Phys* 307, 269-283.
 20. Zagrovic, B., and Pande, V. S. (2003) Structural correspondence between the alpha-helix and the random-flight chain resolves how unfolded proteins can have native-like properties, *Nat Struct Biol* 10, 955-961.
 21. Zagrovic, B., Sorin, E. J., and Pande, V. (2001) beta-hairpin folding simulations in atomistic detail using an implicit solvent model, *J Mol Biol* 313, 151-169.
 22. Zhou, R. H., and Berne, B. J. (2002) Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water?, *P Natl Acad Sci USA* 99, 12777-12782.
 23. Zhou, R. H., Berne, B. J., and Germain, R. (2001) The free energy landscape for beta hairpin folding in explicit water, *P Natl Acad Sci USA* 98, 14931-14936.
 24. Zhu, J. A., Shi, Y. Y., and Liu, H. Y. (2002) Parametrization of a generalized born/solvent-accessible surface area model and applications to the simulation of protein dynamics, *J Phys Chem B* 106, 4844-4853.
 25. Cochran, A. G., Skelton, N. J., and Starovasnik, M. A. (2001) Tryptophan zippers: Stable, monomeric beta-hairpins, *P Natl Acad Sci USA* 98, 5578-5583.
 26. Du, D. G., Zhu, Y. J., Huang, C. Y., and Gai, F. (2004) Understanding the key factors that control the rate of beta-hairpin folding, *P Natl Acad Sci USA* 101, 15915-15920.
 27. Guvench, O., and Brooks, C. L. (2005) Tryptophan side chain electrostatic interactions determine edge-to-face vs parallel-displaced tryptophan side chain geometries in the designed beta-hairpin "trpzip2", *J Am Chem Soc* 127, 4668-4674.
 28. Ulmschneider, J. P., and Jorgensen, W. L. (2004) Polypeptide folding using Monte Carlo sampling, concerted rotation, and continuum solvation, *J Am Chem Soc* 126, 1849-1857.
 29. Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R., and Pande, V. S. (2002) Native-like mean structure in the unfolded ensemble of small proteins, *J Mol Biol* 323, 153-164.
 30. Pitera, J. W., Haque, I., and Swope, W. C. (2006) Absence of reptation in the high-temperature folding of the trpzip2 beta-hairpin peptide, *J Chem Phys* 124, -.
 31. Zhang, J., Qin, M., and Wang, W. (2006) Folding mechanism of beta-hairpins studied by replica exchange molecular simulations, *Proteins-Structure Function and Bioinformatics* 62, 672-685.
 32. Chen, J. H., Im, W. P., and Brooks, C. L. (2006) Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field, *J Am Chem Soc* 128, 3728-3736.

33. Arunachalam, J., Kanagasabai, V., and Gautham, N. (2006) Protein structure prediction using mutually orthogonal Latin squares and a genetic algorithm, *Biochem Bioph Res Co* 342, 424-433.
34. Singhal, N., and Pande, V. S. (2005) Error analysis and efficient sampling in Markovian state models for molecular dynamics, *J Chem Phys* 123, -.
35. Singhal, N., Snow, C. D., and Pande, V. S. (2004) 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.
36. Smith, A. W., Chung, H. S., Ganim, Z., and Tokmakoff, A. (2005) Residual native structure in a thermally denatured beta-hairpin, *J Phys Chem B* 109, 17025-17027.
37. Streicher, W. W., and Makhatadze, G. I. (2006) Calorimetric evidence for a two-state unfolding of the beta-hairpin peptide trzip4, *J Am Chem Soc* 128, 30-31.
38. Yang, W. Y., and Gruebele, M. (2004) Detection-dependent kinetics as a probe of folding landscape microstructure, *J Am Chem Soc* 126, 7758-7759.
39. Huang, Z. (2005) Design of new molecular dynamics global minimum search protocols for mapping energy landscapes and conformations of folded polypeptides and mini-proteins, in *Chemistry and Biochemistry*, p 282, The University of Oklahoma, Norman.
40. Dempsey, C. E., and Mason, P. E. (2006) Insight into indole interactions from alkali metal chloride effects on a tryptophan zipper beta-hairpin peptide, *J Am Chem Soc* 128, 2762-2763.
41. Du, D. G., Tucker, M. J., and Gai, F. (2006) Understanding the mechanism of beta-hairpin folding via phi-value analysis, *Biochemistry-Us* 45, 2668-2678.
42. Wang, J. P., Chen, J. X., and Hochstrasser, R. M. (2006) Local structure of beta-hairpin isotopomers by FTIR, 2D IR, and ab initio theory, *J Phys Chem B* 110, 7545-7555.
43. Huang, Z. N., and Wheeler, R. A. Replica Microcanonical Search Methods for Mapping Potential Energy Landscapes and Conformations of miniproteins, *In preparation*.
44. Stultz, C. M., and Karplus, M. (1998) On the potential surface of the locally enhanced sampling approximation, *J Chem Phys* 109, 8809-8815.
45. Case, D. A., Darden, T. A., Cheatham, I., T.E. , Simmerling, C. L., Wang, J., Duke, R. R. E., Merz, L., K.M. , Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) AMBER 8, 8 ed., University of California, San Francisco, San Francisco.
46. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular Simulations, *Biopolymers* 56, 275-291.
47. Feig, M., Karanicolas, J., and Brooks, C. L. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology, *J Mol Graph Model* 22, 377-395.
48. Hinsen, K. (2000) The molecular modeling toolkit: A new approach to molecular simulations, *J Comput Chem* 21, 79-85.
49. Andersen, H. C. (1983) Rattle: a Velocity Version of the Shake Algorithm for Molecular Dynamics Calculations, *J Comput Phys* 52, 24-34.

50. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes, *J Comput Phys* 23, 327-341.
51. Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992) Nosé-Hoover Chains: the Canonical Ensemble via Continuous Dynamics, *J Chem Phys* 97, 2635-2643.
52. Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A Computer-Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *J Chem Phys* 76, 637-649.
53. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins* 23, 566-579.
54. Kao, L. F., and Barfield, M. (1985) Conformational Dependencies of Vicinal C-13(O)-N-C-Alpha-C-13 and C-13(O)-N-C-Alpha-H-1 Coupling Constants in Compounds Which Model the Peptide Backbone, *J Am Chem Soc* 107, 2323-2330.
55. Ludvigsen, S., Andersen, K. V., and Poulsen, F. M. (1991) Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of phi-angles, *J Mol Biol* 217, 731-736.
56. McGaughey, G. B., Gagne, M., and Rappe, A. K. (1998) pi-stacking interactions - Alive and well in proteins, *J Biol Chem* 273, 15458-15463.
57. Lloyd, S. P. (1982) Least-Squares Quantization in Pcm, *Ieee T Inform Theory* 28, 129-137.
58. Im, W., Feig, M., and Brooks, C. L. (2003) An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins, *Biophys J* 85, 2900-2918.
59. Simonson, T., Archontis, G., and Karplus, M. (2002) Free energy simulations come of age: Protein-ligand recognition, *Accounts of Chemical Research* 35, 430-437.
60. Feig, M., MacKerell, A. D., and Brooks, C. L. (2003) Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations, *J Phys Chem B* 107, 2831-2836.
61. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins-Structure Function and Bioinformatics* 65, 712-725.
62. Hu, H., Elstner, M., and Hermans, J. (2003) Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution, *Proteins* 50, 451-463.
63. Simmerling, C., Strockbine, B., and Roitberg, A. E. (2002) All-atom structure prediction and folding simulations of a stable protein, *J Am Chem Soc* 124, 11258-11259.
64. Sorin, E. J., and Pande, V. S. (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations, *Biophys J* 88, 2472-2493.

65. Wang, T., and Wade, R. C. (2006) Force field effects on a beta-sheet protein domain structure in thermal unfolding simulations, *J Chem Theory Comput* 2, 140-148.
66. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G. M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. M., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J Comput Chem* 24, 1999-2012.
67. Doye, J. P. K., and Wales, D. J. (1996) On potential energy surfaces and relaxation to the global minimum, *J Chem Phys* 105, 8428-8445.
68. Wales, D. J., and Bogdan, T. V. (2006) Potential energy and free energy landscapes, *J Phys Chem B* 110, 20765-20776.
69. Mortenson, P. N., and Wales, D. J. (2001) Energy landscapes, global optimization and dynamics of the polyalanine Ac(ala)(8)NHMe, *J Chem Phys* 114, 6443-6454.
70. Muñoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of beta-hairpin formation, *Nature* 390, 196-199.

4. Structure of the B1 domain(41–56) of protein G as a β hairpin

4.1. Introduction

Predicting a secondary or tertiary structure from the primary amino acid sequence of a protein is a goal of biochemists. α helices have been studied for several years now, but β -sheets have been harder to simulate. Both encounter entropic loss when folded, but α helices compensate by more extensive hydrogen bond formation (1). One reason that β -sheets have not been studied so extensively is because of force fields, which tend to favor helical conformations (2-9). However, recent developments in force fields allow for the determination of β -sheets and β hairpin structures with less bias towards helices (10, 11).

Residues 41–56 (GEWTYDDATKTFTVTE) of the B1 domain of Streptococcal protein G (peptide G) form a naturally-occurring β hairpin (Figure 4.1), which has undergone extensive study (1, 4, 12-46). The actual B1 domain of protein G is a 56-residue protein with a $\beta\beta\alpha\beta\beta$ motif where the helix forms a crossover between the two hairpins (20). The protein is presumed stable because the hydrophobic areas of the

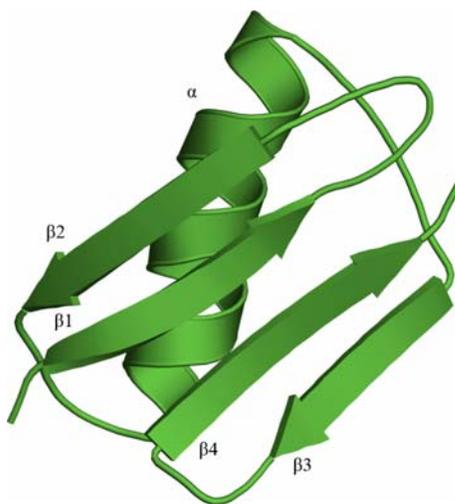


Figure 4.1 Full B1 domain of Streptococcal protein G

hairpins are buried near the helix, exposing the hydrophilic residues to the solution. Blanco et al. discovered that residues 41–56 (peptide G) retain a hairpin motif in water (30–40% hairpin motif) (1, 13, 24, 44)..

The β 3 (42–47) and β 4 (50–56) segments of protein G in water were found to be stable as β hairpins by both experimental (1, 13, 14, 22, 24, 30) and computational (12, 15, 19, 25, 27, 29, 31, 32, 34, 39, 40, 47, 48) studies. Residues 42–47 and 50–56 are antiparallel strands connected by a 6-residue turn (residues 46–51). Several factors have been deemed responsible for the stability of the β hairpin including the turn region, main-chain hydrogen bonds, salt bridges, and the hydrophobic core. The turn region has been shown to be the initiating factor for hairpin formation (1, 24, 29, 40). The turn region is stabilized by the close contact between the backbone carbonyl of Asp46 and the backbone amino group of Ala48 as well as by the hydrogen bonding between Thr49 and Thr51. Upon formation of the turn region, the hairpin is born out of interactions between the three aromatic rings of Trp3, Tyr5, and Phe12 with Val54. A “zipper” effect follows in which the main-chain hydrogen bonds form beginning from the turn region. Salt bridges further stabilize the hairpin. In addition, the β 3 and β 4 interaction is believed stabilized by a hydrophobic core including interactions between Trp43···Val54 and Tyr45···Phe52 pairs. Huyghues-Despointes noted that the terminal charges of the hairpin also enhance its stability (45).

β hairpins tend to be harder to simulate than α helices. One major reason is the bias of several force fields favoring helical conformations, including the popular Amber94 and Amber99 force fields (2, 6, 10, 11, 42, 49-53). Muñoz (1) states that helix formation is easier than hairpin formation because of energy compensation. Formation of

both an α helix and a β hairpin costs entropy, but helices compensate more for this loss by hydrogen bond formation. β hairpins, on the other hand, must compensate by turn formation, side chain interactions, and main chain hydrogen bond stabilization. Krivov and Karplus (26) note that the α helical conformation of peptide G is lower in free energy (FE) than the native β hairpin and that the non-native conformation is stabilized by its higher entropy. However, non-native conformations lie within local free energy minima surrounded by high free energy barriers. Garcia (19), Zagrovic (37), Zhou (4), and Gallicchio (18) notice some helical content within peptide G on the basis of different simulation conditions. Khavrutskii (46) even states that the native β hairpin of peptide G may not be the lowest energy conformation but, in fact, that several conformations, including the β hairpin and α helices, may coexist because they have similar energies. The Pak group (47) confirms this by finding semi- α -helical conformations that are 11 kcal/mol lower in FE than the experimental conformation. Gallicchio et al. (18), on the other hand, suggests that α helices may be metastable conformations en route to the thermodynamically-favored β hairpin. Helical conformations are possible even when they are substantially higher energy than the experimental conformation because, as for the β hairpin, local contacts between backbone atoms stabilize the helix and the side chain interactions do not strongly disfavor a helix (18). Multiple free energy minima between helices and β hairpins are also energetically competitive because main-chain interactions offer similar stabilization. The side-chain interactions, therefore, determine which of the two motifs is favored thermodynamically.

Recently, two protocols for geometry optimization were developed within our group (54, 55). The two protocols, disrupted velocity (DIVE) and divergent path (DIP)

search protocols allow a polypeptide to traverse the potential energy (PE) surface in search of its global PE minimum. The PE surface is searched compared with the free energy surface for two reasons: (1) the potential energy surface is temperature-independent and therefore easier to determine, and (2) the potential energy frequently represents the largest contribution to the free energy. DIVE searches the PE surface by disrupting the kinetic energy and therefore the atomic velocities of the simulation. DIP simply allows multiple copies of a polypeptide, all at the same temperature to travel in multiple directions simultaneously. We have shown that when the two protocols are used together in succession, the experimental conformation can be reproduced along with a number of low-energy, non-native conformations (54, 55).

We consider peptide G an excellent case study for our continuing tests of the protocols because of the delicate energetic balance between an α helix and a β hairpin. Therefore, we use the two protocols in conjunction with each other to determine whether the β hairpin is truly at the lowest PE near physiological temperature. This chapter contains three major sections. We first describe the protocols used for simulation and for analysis. We follow this by listing and discussing the results of our simulations. We then present closing remarks.

4.2. Simulation protocol

4.2.1. Protocol description and simulation setup

Two protocols are implemented for our study. Disrupted velocity (DIVE) and divergent path (DIP) (54, 55) are discussed in previous literature. The DIVE protocol, as the name implies, perturbs the atomic velocities (including the directions and magnitudes) of the system allowing the conformation either to overcome or to circumvent

PE barriers. Essentially, the protocol allows a conformation to locate PE minima near 0 K. DIP, on the other hand, allows a conformation to traverse across the PE surface at a constant temperature in search of PE minima. Multiple conventional MD simulations are run simultaneously by assigning atomic velocities of identical atoms in different simulations to different directions, allowing more of the PE surface to be explored.

Residues 41–56 of the B1 domain of protein G (peptide G, GEWTYDDATKTFTVTE) form a β hairpin with a type I turn. Both termini are charged, and each acidic or basic amino acid has its side chains deprotonated or protonated, respectively. Three initial conformations are simulated using the two protocols: fully-extended, β hairpin, and α -helical conformations. The β hairpin is taken from the NMR model (PDB 2GB1) (20). The ϕ and ψ angles are selected as -180° and $+180^\circ$ for the fully-extended conformation and -60° and -40° for the α -helical conformation using the leap feature in AMBER 8 (56). Upon completion of the DIVE simulations, the conformation found at the lowest PE is then simulated on the basis of our analysis using both protocols. For the DIVE protocol, six conformations are simulated whereas ten conformations are simulated using the DIP protocol. Conformations are designated by the derivation from the initial conformation. For instance, the second round of DIVE simulations obtained from the fully-extended conformation is assigned the name “ext””. Before simulations begin, the PEs of each conformation are minimized for 100 steps using steepest descent. Minimization is done using the Multiscale Modeling Tools for Structural Biology program (MMTSB) (57) in a Generalized Born/surface area (GB/SA) implicit solvent model (58). After minimization, the Molecular Modeling Toolkit (MMTK) (59) is then used to convert the coordinate files into files suitable for our suite

of programs. The atomic velocities are scaled for DIP simulations to give an initial temperature of 300 K.

Other parameters are incorporated into the simulations. A modified version of Amber99 (10) is used during minimization and data collection. Constant temperature is maintained by the Nosé-Hoover Chain method (60), and the equations of motion are integrated by using the velocity-Verlet method (61). Distances between covalent bonds involving hydrogen are constrained by the SHAKE (62) algorithm. All simulations are run in a Generalized Born/surface area (GB/SA) implicit solvent model (58), which is defined by

$$G_{pol} = -\frac{1}{2} \left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w} \right) \sum_{ij} \frac{q_i q_j}{f_{gb}}$$

$$f_{gb} = \left[r_{ij}^2 + \alpha_i \alpha_j \exp \left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j} \right) \right]^{\frac{1}{2}}$$

where G_{pol} is the solvation free energy of the solute-solvent electrostatic polarization term, ϵ_p is the dielectric value within of the protein, ϵ_w is the solvent dielectric constant, r_{ij} is the separation distance between atoms i and j , q_i and q_j are the partial atomic charges, α_i and α_j are the corresponding effective Born radii, and f_{gb} is a complex function of r_{ij} , α_i and α_j (63). The effective Born radius determines the charge distance between the solute and the continuum dielectric boundary. All implicit solvent simulations have an external dielectric constant of 78.5 for water, an internal dielectric constant of 1.0 for proteins or peptides, surface tension at 0.005 kcal/mol-Å², and an offset of 0.9 Å. Electrostatic and Lennard-Jones cutoffs are set to 999 Å to represent an infinite cutoff.

The initial atomic velocities for each trajectory are assigned different atomic velocity magnitudes in the DIVE protocol. The different magnitudes correspond with different initial kinetic energies, thus different initial temperatures. The trajectories evolve for an assigned period before the atomic velocities are reassigned according to $\mathbf{p}' \equiv \sigma^{1/2} \mathbf{p}$ where \mathbf{p} and \mathbf{p}' are the momenta of the particles before and after atomic velocity reassignment, respectively. σ is a scaling parameter for the magnitude of the simulation's kinetic energy after atomic velocity reassignment. When the atomic velocities are reassigned, both the magnitude is rescaled and the direction is changed.

When $\sigma > 1$, the scaling parameter increases kinetic energy, but when $0 < \sigma < 1$, the parameter lowers kinetic energy. The scaling parameter σ may be chosen prior to starting the simulation or the scaling parameter σ may be calculated on-the-fly by $\sigma = |T - \Delta T|/T$, where ΔT is defined as the difference in temperature before and after atomic velocity rescaling. The parameter may alternatively be calculated from a target temperature T_{target} by $\sigma = T_{\text{target}}/T$.

A threshold temperature is also defined for the DIVE protocol. When simulation temperatures fall below the threshold temperature, kinetic energy is added to increase the temperature. If the simulation temperatures rise above the threshold temperature, especially after velocity rescaling, kinetic energy is removed thus lowering the simulation temperature. Typically during a DIVE simulation, heating occurs once whereas cooling occurs multiple times. By having these heating and cooling cycles, the trajectories are able to sample multiple potential energy minima near 0 K.

Each simulation is run for 4 million steps per trajectory with a 2 fs time step (4×10^6 steps/copy \times 2 fs/step \times 6 copies/simulation = 48 ns total simulation time).

Trajectory, energy, and checkpoint files are output every 250 steps (0.5 ps). Each trajectory in the DIVE protocol is assigned an initial temperature (10 K, 50 K, 100 K, 300 K, 600 K, and 1000 K). Atomic velocity reassignment for each trajectory occurs after 20,000 steps. The scaling parameter for heating is calculated from a target temperature, and the scaling parameter for cooling is set at 0.25 K/step. The threshold temperature is 10 K. Simulations are heated to a target temperature of 1000 K when they fell below the threshold temperature.

All simulations using the DIP protocol are run at a constant temperature of 300 ± 20 K. All initial conformations are assigned atomic velocities with the same magnitude but with six different directions. A different path over the PE surface is followed by each trajectory.

4.2.2. Data analysis

Backbone rmsd's, hydrogen bonds (H-bonds), distances for possible salt bridge formation, and aromatic ring geometries are analyzed using AMBER 8's analysis module ptraj (56), and the secondary structures are determined by the program STRIDE (64), which assigns the secondary structure on the basis of backbone dihedral angles and backbone-backbone H-bonding interactions. The backbone atom rmsd's are calculated in comparison to the NMR model. H-bonds are defined conventionally by a range of $180.0^\circ \pm 60.0^\circ$ for the $X-H \cdots X_1$ angle where X and X_1 are heavy atoms, with a 3.5 Å distance between heavy atoms and with an appearance of more than 5% in the simulations. α helices are defined by H-bonds between residues i and $i+4$ whereas 3_{10} and π helices have H-bonds between residues i and $i+3$ and between residues i and $i+5$, respectively. For DIVE simulations, each conformation of minimum potential energy was analyzed for the

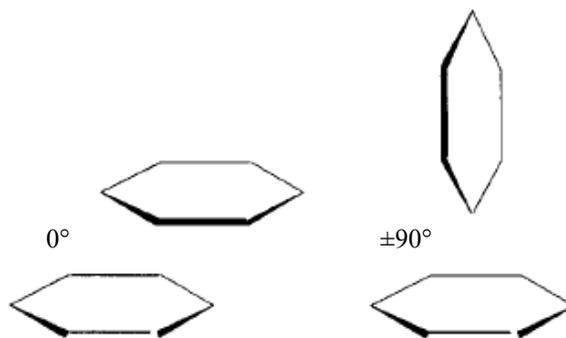


Figure 4.2 Geometric orientation of aromatic rings for possible π -stacking. The left figure shows aromatic rings parallel to each other and slightly out of phase (coplanar), whereas the right figure represents two aromatic rings perpendicular to one another for π -stacking (T-shaped). Figure from ref. (65).

presence of H-bonding. H-bonds are determined for each low PE conformation within the DIP simulations.

Distances are calculated for the following five possible side chain-side chain interactions: Asp46 \cdots Lys50, Asp47 \cdots Lys50, Glu56 \cdots Lys50, and Gly41 \cdots Glu56 terminal salt bridge. The pairs of oppositely charged side chains within 3.5 Å between the terminal heavy atoms of the side chains are considered to be in close proximity and are considered to indicate the possible formation of salt bridges. When a distance between aromatic rings (Tyr45 \cdots Phe52, Trp49 \cdots Phe52, and Tyr45 \cdots Trp49) is less than 6 Å with a dihedral angle between $\pm 90^\circ$ (Figure 2.2), π -stacking is deemed possible (65). The heavy atoms within the aromatic ring determine the ring distances, and the improper dihedral angle is based on $C\delta_i-C\epsilon_i\cdots C\epsilon_j-C\delta_j$ for each aromatic side chain.

$^3J_{\text{HN-H}\alpha}$ is determined from the H-N-C α -H α dihedral angle, and $^3J_\phi$ is determined from the ϕ (C $_{n-1}$ -N-C α -C) dihedral angle. The coupling constants are calculated by using the Karplus equations (66, 67) (eqs (4.1) and (4.2)).

$${}^3J_{\text{HN-H}\alpha} = 6.7\cos^2\theta - 1.3\cos\theta + 1.5 \quad (4.1)$$

$${}^3J_{\phi} = 1.8\cos^2\theta - 0.2\cos\theta + 0.5 \quad (4.2)$$

where the angle θ is the dihedral angle for H–N–C α –H α or for C–N–C α –C. The ranges of ${}^3J_{\text{HN-H}\alpha}$ and ${}^3J_{\phi}$ coupling constants are calculated from the respective dihedral angles of all NMR models. The maximum and minimum coupling constants reported in the tables (Table 4.1) are found from the 24 NMR models of peptide G, and the percentage of ${}^3J_{\text{HN-H}\alpha}$ and ${}^3J_{\phi}$ coupling constants are determined from the fifteen coupling constants using $\%{}^3J = \sum J_x / J_{\text{total}}$, where J_x is assigned either 1 or 0 depending upon whether the calculated coupling constant is within range (1 = ‘in range’, 0 = ‘not in range’) and J_{total} is the total number of available coupling constants. For both ${}^3J_{\text{HN-H}\alpha}$ and ${}^3J_{\phi}$ coupling constants, $J_{\text{total}} = 16$.

Cluster analysis groups conformations together in families, or clusters on the basis of rmsd comparisons between conformations. Cluster analysis for both DIVE and DIP

Table 4.1 Experimental range of coupling constants for both ${}^3J_{\text{HN-H}\alpha}$ and ${}^3J_{\phi}$

		${}^3J_{\text{HN-H}\alpha}$							
		GLU	TRP	THR	TYR	ASP	ASP	ALA	THR
Min		1.45	7.33	3.03	4.03	1.44	2.59	8.40	1.50
Max		6.88	7.92	6.90	4.29	2.07	9.50	9.47	3.39
		LYS	THR	PHE	THR	VAL	THR	GLU	
Min		7.82	1.68	8.99	3.62	5.65	2.00	3.01	
Max		9.50	9.45	9.37	6.26	6.17	5.58	9.49	
		${}^3J_{\phi}$							
		GLU	TRP	THR	TYR	ASP	ASP	ALA	THR
Min		1.08	1.22	1.46	1.12	0.5	0.96	0.56	1.5
Max		1.77	1.42	2.16	1.2	1.07	1.34	0.93	1.8
		LYS	THR	PHE	THR	VAL	THR	GLU	
Min		1.22	0.78	1.99	1.19	1.67	0.79	0.5	
Max		1.78	0.92	2.09	1.6	1.8	1.08	2.03	

simulations is done using the MMTSB software (57). Conformations are classified for residues ranging from Glu2 to Thr15. An iterative process with an error tolerance of 0.5, a least-squares fit based on rmsd comparisons between conformations and a centroid method are used in the cluster analysis (68). A centroid is defined as a central, average conformation that lies at the center of each cluster. Each cluster encompasses structures within a 3.0 Å variation from the centroid. The number of conformations in each cluster, information about each cluster, and a representative conformation of each cluster are determined during the analysis.

4.3. Simulation results and discussion

By using the disrupted velocity (DIVE) protocol in conjunction with the divergent path (DIP) protocol, a β hairpin similar to the experimental conformation was reproduced. Other low-PE, non-native conformations are also found by using the two protocols. Analysis of the conformations determine whether the non-native conformations are prevalent compared with the β -hairpin conformation near 0 K and near physiological temperatures at 300 K.

4.3.1. Divergent path (DIP) results and discussion

The native β hairpin (Figure 4.3b) was equilibrated near physiological temperatures (300 K) to determine whether it will retain its experimental conformation. The calculated conformation is a β hairpin with strands from Trp43 to Asp46 and from Thr51 to Val54, whereas the experimental conformation has β strands from Glu42 to Asp47 and from Lys50 to Glu56. A type I turn proceeded from Asp47 to Lys50, which is similar to the experimental conformation. The H-bond backbone-backbone interactions and the backbone atom rmsd confirm the stability of the hairpin. The side chains,

however, tend to be more flexible as evidenced by the heavy-atom rmsd. Distances appropriate for one non-native and one native salt bridge are achieved during the simulation: Glu47...Lys50 (26.3% occurrence) and Gly41...Glu56 (12.4% occurrence). The geometry of the aromatic side chains to form π -stacking arises between both Trp43...Phe52 and Tyr45...Phe52; the native peptide G, however has only Tyr45 and Phe52 within the appropriate distance for possible π -stacking. Unexpectedly though, the percentage of 3J coupling constants are less than 30% compared with the NMR models. The lower-than-expected percentage of coupling constants stem partially from slight differences in the dihedral angles; because the Karplus equations use transcendental functions, a slight shift in the dihedral angle can have a larger effect on the calculation.

The conformation labeled β^* (Figure 4.3j) represents the lowest PE β hairpin and is simulated near physiological temperatures to confirm its stability. The average conformation is slightly altered from its starting conformation. Compared with the experimental conformation, the β strands shorten by two residues, but the type I turn remains. The shift is further evidenced by the increase in the backbone atom rmsd, and because of the dynamic nature of the simulation, the side chains fluctuate in position as shown by the increase in heavy-atom rmsd. The side chains of residues Glu47 and Lys50 remain in close proximity for possible formation of a non-native salt bridge during 31.0% of the equilibrated simulation although the position of the terminal residues is less frequently within appropriate distance (7.5% throughout the simulation). π -Stacking interactions may occur because Trp43...Phe52 and Tyr45...Phe52 are within close proximity and oriented correctly for π -stacking. The percentage of coupling constants, which fall within the calculated range, is still relatively low, but as stated before, slight

variations within the dihedral angles can alter the percentage of coupling constants significantly.

The two β hairpin structures are both similar to the experimental conformation but also differ from each other. β^* has lower PE than the equilibrated experimental conformation by 5.8 kcal/mol. The two β hairpins are both stable near physiological temperatures at 300 K.

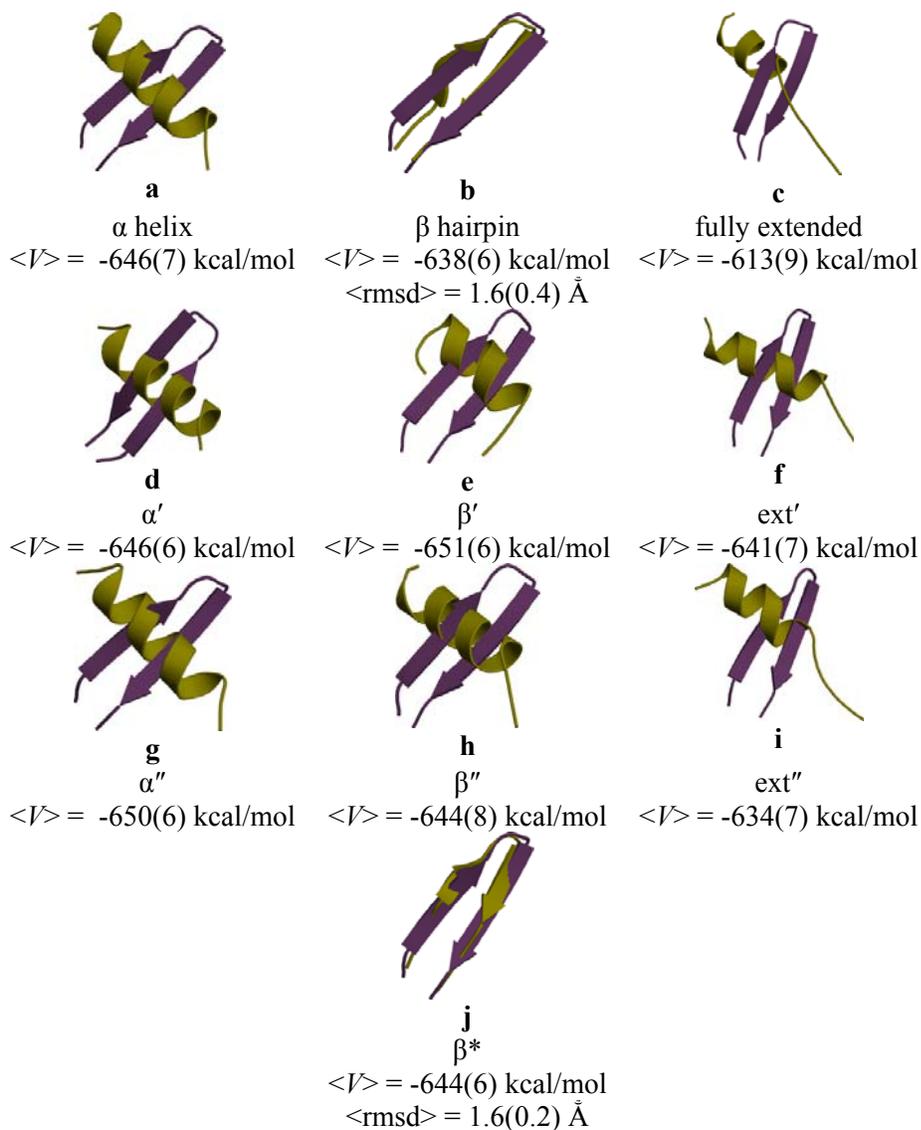


Figure 4.3 Conformation (yellow) with the lowest PE from each DIP simulation (yellow) compared to the NMR model (magenta). β^* represents a β hairpin found within the simulations. Standard deviations of the average PEs and rmsd's are in parentheses.

The remaining conformations determined by using the DIP protocol are non-native α -helical conformations. The α helices are prevalent compared with the 3_{10} helices, which are observed during less than 30% of the simulations. All of the non-native conformations have charged polar side chains within distance for possible salt bridge formation. The majority of helical conformations have Glu46 and Lys50 within range; one conformation, however, displays possible salt bridge interactions between Glu47 and Lys50. None of the aromatic side chains are aligned to imply π -stacking. 25–40% of calculated $^3J_{\text{HN-H}\alpha}$ coupling constants agree with experimental coupling constants, and the percentage of $^3J_{\phi}$ values within the range of experimental values is 20–25%.

The fully-extended conformation, ext, collapses into an average conformation of an α helix extending from Glu42 to Thr49 (Figure 4.3c), whereas an initial α helix equilibrates into an α helix extending from Glu42 to Val54 (Figure 4.3a). Ext'' consists of an α helix from Glu42 to Thr51 (Figure 4.3i). α' (Figure 4.3d) and α'' (Figure 4.3g) are similar in conformation with an α helix extending from Glu42 to Val54. An α helix (Thr44–Thr53) comprises the average conformation of β' (Figure 4.3e), whereas β'' (Figure 4.3h) is 4.6 kcal/mol higher PE than β' and has an average conformation composed of an α helix from Glu42 to Thr53.

The use of cluster analysis allows the conformations found within our simulations to be classified according to their conformations, independent of their PE. The number of structures within each cluster also offers a qualitative view of a cluster's entropic contribution to free energy of the cluster's conformation. Cluster analysis is performed on the trajectories from the DIP simulations. One cluster with 3428 conformations resembles the native β hairpin. The β hairpin extends from Thr44 to Thr53 with a type I

turn between Asp47 and Lys50, whereas the experimentally-determined β hairpin includes Glu42–Glu56 with a type I turn from Asp47 to Lys50.

Four other clusters contain more than 1000 non-native conformations. α helices of varying residue lengths comprise these clusters. The largest family of conformations consisted of 4956 conformations with a representative α helix extending from Glu42 to Val54. The next family of conformations includes 4211 conformations with an α helical conformation similar to the most highly populated cluster (Glu42–Thr53). Another cluster, which contains 1483 conformations, is similar in secondary structure to the other two clusters of helical structures: an α helix extending from Thr44 to Thr53. The fourth cluster of helical conformations surrounds 1777 conformations and contains an α helix extending from Glu42 to Thr51. Two clusters exhibit more non-native conformations than the β hairpin-like cluster. Clusters with several non-native conformations implies that the native β hairpin may not be entropically favored, which qualitatively agrees with work done by the Eaton group (*1*).

Simulations just described indicate that peptide G is stable as a β hairpin near physiological temperatures by performing simulations (near 300 K). The stabilization can be inferred to result from three main factors. With the distance between N_{Gly41^+} and O_{Glu16^-} (the terminal salt bridge) less than 3.5 Å, the β hairpin is possibly prevented from completely unfolding and refolding into another conformation. A side chain distance appropriate for salt bridge formation between Asp47 and Lys50 is observed in our simulations, but a salt bridge is instead present between Asp46 and Lys50, according to NMR data (*13, 14, 20*). π -Stacking within the hydrophobic core is also believed to contribute to β hairpin stability. The aromatic side chains in Tyr45 and Phe52 are within

the appropriate geometrical configuration for π -stacking; yet the flexibility of the side chains allow the aromatic rings to fluctuate between parallel and T-shaped configurations. In our simulations, the hydrophobic interaction between Trp43 and Val54 does not play the significant role proposed by Kobayashi (24).

Hydrogen bonds, although weaker than salt bridges, are a major factor in the stabilization of the β hairpin. With the backbone aligned in an antiparallel motif, the β -hairpin structure is similar to the original NMR model. β and β^* exhibit more than ten main-chain hydrogen bonds, which agree with those of Gronenborn, et al. (20) in that a hydrogen bond between Thr49 and Thr51 is present. However, $O_{\text{asp46}^+} \cdots \text{NH}_{\text{Ala48}}$ is not present within our simulations of the β -hairpin structure. $O_{\text{Thr44}} \cdots \text{NH}_{\text{Thr53}}$, on the other hand, has more than 98% occurrence during both β and β^* DIP simulations. The simulations agree with Ma, et al. (29) in that we observe the stabilization effects by Thr44, Asp46, Thr53, and Thr55 with other residues.

We further note that the average PEs for the conformations described, excluding the equilibrated fully-extended simulation, are within 20 kcal/mol of each other (Figure 4.4). As stated within our work on C-peptide of RNase A (chapter 1) and tryptophan zipper 2 (chapter 3), this shows that the PE surface is rough with some PE barriers preventing various conformational transitions. The small range in PEs also indicates that the modified Amber99 force field (10) lacks the ability to distinguish between α helices and β hairpins, but brings the two secondary structural elements closer in energy than the original Amber99 force field (10). We note, however, that an energy surface devoid of a discriminatory feature is still imperfect. On the basis of the work by Krivov and Karplus (26), Garcia and Sanbonmatsu (19), Pande et al. (37), Zhou and Berne (4), Levy et al.

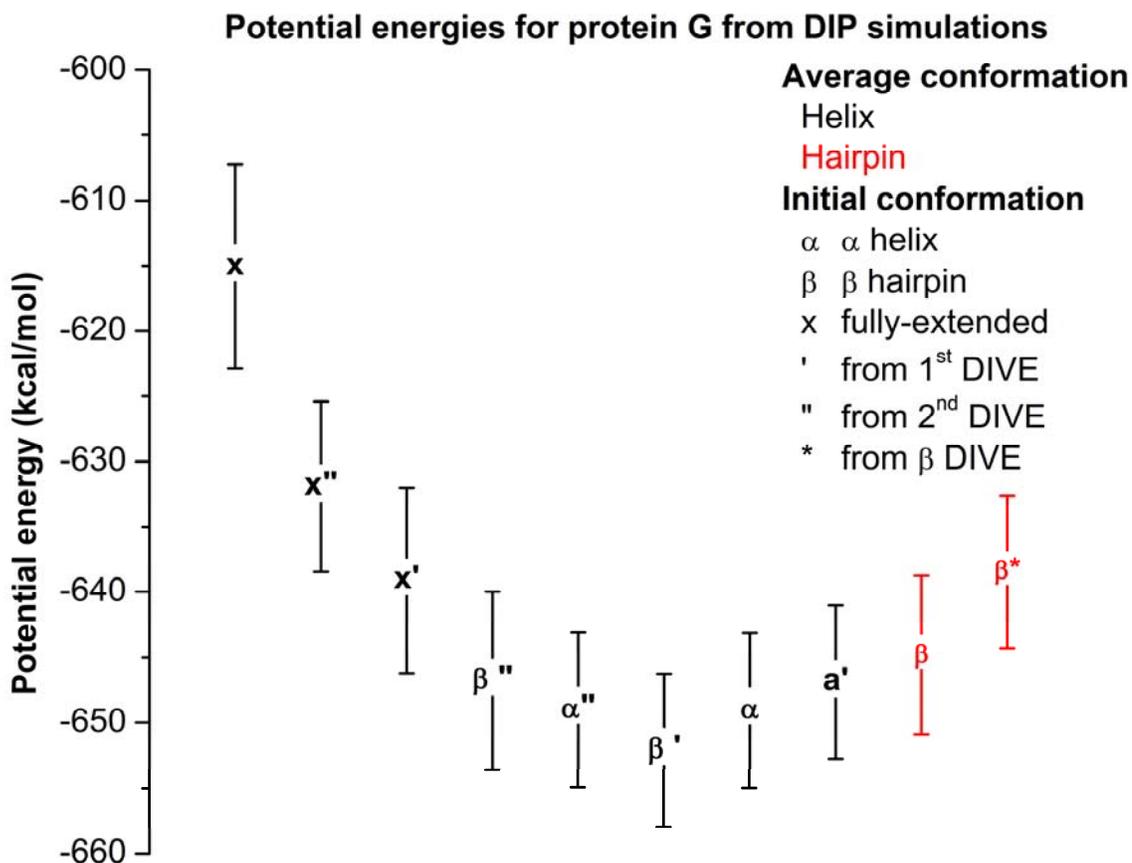


Figure 4.4 Average PE with standard deviation bars for the equilibrated conformations from the divergent path (DIP) simulations.

(18) and our own cluster analysis, we anticipate that the β hairpin, although the experimental conformation, may not necessarily correspond to structure of the lowest PE.

If, on the other hand, we consider work done by the Wales group (17, 48, 69-73), another plausible explanation for our simulation results arises. Above, we assume that the PE surface is funnel-shaped in which case the β hairpin is higher in average PE than the helical conformations. We also note that the standard deviation between energy minima overlaps, which should allow for rapid transitions between conformations. The rapid transition does not occur as expected.

We therefore suggest that the minima found during our simulations at 300 K are probably within a ‘weeping willow’ PE landscape (Figure 4.5) (69, 71). The ‘weeping willow’ model suggests that a distinct global energy minimum exists along with several local minima but the energy minima are separated by high-energy barriers. This hypothesis seems to explain why the conformations are not rapidly transitioning into other conformations or converging into a single conformation. The differences between PE minima may be small, but the amount of kinetic energy necessary to overcome the PE barrier prevents the transition. Therefore, the β' conformation resides in the global PE

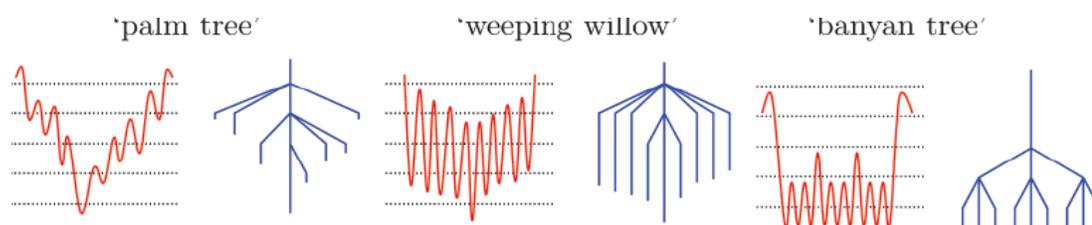


Figure 4.5 Three pairs of PE landscapes (left) and their corresponding disconnectivity graphs (right). The graphs are drawn as PE (vertical axis) relative to arbitrary coordinates (horizontal axis). The endpoints of the disconnectivity graphs represent PE minima, and the points where the branches are joined correspond to a common PE “superbasin”. For the PE landscape, the wells represent the minima of a system surrounded by PE barriers. Reprinted by permission from ref. (71). Copyright 2006 American Chemical Society

minimum 7.29 kcal/mol lower than a native-like conformation.

Another plausible explanation for the “weeping willow” PE surface model may stem from the distance-dependent dielectric constant used by the generalized Born implicit solvent model (74). The distance-dependent dielectric constant seems to alter the PE surface by flattening an expected “funnel-shaped” surface. Mortenson and Wales found that a distant-dependent dielectric constant gave the preferred “funnel-shaped” surface described by the “palm tree” disconnectivity graph. They, however, use Cornell et al.’s Amber94 (75) force field to study on an α helix-forming conformation. Amber94 has been shown to bias simulations toward α helices (2, 4, 8-11).

4.3.2. Disrupted velocity (DIVE) results and discussion

When the disrupted velocity (DIVE) protocol is used in conjunction with the divergent path (DIP) protocol, the experimental conformation is reproduced along with other conformations. A β -hairpin-like conformation shall be described first followed by alternative conformations that are found using the DIVE protocol.

The β hairpin (β^* , Figure 4.6d) consists of two β strands from Glu42 to Asp46 and from Thr51 to Thr55 with a type I turn from Asp47 to Lys50. In contrast, the experimental conformation has longer β strands from Glu42 to Asp47 and from Lys50 to Glu56, with a type I turn. The backbone atom rmsd of β^* is 0.9 Å relative to the experimental conformation, but the side chains are not fully aligned as noted by the heavy-atom rmsd of 1.8 Å. The only distances indicative of salt bridges are between the terminal amino acids Gly41 and Glu56, similar to those of the native β hairpin. The two aromatic residues Tyr45 and Phe52 are in the geometric configuration to fit the π -stacking definition (65). Less than 30% of the NMR coupling constants agree with those

calculated from the experimental conformation. Once again, this is not surprising because a slight change in the dihedral angles can greatly change the coupling constants.

The following conformations are lower in PE than β^* , but non-native secondary structures (α helices) comprise these conformations. Only one conformation has charged side chains that are less than 3.5 Å apart allowing for possible side chain interactions, but the inferred salt bridges are non-native. The aromatic side chains are not in a geometric configuration indicating that π -stacking may be unavailable for any of the following conformations. The percentage of $^3J_{\text{HN-H}\alpha}$ coupling constants within the range of constants calculated from the NMR model ranges from 30–50%, and the percentage of $^3J_{\phi}$ coupling constants range from 20% to 33%.

Ext' (Figure 4.6c) and ext'' (Figure 4.6g) are composed of α helices between Trp43 and Thr53. The α helices are evidenced by backbone conformational angles and the backbone-backbone H-bonding interactions between amino acids indexed i and $i+4$.

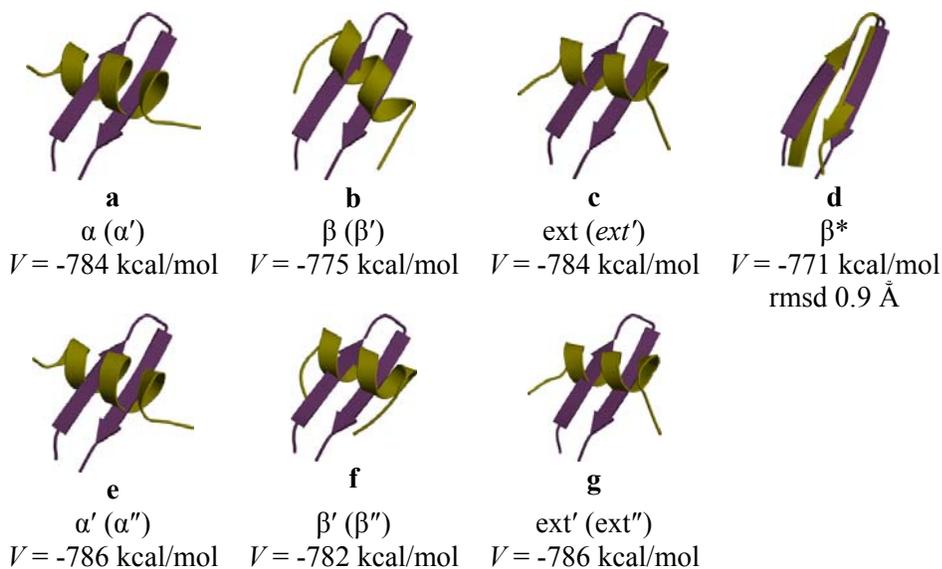


Figure 4.6 Conformation (yellow) with the lowest PE from each DIVE simulation (yellow) compared to the NMR model (magenta). β^* represents a β hairpin found within the simulations. Both the initial simulation name and conformation name (in parenthesis) are listed with the PE and the backbone atom rmsd's below each name.

β' (Figure 4.6b) and β'' (Figure 4.6f) consist of α helices extending from Tyr45 to Thr53 and between Thr44 and Thr53, respectively. α' (Figure 4.6a) and α'' (Figure 4.6e) have conformations consisting of α helices from Glu42 to Asp47 and π helices from Ala48 to Phe52, respectively.

For the DIVE simulations, all conformations representing PE minima found at temperatures below 10 K 489 conformations are analyzed, and eight clusters contained ten or more conformations. One cluster has a β hairpin as its representative conformation (10 conformations) whereas the majority of the clusters have α helices in some form as the representative conformation. The representative β hairpin extends from Glu42 to Thr55 with a type I β turn proceeding from Asp47 to Lys50, whereas the experimentally-determined conformation exhibits a β hairpin from Glu42 to Glu56 with a type I β turn between Asp47 and Lys50. Not all conformations within the cluster are derived from the experimental conformation, which implies that DIVE can locate β hairpins from different starting conformations. Because the β -hairpin structure is found in simulations initialized with both an α -helix and fully-extended structures, we surmise that DIVE can sample a broad range of PEs and find the desired secondary structure. The other clusters consist of varying lengths of an α helix which gives further credence to the idea that the free energy of the experimental conformation near 0 K has a less favorable entropic contribution than non-native conformations.

The PE for β^* is higher than the PE for helical conformations. One might expect the opposite to be true because other simulations have shown that the β hairpin lies within the global *free energy* minimum (15, 19, 25, 27, 29, 31, 32, 34, 39, 40). With the continual perturbation of kinetic energy and atomic velocity direction in the DIVE

method, one should expect more conformations to be located in a simple molecular dynamics simulation. The DIP protocol then searches the PE surface near physiological temperatures (around 300 K) seeking PE minima. Using multiple initial conformations further increases the number of the PE minima found. In fact, Huang et al. (54, 55) have shown that the DIVE and DIP protocols allow a wide range of conformations to be observed. Thus, it is plausible that the peptide G may represent a local PE minimum separated from other basins by high-PE barriers. This would suggest that the experimental conformation is not necessarily the global PE minimum for the force field used in this study.

We cannot discount the fact that the simulations are completed in an implicit water solvent. Garcia (19) used an explicit solvent model, which allows for interactions with explicit water molecules, whereas both Zagrovic (37) and Zhou (4) performed their simulations in implicit water solvent but with different force fields. Zagrovic (37) stated that semi-helical intermediates are possible as traps beyond the typical folding pathway whereas Zhou (4) focused more on the differences in force fields. We tend to agree with both of them in their observations because a large population of helical conformations is found by using both the DIVE and DIP protocols. The implicit solvent may overweight the existence of non-native states and in fact, exhibit a different free energy surface compared with explicit solvent (38).

The DIVE protocol can sample several conformations for peptide G including the β -hairpin-like conformations within 1.0 Å of experiment. The DIVE protocol located β hairpins from different starting conformations, which indicates that peptide G can fold into a β hairpin from other conformations. Near 0 K, the β hairpin, however, is 15.3

kcal/mol higher in PE than the α -helical ext" conformation, which corresponds to the global PE minimum. On the basis of the conformations found near 0 K, we assert that the global PE minimum consists of an α helix extending from Trp3 to Thr13. Around 300 K, the difference in PEs between α helices and β hairpins decreases, but a non-native conformation is still lower in PE than the native-like conformation.

4.4. Conclusions

We have tested both the DIVE and DIP protocols for their abilities to reproduce the observed β -hairpin structure of the B1 domain spanning residues 41 to 56 of Streptococcal protein G near 0 K (DIVE simulations) and near 300 K (DIP simulations). Our simulations show that the experimental conformation can be reproduced within 0.9 Å backbone atom rmsd of the NMR model by using DIVE simulations of an initial β -hairpin structure and within 1.4 Å backbone atom rmsd of the NMR model by using DIP simulations, starting from a β -hairpin structure. Cluster analysis of the DIP simulations shows that the two most populous clusters each have an rmsd less than 2.0 Å from the NMR structure and together contain 14% of the structures analyzed. Similarly, cluster analysis of the DIVE simulations shows that the two largest clusters display α -helical conformations and contain 2% of the structures analyzed. Although the structure representing the global potential energy (PE) minimum found by using each of the two methods is an α helix rather than a β hairpin, the experimental β -hairpin conformation is only 7–14 kcal/mol higher in energy than the global minimum, and the β -hairpin conformation folds from non-native conformations using the DIVE protocol. We further propose, on the basis of the cluster analysis that the experimental conformation lies within a wide-basin local PE minimum surrounded by high PE barriers, which prevent

simulated structures from refolding into the conformation residing in the global PE minimum.

We have shown that the disrupted velocity (DIVE) protocol can sample a large region of the potential energy (PE) surface in search of a global PE minimum. We also find that DIVE can locate native β -hairpin conformations starting from an α helix or an extended conformation. When DIVE is used along with DIP, we demonstrate that a β hairpin can be located on the PE surface at temperatures near 300 K. Because the β hairpin does not undergo conformational change, the modified Amber99 force field (10) indeed offers a reasonable potential energy function that is less biased towards α helices than previous Amber-type force fields (2, 5, 8, 11, 42, 52, 76).

Standard deviations of the PE's for minima of both the hairpin and the helical conformations overlap, but we propose that the conformations are separated by high PE barriers, which prevent the refolding of conformations. The hairpin, as noted, does not correspond to the lowest PE in our simulations. Instead, a helical conformation appears to have a lower PE. The conformation found within the global PE minimum may not necessarily correspond to the experimental conformation, but DIVE allows a system to sample more of the PE surface near 0 K than conventional MD simulations while overcoming PE barriers by reassigning atomic velocity direction and increasing kinetic energy. In turn, the DIP simulations allow the polypeptide an opportunity to search for the PE minima at a constant, higher temperature. We therefore propose that the β hairpin located using the modified Amber99 force field (10) corresponds to a local PE minimum surrounded by high PE barriers.

4.5. References

1. Muñoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of beta-hairpin formation, *Nature* 390, 196-199.
2. Yoda, T., Sugita, Y., and Okamoto, Y. (2004) Comparisons of force fields for proteins by generalized-ensemble simulations, *Chem Phys Lett* 386, 460-467.
3. Mackerell, A. D., Feig, M., and Brooks, C. L. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *J Comput Chem* 25, 1400-1415.
4. Zhou, R. H. (2003) Free energy landscape of protein folding in water: Explicit vs. implicit solvent, *Proteins* 53, 148-161.
5. Mu, Y. G., Kosov, D. S., and Stock, G. (2003) Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments, *J Phys Chem B* 107, 5064-5073.
6. Hu, H., Elstner, M., and Hermans, J. (2003) Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution, *Proteins* 50, 451-463.
7. Wang, J. M., Cieplak, P., and Kollman, P. A. (2000) 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.
8. Reddy, M. R., Erion, M. D., Agarwal, A., Viswanadhan, V. N., McDonald, D. Q., and Still, W. C. (1998) Solvation free energies calculated using the GB/SA model: Sensitivity of results on charge sets, protocols, and force fields, *J Comput Chem* 19, 769-780.
9. Dixit, S. B., Bhasin, R., Rajasekaran, E., and Jayaram, B. (1997) Solvation thermodynamics of amino acids - Assessment of the electrostatic contribution and force-field dependence, *J Chem Soc Faraday T* 93, 1105-1113.
10. Okur, A., Strockbine, B., Hornak, V., and Simmerling, C. (2003) Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins, *J Comput Chem* 24, 21-31.
11. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins-Structure Function and Bioinformatics* 65, 712-725.
12. Andrec, M., Felts, A. K., Gallicchio, E., and Levy, R. M. (2005) Protein folding pathways from replica exchange simulations and a kinetic network model, *P Natl Acad Sci USA* 102, 6801-6806.
13. Blanco, F. J., Rivas, G., and Serrano, L. (1994) A Short Linear Peptide That Folds into a Native Stable Beta-Hairpin in Aqueous-Solution, *Nat Struct Biol* 1, 584-590.

14. Blanco, F. J., and Serrano, L. (1995) Folding of Protein-G B1 Domain Studied by the Conformational Characterization of Fragments Comprising Its Secondary Structure Elements, *Eur J Biochem* 230, 634-649.
15. Bolhuis, P. G. (2003) Transition-path sampling of beta-hairpin folding, *P Natl Acad Sci USA* 100, 12129-12134.
16. Dinner, A. R., Lazaridis, T., and Karplus, M. (1999) Understanding beta-hairpin formation, *P Natl Acad Sci USA* 96, 9068-9073.
17. Evans, D. A., and Wales, D. J. (2004) Folding of the GB1 hairpin peptide from discrete path sampling, *J Chem Phys* 121, 1080-1090.
18. Gallicchio, E., Andrec, M., Felts, A. K., and Levy, R. M. (2005) Temperature weighted histogram analysis method, replica exchange, and transition paths, *J Phys Chem B* 109, 6722-6731.
19. Garcia, A. E., and Sanbonmatsu, K. Y. (2001) Exploring the energy landscape of a beta hairpin in explicit solvent, *Proteins* 42, 345-354.
20. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G, *Science* 253, 657-661.
21. Guo, C. L., Levine, H., and Kessler, D. A. (2000) How does a beta-hairpin fold/unfold? Competition between topology and heterogeneity in a solvable model, *P Natl Acad Sci USA* 97, 10775-10779.
22. Honda, S., Kobayashi, N., and MuneKata, E. (2000) Thermodynamics of a beta-hairpin structure: Evidence for cooperative formation of folding nucleus, *J Mol Biol* 295, 269-278.
23. Klimov, D. K., and Thirumalai, D. (2000) Mechanisms and kinetics of beta-hairpin formation, *P Natl Acad Sci USA* 97, 2544-2549.
24. Kobayashi, N., Honda, S., Yoshii, H., and MuneKata, E. (2000) Role of side-chains in the cooperative beta-hairpin folding of the short C-terminal fragment derived from streptococcal protein G, *Biochemistry-US* 39, 6564-6571.
25. Kolinski, A., Ilkowski, B., and Skolnick, J. (1999) Dynamics and thermodynamics of beta-hairpin assembly: Insights from various simulation techniques, *Biophys J* 77, 2942-2952.
26. Krivov, S. V., and Karplus, M. (2004) Hidden complexity of free energy surfaces for peptide (protein) folding, *P Natl Acad Sci USA* 101, 14766-14770.
27. Lee, J., and Shin, S. (2002) Two-dimensional correlation analysis of peptide unfolding: Molecular dynamics simulations of beta hairpins, *J Phys Chem B* 106, 8796-8802.
28. Lee, J., and Shin, S. M. (2001) Understanding beta-hairpin formation by molecular dynamics simulations of unfolding, *Biophys J* 81, 2507-2516.
29. Ma, B. Y., and Nussinov, R. (2000) Molecular dynamics simulations of a beta-hairpin fragment of protein G: Balance between side-chain and backbone forces, *J Mol Biol* 296, 1091-1104.
30. Muñoz, V., Henry, E. R., Hofrichter, J., and Eaton, W. A. (1998) A statistical mechanical model for beta-hairpin kinetics, *P Natl Acad Sci USA* 95, 5872-5879.

31. Pande, V. S., and Rokhsar, D. S. (1999) Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G, *P Natl Acad Sci USA* 96, 9062-9067.
32. Roccatano, D., Amadei, A., Di Nola, A., and Berendsen, H. J. C. (1999) A molecular dynamics study of the 41-56 beta-hairpin from B1 domain of protein G, *Protein Sci* 8, 2130-2143.
33. Swope, W. C., Pitera, J. W., Suits, F., Pitman, M., Eleftheriou, M., Fitch, B. G., Germain, R. S., Rayshubski, A., Ward, T. J. C., Zhestkov, Y., and Zhou, R. (2004) Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and beta-hairpin peptide, *J Phys Chem B* 108, 6582-6594.
34. Tsai, J., and Levitt, M. (2002) Evidence of turn and salt bridge contributions to beta-hairpin stability: MD simulations of C-terminal fragment from the B1 domain of protein G, *Biophys Chem* 101, 187-201.
35. Wei, G. H., Derreumaux, P., and Mousseau, N. (2003) Sampling the complex energy landscape of a simple beta-hairpin, *J Chem Phys* 119, 6403-6406.
36. Wei, G. H., Mousseau, N., and Derreumaux, P. (2004) Complex folding pathways in a simple beta-hairpin, *Proteins-Structure Function and Bioinformatics* 56, 464-474.
37. Zagrovic, B., Sorin, E. J., and Pande, V. (2001) beta-hairpin folding simulations in atomistic detail using an implicit solvent model, *J Mol Biol* 313, 151-169.
38. Zhou, R. H., and Berne, B. J. (2002) Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water?, *P Natl Acad Sci USA* 99, 12777-12782.
39. Zhou, R. H., Berne, B. J., and Germain, R. (2001) The free energy landscape for beta hairpin folding in explicit water, *P Natl Acad Sci USA* 98, 14931-14936.
40. Zhou, Y. Q., and Linhananta, A. (2002) Role of hydrophilic and hydrophobic contacts in folding of the second beta-hairpin fragment of protein G: Molecular dynamics simulation studies of an all-atom model, *Proteins* 47, 154-162.
41. Felts, A. K., Harano, Y., Gallicchio, E., and Levy, R. M. (2004) Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model, *Proteins-Structure Function and Bioinformatics* 56, 310-321.
42. Yoda, T., Sugita, Y., and Okamoto, Y. (2004) Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations, *Chem Phys* 307, 269-283.
43. Chen, J. H., Im, W. P., and Brooks, C. L. (2006) Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field, *J Am Chem Soc* 128, 3728-3736.
44. Fesinmeyer, R. M., Hudson, F. M., and Andersen, N. H. (2004) Enhanced hairpin stability through loop design: The case of the protein G B1 domain hairpin, *J Am Chem Soc* 126, 7238-7243.
45. Huyghues-Despointes, B. M. P., Qu, X. T., Tsai, J., and Scholtz, J. M. (2006) Terminal ion pairs stabilize the second beta-hairpin of the B1 domain of protein G, *Proteins-Structure Function and Bioinformatics* 63, 1005-1017.

46. Khavrutskii, I. V., Byrd, R. H., and Brooks, C. L. (2006) A line integral reaction path approximation for large systems via nonlinear constrained optimization: Application to alanine dipeptide and the beta hairpin of protein G, *J Chem Phys* 124.
47. Jang, S., Shin, S., and Pak, Y. (2002) Molecular dynamics study of peptides in implicit water: Ab initio folding of beta-hairpin, beta-sheet, and beta beta alpha-motif, *J Am Chem Soc* 124, 4976-4977.
48. Carr, J. M., Trygubenko, S. A., and Wales, D. J. (2005) Finding pathways between distant local minima, *J Chem Phys* 122.
49. Feig, M., MacKerell, A. D., and Brooks, C. L. (2003) Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations, *J Phys Chem B* 107, 2831-2836.
50. Simmerling, C., Strockbine, B., and Roitberg, A. E. (2002) All-atom structure prediction and folding simulations of a stable protein, *J Am Chem Soc* 124, 11258-11259.
51. Sorin, E. J., and Pande, V. S. (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations, *Biophys J* 88, 2472-2493.
52. Wang, T., and Wade, R. C. (2006) Force field effects on a beta-sheet protein domain structure in thermal unfolding simulations, *J Chem Theory Comput* 2, 140-148.
53. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G. M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. M., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J Comput Chem* 24, 1999-2012.
54. Huang, Z. N., and Wheeler, R. A. Replica Microcanonical Search Methods for Mapping Potential Energy Landscapes and Conformations of miniproteins, *In preparation*.
55. Huang, Z. (2005) Design of new molecular dynamics global minimum search protocols for mapping energy landscapes and conformations of folded polypeptides and mini-proteins, in *Chemistry and Biochemistry*, p 282, The University of Oklahoma, Norman.
56. Case, D. A., Darden, T. A., Cheatham, I., T.E. , Simmerling, C. L., Wang, J., Duke, R. R. E., Merz, L., K.M. , Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) AMBER 8, 8 ed., University of California, San Francisco, San Francisco.
57. Feig, M., Karanicolas, J., and Brooks, C. L. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology, *J Mol Graph Model* 22, 377-395.
58. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular Simulations, *Biopolymers* 56, 275-291.
59. Hinsen, K. (2000) The molecular modeling toolkit: A new approach to molecular simulations, *J Comput Chem* 21, 79-85.
60. Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992) Nosé-Hoover Chains: the Canonical Ensemble via Continuous Dynamics, *J Chem Phys* 97, 2635-2643.

61. Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A Computer-Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *J Chem Phys* 76, 637-649.
62. Andersen, H. C. (1983) Rattle: a Velocity Version of the Shake Algorithm for Molecular Dynamics Calculations, *J Comput Phys* 52, 24-34.
63. Born, M., and von Karman, T. (1912) On fluctuations in spatial grids., *Phys Z* 13, 297-309.
64. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins* 23, 566-579.
65. McGaughey, G. B., Gagne, M., and Rappe, A. K. (1998) pi-stacking interactions - Alive and well in proteins, *J Biol Chem* 273, 15458-15463.
66. Kao, L. F., and Barfield, M. (1985) Conformational Dependencies of Vicinal C-13(O)-N-C-Alpha-C-13 and C-13(O)-N-C-Alpha-H-1 Coupling Constants in Compounds Which Model the Peptide Backbone, *J Am Chem Soc* 107, 2323-2330.
67. Ludvigsen, S., Andersen, K. V., and Poulsen, F. M. (1991) Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of phi-angles, *J Mol Biol* 217, 731-736.
68. Lloyd, S. P. (1982) Least-Squares Quantization in Pcm, *Ieee T Inform Theory* 28, 129-137.
69. Doye, J. P. K., and Wales, D. J. (1996) On potential energy surfaces and relaxation to the global minimum, *J Chem Phys* 105, 8428-8445.
70. Wales, D. J., and Scheraga, H. A. (1999) Review: Chemistry - Global optimization of clusters, crystals, and biomolecules, *Science* 285, 1368-1372.
71. Wales, D. J., and Bogdan, T. V. (2006) Potential energy and free energy landscapes, *J Phys Chem B* 110, 20765-20776.
72. Komatsuzaki, T., Hoshino, K., Matsunaga, Y., Rylance, G. J., Johnston, R. L., and Wales, D. J. (2005) How many dimensions are required to approximate the potential energy landscape of a model protein?, *J Chem Phys* 122, -.
73. Wales, D. J. (2005) The energy landscape as a unifying theme in molecular science, *Philos T Roy Soc A* 363, 357-375.
74. Mortenson, P. N., and Wales, D. J. (2001) Energy landscapes, global optimization and dynamics of the polyaniline Ac(ala)(8)NHMe, *J Chem Phys* 114, 6443-6454.
75. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J Am Chem Soc* 117, 5179-5197.
76. Mackerell, A. D. (2004) Empirical force fields for biological macromolecules: Overview and issues, *J Comput Chem* 25, 1584-1604.

5. Exploration of the potential energy surface of $\beta\beta\alpha 5$

5.1. Introduction

Recently, our group developed two molecular dynamics protocols that aid in geometry optimization (1, 2). The disrupted velocity (DIVE) protocol allows a polypeptide to locate multiple conformations on the potential energy (PE) surface by perturbing both the atomic velocity directions and magnitudes (kinetic energy). The perturbations permit the system to circumvent or to overcome PE barriers. With the divergent path (DIP) protocol, multiple, independent copies of a system search for multiple PE minima by generating multiple, independent copies simultaneously. Copies of the system travel across the PE surface in different directions at a constant kinetic energy and temperature.

Prior to this study, the DIVE and DIP protocols were validated on α helices (1, 2) (chapter 1) and β hairpins (chapters 3 and 4). The α helices folds to within 1.0 Å backbone atom rmsd of experiment and are close to the PE minimum at 300 K. Results varied for the β hairpins, on the other hand. The tryptophan zipper (trpzip) folds correctly with the hairpin conformation corresponding to the lowest PE minimum. For the B1 domain(41–56) of Streptococcal protein G, the global PE minimum contains helical conformations whereas the native β hairpin is higher in PE. According to Klimov and Karplus (3), helical conformations could be the global free energy minimum based upon entropic contributions.

In this chapter, we present our work to model the structure of a 23-residue $\beta\beta\alpha$ motif ($\beta\beta\alpha 5$, Figure 5.1) (4). The $\beta\beta\alpha$ motif has, as its name implies, two β strands in a hairpin formation (residues 1–8) with a type II' turn at the N-terminus followed by a loop region (9–12) and ends with an α helix (13–20) at the C-terminus. The $\beta\beta\alpha 5$ protein contains only one non-natural amino acid, a D-proline, at residue 4, which helps to maintain the type II' β turn. This is contrasted with the original $\beta\beta\alpha$ motif ($\beta\beta\alpha 1$), which has both the D-proline at residue 4 and a 3-(1,10-phenanthrol-2-yl)-L-alanine at residue 6 to stabilize the hydrophobic core between the hairpin and the helix (5). In $\beta\beta\alpha 5$, Tyr6, Phe8, Leu14, and Leu17 make a hydrophobic core for the hairpin and helix packing, and the polar side chains of Arg2⁺ and Asp7⁻ are within close enough distance for possible salt bridge formation. Other simulations of the $\beta\beta\alpha$ motif have identified the experimental conformation as the global PE minimum conformation (6-12), but Dill et. al (13) find the experimental conformation to be higher in free energy than a non-native conformation. The Pak group (14) finds that, around 430 K and 450 K, the experimental conformation is 0.3 kcal/mol higher than a helical conformation. Using a different modified Amber force field from the one used here, Jang states that $\beta\beta\alpha 5$ has a single

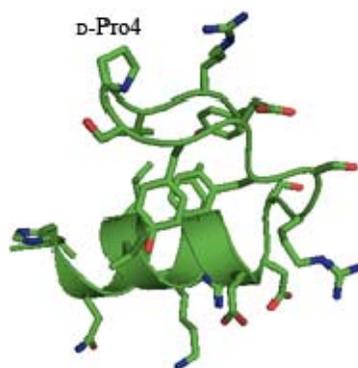


Figure 5.1 NMR model structure for $\beta\beta\alpha 5$ with the non-natural amino acid labeled

free energy well (6) in agreement with other simulations (7, 8, 11, 12, 14). $\beta\beta\alpha 5$ presents a challenge for our protocols because it is the first test of the ability of DIVE and DIP to reproduce the tertiary structure of peptides. The next section documents the protocols and the analysis techniques. Simulation results and our thoughts on the results are then offered. Finally, closing remarks are presented.

5.2. Simulation details

We implement two protocols for our study. The disrupted velocity (DIVE) and divergent path (DIP) protocols are discussed in previous literature (1, 2). The DIVE protocol, as the name implies, perturbs the atomic velocity of the system allowing the conformation either to overcome or to circumvent PE barriers. Essentially, the protocol allows a conformation to locate PE minima near 0 K using a microcanonical simulation (constant number of atoms, volume, and total energy). DIP, on the other hand, allows a conformation to traverse the PE surface at a constant temperature by allowing an initial conformation to follow multiple, independent paths. Multiple conventional MD simulations are run simultaneously by assigning atomic velocities of identical atoms in different simulations to different directions, allowing more of the PE surface to be explored.

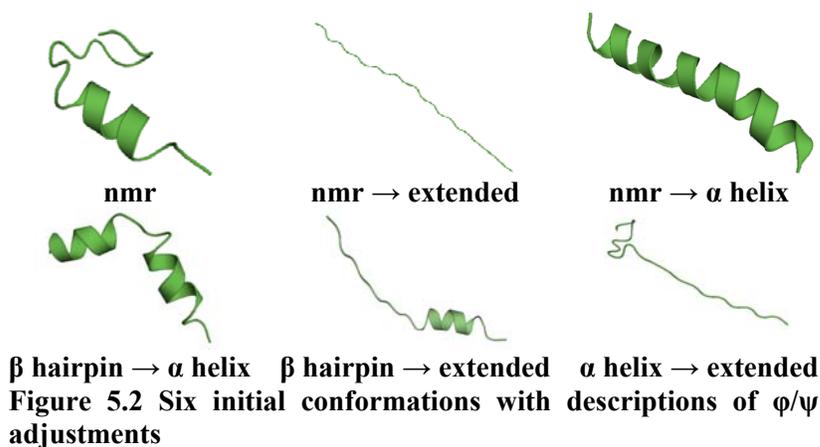
For the DIVE protocol, the initial atomic velocities for each copy are assigned different atomic velocity magnitudes. The different magnitudes correspond with different initial kinetic energies (i.e., different initial temperatures). The conformations evolve for an assigned time period before the atomic velocities are reassigned according to $\mathbf{p}' \equiv \sigma^{1/2} \mathbf{p}$ where \mathbf{p} and \mathbf{p}' are the momenta of the particles before and after atomic

velocity reassignment, respectively. When the atomic velocities are reassigned, both the magnitude is rescaled and the direction is changed.

σ is a scaling parameter for the magnitude of the simulation's kinetic energy after atomic velocity reassignment. When $\sigma > 1$, kinetic energy is increased, but when $0 < \sigma < 1$, kinetic energy is removed. The scaling parameter σ may be chosen before starting the simulation or it calculated on-the-fly by $\sigma = |T - \Delta T|/T$, where ΔT is defined as the temperature difference before and after atomic velocity rescaling. The parameter may alternatively be calculated from a target temperature T_{target} by $\sigma = T_{\text{target}}/T$.

We also define a threshold temperature for the DIVE protocol. When simulation temperatures fall below the threshold temperature, kinetic energy is added to increase the temperature. If the simulation temperatures rise above the threshold temperature, especially after velocity rescaling, kinetic energy is removed thus decreasing the simulation temperature. Typically during a DIVE simulation, heating occurs once whereas cooling occurs multiple times during each heating/cooling cycle. By using multiple heating and cooling cycles, the simulations are able to sample multiple potential energy minima near 0 K.

We selected $\beta\beta\alpha 5$ (pdb code: 1T8J, sequence: Ace-YRV_D-PSYDFSR_SDELAKL-



LRQHAG-NH₂) to test the ability of DIVE and DIP to reproduce peptide tertiary structure. Residues Tyr1–Phe8 represent a β -hairpin motif with a type II' β -turn (D-Pro4 and Ser5). Glu13–Gln20 conform to an α helix whereas Ser9–Phe12 represent a flexible loop region. Each acidic or basic amino acid has its side chains deprotonated or protonated, respectively. To avoid bias, six initial conformations are selected for simulation (**Figure 5.2**). The five non-native conformations are made by adjusting the ϕ and ψ angles of the experimental conformation. α helices have ϕ and ψ angles of -60° and -40° , respectively, and the fully extended conformation is assigned $\phi = -180^\circ$ and $\psi = 180^\circ$.

For all simulations, the PEs of the conformation are minimized for 100 steps with steepest descent in implicit solvent (15) by using the Multiscale Modeling Tools for Structural Biology (MMTSB) (16) program. The Molecular Modeling Toolkit (MMTK) (17) is used to convert the coordinates into files used by our programs. For DIP simulations, atomic velocities are scaled to give an initial temperature of 300 K.

Other parameters are incorporated into the MD simulations. A modified version of the Amber99 (18) force field is used during minimization and simulations used for data collection. Distances to covalent bonds involving hydrogen are constrained by using the SHAKE (19) algorithm. The equations of motion are integrated with the velocity-Verlet method (20). Constant temperature is maintained using the Nosé-Hoover Chain method (21). The simulations are run with the Generalized Born/surface area (GB/SA) implicit solvent model (15), which is defined by

$$G_{pol} = -\frac{1}{2} \left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w} \right) \sum_{ij} \frac{q_i q_j}{f_{gb}}$$

$$f_{gb} = \left[r_{ij}^2 + \alpha_i \alpha_j \exp \left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j} \right) \right]^{\frac{1}{2}}$$

where G_{pol} is the solvation free energy of the solute-solvent electrostatic polarization term, and ϵ_p is the dielectric value within of the protein. ϵ_w is the solvent dielectric constant and r_{ij} is the separation distance of particles i and j . q_i and q_j are the atomic charges, α_i and α_j are the corresponding effective Born radii (22), and f_{gb} is a complex function of r_{ij} , α_i and α_j . The charge distance between the solute and the continuum dielectric boundary is determined by the effective Born radius. All implicit solvent simulations have an external dielectric constant of 78.5, an internal dielectric constant of 1.0, surface tension at 0.005 kcal/mol-Å², and an offset of 0.9 Å. Periodic boundaries are not used; instead, electrostatic and Lennard-Jones cutoffs are set to 999 Å to represent an infinite cutoff. A modified version of Amber99 (18) is used during minimization and data collection.

Each simulation is run for 4 million steps per simulation with a 2 fs time step and with six copies per simulation (4×10^6 steps/copy \times 2 fs/step \times 6 copies/simulation = 48 ns total simulation time). Data is output every 250 steps (0.5 ps). Each copy involved in the DIVE protocol is assigned an initial temperature (10 K, 50 K, 100 K, 300 K, 600 K, and 1000 K). Atomic velocity reassignment for each simulation occurs after 20,000 steps. The scaling parameter for heating is calculated from a target temperature and the scaling parameter for cooling is set at 0.25 K/step. The threshold temperature is 10 K. Simulations are heated to a target temperature of 1000 K when they fell below the

threshold temperature. Although the target temperature is assigned at 1000 K, 500 K is the maximum temperature achieved because of rapid energy redistribution.

For the DIP protocol, all simulations are run at a constant temperature of 300 ± 20 K. All initial conformations are assigned atomic velocities with the same magnitude but identical atoms in each of the six different copies were assigned with six different atomic velocity directions. Each copy follows a different path over the PE surface.

Figure 5.3 depicts how conformations are selected. One of the six initial conformations (e.g., nmr) is simulated with the DIVE protocol. Conformations are designated by the derivation from the initial conformation. For instance, the second round of DIVE simulations obtained from the fully-extended conformation is assigned the name “ext”. The conformation corresponding to the lowest PE (e.g., nmr’) is simulated in a second round by using the DIVE protocol. The corresponding conformation (e.g., nmr’’) is then equilibrated utilizing the DIP protocol. The other two conformations (nmr and nmr’) are also simulated with the DIP protocol.

Backbone rmsd’s, hydrogen bonds (H-bonds), distances between ionic side

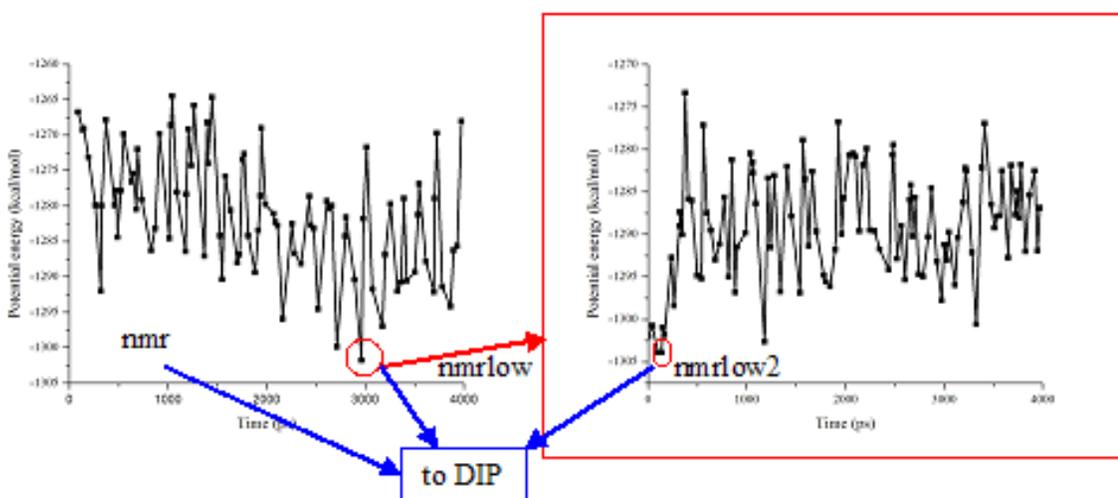


Figure 5.3 How conformations are selected during DIVE simulations for further simulation either with the DIVE or the DIP protocol. Red represents DIVE simulations whereas the blue indicates a DIP simulation.

chains, distances and angles between aromatic side chains, and hydrophobic distances are calculated using AMBER 8's analysis module ptraj (23). The backbone rmsd's are calculated relative to the experimental conformation. Backbone rmsd's are determined for the entire tertiary structure, the β hairpin, the α helix, and the tertiary structure excluding the loop region. H-bonds are conventionally defined by a range of $180.0^\circ \pm 60.0^\circ$ for the $X-H \cdots X_1$ angle where X and X_1 are polar heavy atoms, with a 3.5 Å distance between polar heavy atoms and with an appearance in more than 5% in the simulations. α helices are defined by H-bonds between residues i and $i+4$ whereas 3_{10} and π helices have H-bonds between residues i and $i+3$ and between residues i and $i+5$, respectively.

Sixteen distances are calculated between pairs of oppositely charged side chains: R2D7, R2D12, R2E13, R2H21, D7R10, R10D12, R10E13, R10H21, D7K16, D12K16, E13K16, K16H21, D7R19, D12R19, E13R19, and R19H21. When the terminal heavy atoms of oppositely charged side chains are less than 3.5 Å apart, the formation of salt bridges is likely possible. In the native tertiary structure, Tyr6, Phe8, Leu14, and Leu17 (4) are in close proximity. We calculate the distances between the following side chain combinations: Tyr6 and Phe8, Phe8 and Leu14, and Leu14 \cdots Leu17. If the terminal heavy atoms of the side chain pairs are less than 6 Å apart, the two residues are considered within range. When the geometric configuration between all the heavy atoms of the aromatic side chains of Tyr6 and Phe8 are less than 6 Å with a dihedral angle ($C\delta_i-C\epsilon_i \cdots C\epsilon_j-C\delta_j$) between $\pm 90^\circ$, π -stacking is deemed possible (24).

An alternative approach to describing conformations is via cluster analysis. Cluster analysis is independent of the PE surface and allows conformations to be

classified on the basis of rmsd's between Cartesian coordinates of the conformations analyzed. Cluster analysis also permits a qualitative analysis of the entropic contribution to free energy; if a cluster contains a high number of conformations, the entropy may contribute more to the folding of the particular conformation. A single representative conformation of each cluster is listed because other details will vary with the individual conformations within the cluster.

Cluster analysis allows conformations to be classified on the basis of structural similarities independent of PE. Conformations are grouped into families, or clusters, with similar secondary and tertiary structures by rmsd comparisons of the conformations. Cluster analysis for both DIVE and DIP simulations is determined using the MMTSB software (16). Conformations are classified by comparison of the β hairpin and loop regions (Tyr1–Asp12) because the α helical region (Glu13–Gln20) is typically within 1.0 Å of experiment. The analysis consists of an iterative process with an error tolerance of 0.5, a least-squares fit, and a centroid method in which a centroid is defined as the average conformation representing the cluster (25). Clusters are defined as having a radius of 3.0 Å from the defined average conformation. The number of conformations in each cluster, information about each cluster, and a representative conformation of each cluster are determined during the analysis.

5.3. Simulation results and discussion

5.3.1. Divergent path (DIP) simulations

As in the previous chapters, DIVE is used in conjunction with DIP to determine possible conformations of the primary sequence. The primary sequence folds into a conformation similar to the experimental conformation, but alternative non-native

conformations are present as well. This section presents the equilibrated data near 300 K first, followed by the conformations that are found at low temperatures.

The simulated experimental conformation (Figure 5.4a) has an α helix from Arg10 to Ala22 and a β hairpin from Arg2 to Asp7 and a β hairpin from Tyr1 to Asp7 with a type II' turn. The individual secondary structures agree within 2.0 Å of experiment (α helix 0.7 ± 0.2 Å, β hairpin 1.7 ± 0.8 Å). The equilibrated conformation has six possible noncovalent interactions between the polar side chains (R2D7, R2D12, R2E13, D7R10, R10E13, and D12K16). In the NMR model structures, Arg2 and Asp7 are the only two polar side chains in close proximity to each other, which are experimentally expressed (4), and they appear within the proper distance range during 32% of the simulation. A 3_{10} helix occurs during the simulation but is found in only 10%

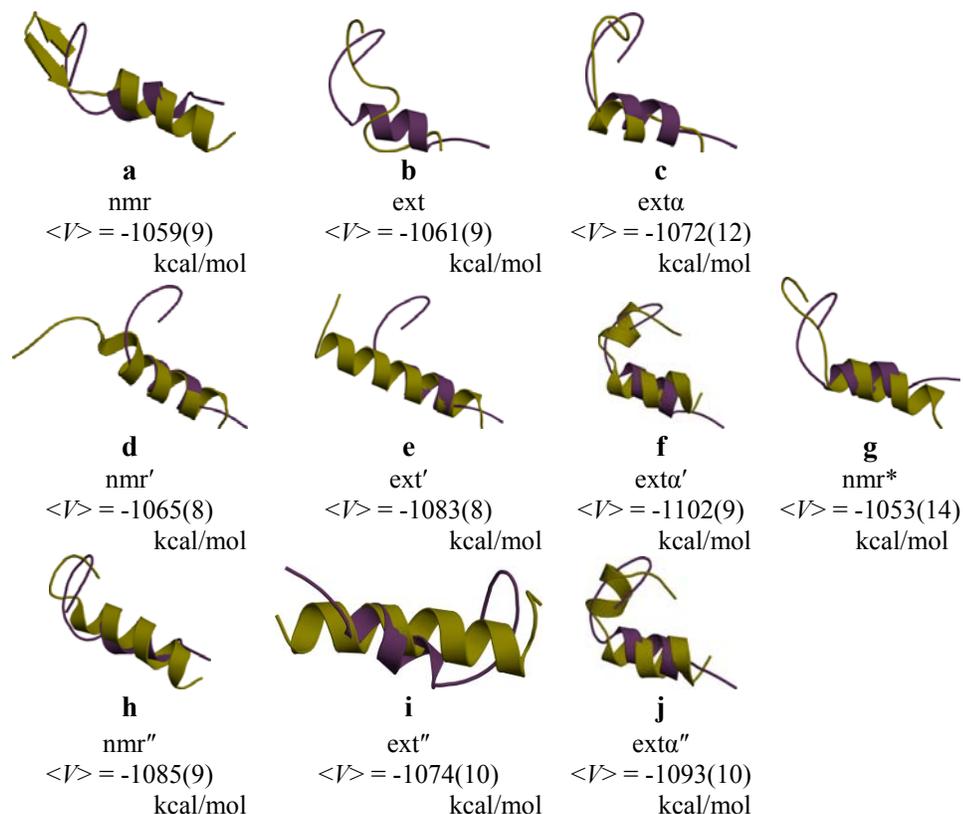


Figure 5.4 Average conformations from the various DIP simulations (yellow) aligned with the NMR model (magenta). PE's for each conformation are given, with standard deviations in parentheses.

of structures analyzed, compared to a 95% occurrence of α -helical structures. Thus, secondary structural elements of $\beta\beta\alpha 5$, and the β hairpin, are retained during DIP simulations started from the results of DIVE simulations initiated from the native structure.

Although the DIVE protocol retains secondary structural elements of $\beta\beta\alpha 5$, some of the protein's tertiary structure is lost. In fact, amino acids located within different secondary structural elements that are close to each other in the native structure are more remote after the simulations. Leu14 and Leu17, both located in the α helix, are in the proper geometric configuration (6 Å between heavy atoms of the side chains) for possible formation of the hydrophobic core whereas the side chains of Tyr6 and Phe8 of the β hairpin are nowhere near Leu14 and Leu17. Experimentally, the side chains of Tyr6 and Phe8 are supposed to be in close proximity to the side chains of Leu14 and Leu17 allowing the α helix and the β hairpin to stay together. In our simulations, the loop region is flexible and moves the β hairpin away from the α helix. Tyr1 and Tyr6 are within an acceptable geometric configuration (distance between the heavy atoms of the aromatic rings < 6.0 Å, angle between the aromatic rings < $\pm 120^\circ$) for possible noncovalent interactions. However, the possible polar side chain interactions (R2D12 and R2E13) reveal that the α helix and the β hairpin do, in fact, come within proximity occasionally during the simulation, but the majority of the simulation is spent with the α helix and the β hairpin apart.

When nmr* (Figure 5.4g) is simulated at 300 ± 20 K, the individual secondary structures remain within 2.0 \AA of the experimental secondary structural elements. Two β strands are found in the equilibrated nmr* structure (Arg2–Val3 and Tyr6–Asp7) with a type II' turn between Val3 and Tyr6, and the α helix extends from Ser11 to Ala22. Within the β hairpin, the polar side chains of Arg2 and Glu7 are within close proximity during the simulation, but unlike the equilibrated experimental conformation, the polar side chains appear at a distance of less than 3.5 \AA during 83% of the simulation. Four additional polar side chains (Arg10, Asp12, Glu13, and Lys16) within the α helix are at distances less than 3.5 \AA implying possible salt bridge formation. Leu14 and Leu17 in

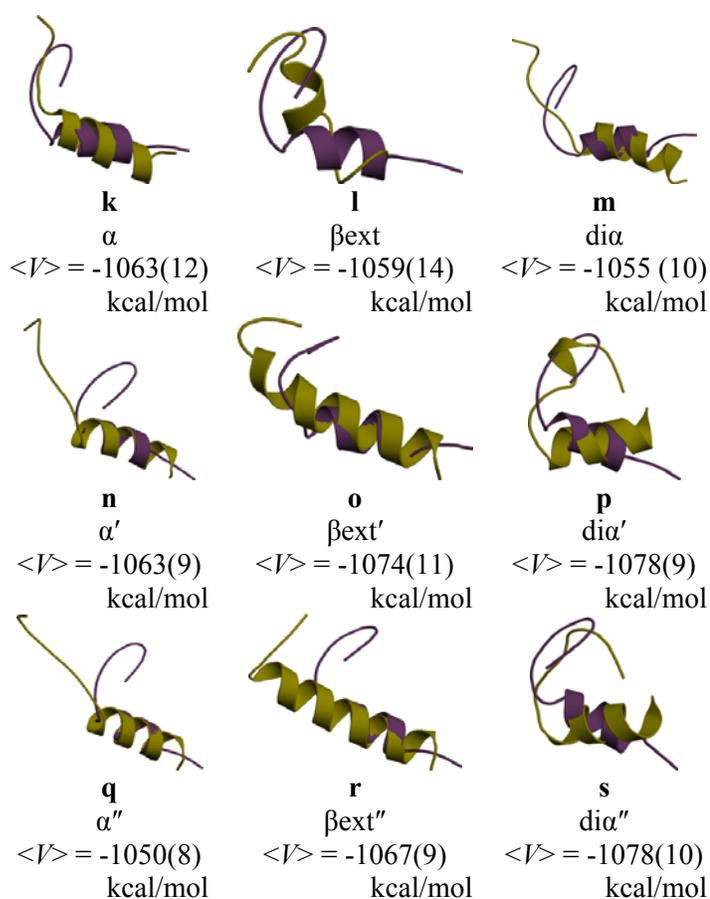


Figure 5.5 More average conformations from the various DIP simulations (yellow) aligned with the NMR model (magenta). PEs are listed with their standard deviations in parenthesis

addition to Tyr6 and Phe8 are aligned in a proper geometric configuration for proper packing of the α helix and β hairpin, but Phe8 and Leu14 are further apart indicating that the loop region is quite flexible, allowing the β hairpin and the α helix to drift apart. Tyr6 and Phe8 fluctuate in the proper geometric alignment for possible interaction, however the proper orientation of the α helix and β hairpin is not maintained.

Compared with the above simulations, we observe striking conformational and structural differences in the remaining simulations. First, the α helical region from Asp12 to Arg19 has a backbone atom rmsd of less than 1.5 Å compared with the experimental conformation for all but two simulations and fluctuates between an α helix and a 3_{10} helix. The β hairpin region from Tyr1 to Asp7, however, has a backbone atom rmsd of more than 2.0 Å compared with the experimental conformation for all but three simulations. The side chains of Leu14 and Leu17, which are apparently involved in a hydrophobic core of the experimental conformation, are aligned more closely than the nonpolar side chains of Tyr6 and Phe8, but Phe8 and Leu14 are never within close contact. The two charged polar side chains implicated in stabilizing the β hairpin are in close proximity during seven simulations; other polar side chain pairs fluctuate within 3.5 Å or less during the various simulations. In six simulations, three aromatic side chain pairs are found to be less than 6 Å apart and to have angles less than 120°; the Tyr6 and Phe8 aromatic ring pair is the only one that is deemed to form a π -stacking interaction in the experimental conformation.

Ext (Figure 5.4b) lacks any classifiable secondary structures. Yet, some folding and unfolding of helical turns occurs on the basis of H-bond backbone-backbone interactions, but the majority of backbone-backbone interactions are between residues i

and $i+2$. Calculations of the individual secondary structure rmsd's confirm the lack of typical secondary or tertiary structures. Contacts between three polar side chain pairs (R2E13, D12K16, and K16H21) fluctuate throughout the equilibrated simulation. The only π -stacking interaction is between Tyr1 and Tyr6, whose aromatic rings occasionally align in the proper geometric configuration.

Ext'' (Figure 5.4i) has an α helix from Tyr6 to Ala22 and contains flexible termini whereas β ext'' (Figure 5.5r), consists of flexible termini and an α helix from Ser5 to Ala22. The Tyr1 and Tyr6 aromatic ring pair for the ext'' simulation fluctuate only very briefly into between a proper geometric configuration for π -stacking. Therefore, β ext'' apparently does not have any aromatic ring pairs in alignment displaying π -stacking.

The α helix within the equilibrated α conformation (Figure 5.5k) spans the amino acid sequence from Ser11 to Leu17 whereas ext α '' (Figure 5.4j) is composed of an α helix from Ser5 to Leu18. None of the aromatic side chains are within an acceptable geometric configuration for π -stacking during the ext α '' simulation, but the Tyr1 and Phe8 pair is aligned during the ext α '' simulation.

The β hairpin in β ext (Figure 5.5l) refolds into a 3_{10} helix (D-Pro4–Tyr6), and Ser9–Asp12 of the fully-extended region folds into an α helix, whereas the β ext' conformation (Figure 5.5o) consists of a 3_{10} helix from D-Pro4 to Tyr6 and an α helix extending from Ser9 to Ala22. Surprisingly, the β hairpin region of the β ext' conformation agrees within 1.7 ± 0.7 Å with the native β hairpin region, simply indicating that the regions can overlay.

The initial dia helix (Figure 5.5m) and the initial α helix both fold into an α helix extending from Ser11 to Ala22 but differ in average PE by 8.1 kcal/mol. For the

equilibrated α conformation, the Tyr1 and Phe8 aromatic side chains align in the proper geometry for possible interaction. The equilibrated $di\alpha$ conformation, on the other hand, lacks any pairings of aromatic residues.

$Di\alpha'$ (Figure 5.5p) and $di\alpha''$ (Figure 5.5s) are similar in conformation and differ in average PE by 0.2 kcal/mol. Both are composed of two α helices between Ser5 and Phe8 and between Glu13 and Ala22, with flexible termini. The Tyr1 and Phe8 aromatic side chain pair is the only aromatic residues to align properly for π -stacking. The π -stacked configuration is more prevalent in $di\alpha''$ compared with $di\alpha'$.

An α helix from Ser9 to Ala22 is evident in both nmr' (Figure 5.4d), and nmr'' (Figure 5.4h), but the two conformations have a 19 kcal/mol PE difference. Ext' (Figure 5.4e) and $ext\alpha'$ (Figure 5.4f) are composed of α helices from Pro4 to Ala22 and from Ser5 to Arg19, respectively. α' (Figure 5.5n) and α'' (Figure 5.5q) both have α helices extending from Ser11 to Ala22 while having a 19 kcal/mol difference. $Ext\alpha'$ is similar to $ext\alpha''$ in conformation, but $ext\alpha''$ is 9 kcal/mol lower in PE than $ext\alpha'$. $Ext\alpha'$, α'' , nmr' , and nmr'' maintain a conformation similar to their initial DIVE conformations with the typical transitions between helical motifs. Ext' , however, deviates from its initial conformation (Figure 5.6) since its α helix straightens from its slightly bent initial

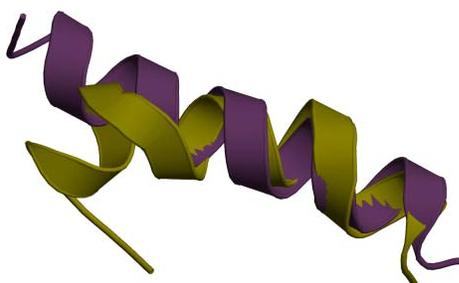


Figure 5.6 Overlay of initial ext' conformation (yellow) from the DIVE simulation with the average, equilibrated conformation from DIP (magenta)

conformation. The α helix of α' extended by two residues (from Gln20 to Ala22) compared with its initial conformation.

Cluster analysis is run for the nineteen simulations to group the many structures into conformational families. Several of the conformations are found in seven clusters (approximately 53% of the conformations). One family of conformations (1714 conformations) consists of a β hairpin (Arg2–Asp7) with a type II' turn and an α helix (Ser11–Ala22) and compares well with the experimental secondary structure (β hairpin: Tyr1–Phe8; α helix: Glu13–Gln20). The β hairpin is within 1.1 Å (backbone atom rmsd) of the native hairpin. In contrast, the most highly-occupied cluster (4004 conformations) has a representative conformation with α helices from Ser5 to Phe8 and from Glu13 to Ala22. The second most populous cluster (3960 conformations) exhibits an α helix extending from Ser5 to Arg19. The third largest cluster (3277 conformations) is represented by an α helix from D-Pro4 to Asp12, and the fourth cluster (3089 conformations) has an α helix from Ser5 to Ala22. The fifth most populous cluster (2514 conformations) is also characterized by an α helix (Tyr6–Ala22). Finally, the seventh most highly-populated cluster (1682 conformations) is represented by a conformation similar to that of the sixth cluster (described at the beginning of this paragraph); however, the backbone atom rmsd is slightly higher for the β hairpin (1.2 Å).

As with our previous work, we find that the two conformations which best resemble the experimental conformation are maintained at 300 K. Because the individual secondary structures are retained, during the simulations, the modified Amber force field (18) seems to reproduce both secondary structure motifs quite well. Unfortunately, the experimentally determined tertiary structure, packing of the α -helix relative to the β

hairpin, was not well reproduced. Thus, the loop region is rather dynamic, and although the two secondary structures may come within close proximity, the inferred hydrophobic core seems not to play a prevalent role in forming the tertiary structure during our calculations.

We note, however, that the two conformations containing both α -helical and β -hairpin secondary structures are higher in PE than other conformations.. Other simulations (6-12) suggest that the experimentally observed secondary and tertiary structures of the $\beta\beta\alpha$ motif represents the global free energy minimum. However, Dill (13) and Jang (14) suggest that the experimental conformation is actually higher in free energy than non-native conformations like helix-bundles. Our results agree more with the idea that the experimental conformation, although attainable, does not reside in the global PE minimum.

We started several simulations from conformations that are not in a native-like conformation, which offers us a statistical advantage over groups that simply start from the experimental conformation or from a fully-extended conformation. Figure 5.7 displays several possible PE minima and illustrates the roughness of the PE surface. Some of the minima overlap because of fluctuations in the PE throughout a simulation; yet, the standard error of the mean for each PE quantifies the accuracy of the average PE within ± 0.1 kcal/mol (data not shown), which indicates that the PE for each equilibrated conformation is fairly accurate. The two native-like conformations, nmr and nmr*, are higher in PE than ext α' (42.6 kcal/mol and 48.9 kcal/mol, respectively).

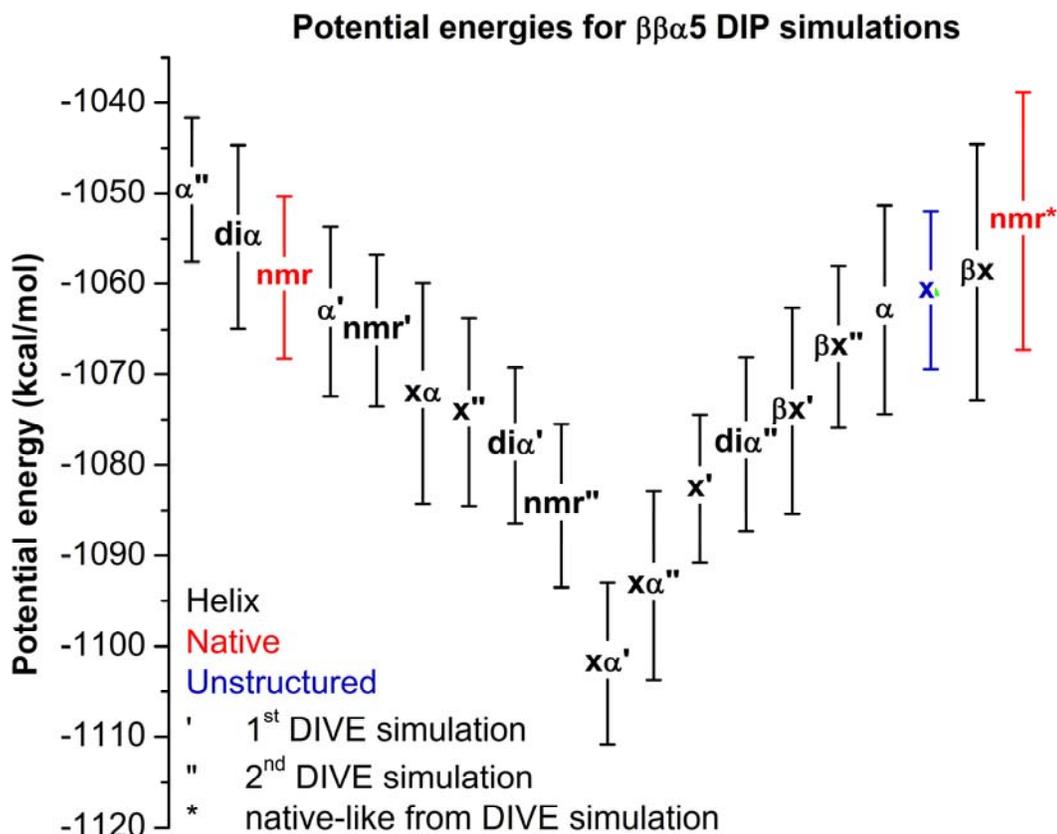


Figure 5.7 Average potential energies and standard deviation bars for $\beta\beta\alpha 5$ simulations

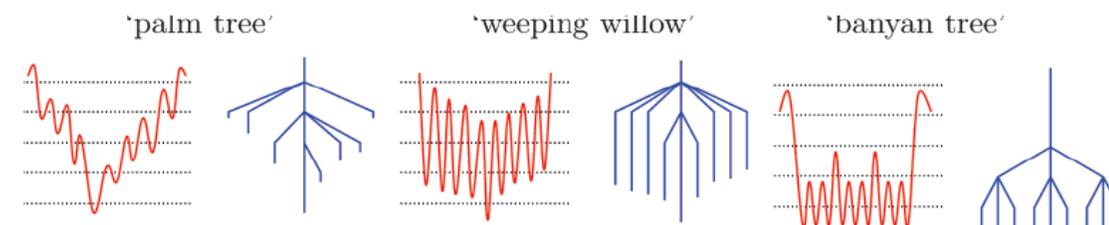


Figure 5.8 Disconnectivity graph (right) representing the PE landscape (left). The graphs are drawn as PE (vertical axis) relative to arbitrary units (horizontal axis). The endpoints of the disconnectivity graphs represent PE minima, and the points where the branches are joined correspond to a common PE “superbasin”. For the PE landscape, the wells represent the minima of a system surrounded by PE barriers. Reprinted with permission from ref. (26). Copyright 2006 American Chemical Society

Because the conformations do not rapidly transition to the conformation corresponding to the lowest PE, we surmise that the conformations are surrounded by high PE barriers. We suggest that the PE surface is, by the models set forth by the Wales group (26), more like a ‘weeping willow’ than a ‘palm tree’ (Figure 5.8). The ‘palm tree’

PE surface assumes that several local PE minima exist with low PE barriers whereas the ‘weeping willow’ PE surface is characterized by local PE minima separated by high PE barriers. The conformations in the local PE minima of the ‘weeping willow’ model are not as likely to fall into the global PE minimum without the aid of increased kinetic energy.

5.3.2. Disrupted velocity (DIVE) simulations

Of the twelve simulations, none of conformations corresponding to the lowest PEs reproduce the experimental conformation. A conformation (Figure 5.9g) with a higher PE is found in the simulation derived from the NMR model and has a backbone atom rmsd less than 2.0 Å with a secondary structure rmsd of 1.2 Å for the β hairpin region (Tyr1–Phe8) and with a secondary structure rmsd of 0.6 Å for the α helical region (Glu13–Gln20). The backbone atom rmsd of the tertiary structure excluding the loop region is 1.2 Å. Four pairs of polar side chains are within close proximity: R2D7, R10D12, R10E13, and D12K16. The Arg2 and Glu7 polar side chain pair is the only salt bridge inferred from the NMR model. Of the four amino acid pairs implicated in the hydrophobic core, the hydrophobic side chains of Leu14 and Leu17 are the only two side chains within close proximity of the four implicated in the hydrophobic core; these residues are both in the α helical region. None of the aromatic side chains align in the proper geometric configuration for possible π -stacking (distance between the heavy atoms of the aromatic rings < 6.0 Å, angle between the aromatic rings $< \pm 120^\circ$).

The following conformations do not resemble a native-like conformation. The β hairpin region (Tyr1–Phe8) for most conformations has a backbone atom rmsd relative to the NMR model of more than 2.0 Å whereas the α helical region (Glu13–Gln20) for all conformations has a backbone atom rmsd of less than 1.0 Å relative to the experimental

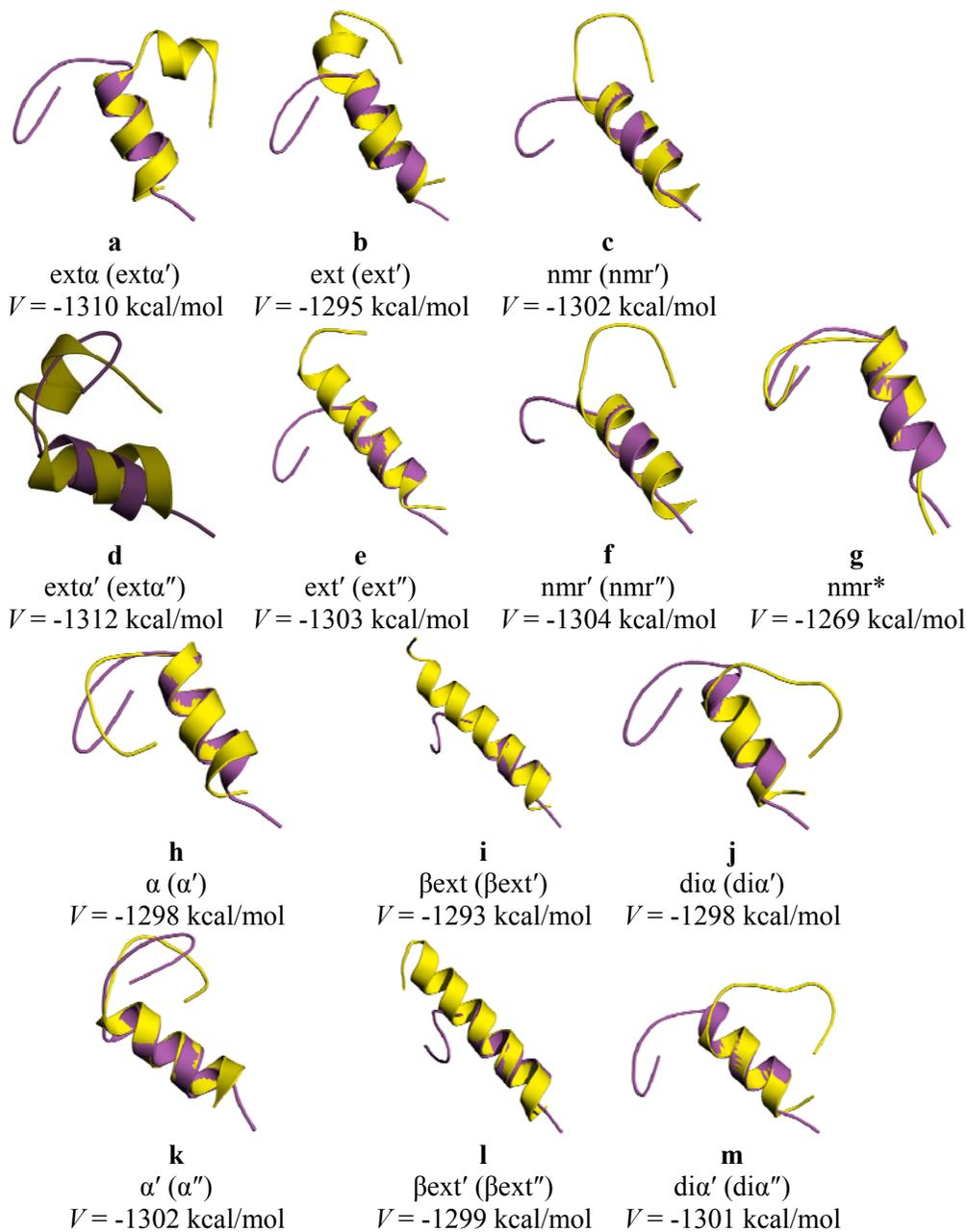


Figure 5.9 Simulated conformations (yellow) overlaying the NMR model (magenta) from the DIVE simulations. Names are listed by initial conformation and the conformation name used in the text.

conformation. Six of the sixteen polar side chain pairs appear in the conformations (R2D12, R2D13, D7R10, D7R19, R10E13, D12K16, and E13K16), and eight of the twelve conformations have the hydrophobic side chains of Leu14 and Leu17 within a distance of less than 6.0 Å. The aromatic side chains of Tyr1 and Tyr6 are aligned properly implying possible π -stacking for two conformations, and the aromatic side chains of Tyr1 and Phe8 are also in a proper geometric configuration.

Of the remaining conformations, $\text{ext}\alpha''$ (Figure 5.9f) exhibits the lowest PE and is composed of an α helix (Ser5–Arg19) whereas $\beta\text{ext}'$ (Figure 5.9i) consists of an α helix extending from Ser5 to Ala22. Ext'' (Figure 5.9e) and $\beta\text{ext}''$ (Figure 5.9l) are similar in their secondary structure because both are composed of α helices from Ser5 either to His21 or to Ala22, respectively. $\text{Di}\alpha'$ (Figure 5.9j), and $\text{di}\alpha''$ (Figure 5.9m) both contain α helices from Ser5 to Phe8 and from Glu13 to Ala22.

$\text{Ext}\alpha'$ (Figure 5.9c), ext' (Figure 5.9b), nmr' (Figure 5.9a), nmr'' (Figure 5.9d), α' (Figure 5.9h), and α'' (Figure 5.9k) have α helical conformations of varying lengths. Nmr' and nmr'' share nearly identical conformation (< 0.4 Å backbone difference). An α helix extending from Ser5 to Arg19 comprises the dominant feature of the $\text{ext}\alpha'$ conformation. The ext' conformation exhibits an α helical secondary structure from D-Pro4 to Ala22. α' and α'' have α helices extending from Ser11 to Gln20 and from Ser11 to Ala22, respectively.

When cluster analysis is performed for the twelve DIVE simulations, 1010 conformations are classified. One cluster has 24 conformations which are composed of both secondary structures from the NMR model (β hairpin with type II' turn: Arg2–Asp7; α helix: Phe8–Gln20). Several of the conformations are not identical to that of the NMR

model. The representative β hairpin is short by two residues, Tyr1 and Phe8, and the α helix extends beyond the experimental conformation (Glu13–Gln20). Yet, the analysis shows that DIVE samples conformations which exhibit both secondary structures without necessarily being derived from the experimental conformation.

Four other clusters have more than 20 conformations within them. The cluster with the largest number of conformations (42 conformations) has a general conformation consisting of an α helix from Ser9 to Leu17 and a π helix from Leu17 -Ala22. The next cluster (35 conformations) has two α helices separated by a γ -turn. 22 conformations are found in the fourth-largest cluster and are composed of an α helix (Asp7–Leu17). The final cluster encompasses 22 conformations but exhibits no noticeable secondary or tertiary structures.

In summary, a native-like conformation is reproduced within the DIVE simulations at a higher PE than the non-native conformations,. The tertiary structure of the experimentally-determined conformation is not reproduced within 2.0 Å of experiment. Although the experimentally-determined tertiary structure is not found, DIVE is able to sample some secondary structures as evidenced by cluster analysis.

The inability to locate the tertiary structure may be due to using the implicit solvent model of Tsui and Case (15), which has been shown to alter the free energy surface compared with explicit solvent (27). In addition, the internal dielectric constant represents a gas-phase environment instead of the hydrophobic environment encountered in proteins. The expected hydrophobic interactions are not as stable because of the unnatural internal environment.

Protein folding typically is expected to occur on a funnel-like energy surface where the system collapses into the native state (28-30). Yet, we have shown in previous chapters and in this chapter that the PE surface may be rough, with deep minima separated by large PE barriers. Considering work done by the Wales group (26), we suggest that, by usage of disconnectivity graphs, the PE minima are probably representative of the ‘weeping willow’ model of disconnectivity graphs (Figure 5.8). This model involves a distinct global PE minimum with high PE barriers preventing conformations from converging into a single conformation. This is compared with the ‘palm tree’ model, which also assumes a global PE minimum with low PE barriers allowing convergence of conformations as they move towards the global PE minimum.

The DIVE simulations present a possible global PE minimum, which should be representative of the global free energy minimum if the PEs are considered to contribute more to the free energy than entropy. The described conformations are probably surrounded by high PE barriers, and high therefore free energy, barriers. A number of PE minima have been mapped because the DIVE protocol allows conformations either to overcome or to circumvent the PE barriers.

By simulating the conformations found within possible PE minima near 300 K, we determine whether the conformations reside within local PE minima and free energy minima at a higher temperature. Near 0 K, the conformation designated as $\text{ext}\alpha''$ (see Figure 5.4j) resides within the lowest PE minimum, but near 300 K, the conformation designated as $\text{ext}\alpha'$ (see Figure 5.4f) resides within the lowest PE minimum.

5.4. Conclusions

The DIVE and DIP protocols have been tested for their abilities to reproduce both secondary and tertiary structures of the $\beta\beta\alpha$ motif near 0 K and 300 K, respectively. The DIVE simulations can also reproduce the secondary structures of both α helices and β hairpins, within 0.6 Å and 1.2 Å backbone atom rmsd's, respectively, relative to the NMR model. In comparison, the DIP simulations can reproduce both α helices and β hairpins within 0.5 Å and 1.5 Å backbone atom rmsd's relative to the NMR model when simulated from a native-like conformation. Cluster analysis of the DIP simulations finds one cluster that contains both secondary structures less than 2.0 Å backbone atom rmsd compared to the NMR model, and the cluster contains 12% of the conformations simulated. Similarly, cluster analysis of the conformations found in DIVE displays a cluster that contains both secondary structures; however, the cluster contains only 2% of the conformations. The structure representing the global PE minimum consist of α helices from Ser5 to Phe8 and from Glu13 to Ala22 for the cluster analysis performed with the DIP simulations and composed of an α helix from Ser9 to Leu17 and a π helix from Leu17 -Ala22 for cluster analysis performed with the DIVE simulations.

Although the overall $\beta\beta\alpha$ tertiary structure is not maintained during our simulations, the individual secondary structures are retained. This suggests that the loop region is flexible but that the β hairpin and the α helix are more ordered. Because the β hairpin does not collapse into an α helix while the experimental α helix is retained, the modified Amber force field (18) appears capable of modeling both secondary structure types. We also note that the calculated structures containing secondary structural

elements displayed by the native structure are not the global PE minimum in our simulations.

We also find several conformations that do not maintain the β -hairpin with PEs lower than the native-like conformations, but within 1–2 kcal/mol of the native-like conformations. We propose that the $\beta\beta\alpha$ motif can fold into both the β hairpin and the α helix without the tertiary structure, but the $\beta\beta\alpha$ motif is more likely to fold into secondary structures dominated by α helices. Considering cluster analysis, we further propose that the native conformation lies within a narrow-basin local PE minimum surrounded by high PE barriers, which prevent simulated structures from refolding into the conformation residing in the global PE minimum.

Because several conformations have PEs close to each other, but they do not interconvert, we suspect that the PE surface may be rough with several minima separated by high PE barriers. By using both the DIVE and DIP protocols, the PE surface may be mapped more thoroughly than by using conventional MD simulations. DIVE allows a simulation to map the PE surface of a system near 0 K. DIP, on the other hand, finds PE minima near physiological temperatures (simulation temperatures near 300 K). Finally, we suggest that the experimental conformation may not necessarily correspond to the global PE or free energy minimum for the modified Amber force field (18) used here, but merely resides in a local energy minimum.

5.5. References

1. Huang, Z. N., and Wheeler, R. A. Replica Microcanonical Search Methods for Mapping Potential Energy Landscapes and Conformations of miniproteins, *In preparation*.
2. Huang, Z. (2005) Design of new molecular dynamics global minimum search protocols for mapping energy landscapes and conformations of folded polypeptides and mini-proteins, in *Chemistry and Biochemistry*, p 282, The University of Oklahoma, Norman.
3. Klimov, D. K., and Thirumalai, D. (2000) Mechanisms and kinetics of beta-hairpin formation, *P Natl Acad Sci USA* 97, 2544-2549.
4. Struthers, M., Ottesen, J. J., and Imperiali, B. (1998) Design and NMR analyses of compact, independently folded BBA motifs, *Fold Des* 3, 95-103.
5. Struthers, M. D., Cheng, R. P., and Imperiali, B. (1996) Design of a monomeric 23-residue polypeptide with defined tertiary structure, *Science* 271, 342-345.
6. Jang, S., Kim, E., and Pak, Y. (2006) Free energy surfaces of miniproteins with a beta beta alpha motif: Replica exchange molecular dynamics simulation with an implicit solvation model, *Proteins-Structure Function and Bioinformatics* 62, 663-671.
7. Rhee, Y. M., Sorin, E. J., Jayachandran, G., Lindahl, E., and Pande, V. S. (2004) Simulations of the role of water in the protein-folding mechanism, *P Natl Acad Sci USA* 101, 6456-6461.
8. Snow, C. D., Nguyen, N., Pande, V. S., and Gruebele, M. (2002) Absolute comparison of simulated and experimental protein-folding dynamics, *Nature* 420, 102-106.
9. Totrov, M., and Abagyan, R. (2001) Rapid boundary element solvation electrostatics calculations in folding simulations: Successful folding of a 23-residue peptide, *Biopolymers* 60, 124-133.
10. Wang, L., Duan, Y., Shortle, R., Imperiali, B., and Kollman, P. A. (1999) Study of the stability and unfolding mechanism of BBA1 by molecular dynamics simulations at different temperatures, *Protein Sci* 8, 1292-1304.
11. Wen, E. Z., Hsieh, M. J., Kollman, P. A., and Luo, R. (2004) Enhanced ab initio protein folding simulations in Poisson-Boltzmann molecular dynamics with self-guiding forces, *J Mol Graph Model* 22, 415-424.
12. Wen, E. Z., and Luo, R. (2004) Interplay of secondary structures and side-chain contacts in the denatured state of BBA1, *J Chem Phys* 121, 2412-2421.
13. Dill, K. A., Phillips, A. T., and Rosen, J. B. (1997) Protein structure and energy landscape dependence on sequence using a continuous energy function, *J Comput Biol* 4, 227-239.
14. Jang, S., Shin, S., and Pak, Y. (2002) Molecular dynamics study of peptides in implicit water: Ab initio folding of beta-hairpin, beta-sheet, and beta beta alpha-motif, *J Am Chem Soc* 124, 4976-4977.
15. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular Simulations, *Biopolymers* 56, 275-291.

16. Feig, M., Karanicolas, J., and Brooks, C. L. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology, *J Mol Graph Model* 22, 377-395.
17. Hinsen, K. (2000) The molecular modeling toolkit: A new approach to molecular simulations, *J Comput Chem* 21, 79-85.
18. Okur, A., Strockbine, B., Hornak, V., and Simmerling, C. (2003) Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins, *J Comput Chem* 24, 21-31.
19. Andersen, H. C. (1983) Rattle: a Velocity Version of the Shake Algorithm for Molecular Dynamics Calculations, *J Comput Phys* 52, 24-34.
20. Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A Computer-Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *J Chem Phys* 76, 637-649.
21. Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992) Nosé-Hoover Chains: the Canonical Ensemble via Continuous Dynamics, *J Chem Phys* 97, 2635-2643.
22. Born, M., and von Karman, T. (1912) On fluctuations in spatial grids., *Phys Z* 13, 297-309.
23. Case, D. A., Darden, T. A., Cheatham, I., T.E. , Simmerling, C. L., Wang, J., Duke, R. R. E., Merz, L., K.M. , Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) AMBER 8, 8 ed., University of California, San Francisco, San Francisco.
24. McGaughey, G. B., Gagne, M., and Rappe, A. K. (1998) pi-stacking interactions - Alive and well in proteins, *J Biol Chem* 273, 15458-15463.
25. Lloyd, S. P. (1982) Least-Squares Quantization in Pcm, *Ieee T Inform Theory* 28, 129-137.
26. Wales, D. J., and Bogdan, T. V. (2006) Potential energy and free energy landscapes, *J Phys Chem B* 110, 20765-20776.
27. Zhou, R. H., and Berne, B. J. (2002) Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water?, *P Natl Acad Sci USA* 99, 12777-12782.
28. Anfinsen, C. B. (1973) Principles That Govern Folding of Protein Chains, *Science* 181, 223-230.
29. Leopold, P. E., Montal, M., and Onuchic, J. N. (1992) Protein Folding Funnels - a Kinetic Approach to the Sequence Structure Relationship, *P Natl Acad Sci USA* 89, 8721-8725.
30. Zwanzig, R., Szabo, A., and Bagchi, B. (1992) Levinthals Paradox, *P Natl Acad Sci USA* 89, 20-22.

6. Prediction of a protein's unknown conformation

6.1. Introduction

Nitrogen catabolism is regulated in *Saccharomyces cerevisiae* by Ure2p, a two-domain protein (1-8). Residues 1–91 are contained in the N-domain, and residues 92-354 are included in the C-domain. The C-domain is enzymatically active during nitrogen regulation; no enzymatic functionality is known to occur within the N-domain. However, the N-domain can misfold into a prion form, [URE3] (9-32).

The C-domain is crystallized and its structure determined by Umland, et al., and its structure was determined at a resolution of 2.3 Å (8) (Figure 6.1) and by Bousset, et al. at a resolution of 2.5 Å (33). Two domains are within the crystallized dimer. The N-terminal domain is represented by Glu112 to Gly197, which has a $\beta\alpha\beta\alpha\beta\alpha$ motif. A linker between the two domains is formed by Asn198 to Asp204. The C-terminal domain (Asp205 to Glu354) has six α helices with a single 3_{10} helix turn. The cleft between the C-domain and the N-domain of Ure2p(97–354) may be used to bind various substrates involved in the nitrogen regulation. Within the dimer, each monomer interacts between $\alpha 5$ and $\beta 4/\alpha 4$ on the N-terminal region and $\alpha 5/\alpha 6$ on the C-terminal side. The dimer is necessary for regulatory activity.



Figure 6.1 Dimer of Ure2p(97-354), pdb code: 1HQO (8). Monomer A is yellow and monomer B is magenta.

When the N-domain misfolds into [URE3], the protein is formed into fibrils and enzymatic activity is inhibited (6). [URE3] is a non-Mendelian mutation that is phenotypically similar to a *ure2* gene mutation (5, 9, 34). [URE3] is classified as a prion protein on the basis of three factors: reversible curability by guanidine HCl, dependence upon the chromosomal *ure2* gene, and induction by overexpression of Ure2p (11).

[URE3] is either aggregated or formed into fibrils (15, 16, 22-28, 35, 36). The formation of fibrils or aggregates is indicative of the hydrophobicity of the N-domain. Because the N-domain is insoluble in aqueous solution, any structural characterization has so far been experimentally impossible (14, 15, 21, 23, 26). The prion region is rich in asparagine and glutamine (13). Through computational studies, we present several possible secondary structures of residues 1–64.

We attempt to predict potential secondary and tertiary structures using two protocols developed within our group: the disrupted velocity (DIVE) and the divergent path (DIP) search protocols (37, 38). We have presented validation studies in the previous chapters. Through our studies, we find that both α helices and β hairpins both separately and together can be located on a potential energy (PE) surface using these two protocols in conjunction. We now stretch the two protocols to the limit by predicting conformations based solely on the primary sequence.

6.2. Simulation procedure and analysis

Because the secondary and tertiary structures of the N-domain of Ure2p have, to our knowledge, not been determined, Ure2p(1–64) is simulated using the disrupted velocity (DIVE) and divergent path (DIP) protocols (37, 38). Ure2p(1–64) consists of MMNNGNQVSNLSNALRQVNIGNRNSNYYYDQSNINFDfsygvnnnnnnnss

NNNNVQNNNSG. Two initial conformations are selected because the actual conformation is unknown. A fully-extended conformation ($\varphi = -180^\circ$, $\psi = +180^\circ$) and an α -helical conformation ($\varphi = -60^\circ$, $\psi = -40^\circ$) are simulated. After the initial DIVE simulations, three conformations are selected for further simulation using both protocols. The conformations corresponding to the three lowest PEs are selected for further simulation. In total, eight conformations are simulated. Conformations are designated by the derivation from the initial conformation. For instance, the second round of DIVE simulations obtained from the fully-extended conformation is assigned the name “extlow2”.

For all simulations, the PE for the initial conformations is minimized for 100 steps with steepest descent in implicit water solvent (39) by the Multiscale Modeling Tools for Structural Biology (MMTSB) (40) program. The Molecular Modeling Toolkit (MMTK) (41) is used to convert the coordinates into files used by our programs. For DIP simulations, velocities are scaled to an initial temperature of 300 K.

Several parameters are included in the simulations. Covalent bonds to hydrogen are constrained by the SHAKE (42) algorithm. All simulations are run in a Generalized Born/surface area (GB/SA) implicit solvent model (39) with an external dielectric constant of 78.5, an internal dielectric constant of 1.0, surface tension at 0.005 kcal/mol- \AA^2 , and an offset of 0.9 \AA . A modified version of Amber99 (43) is used during minimization and data collection. Instead of periodic boundaries, electrostatic and Lennard-Jones cutoffs are set to 999 \AA to represent an infinite cutoff. Constant temperature is maintained using the Nosé-Hoover Chain method (44), and the equations of motion are integrated with the velocity-Verlet method (45). Both protocols have six

simultaneous, yet independent, copies transpiring within each simulation. Each copy is initialized with the same atomic velocity magnitude to maintain constant temperature but different atomic velocity directions.

Each simulation is run for 4 million steps with a 2 fs time step (4×10^6 steps/copy \times 2 fs/step \times 6 copies/simulation = 48 ns total simulation time), and data is output every 250 steps (0.5 ps). Each copy in the DIVE protocol is initially assigned different temperatures (10 K, 50 K, 100 K, 300 K, 600 K, and 1000 K). Atomic velocity reassignment for each copy occurs after 20,000 steps. The scaling parameter for heating is calculated from a target temperature ($T_{\text{target}} = 1000$ K), and the scaling parameter for cooling is set at 0.25. The threshold temperature is 10 K. Above the threshold temperature, kinetic energy is removed by decreasing atomic velocities, and below the threshold temperature, kinetic energy is added by increasing atomic velocities. Although the target temperature is assigned at 1000 K, 500 K is the maximum temperature achieved because of rapid energy redistribution.

For the DIP protocol, all simulations are run at a constant temperature of 300 ± 20 K. All copies are initially assigned the same atomic velocity magnitude but different atomic velocity directions. Each copy is allowed to traverse the PE surface in a canonical simulation (constant number of atoms, volume, and temperature) in which the temperatures are within a specified but limited range.

Hydrogen bond (H-bond) distances between charged polar side chains, and distances between aromatic side chains are calculated using the analysis module ptraj of AMBER 8 (46). H-bonds are conventionally defined by a range of $180.0^\circ \pm 60.0^\circ$ for the $X-H \cdots X_1$ angle where X and X_1 are polar heavy atoms, with a distance less than 3.5 Å

distance between polar heavy atoms and with an appearance in more than 5% of the simulations.. For DIP, H-bonds are determined for each low energy conformation, and for DIVE, H-bonds are determined for each initial simulated conformation. Salt bridges are deemed possible if the distances between the terminal atoms of the charged polar side chains are less than 3.5 Å. Five possible ionic, interactions are defined between R24D31, R17E38, R24D31, R24E38 and the terminal M1G64. Conventionally, two aromatic side chains are considered in proper alignment when the distances between any heavy atom in the aromatic rings is less than 6 Å apart and when the dihedral angle ($C\delta_{i1}-C\delta_{i2}\cdots C\delta_{j2}-C\delta_{j1}$) between the two aromatic rings is less than $\pm 90^\circ$. In Ure2p(1–64), eleven aromatic side chain pairs can possibly align in the proper geometry for π -stacking: Y28Y29, Y29Y30, Y28F37, Y29F37, Y30F37, F37F39, F37Y41, Y28F39, Y29F39, Y30F39, and F39Y41. Distances are calculated between the heavy atoms of the aromatic rings, and the torsion angle is calculated for $C\delta_{i1}-C\delta_{i2}\cdots C\delta_{j2}-C\delta_{j1}$. STRIDE³³, which classifies secondary structures on the basis of backbone dihedral angles and H-bond interactions, is used to determine the secondary and tertiary structures of conformations used in DIVE and DIP simulations.

An alternative analytical tool is cluster analysis, which groups conformations on the basis of secondary and tertiary structures. Cluster analysis can provide a qualitative indication of the entropic contribution to the free energy from a particular family of conformations. Cluster analysis for both DIVE and DIP simulations is determined by MMTSB (40). Cluster analysis consists of an iterative process with an error tolerance of 0.5 in the rmsd. The conformations are aligned by a least-squares fit rmsd comparison of the conformations and are classified by a centroid method (47), which groups

conformations together on the basis of an average conformation. Centroid clusters are defined as having a radius of 3.0 Å. All residues are used for the classification of clusters.

6.3. Simulation results and discussion

Because the conformation of the N-domain of Ure2p is unknown at this time, two conformations, a fully-extended and an α helix conformations, are initially run using DIVE, and the conformations corresponding to the three lowest PEs are further simulated (designated $\alpha.1'$, $\alpha.2'$, $\alpha.3'$, ext.1', ext.2', and ext.3') with both DIVE and DIP. Upon completion of the second round of DIVE simulations, the lowest PE conformations are equilibrated using DIP and are designated with a 'b' following the initial definition (e.g., ext.3'').

6.3.1. Divergent path (DIP) results and discussion

The divergent path (DIP) simulations reveal several conformations near physiological temperatures. The average conformation from each equilibrated simulation is described with information about the secondary and tertiary structures available. Afterwards, a description of the results and a PE-independent analysis of the conformations are mentioned.

The fully-extended and the fully α -helical conformations are simulated near physiological temperatures. The initial α helix (Figure 6.4a) refolds into two α helices (Met2–Ser53 and Asn57–Asn62). The two aromatic side chains of Phe37 and Tyr41 are aligned properly according to the $i, i+4$ definition of an α helix. The fully-extended conformation (Figure 6.4b) folds into a unique conformation — two β strands (Arg17–Asn20 and Asp31–Asn34) and an α helix (Asn47–Ser51).

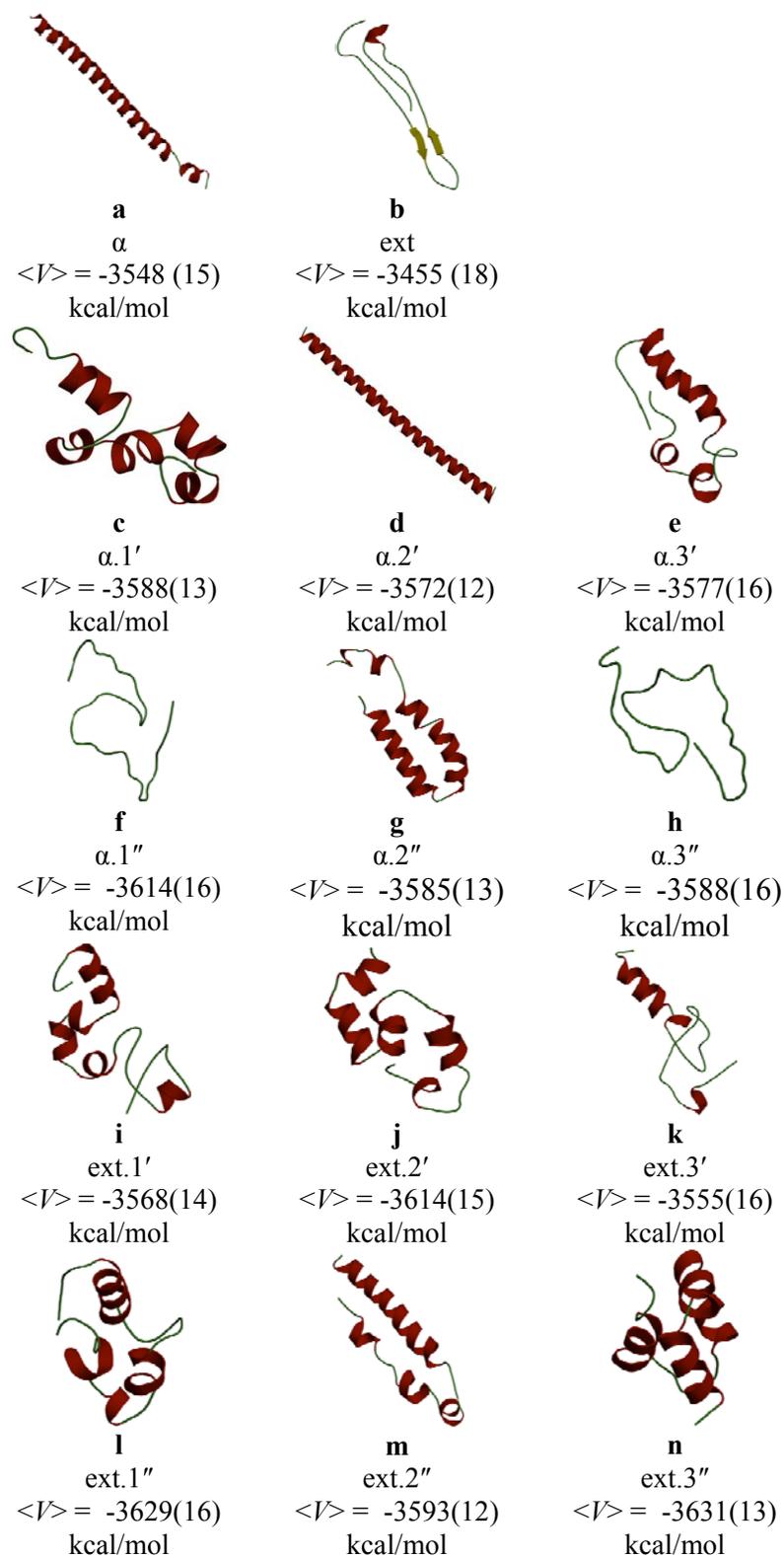


Figure 6.2 Average equilibrated conformations of ten conformations originating from the DIVE simulations in addition to the fully-extended and α helix conformations. α helices are colored red, loops are green, and β -sheets are yellow.

The $\alpha.1'$ (Figure 6.4c) simulation consists of an average of four α helices: Leu12–Ile21, Asn27–Glu38, Asn50–Ser51, and Ser53–Val58. The side chain interactions between R17D31, R17E38, and R24E38 and the close proximity of Y29F37 and Y30F37 appear to stabilize the last three helices. On the other hand, $\alpha.1''$ (Figure 6.4f) contains no noticeable secondary structure with one possible interaction between Tyr30 and Phe37.

The average conformation of the $\alpha.2'$ (Figure 6.4d) simulation is an α helix extending from Met2–Ser63, and the aromatic side chains of Tyr37 and Tyr41 are positioned in proximity because of α helix ($i, i+4$). The $\alpha.2''$ simulation (Figure 6.4g), on the other hand, exhibits four α helices (α_1 : Asn4–Asn20, α_2 : Asn27–Ser40, α_3 : Val43–Val48, and α_4 : Ser51–Asn56) and a 3_{10} helix (Val58–Asn60). The interaction between Arg17 and Glu31 appears to maintain the close proximity of α_1 and α_2 , whereas interactions between Tyr29 and Phe37 and between Phe37 and Tyr41 seem to maintain α_2 .

The $\alpha.3'$ (Figure 6.4e) average conformation has three helices (Gln8–Ser13, Leu16–Asn20, and Ile35–Ser52). The tertiary structure is maintained by the side chain interactions between Arg17 and Asp31 and between Arg24 and Glu38. $\alpha.3''$ (Figure 6.4h), on the other hand, is composed of no secondary structural elements, but α helices may continue to fluctuate between Asn11 and Arg17 and between Tyr41 and Gln59.

The ext.1' (Figure 6.4i) simulation exhibits an average conformation consisting of five α helices: Gln8–Arg17, Val19–Gly22, Arg24–Tyr29, Asp31–Asn36, and Asn48–Ser52. The close proximity between the Arg17 and Glu38 side chains appears to stabilize a loop region containing three of the α helices. On the other hand, ext.1'' (Figure 6.4j) is composed of four α helices (Gln8–Asn20, Gln32–Phe37, Val43–Asn47, and

Asn50–Asn54). Three charged polar side chain pairs are within distance for possible interaction (R17D31, R24D31, and R24E38) and the interaction between the terminal residues Met1 and Gly64 is also present; the four potential interactions allow for maintenance of the overall tertiary structure.

The average ext.2' (Figure 6.4j) conformation exhibits five α helices; the first helix extends from Asn4 to Val9. Leu12 to Ile21, Asn23–Tyr30, Phe37–Tyr41, and Asn55–Val58 complete the secondary structures present, whereas the average ext.2" (Figure 6.4m) is composed of three α helices (Asn4–Ser10, Leu16–Asn20, Asn25–Tyr30) and an additional π helix extending from Glu38 to Gly42 followed by an α helix from Val43 to Ser63. The aromatic side chains of Phe37 and Tyr41 are within a proper geometric configuration attributed to the α helical conformation during the ext.2' simulation. During the ext.2" simulation, the distance between the Arg24 and Asp31 side chains seem to help stabilize the α helix in which they are located, whereas the side chain interactions between Arg17 and Glu38 and between Arg24 and Glu38 appear to help maintain a tertiary structure between the second and fourth helices and the third and fourth helices, respectively.

During the ext.3' (Figure 6.4k) simulations, a π helix extends between Asn11 and Arg17 followed by a 3_{10} helix from Val19 to Gly22; in addition, α helices between Asn4 and Ser10 and between Asn56 and Asn60 are present. Ext.3" (Figure 6.4n) is composed of four α helices (Gln8–Leu16, Tyr28–Ile35, Phe39–Val43, and Ser51–Ser63) and a π helix (Asn44–Asn48). The side chain interaction between Arg17 and Asp31 appears to help maintain a close proximity of the first two helices.

The distances between charged polar side chains as well as between aromatic rings play a role in the secondary and tertiary structure of Ure2p(1–64). Of the five possible interactions between charged polar side chains, six simulations displayed the pair interaction R17E38, whereas R24D31 is found only during two simulations. R17D31 and R24E38 are included in three and four simulations, respectively. Interestingly, three simulations ($\alpha.1''$, $\alpha.3''$, ext.1'', ext.2'') have the terminal residues in close contact. Fourteen possible aromatic side chain interactions were investigated, and Y29Y41, Y30Y41, Y28F37, Y28f39, Y29F39, Y30F39, and F37F39 do not appear during any of the simulations. The alignment for π -stacking between Tyr28 and Tyr29 occurs during eight of the simulations, and the other aromatic side chain interactions appear less often. When Phe37 and Tyr41 are in close proximity, they are typically found in an α helix.

Cluster analysis is another way to view the sampling ability of the simulations. Cluster analysis classifies conformations on the basis of similarities. This analysis technique is independent of PE and offers a qualitative determination of the entropic contribution to a conformational family's free energy. A large number of conformations that comprising a cluster indicates a higher entropic contribution.

Table 6.1 Description of secondary structure for representative conformation within the seven highly populated clusters

Number of conformations within a cluster	Secondary structures of representative conformation	Simulation represented by conformation
993	Globular	$\alpha.1''$
961	α helices (Gln8–Val19, Gln32–Phe37, Val43–Asn46) π helices (Asn50–Asn54)	ext.1''
879	α helices (Gln8–Asn36)	ext.1'
867	α helices (Asn4–Tyr29, Phe37–Tyr41, Asn55–Val58)	ext.2'
713	α helices (Met2–Ser63)	$\alpha.2'$
707	α helices (Asn5–Asn11, Val19–Gly22, Asn56–Gln59) π helices (Leu12–Leu16)	ext.3'
631	α helices (Gln8–Leu16, Tyr28–Ile35, Phe39–Val43, Ser51–Ser63) π helices (Asn44–Asn48)	ext.3''
596	α helices (Asn5–Val9, Ala15–Ile21, Asn27–Tyr41, Asn44–Asn50, Ser53–Val58)	$\alpha.1'$
510	α helices (Asn4–Val9, Leu12–Leu16, Asn25–Tyr30, Ile35–Ser63) π helices (Arg17–Ile21)	ext.2''

Nine clusters, or conformation families, are found with significant populations (more than 500 conformations per family). The secondary structures of each representative conformation are described in Table 6.1. The majority of the representative conformations are composed of α helices, but the cluster containing the most conformations is globular. The conformations are merely representatives of a larger family of conformations, and the representatives offer us a glimpse into the possible entropic contributions of the overall family. Ext.3'' may reside within the lowest average PE of our simulations, but according to cluster analysis, ext.3'' may not experience the highest entropy contribution to free energy; instead, the representative conformation fits the average conformation of $\alpha.1''$ (data not shown). We can conclude, however, that the

cluster analysis alone cannot be the overall determinant of the most-favored conformation. This analytical tool merely groups conformations together on the basis of similarities within the secondary and tertiary structures.

The conformations found within the simulations may match the hypothesis of Perrett et al., (14), Boussett et al. (24), or Baxa et al. (23) Two of the groups indicate that the N-domain of Ure2p may not be structured near physiological temperatures (14, 23), which we find (Figure 6.2f,h). On the other hand, Boussett et al. (24) suggest that the N-domain, when in the prion form [URE3], is predominately α -helical in conformation instead of the expected β -sheet conformation. The majority of our conformations are, in fact, α -helical with the exception of the two globular conformations and a conformation which contains two antiparallel β strands (Figure 6.2b).

On the basis of the PE observations (Figure 6.3), one might conclude that none of the conformations mentioned is the definitive global PE conformation. The standard deviation, however, conveys the expected fluctuations during the simulation, but the standard error of the mean, which measures the accuracy of the average PE, ranges between 0.2 kcal/mol and 0.3 kcal/mol suggesting that the PEs are indeed minima. With this in mind, we propose the average conformation designated as ‘ext.3’ resides in the global PE minimum. The lack of conformational convergence into a single minimum is probably represented by the ‘weeping willow’ model as defined by the Wales group (48-54) because Ure2p(1–64) appears to have a distinctive global PE minimum with several local PE minima separated by relatively high barriers.

Biochemically, the high PE barriers between conformations is reflected in the high kinetic energy required to break and to re-form noncovalent interactions. The PE

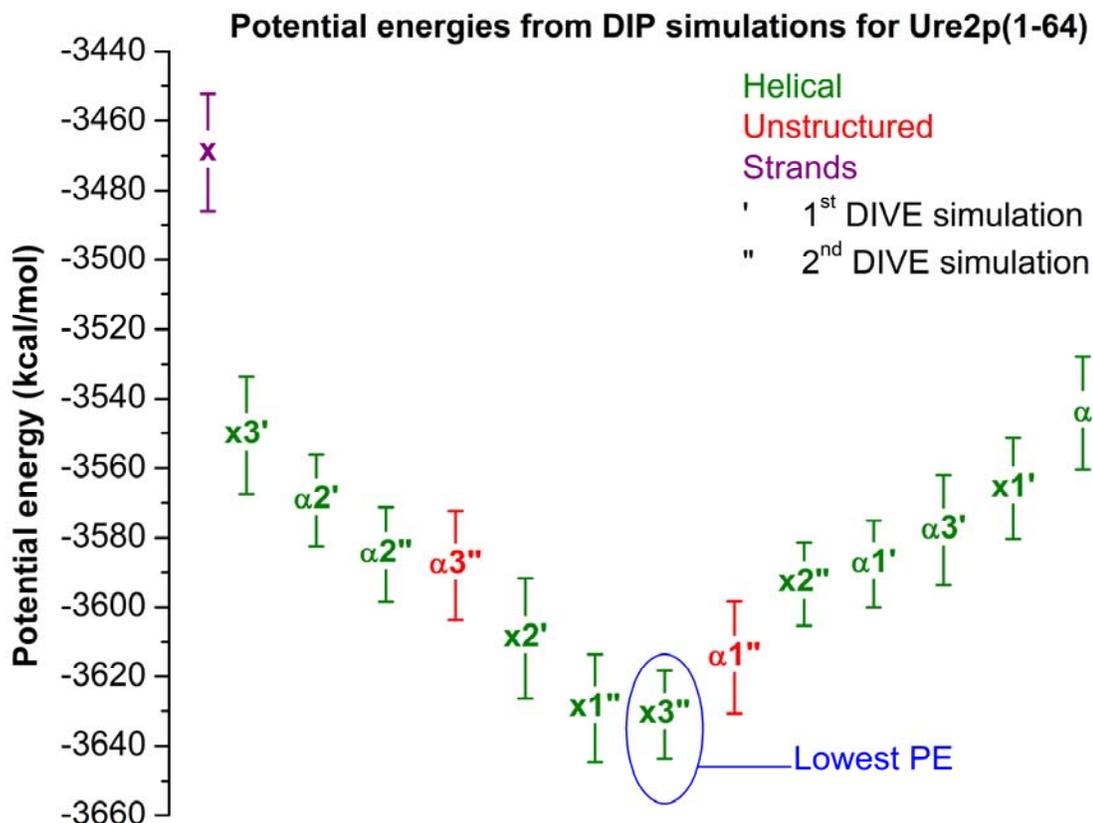


Figure 6.3 Overlapping PEs (PE) of the equilibrated conformations from the divergent path (DIP) simulations

minima may appear similar but without the necessary kinetic energy, conformational changes are not likely. The DIVE protocol allows the barriers to be crossed, but when the various conformations are simulated with the DIP protocol, the kinetic energy is fixed disfavoring PE barrier crossing within the limited time for the simulations.

Yet, one cannot overlook the detail that the majority of the equilibrated conformations exhibit some helical conformation. We suspect that the native conformation itself should maintain this particular secondary structure. However, the secondary structure should, by all accounts, fold into a β -hairpin-rich motif in water (12, 14, 31, 35, 55, 56). Because a single protein is simulated, the conformation(s) may not be representative of a typical aggregate.

6.3.2. Disrupted velocity (DIVE) results and discussion

The disrupted velocity (DIVE) protocol sampled 623 conformations and, of those, twelve conformations are listed below. Following the listed conformations, an analysis of all the conformations is performed allowing the conformations to be grouped according to similarities on the basis of secondary structure.

Six conformations are derived from a full-length α helix. $\alpha.1'$ (Figure 6.4a) contains five α helices (Asn7–Asn20, Asn27–Ser33, Asn36–Tyr41, Asn44–Asn49, and Ser53–Val59). $\alpha.1''$ (Figure 6.4d) consists of five α helices: Asn7–Asn20, Asn27–Ser33, Asn36–Tyr41, Asn44–Asn49, and Ser53–Val58. Five α helices and a π helix comprise $\alpha.2'$ (Figure 6.4b): Asn3–Asn11, Leu12–Ile21 (π helix), Asn23–Asp31, Ser33–Tyr41, Asn46–Ser51, and Ser53–Asn63, but $\alpha.2''$ (Figure 6.4e) consists of an α helix and π helix combination (Gly6–Gln18 and Val19–Asn23, Tyr28–Phe37 and Glu38–Gly42, and Ser51–Asn56). $\alpha.3'$ (Figure 6.4c) has the secondary structure consisting of three α helices compared with $\alpha.3''$ (Figure 6.4f), which is composed of five α helices (Asn4–Asn8, Asn11–Asn17, Tyr29–Tyr33, Tyr41–Asn50) and a π helix (Ile35–Phe39).

Similar to the six conformations derived from a full-length α helix, six conformations are simulated from a fully-extended conformation. Ext.1' (Figure 6.4g) is composed of two α helices from Asn7 to Arg17 and from Gln32 to Glu38. On the other hand, ext.1'' (Figure 6.4j) is composed of α helices extending from Gln8 to Asn20 and from Gln32 to Phe37 and has a π helix extending from Asn50 to Asn54. Ext.2' (Figure 6.4h) contains three α helices (Asn4–Val9, Leu12–Gly22, and Phe37–Gly42). According to STRIDE's (57) H-bond classification, however, Asn27–Tyr30 is not included as an α helix, but the backbone dihedral angles permit the region to be classified as such. Ext.2''

(Figure 6.4k) is composed of four α helices (Asn4–Ser10, Leu12–Ile21, Asn25–Tyr30, and Phe37–Gln59). Ext.3' (Figure 6.4i) contains only three α helices: Gln8–Leu16, Phe39–Val43, and Asn56–Asn60. On the other hand, ext.3'' (Figure 6.4l) is composed of four α -helical segments (Leu12–Arg17, Tyr28–Ile35, Phe39–Val43, and Ser51–Asn62).

The Ure2p(1–64) exhibits several noncovalent interactions including charged polar side chain interactions, which may form salt bridges, and aromatic side chain pairs that align for possible π -stacking. The side chains of Arg17 and Glu38 are within less than 3.5 Å in seven of the conformations, and the other three charged polar side chain pairs are less common. The four salt bridges keep secondary structures in close proximity. Only two aromatic side chain pairs (F37Y41 and F37F39) play a significant role in the secondary structures of the various conformations.

Cluster analysis is performed on the conformations found near 0 K. 590 clusters are found, and six clusters, or families of conformations, contain more than five conformations. The three largest clusters contain seven conformations with either three or four α helices. The other three clusters contain six conformations and vary between either two or three α helices. We cannot determine much from the cluster analysis because the conformations seem to be evenly distributed. If a smaller region of Ure2p(1–64) was used for cluster analysis, a different picture might emerge, but we wanted to classify the conformations on the basis of the whole protein..

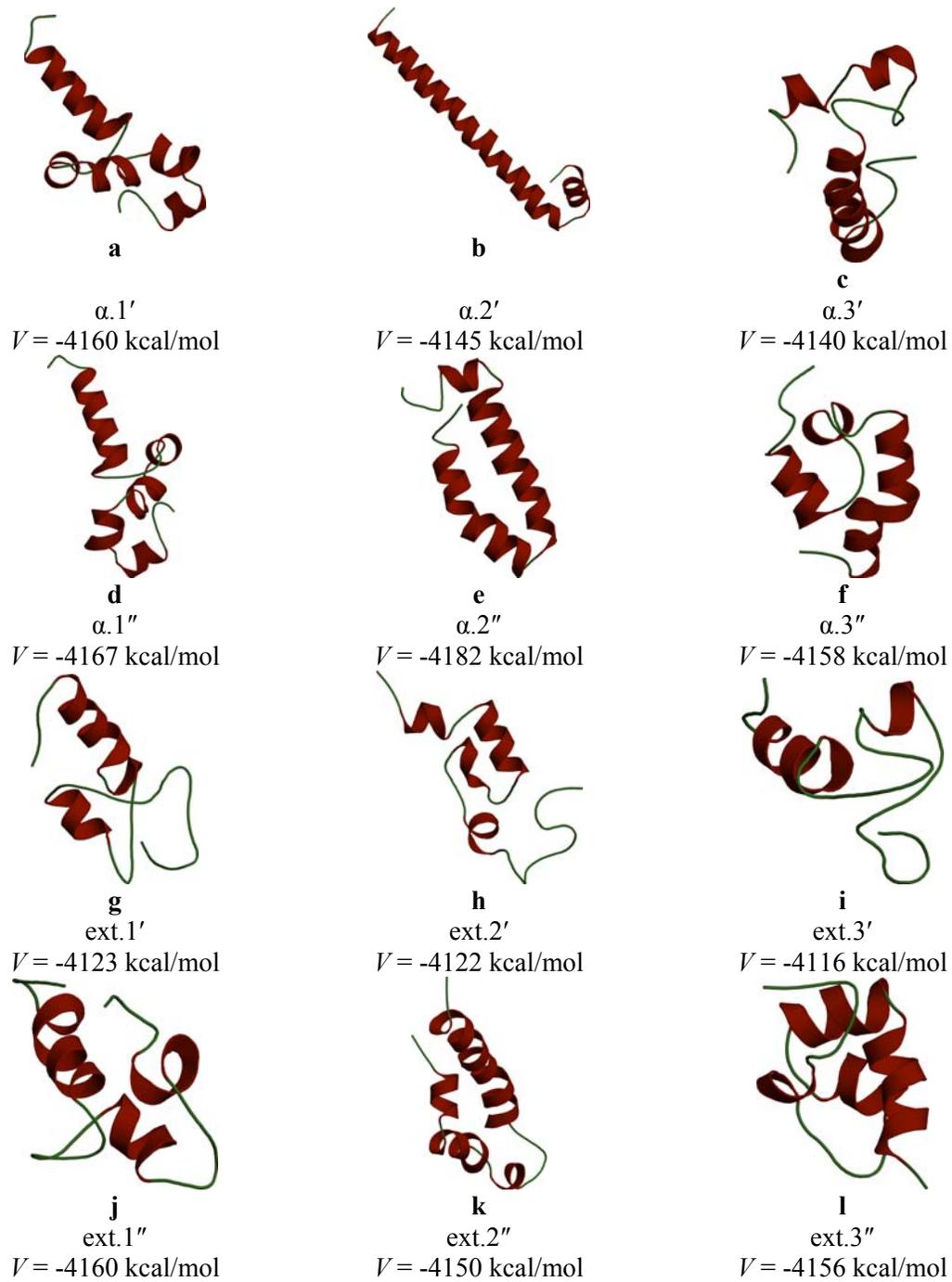


Figure 6.4 The twelve conformations corresponding to the lowest PE simulated using DIVE. The conformations are designated based upon their derivation from either the fully-extended or the α -helical conformation. α helices are colored red, and loops are colored green.

From the simulations near 0 K, we have sampled several conformations. On the basis of PE only, $\alpha.2''$ falls into the category of a possible global PE minimum. $\alpha.1''$ has a PE that is 15 kcal/mol higher than the PE of $\alpha.2''$. However, the conformations mentioned above do not reveal a complete picture of the PE surface, and therefore, our results determining the conformation corresponding to the lowest PE is inconclusive.

6.4. Conclusions

Overall, we have shown that both the DIVE and DIP protocols can be used individually or together for prediction of conformations. The information presented is limited and further work is needed. The DIVE protocol allows several conformations to be sampled near 0 K, whereas DIP equilibrates the conformations near 300 K. Both the computational and the real time for each simulation are limiting factors. However, we offer the possibility that the two protocols have indeed found several alternative conformations, which agree with previous work by other groups.

Biochemically, the results indicate that several possible conformations exist within a narrow PE range. The inability of the conformations to converge or to transition amongst themselves rapidly is due to high PE barriers. Kinetic energy is required to break and re-form the noncovalent bonds; the DIVE protocol inputs enough kinetic energy to permit conformational changes, but the DIP protocol maintains constant kinetic energy disallowing rapid conformational changes. We propose that the global PE minimum for Ure2p(1–64) has a high degree of α -helical content, similar to the extlow.3b average conformation..

Until the actual conformation of Ure2p(1–64) is experimentally determined, we are uncertain which conformation is correct. We have presented several possible

conformations that may aid in future experimental studies of the protein. Yet, we must conclude that more simulations may be necessary, some perhaps starting from β -hairpin conformations, to offer alternative conformations.

6.5. References

1. Beck, T., and Hall, M. N. (1999) The TOR signaling pathway controls nuclear localization of nutrient-regulated transcription factors, *Nature* 402, 689-692.
2. Blinder, D., Coschigano, P. W., and Magasanik, B. (1996) Interaction of the GATA factor Gln3p with the nitrogen regulator Ure2p in *Saccharomyces cerevisiae*, *J Bacteriol* 178, 4734-4736.
3. Coschigano, P. W., and Magasanik, B. (1991) The Ure2 Gene Product of *Saccharomyces Cerevisiae* Plays an Important Role in the Cellular-Response to the Nitrogen-Source and Has Homology to Glutathione S-Transferases, *Mol Cell Biol* 11, 822-832.
4. Courchesne, W. E., and Magasanik, B. (1988) Regulation of Nitrogen Assimilation in *Saccharomyces Cerevisiae*: Roles of the Ure2 and Gln3 Genes, *J Bacteriol* 170, 708-713.
5. Drillien, R., Aigle, M., and Lacroute, F. (1973) Yeast Mutants Pleiotropically Impaired in Regulation of 2 Glutamate Dehydrogenases, *Biochem Bioph Res Co* 53, 367-372.
6. Drillien, R., and Lacroute, F. (1972) Ureidosuccinic Acid Uptake in Yeast and Some Aspects of Its Regulation, *J Bacteriol* 109, 203-&.
7. Edskes, H. K., Hanover, J. A., and Wickner, R. B. (1999) Mks1p is a regulator of nitrogen catabolism upstream of Ure2p in *Saccharomyces cerevisiae*, *Genetics* 153, 585-594.
8. Umland, T. C., Taylor, K. L., Rhee, S., Wickner, R. B., and Davies, D. R. (2001) The crystal structure of the nitrogen regulation fragment of the yeast prion protein Ure2p, *P Natl Acad Sci USA* 98, 1459-1464.
9. Wickner, R. B. (1994) [Ure3] as an Altered Ure2 Protein: Evidence for a Prion Analog in *Saccharomyces Cerevisiae*, *Science* 264, 566-569.
10. Masison, D. C., and Wickner, R. B. (1995) Prion-Inducing Domain of Yeast Ure2p and Protease Resistance of Ure2p Prion-Containing Cells, *Science* 270, 93-95.
11. Masison, D. C., Maddelein, M. L., and Wickner, R. B. (1997) The prion model for [URE3] of yeast: Spontaneous generation and requirements for propagation, *P Natl Acad Sci USA* 94, 12503-12508.
12. Fernandez-Bellot, E., Guillemet, E., Baudin-Baillieu, A., Gaumer, S., Komar, A. A., and Cullin, C. (1999) Characterization of the interaction domains of Ure2p, a prion-like protein of yeast, *Biochem J* 338, 403-407.
13. Maddelein, M. L., and Wickner, R. B. (1999) Two prion-inducing regions of Ure2p are nonoverlapping, *Mol Cell Biol* 19, 4516-4524.
14. Perrett, S., Freeman, S. J., Butler, P. J. G., and Fersht, A. R. (1999) Equilibrium folding properties of the yeast prion protein determinant Ure2, *J Mol Biol* 290, 331-345.
15. Taylor, K. L., Cheng, N. Q., Williams, R. W., Steven, A. C., and Wickner, R. B. (1999) Prion domain initiation of amyloid formation in vitro from native Ure2p, *Science* 283, 1339-1343.

16. Thual, C., Komar, A. A., Bousset, L., Fernandez-Bellot, E., Cullin, C., and Melki, R. (1999) Structural characterization of *Saccharomyces cerevisiae* prion-like protein Ure2., *Faseb J* 13, A1485-A1485.
17. Bousset, L., Thomson, N. H., Radford, S. E., and Melki, R. (2002) The yeast prion Ure2p retains its native alpha-helical conformation upon assembly into protein fibrils in vitro, *Embo J* 21, 2903-2911.
18. Edskes, H. K., and Wickner, R. B. (2002) Conservation of a portion of the *S-cerevisiae* Ure2p prion domain that interacts with the full-length protein, *P Natl Acad Sci USA* 99, 16384-16391.
19. Mornon, J. P., Prat, K., Dupuis, F., and Callebaut, I. (2002) Structural features of prions explored by sequence analysis I. Sequence data, *Cell Mol Life Sci* 59, 1366-1376.
20. Baudin-Baillieu, A., Fernandez-Bellot, E., Reine, F., Coissac, E., and Cullin, C. (2003) Conservation of the prion properties of Ure2p through evolution, *Mol Biol Cell* 14, 3449-3458.
21. Bousset, L., Briki, F., Doucet, J., and Melki, R. (2003) The native-like conformation of Ure2p in fibrils assembled under physiologically relevant conditions switches to an amyloid-like conformation upon heat-treatment of the fibrils, *J Struct Biol* 141, 132-142.
22. Zhu, L., Zhang, X. J., Wang, L. Y., Zhou, J. M., and Perrett, S. (2003) Relationship between stability of folding intermediates and amyloid formation for the yeast prion Ure2p: A quantitative analysis of the effects of pH and buffer system, *J Mol Biol* 328, 235-254.
23. Baxa, U., Ross, P. D., Wickner, R. B., and Steven, A. C. (2004) The N-terminal prion domain of Ure2p converts from an unfolded to a thermally resistant conformation upon filament formation, *J Mol Biol* 339, 259-264.
24. Bousset, L., Redeker, V., Decottignies, P., Dubois, S., Le Marechal, P., and Melki, R. (2004) Structural characterization of the fibrillar form of the yeast *Saccharomyces cerevisiae* prion Ure2p, *Biochemistry-Us* 43, 5022-5032.
25. Ripaud, L., Maillet, L., Immel-Torterotot, F., Durand, F., and Cullin, C. (2004) The [URE3] yeast prion results from protein aggregates that differ from amyloid filaments formed in vitro, *J Biol Chem* 279, 50962-50968.
26. Baxa, U., Cheng, N. Q., Winkler, D. C., Chiu, T. K., Davies, D. R., Sharma, D., Inouye, H., Kirschner, D. A., Wickner, R. B., and Steven, A. C. (2005) Filaments of the Ure2p prion protein have a cross-beta core structure, *J Struct Biol* 150, 170-179.
27. Fay, N., Redeker, V., Savistchenko, J., Dubois, S., Bousset, L., and Melki, R. (2005) Structure of the prion Ure2p in protein fibrils assembled in vitro, *J Biol Chem* 280, 37149-37158.
28. Pierce, M. M., Baxa, U., Steven, A. C., Bax, A., and Wickner, R. B. (2005) Is the prion domain of soluble Ure2p unstructured?, *Biochemistry-Us* 44, 321-328.
29. Talarek, N., Maillet, L., Cullin, C., and Aigle, M. (2005) The [URE3] prion is not conserved among *Saccharomyces* species, *Genetics* 171, 23-34.
30. Bach, S., Tribouillard, D., Talarek, N., Desban, N., Gug, F., Galons, H., and Blondel, M. (2006) A yeast-based assay to isolate drugs active against mammalian prions, *Methods* 39, 72-77.

31. Cecchini, M., Curcio, R., Pappalardo, M., Melki, R., and Caflisch, A. (2006) A molecular dynamics approach to the structural characterization of amyloid aggregation, *J Mol Biol* 357, 1306-1321.
32. Edskes, H. K., Naglieri, B. M., and Wickner, R. B. (2006) Nitrogen source and the retrograde signalling pathway affect detection, not generation, of the [URE3] prion, *Yeast* 23, 833-840.
33. Bousset, L., Belrhali, H., Janin, J., Melki, R., and Morera, S. (2001) Structure of the globular region of the prion protein Ure2 from the yeast *Saccharomyces cerevisiae*, *Structure* 9, 39-46.
34. Aigle, M., and Lacroute, F. (1975) Genetic Aspects of [Ure3], a Non-Mitochondrial, Cytoplasmically Inherited Mutation in Yeast, *Mol Gen Genet* 136, 327-335.
35. Ripaud, L., Maillet, L., and Cullin, C. (2003) The mechanisms of [URE3] prion elimination demonstrate that large aggregates of Ure2p are dead-end products, *Embo J* 22, 5251-5259.
36. Redeker, V., Halgand, F., Le Caer, J.-P., Bousset, L., Laprevote, O., and Melki, R. (2007) Hydrogen/Deuterium Exchange Mass Spectrometric Analysis of Conformational Changes Accompanying the Assembly of the Yeast Prion Ure2p into Protein Fibrils, *J Mol Biol* 369, 1113-1125.
37. Huang, Z. N., and Wheeler, R. A. Replica Microcanonical Search Methods for Mapping Potential Energy Landscapes and Conformations of miniproteins, *In preparation*.
38. Huang, Z. (2005) Design of new molecular dynamics global minimum search protocols for mapping energy landscapes and conformations of folded polypeptides and mini-proteins, in *Chemistry and Biochemistry*, p 282, The University of Oklahoma, Norman.
39. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular Simulations, *Biopolymers* 56, 275-291.
40. Feig, M., Karanicolas, J., and Brooks, C. L. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology, *J Mol Graph Model* 22, 377-395.
41. Hinsen, K. (2000) The molecular modeling toolkit: A new approach to molecular simulations, *J Comput Chem* 21, 79-85.
42. Andersen, H. C. (1983) Rattle: a Velocity Version of the Shake Algorithm for Molecular Dynamics Calculations, *J Comput Phys* 52, 24-34.
43. Okur, A., Strockbine, B., Hornak, V., and Simmerling, C. (2003) Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins, *J Comput Chem* 24, 21-31.
44. Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992) Nosé-Hoover Chains: the Canonical Ensemble via Continuous Dynamics, *J Chem Phys* 97, 2635-2643.
45. Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A Computer-Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *J Chem Phys* 76, 637-649.
46. Case, D. A., Darden, T. A., Cheatham, I., T.E. , Simmerling, C. L., Wang, J., Duke, R. R. E., Merz, L., K.M. , Wang, B., Pearlman, D. A., Crowley, M.,

- Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) AMBER 8, 8 ed., University of California, San Francisco, San Francisco.
47. Lloyd, S. P. (1982) Least-Squares Quantization in Pcm, *Ieee T Inform Theory* 28, 129-137.
 48. Doye, J. P. K., and Wales, D. J. (1996) On potential energy surfaces and relaxation to the global minimum, *J Chem Phys* 105, 8428-8445.
 49. Wales, D. J., and Scheraga, H. A. (1999) Review: Chemistry - Global optimization of clusters, crystals, and biomolecules, *Science* 285, 1368-1372.
 50. Carr, J. M., Trygubenko, S. A., and Wales, D. J. (2005) Finding pathways between distant local minima, *J Chem Phys* 122.
 51. Evans, D. A., and Wales, D. J. (2004) Folding of the GB1 hairpin peptide from discrete path sampling, *J Chem Phys* 121, 1080-1090.
 52. Wales, D. J., and Bogdan, T. V. (2006) Potential energy and free energy landscapes, *J Phys Chem B* 110, 20765-20776.
 53. Komatsuzaki, T., Hoshino, K., Matsunaga, Y., Rylance, G. J., Johnston, R. L., and Wales, D. J. (2005) How many dimensions are required to approximate the potential energy landscape of a model protein?, *J Chem Phys* 122, -.
 54. Wales, D. J. (2005) The energy landscape as a unifying theme in molecular science, *Philos T Roy Soc A* 363, 357-375.
 55. Fernandez-Bellot, E., and Cullin, C. (2001) The protein-only theory and the yeast *Saccharomyces cerevisiae*: the prions and the propagons, *Cell Mol Life Sci* 58, 1857-1878.
 56. Pieri, L., Bucciantini, M., Nosi, D., Formigli, L., Savistchenko, J., Melki, R., and Stefani, M. (2006) The yeast prion Ure2p native-like assemblies are toxic to mammalian cells regardless of their aggregation state, *J Biol Chem* 281, 15337-15344.
 57. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins* 23, 566-579.

7. Conclusions

7.1. Closing thoughts

The previous five chapters stepped through the testing process of both the disrupted velocity (DIVE) and divergent path (DIP) protocols and culminated with the prediction of several conformations for Ure2p(1–64). The research has shown that both protocols are viable for structural predictions. The initial work on an α -helical conformation verifies that the two protocols are capable of locating the α -helical secondary structure. Simulating two β hairpins further extended the testing process. Both simulations reproduce the native conformation. The β hairpin of tryptophan zipper is proposed to correspond to the global potential energy (PE) minimum, but the β hairpin of the B1 domain(41–56) of protein G does not correspond to the global PE minimum. The final test involves a more complex system whose native structure is composed of both α -helical and β -hairpin secondary structures and tertiary structure. Simulations maintained the secondary structures but the tertiary structure is lost. Additionally, the global PE minimum does not correspond to any native-like conformation.

We recognize that the native conformation of a polypeptide or protein should correspond to the global free energy minimum and may not necessarily reside in the global PE minimum. We propose that polypeptides and proteins have a PE surface corresponding to the ‘weeping willow’ model (1-5) (Figure 7.1). The PE surface has a distinct global minimum with several local minima, but the local minima are surrounded by high PE barriers, which prevent conformations from converging to the corresponding global PE minimum.

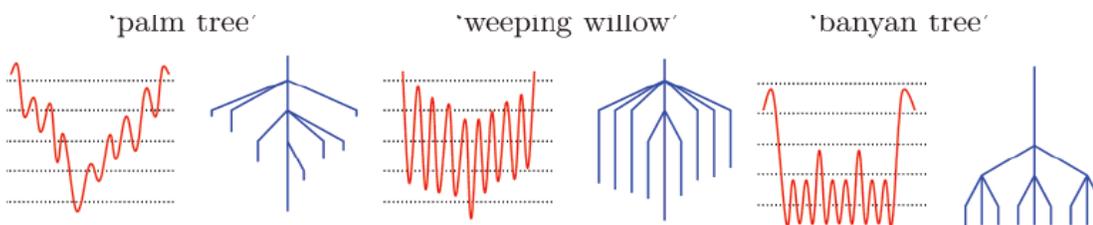


Figure 7.1 Disconnectivity graph (right) representing the PE landscape (left). The graphs are drawn as PE (vertical axis) relative to arbitrary units (horizontal axis). The endpoints of the disconnectivity graphs represent PE minima, and the points where the branches are joined correspond to a common PE “superbasin”. For the PE landscape, the wells represent the minima of a system surrounded by PE barriers. Reprinted with permission from ref. (5). Copyright 2006 American Chemical Society

Finally, we predict several possible conformations for Ure2p(1–64). The previous work indicates that both helices and hairpins can be found during a simulation and that hairpins can certainly be maintained at 300 ± 20 K. Because the tertiary structure of $\beta\beta\alpha 5$ was not maintained, we can only offer a possible secondary structural composition of Ure2p(1–64). We propose that Ure2p(1–64) is composed mostly of α helices. We anticipate that one of our structures may, in fact, correlate to the native structure, but until the structure of Ure2p(1–64) is determined experimentally, we will not know.

Our work highlights three main points. First, DIVE and DIP remain promising protocols for generating low PE secondary structures of peptides, but not tertiary structures. Second, our work implies that the PE of experimental structures may not necessarily correspond to the global PE minimum. Finally, we suggest that sampling is no longer a major issue; instead, the force field needs to be rethought since secondary structures can be reproduced and maintained but tertiary structure cannot be reproduced by the protocols tested here.

7.2. References

1. Mortenson, P. N., and Wales, D. J. (2001) Energy landscapes, global optimization and dynamics of the polyalanine Ac(ala)(8)NHMe, *J Chem Phys* 114, 6443-6454.
2. Evans, D. A., and Wales, D. J. (2004) Folding of the GB1 hairpin peptide from discrete path sampling, *J Chem Phys* 121, 1080-1090.
3. Carr, J. M., Trygubenko, S. A., and Wales, D. J. (2005) Finding pathways between distant local minima, *J Chem Phys* 122.
4. Wales, D. J. (2005) The energy landscape as a unifying theme in molecular science, *Philos T Roy Soc A* 363, 357-375.
5. Wales, D. J., and Bogdan, T. V. (2006) Potential energy and free energy landscapes, *J Phys Chem B* 110, 20765-20776.