# Minimal Models of Biological Structures

by

Josh Paul Kemp

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2001

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

# Abstract

This thesis focuses on the use of minimal models to study biological molecules such as proteins. The major aim of this work was the design and study of a new minimal model that could be used to study the statistical properties of the anisotropic interactions in helical forming segments. We discuss in detail what effect the potential energy form has on these various states by systematically varying the potential from strongly anisotropic to isotropic. The data demonstrates that the foldability of a helix is strongly related to the anisotropic nature of the potential.

The model is further modified to examine the folding process of these segments with emphasis on how nucleation and anisotropy affect folding. The folding times of different helical systems are examined. The results demonstrate that the foldability of the helix segments are strongly correlated to the interplay between nucleation and propagation. This interplay not only affects the foldability of the structure, but significantly affects the scaling behavior of the folding times.

Finally, a classic minimal lattice model is implemented to study the folding properties of prion-like sequences. With this simplified model we attempt to find sequences that exhibit prion-like behavior. The sequences that are identified as prion-like are further studied by analyzing the native state structures for the features that give these sequences their unique properties.

# Acknowledgements

I would like to thank my superivor Dr. Jeff Chen for all of his support and valuable direction throughout my Ph.D. I would also like to thank my thesis committee of Dr. Bernie Nickel, Dr. Don Sullivan, and Dr. Ming Li, who have engaged me in insightful conversations on the helix models in this work. Further, I would like to recognize the support of my collegues in the Condensed Matter Theory Group. In particular, scientific disucussion with Dr. Byron den Hertog and Dr. Michel Gingras have been an invaluable resource.

I would also like to acknowledge the contribution of Howard Swerdfeger to the prion project in chapter 5. Howard was responsible for implementing the 27-mer model, and for determining the foldability and mean first passage times of the sequences studied. (His work on this project will also constitute part of his course work for course 437).

Finally, I would like to thank the Natural Science and Engineering Research Council of Canada for the finanical support they have provided over the duration of my Ph.D.

# Contents

# List of Tables

# List of Figures

# Introduction

The human genome project has produced a tremendous amount of information about the genetic sequences found in the human body. Although this data tells us all the sequences of possible proteins, it tells us nothing about the three dimensional (3D) structure of an individual protein. Determining the 3D-structure of a single protein can be an extremely difficult problem, and some protein structures cannot be determined with modern day techniques. For this reason, understanding proteins presents one of the most challenging problems in the 21st century.

Although biologists have made significant advances in understanding individual proteins, there were no concentrated efforts to understand the general properties of proteins as a whole until recently. In general, it appears that a protein sequence is just an atypical random polymer chain, which for some unknown reason repeatedly seeks out the same structure each time it is folded from any random configuration. At first glance, there appear to be no similarities among the sequences of proteins that fold to unique, biologically useful structures.

Perhaps the most significant breakthroughs in understanding the atypical properties of proteins have been made by those researchers who have taken the simplest

yet most direct approach to the problem. As is the approach with many problems that involve systems too complex to understand, it is advantageous to construct simple idealist models that examine only a few important features of the system. The only disadvantage to this method is that one must be careful of the conclusions drawn from the model, otherwise they can shed light on problems of enormous complexity. The models that embody this approach are often referred to as "Minimal Models", as these models only contain a limited number of adjustable parameters to simplify the system as much as possible.

In the protein problem, the minimal models which present the simplest view of the system are those which come from polymer physics. Proteins and many other biological structures can be represented at the most basic level by the models used to describe generic polymers. For example, proteins are linear random heteropolymers, while more complicated structures such as transfer-RNA can be represented by a randomly branched polymer structure. Thus, for polymer physicists, the complicated systems of the human body are merely specific cases of their simple generic models. Therefore, these systems can be studied by examining appropriate models of molecules of interest.

This is by far the most difficult part of the problem; designing a model that represents the features of the biological system. In the past, numerous minimal models of protein-like systems have been proposed to study various aspects of these complex systems. The earliest and most successful models tackle the problem with the use of lattice polymers, with simple binary potentials[2], or random energy interactions[1]. Although these models have presented significant insight into the

nature of protein systems, the underlying lattice causes key features to be lost that are important to proteins. The most notable of these deficiencies is the lack of secondary structure, such as $\alpha$-helices and $\beta$-sheets, in a lattice model.

In the spirit of the basic minimal models that have been so successful in examining generic protein properties, chapters 3 and 4 of this thesis examine a minimal off-lattice model of helical secondary structures in proteins. The model uses a semi-flexible polymer, often referred to as a worm-like chain, as the backbone structure for mimicking protein segments. A potential is then designed to generate the helical structures that are subsequently studied. Although the model contains none of the specific features of a protein, by examining the effect of adjusting the few parameters in our model we can learn much about the general characteristics of helical structures and the nature of these structures in proteins. In chapter 5, the standard lattices models for examining proteins are used to study a specific class of proteins called prions. These molecules are the exceptions to the rules that characterize the properties of normal proteins. We attempt to understand these proteins by finding a lattice model version of a prion, and characterizing its unique features.

# Chapter 1

# Basic Concepts in Proteins

Proteins are among the most complex molecules in an organism. These molecules are constructed using the genetic information encoded in the cell that is stored within long molecular stands of DNA. Each strand of DNA holds the sequences of thousands of protein molecules in a linear format. Thus, a DNA molecule could be compared to the ingredient list of a recipe; however, the DNA does not contain the information about how to mix it all together. This is an intriguing problem: How does the linear sequence stored in the DNA molecule end up in a 3-dimensional useful structure? The thermodynamic hypothesis that was proposed by Anfinsen[3], suggests that a sequence will obtain the shape that minimizes the free energy of the system. Anfinsen won a Nobel Prize in Chemistry in 1972 for his work in the study of protein folding. There is one catch to this hypothesis. If a sequence of amino acids is constructed randomly, it is not likely to 0find its lowest energy state. This leads to the question, what is special about a protein sequence that allows it to find

its lowest energy state with such remarkable efficiency. This question of whether there is a special pathway by which a protein folds was proposed by Levinthal[4] in 1968. This is referred to as Levinthal's paradox: If a protein has to sample all the possible $\Omega^N$ states available to it to find the lowest energy configuration, then a protein would never fold. Here $\Omega$ represents the number of available configurations per amino acid and $N$ is the number of amino acids. To attempt to understand this paradox, we need to examine the protein system and the factors that influence their properties.

## 1.1 Introduction to Proteins

### 1.1.1 The Protein Sequence

Protein molecules belong to the general class of random heteropolymers. They are considered polymers as they are built from sub-unit molecules called amino acids. This polymer backbone is often referred to as the primary protein structure. There are 20 different types of amino acids from which a protein is built with what appear to be a random ordering of these molecules. All amino acids have the same basic backbone structure that is shown in Fig. 1.1. The symbol 'R' represents a grouping of molecules called the R-group that distinguishes the amino acids from each other. This grouping of molecules is connected to the amino acid at the $\alpha$-carbon. The left side of the $\alpha$-carbon is attached to an amine group, and attached to the right side is a carboxyl group.

To begin to form a protein chain, the backbones of the amino acids are connected

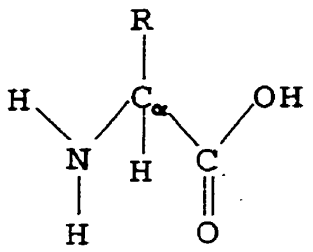


Figure 1.1: The basic structure of an amino acid. The symbol 'R' represents a grouping of molecules that differentiate each amino acid.

via the chemical reaction shown in Fig. 1.2. This figure illustrates the release of a water molecule in the reaction, while a nitrogen carbon bond is formed. This process is continued to form the entire protein sequence. A protein sequence is always labeled with the free amine group to the left and the free carboxyl group to the right. The convention allows for the description of the degrees of freedom of the polymer backbone of the protein. We first note that the double covalent bond associated with the oxygen of the carboxyl group is not stationary and is transient between the oxygen and nitrogen leading to two effective double bonds. The effect of the double covalent bond is the removal of any rotational degrees of freedom, which creates a planar geometry in the amino acid backbone as shown in Fig. 1.3. This leaves only two separate degrees of freedom associated with the backbone bonds of the $\alpha$-carbon, which are referred to as the Ramachandran angles $\phi$ and $\psi$. In addition to the rigidity introduced by the planar structure of the backbone, the R-groups on the amino acid can interact with the backbone. This further reduces the motion by restricting the allowed rotations, creating a rigid protein backbone.

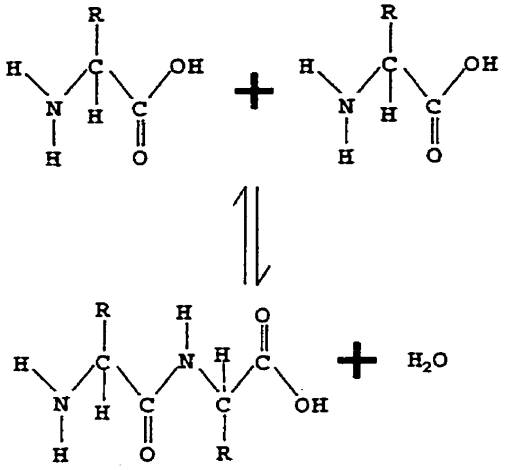

Figure 1.2: Reaction demonstrating the formation of a protein sequence. Water is released as a byproduct of the reaction.

### 1.1.2 Forces within a Protein

Proteins are complex molecules with many interactions between different particles playing important roles in determining the structure. There are the complex particle-particle interactions and the particle-solvent interactions. The later of the two is regarded as the dominant factor in creating the protein structure; however, it is not the only force required.

The most important particle-solvent interaction in the folding of a protein to its three-dimensional structure is the hydrophobicity. This is the preference of the molecule for water as a solvent. For example, a very hydrophobic substance is oil, while any substance that readily dissolves in water is hydrophilic. These materials are often referred to as non-polar and polar respectively. The hydrophobic forces within a protein are not enthalpic, but are generated from an increase in the entropy

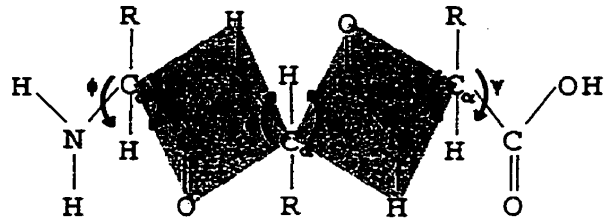

Figure 1.3: Protein chain showing the degrees of freedom about the $\alpha$-carbon. The double bond to the oxygen on the carboxyl carbon becomes transient and oscillates to from a double bond between the nitrogen and carbon (dashed bond). This prohibits rotation about the nitrogen carbon bond, which fixes these atoms in a planar structure.

of the solvent protein system as a whole. When non-polar molecules are placed in water, an ordered cage of water molecules is created around them as water molecules hydrogen bond to encapsulate the non-polar groups. This ordered cage is referred to as a clathrate. When a protein is in an extended conformation, a large clathrate is needed in order to enclose the non-polar residues. In a collapsed conformation, the exposed surface area is much less, thus reducing the number of water molecules that need to be ordered to encapsulate the structure. This reduction in the needed surface area increases the entropy of the entire system making the collapsed state preferred.

All other forces within a protein are enthalpic in nature and can be considered short-ranged forces. These forces can be further subdivided into two groups, isotropic and anisotropic. Dealing first with the isotropic interactions, there are

Van der Waals and ionic bonds, which have the following respective potentials,

$$U_{\text{Ionic}}(r) = \frac{-q_1 q_2}{2\pi\epsilon r} \tag{1.1}$$

$$U_{\text{VDW}}(r) = \frac{a_v}{r^{12}} - \frac{b_v}{r^6} \tag{1.2}$$

where $a$ and $b$ are constants specific to the molecular interaction, $r$ is the distance between particles, $q$ is the particle's charge, and $\epsilon$ is the dielectric constant. The ionic potential is considered short-ranged as the solvent screens the potential over a long range. In addition, the second group of forces which are of an anisotropic nature are those of dipoles and hydrogen bonds. The respective potential forms for these interactions are,

$$U_{\text{Dipole}}(\vec{r}) = \frac{1}{\epsilon}\left[\frac{\vec{\mu}_1 \cdot \vec{\mu}_2}{r^3} - \frac{3(\vec{\mu}_1 \cdot \vec{r})(\vec{\mu}_2 \cdot \vec{r})}{r^5}\right] \tag{1.3}$$

$$U_{\text{Hydrogen}}(r, \theta) = \left[\frac{a_h}{r^{12}} - \frac{b_h}{r^{10}}\right] e^{-\left(\frac{\theta}{\sigma}\right)^2} \tag{1.4}$$

where $\vec{\mu}$ is the dipole moment of the atom, $\theta$ is the angle of alignment between the dipoles of the bonds, and $\sigma$ is a constant related to the angle $\theta$. These interactions, particularly hydrogen bonding, are responsible for the secondary structure formation and stability observed in a protein. It is these forces which are of most interest in this work.

### 1.1.3 Structure of Proteins

The forces at play within a protein yield to complex behavior in terms of the structures formed. Perhaps the most remarkable feature of protein systems is that the same structures are repeatedly observed throughout a variety of different protein types. The three-dimensional structures observed can be broken into two categories, secondary and tertiary.

Secondary structures are the first set of sub-structures that are formed during folding. There are three types of secondary structures: 1) helix, 2) $\beta$-sheet, and 3) random coil. The helix is the most abundant secondary structure observed, and is formed by the hydrogen bonding of amino acids near each other along the chain. Due to atomic constraints on the allowed $\phi$ and $\psi$ angles in most sequence combinations, helices tend to coil with a right-handed preference. The beta sheet is another commonly observed structure, where segments of the protein arrange themselves in extended conformation and bond into a sheet-like structure. Unlike the helices, $\beta$-sheets are formed from hydrogen bonding of amino acids separated by many other amino acids along the chain. The random coil type of secondary structure is the grouping of all remaining configurations. Although random coils do not have any short ranged order, the structures they create usually have a specific role and cannot be replaced by any random sequence arrangement. The randomness in the name applies only to the lack of definite structure, and does not mean that any set of amino acid can replace a random coil segment. This is evident as the random coil segments usually form the functional site of the protein, and replacing it would lead to completely different functionality.

Tertiary structures are the next level of complexity within the proteins overall form, and are built out of the secondary formations discussed above. It is these structures which come together to create the complex proteins observed. Some of the common tertiary structures that are observed in a wide range of proteins are the helix and $\beta$-sheet barrels. There are numerous tertiary structures ranging from simple to complex, and they can contain an active site for biological processes or provide structural support within a membrane.

## 1.2 Application of Minimal Models

As has been demonstrated, a single protein is a very complicated system, making it nearly impossible to understand completely. The problem is further compounded by the fact that there appears to be little correlation between different protein molecules. Yet, there are similar features on another level. The key features of all proteins are: 1) they have a single three-dimensional structural form called the native state, 2) the protein can fold to the native state repeatedly, and 3) the time scale of folding is exceptionally short[1].

There have been numerous models developed to study proteins from the very simple to the extremely complex. We discuss a variety of models that demonstrate some of the key features mentioned above and we also discuss their limitations. We begin with the minimal models, those that model a protein on a lattice, and then move to the more complicated models that attempt to account for all of the possible forces.

### 1.2.1 HP Model

Perhaps the simplest model of those that attempt to model a protein's predominant features is the HP model. This model is considered to be the Ising model of protein systems. The model was first introduced by Lau and Dill[2] in 1989, and makes use of the hydrophobicity believed to dominate force the folding process.

The model is constructed on a lattice to limit the total number of possible configurations. Each amino acid is represented by a spherical monomer and is designated as hydrophobic (H), or hydrophilic (P). The interaction potential is as follows,

$$U_{HP} = \begin{cases} -1 & for\ H-H \\ 0 & for\ P-P \\ -\gamma & for\ H-P \end{cases} \tag{1.5}$$

where $\gamma$ is a variable constant. The hydrophobic interactions (H-H) are attractive in order to mimic the repulsive nature of the solvent monomer interactions that cause these monomers to group together. The interaction is normalized to one to simplify the potential. The hydrophilic interactions (P-P) are set to 0 to simulate the molecule's preference to be dissolved in the solvent. Cross interactions (H-P) have an interaction strength somewhere between these two extremes, and thus, $\gamma$ is set between 0 and 1, and can be varied to study different properties.

Although this model is extremely simplistic, it is very powerful and captures several key features of protein molecules. A study by Li et al.[5] showed that the HP Model has remarkable correlation to a real protein system, by analyzing a matrix of protein interactions. The matrix constructed by Miyazawa and Jernigan[6] con-

tains experimental information approximating the interaction potentials between different amino acids. Li et al. showed that the strongest interaction is that of the hydrophobicity, while the next most dominant force is one of demixing. Demixing is explained in the following way. If in the HP model the interaction parameter $\gamma$ is zero, then it is clear that all the hydrophobic molecules will attempt to group together and will form a central core. The hydrophilic molecules will then surround the central core like a shell. The demixing term, which is characterized by the parameter $\gamma$, allows the hydrophobic molecules to move to the outer shell and the hydrophilic molecules to move to the core, while still resulting in a lower energy. Thus, setting the parameter $\gamma$ to zero is not a physically acceptable condition. Therfore, the HP Model in essence contains two very important features in a very simple model.

This model has been successful in examining some interesting properties of proteins. In another work using this model, Li et al.[7] attempted to show that secondary structure results as a consequence of a sequence having the ability to fold quickly to its native state. More importantly, this model has been used extensively in the inverse protein folding problem, which is the problem of finding the sequence that designs a specific structure[8, 9]. Although this model is successful, it is limited in its scope of questions it can answer. For example, it lacks diversity in sequence energies, as sequences that have degenerate native states frequently occur. These sequences are typically ignored in most studies. In addition, secondary structures are difficult to define on a lattice, which makes it difficult to present conclusive results on this topic[10].

### 1.2.2 27-mer Model

A more advanced variant of the HP model is the 27-mer model. This model was introduced by Šali et al.[1] in 1994. The model introduces an aspect of proteins that the HP model disregards. The HP model only divides the interactions into two types, while in reality there are twenty different amino acids that can have their interactions influenced by their neighbors along the backbone and in the surrounding area. This produces an extensive variety of interaction strengths, which are featured through random energy interactions in the 27-mer model, and it is the only addition to the basic HP model.

The 27-mer model is designed in the same way as the HP model, as it is a lattice model. The difference lies in the choice of interaction energies. The interaction between each set of monomers is chosen randomly from a Gaussian distribution to generate a specific realization of a sequence. The distribution of interaction energies chosen has a standard deviation of 1.0 and is centered about a value $B_0$, which is usually negative. The added advantage of this model is that it will almost always produce a single lowest energy native state, adding a more realistic feature to the basic HP model.

This model has been one of the most successful minimal models in describing proteins. Work by Šali et al.[1, 11] was the first to demonstrate some of the fundamental properties of protein sequences using this model. If one assembles a protein sequence from a random combination of amino acids, the chances are that the sequence will not fold to any particular native state. By generating a large number of random sequences, Šali et al. showed that a small fraction of the se-

quences could fold to the same compact structure when folded numerous times. In addition, these sequences could reach their native state in a relatively short period of computational time. What set these fast-folding sequences apart from the rest? Through analysis of a representative spectrum of the low lying energy states in the system, it was found that the sequences which folded quickly had a native state energy that was much lower than that of any other low energy states. Thus, it was concluded that in order to obtain a fast folding sequence which folds to a single native state, the energy of the native state must be separated by a significant gap from the continuum of low lying energy states.

The 27-mer model has been used to examine numerous other issues in regard to protein folding. It has been used in evolutionary folding studies which study the importance of a central nucleation site that is conserved in a mutation process[12, 13]. This model has also been used to examine the nature of the folding pathway[14, 15]. Does a protein fold via a two-state process of globular collapse then a structural transition, or via a series of intermediate states? These studies along with experimental evidence[16] have shown that the two-state mechanism is the most likely scenario. The model has been so widely used that it has become a standard model for examining the dynamics of protein systems, and we make use of this model later in chapter 5.

### 1.2.3 Gō Type Models

The Gō type model[17, 18] is yet another minimal model of protein folding which is an extension of the 27-mer model. The model was first introduced by Taketomi

et al. in 1975[17]. In this model, the lattice constraints are removed and the protein molecule is simulated off-lattice. The removal of the very restrictive lattice constraints presents numerous other problems. The additional phase space now available to molecule makes finding a sequence that is a good folder very difficult.

The model is usually constructed of a flexible or semi-flexible polymer chain where the interaction strengths are drawn from the random distribution of energies. Typically the model is used in the following way: a) a structure of interest is selected, b) monomers are given random interaction strengths that are adjusted such that the selected structure is the native state of the molecule. The result is a model of a specific structure that has properties similar to that of a protein. The model is ideal for capturing the characteristics of a specific protein, but the results obtained from one structure are generally only transferable to another in a generic way.

The Gō Model has been used in a number of different studies. For example, Zhou et al.[19] studied the structural transitions in a three-helix bundle in a domain of staphylococcus aureus protein. They were able to demonstrate the complex series of structure phase transitions that the bundle undergoes. The model has also been used by Dokholyan et al. [20, 21] to identify the folding nuclei within protein structures and to study the thermodynamic importance of these contacts .

### 1.2.4 All-Atomic Models

A final class of models which are perhaps conceptually the easiest to construct, but technically the most difficult to implement, are those of all-atomic models. This

class of model attempts to model all the atomic details of a protein. They are not minimal models, but are the most used models for studying protein systems. These models are typically used in conjunction with molecular dynamics techniques and are used to examine the motion of various protein structures.

There are wide ranges of all-atomic models that have been created over the years. Perhaps the most notable simulation packages and potentials are the CHARMM[22], AMBER[23], and KNOF90[24]. These packages have been used in numerous studies that typically involve studying the structural behavior of proteins in which the ground state is already known. These all-atomic models present the best method for understanding the internal dynamic behavior of a structure. For example, how hemoglobin captures and releases oxygen[25].

This class of protein models has drawbacks. The first is that although attempts are made to capture the full behavior of the atoms being simulated, no one is fully certain of the correct form of the potential. Although the potentials are very sophisticated, the simulation results will still depend on the potential form, and thus the results may not be as representative of actual systems as would be hoped. However, there has been some suggestion that the potentials are robust, and that significant changes in the parameters produces minimal change in the calculated results[26]. The second and more serious limitation of this class of model is the resources required to run the simulation. The potentials are complex and can be difficult to enumerate. Also, in order to capture the full nature of the system, solvent atoms must be simulated thereby making the total number of atoms that need to be accounted for range into the thousands. This class of models is thus not

ideal for folding simulations, as the time scales involved are orders of magnitude larger than the typical internal motions within protein structures and require too many resources to be simulated.

### 1.2.5 Helices and Zimm-Bragg Theory

As has been discussed, minimal models have been used extensively to understand proteins. An understanding of proteins also requires knowledge of some of the simpler structural components and their behavior. Helices are the most dominant secondary structures observed in all proteins. For this reason many groups have studied them analytically[27, 28, 29, 30, 31, 32] and with all-atomic models[33, 34, 35, 36], but not many have studied them with minimal models[37]. The most notable of the analytical theories is one that was first presented by Zimm and Bragg[27].

Zimm and Bragg's theory treats the system as a quasi one-dimensional Ising system. In this model, each monomer is assumed to exist in one of two states, coil, 'c', or helix, 'h', and two parameters are used to describe the interaction. A parameter $\Delta f_s$ represents the free energy of a monomer existing in a helical state, and as the transition from the coil-to-helix occurs, this parameter goes to zero. The other parameter $\Delta f_\sigma$ describes the loss of free energy due to nucleation of the new helical segment. In the calculation of the relevant quantities, these parameters are represented through the statistical weights of the various states as $s_z = exp[-\Delta f_s/T]$ and $\sigma_z = exp[-\Delta f_\sigma/T] \sim 10^{-4}$. The values $s_z$ and $\sigma_z$ can be thought of as the probability of propagating a helical region and the probability of

nucleating a helical region, respectively.

Thus, in the simplest of Zimm and Bragg's models, weights are assessed to each segment based only on the preceding monomer in the sequence. It is assumed that the beginning of the chain is a coil, and the first three monomers are always in a coil state. In a calculation, random sequences of 'c's and 'h's are considered, and each sequence combination is given a weight based on the following rules:

1) For each 'c' $\Rightarrow 1$

2) For each 'h' which follows a 'h' $\Rightarrow s_z$

3) For each 'h' which follows a 'c'$\Rightarrow \sigma_z s_z$

Rule one implies that the statistical weights are normalized such that all coil segments are given a weight of unity. The second rule states that each helical segment has a free energy different than that of the coil segment. By adjusting the temperature, this energy can be made greater or less than one. The final rule suggests that there should be a loss of free energy when a helical segment is formed due to a reduction in the entropy of the system. The model can be further complicated to add in features of real helical forming protein segments. Changing a single amino acid usually cannot break helical segments. Thus, in order to add the idea that several amino acids must change to a coil state to break a helical segment, a fourth rule is added.

4) For each 'h' which follows less than $\mu$ 'c's $\Rightarrow 0$

This means that segments of less than $\mu$ 'c's do not occur within a segment. Using these rules a matrix describing the system can be written down and solved for the largest eigenvalue. This value can be related to the properties of interest.

For example, the average helicity can be obtained through the following equation,

$$\Theta = d \ln \lambda_0 / d \ln s_z, \tag{1.6}$$

where $\lambda_0$ is the largest eigenvalue, and $\Theta$ is the average helicity. Zimm and Bragg's result for this value demonstrates the strong cooperative nature of transition to a helical state. Zimm and Bragg also showed that $\sigma$ must be zero in order to obtain a true first order phase transition. Therefore, the coil-to-helix transition is not a true phase transition as $\sigma_z$ cannot equal zero. The change to a helical state would occur as a transition from a coil state to a state of alternating helical and coil region, followed by a transition to a nearly perfect helical state with occasional disorder.

Although the helix-coil transition has been studied analytically by other groups, the conclusions are similar while attempting to improve the accuracy of the calculation. For example, Lifson and Roig[28] have introduced a very similar calculation, but remove the directional dependence of the sequence in the Zimm and Bragg method. In addition, the Zimm and Bragg model has also been used as a standard in experimental studies of helical propensities of amino acids[38]. The model does not have any predictive power because $\sigma$ must be determined from experiment, but it does offer a method of comparison of all-atomic simulations to real protein systems [33].

### 1.2.6 Present Work

The extensive use of minimal models has provided a large amount of knowledge about the fundamental nature of proteins. In the study of more complicated issues such as the formation of secondary and tertiary structures, there have been few minimal models constructed that address questions about them.

For helical secondary structures the number of minimal models is limited. One new minimal model has been introduced by Potthast[37], who has constructed a helical model by adding torsion bonding constraints between adjacent neighbors along a polymer backbone. In chapter 3 of this work, we present a new minimal model of a helix that can be used to examine the statistical properties of helical secondary structures. In this model, adjacent neighbor interactions are ignored in favor of studying anisotropic interactions that are typically seen in proteins. In chapter 4, the model of chapter 3 is modified in order to study the dynamic behavior of helices. This is the only minimal model designed to study helical segments.

In the final chapter, prions, which are unique protein sequences, are studied. These sequences have only recently become an area of study. We use the standard 27-mer model to study these structures at their most fundamental level, as previous studies of these structures using minimal models have been limited[39]. This study attempts to show that even the most simplified models can demonstrate the complex behavior of prions, and to provide a method to study the issues relating to prion dynamics.

# Chapter 2

# Simulation Techniques

Before discussing the models in this thesis, an examination of the technical aspects of the simulations is in order. In this chapter, there will be a focus on the theory behind the Monte Carlo technique used and the algorithms used to create the simulation. In addition, a detailed look is taken at the advanced Monte Carlo technique of using a multicanonical ensemble to conduct a simulation as it is employed in chapter 3.

There are numerous techniques to study different aspects of protein systems. Largely, most studies are carried out using molecular dynamics techniques, and to a lesser extent, Monte Carlo techniques. Molecular dynamics techniques focus on the real time motion of particles and attempts to understand the behavior of large molecules under certain conditions. Monte Carlo techniques focus more on examining the statistical properties of these molecules, such as how quantities scale with system size. The Monte Carlo techniques can also be applied to dynamics

models. Despite the fact that the dynamics are no longer directly correlated to real time motion of the particles, the dynamics obtained with this method present useful information about the general behavior of the system.

In simulations of models that involve the inclusion only the most minimal of forces, Monte Carlo simulations seem to dominate due to the decrease in necessary resources. The focus of this research revolves around studies based on Monte Carlo techniques and advanced Monte Carlo techniques, which we will discuss in this section. In addition, some classic polymers models that are also used to study these systems are discussed in detail.

## 2.1 Monte Carlo Method

Monte Carlo has been around for over half a century, but only in the last twenty years has it made its rise to the forefront of scientific research as computers have become more powerful. Monte Carlo is a method of solving complicated integrals based on generating numerous random possible outcomes. For example, the equation

$$I = \int_0^1 dx \exp(-x^2) \ , \tag{2.1}$$

can be solved by generating random numbers between zero and one. The quantity $\exp(-x^2)$ is then simply averaged to obtain the value of the integral. This is perhaps the simplest application of the technique to solve an integral.

### 2.1.1 Metropolis Method

In 1953, Metropolis et al.[40] developed an algorithm to calculate integrals from more complicated systems, such as equations of state. The method, often referred to as importance sampling, chooses random points from a preset distribution based on the previous history of the simulation as to allow the algorithm to sample states in a region that contribute most to the integral. The problem lies in generating a sequence of random states with the proper probability of occurring.

This problem is solved by introducing a Markov process. A Markov process is a sequence of trial states that have two properties: 1) each trial belongs to a finite set of outcomes, and 2) the outcome of each trial depends only on the outcome that preceded it[41]. The transition between any two states is represented by a stochastic matrix, $\pi$, which must satisfy the condition that a limiting distribution can be reached, ie.

$$\rho = \pi\rho \ , \tag{2.2}$$

where $\rho$ is the equilibrium distribution. Thus, any stochastic matrix that obeys Eq. 2.2 and creates the desired distribution could be used in a simulation. A problem arises in determining the matrix that gives the desired distribution. Introducing the condition of detailed balance, which is an unnecessarily strong condition[41], solves this problem. Detailed balance is the condition of microscopic reversibility and is represented by the equation,

$$\rho_m \pi_{mn} = \rho_n \pi_{nm} \ . \tag{2.3}$$

Using this condition, different possible matrices can be generated. Metropolis et al.'s[40] choice for the transition matrix is given below; however, it should be noted that this choice is not unique.

$$\begin{aligned} \pi_{mn} &= \alpha & \rho_n \geq \rho_m \\ \pi_{nm} &= \alpha(\rho_n/\rho_m) & \rho_m < \rho_m \end{aligned} \tag{2.4}$$

where $m \neq n$, and $\alpha$ is a constant usually set to one. In a simulation of a canonical ensemble, the probabilities, $\rho_n$, are the Boltzmann weight functions,

$$\rho_n = e^{\frac{-E_n}{k_B T}} , \tag{2.5}$$

where $E_n$ is the energy of the $n^{th}$ configuration, and $k_B$ is the Boltzmann constant.

The final problem in the metropolis algorithm is generating all the possible trial states. There have been hundreds of methods developed to generate the different states of various systems. The one requirement for these methods is that it must obey ergocity. In other words, the algorithm must be able to generate all possible states, from any initial starting configuration, within a finite amount of time. In polymer physics, there are numerous possible ways to generate the available states; however, we will only discuss three algorithms relevant to this thesis.

### 2.1.2 Non-local Markov Process: Pivot Method

The pivot algorithm is an excellent method for generating the possible configurations of a polymer chain. The method was introduced by Madras and Sokal in

1988[42], and has been used in a variety of studies. In the model, a large cluster of monomers is moved together producing a configuration significantly different than the previous step, and thus, they are not locally related on the energy landscape. This process can produce configurations that are uncorrelated structurally with relatively little computational effort. This makes the technique ideal for studying the conformational properties of polymer systems.

The basic pivot algorithm is described as follows for a polymer chain of $N$ bonds. A bond, $i$, is selected at random from the $N$ available. The remaining $N-i$ bonds are then rotated together using an Euler rotation matrix which is generated randomly such that a uniform distribution is attained in spherical coordinates. The new configuration is the trial state. The energy of the trial state is then numerated and the configuration is accepted or rejected based upon the Metropolis criterion discussed above. These steps can be repeated to generate the entire spectrum of configurations with excellent efficiency while maintaining a constant bond length, which is ideal for many systems where the typical interactions are much weaker than the covalent bonds connecting the monomers.

This algorithm can also be modified to simulate systems that include more constraints. For example, in a simulation where the azimuthal angle has been fixed, a single move is just slightly more complicated. In this case, the bond around which the monomers are rotated must first be rotated to lie along the z-axis using an Euler rotation. A second rotation in which the polar angle is changed is then carried out. This is followed by an inverse rotation of the original rotation to restore the chain so that it lies along its original axis. This ensures that a fixed azimuthal bond angle

is maintained in addition to the fixed bond length.

These methods are particularly effective in studying the statistical properties of polymers in general, and are perhaps the most efficient in cases such as the fixed azimuthal bond angle. Although this algorithm is effective at sampling states at high temperatures, depending on the type of system, it can be very ineffective at sampling the low temperature region where smaller local moves are favored by the energy constraints. For example, sampling the collapsed states of a polymer system with a typical Monte Carlo pivot may not be possible as large movements of a large number of molecules would probably be rejected, especially when the system size is large and the probability of overlapping monomers is high.

### 2.1.3 Local Markov Process: Dynamic Motion of a Polymer

Another class of algorithms which can produce the spectrum of states of the system are those that use only local moves. This is perhaps one of the oldest methods for generating the states of a system. The method differs from those of the non-local class as only one monomer is being moved at a time, which takes considerably longer to generate structural uncorrelated structures; however, it adds the advantage that it simulates the dynamic behavior of the molecule.

The type of local move algorithm used depends greatly on the constraints of the system. If a monomer is not restricted in its movement by such things as a constant bond length, then the algorithm is very simple. In this case, a move consists of randomly selecting one of the $N+1$ monomers, $i$, and displacing it by a small amount to obtain the new trial position. The trial position is then accepted

based on the Metropolis criterion. If the common constraint of a fixed bond length is added, then the approach differs slightly. The trial state is generated by creating an axis between the $i-1$ and $i+1$ monomers and then rotating the $i^{\text{th}}$ monomer about this axis. This maintains the constant bond length constraint. More complicated algorithms can be derived to handle constraints such as maintaining a fixed bond angle, which is handled with the SHAKE algorithm[43].

### 2.1.4 Lattice Markov Process: 3D Cubic Lattice

The final algorithm of interest is one that is used to generate configurations for a polymer chain when it is confined to a lattice. The algorithm is also designed to use local moves so that the Markov process also mimics the dynamic motion of the system. Generating the motion of a polymer chain on a lattice through a series of local moves can be tricky, as some lattices can sometimes introduce problems with ergodicity.

It has been shown that there are four types of moves that need to be included in order to simulate the dynamic motion of a polymer correctly. These moves are depicted in Fig. 2.1. Fig. 2.1a is called a crankshaft move and is a double monomer move that must be included with the other three single monomer moves in order to obtain the correct behavior[44]. To conduct a simulation, the procedure of Šali et al.[11] is used to mimic the folding of a protein. At the start of this procedure it is randomly decided whether a single or double monomer move will be performed, with respective probabilities of 20% and 80%. In a double move, an adjacent monomer pair is selected and the configuration checked to determine if a crankshaft move

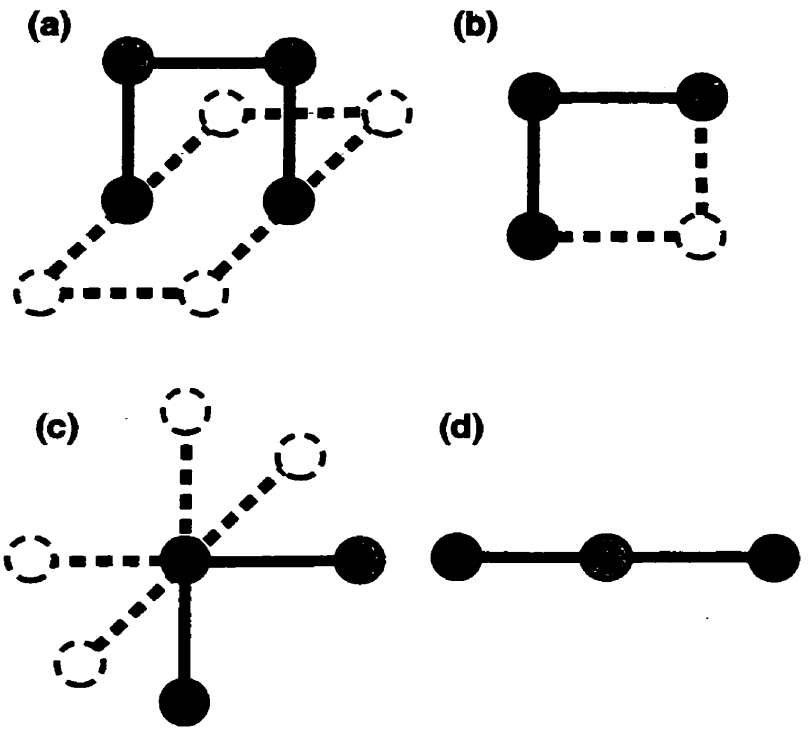


Figure 2.1: The four required moves for a lattice, a) crankshaft, b) corner flip, c) end move, and d) no move.

can be performed. If it cannot the move is rejected and the selection of a single and double monomer move is repeated. In the case of a single monomer move a random monomer is selected and the configuration is checked to determine if one of the three single monomer moves can be performed. Once a double or single monomer move is applied, the energy of the new trial configuration is calculated and the move is further rejected or accepted based on the Metropolis criterion. This procedure is a common method for simulating the lattice dynamics of the 27-mer model discussed in chapter 1.

## 2.2 Multicanonical Technique

The key assumption that is made in a Monte Carlo simulation is that the system is able to sample the representative proportion of states that contribute most to the integral. The Markov sequence generated during the simulation should sample these states. This is not always the case. In solid state systems, temperature plays a key role in the effectiveness of an algorithm to sample the available states. At high temperatures, there are usually no obstacles to prevent the movement between states if the algorithm is ergodic. At low temperatures, energy barriers can play a significant role in reducing the effectiveness of a Metropolis algorithm, as the energy barriers artificially separate regions of configurational space.

In the low temperature regime, a polymer molecule has a high probability of becoming trapped in an energy minimum and not sampling the entire conformational space available to it within a finite amount of computational time. Overcoming this problem has been the focus of numerous simulation-methodology studies for many years. Perhaps the most effective solution to this problem to date is the set of algorithms that stem from the umbrella sampling technique[45]. The basic principle is to find a more effective statistical weight than the conventional Boltzmann weight used in a typical Monte Carlo simulation. The object is to bias the weights in such a way that movement through the low temperature states is favored so that the polymer does not become trapped. This is done by guessing a weight function based on such things as the density of states or the lowest energies of the system. A few of the more recent algorithms of this class include simulated tempering[46],

multicanonical annealing[47], and the multicanonical technique[48, 49, 50], all of which have their particular advantages and disadvantages and are well suited for the systems to which they are applied.

In chapter 3, we use the multicanonical technique to study the statistical properties of our helix model. In the multicanonical procedure, we re-weight the temperature of the system to produce a histogram that is relatively flat over all energies in the model. In other words, the simulation visits each energy state of the system an equal number of times during the production run. This is demonstrated by examining Fig. 2.2 which illustrates two histograms. Fig. 2.2a shows a typical histogram of the probability of visiting a specific energy state under a Boltzmann distribution, while Fig. 2.2b shows a desired histogram of the probability of visiting a specific energy state with our new distribution functions. This is done by making temperature a function of energy, which removes the specific temperature from the Boltzmann weight function and essentially creates a simulation that is performed at all temperatures at the same time. The procedure for creating this new weight function is outlined below.

In this method, we are really interested in modifying the Boltzmann probability of the energy states (Fig. 2.2a), which is given by,

$$P(E) = \frac{n(E)e^{-\beta E}}{Z} \tag{2.6}$$

where $n(E)$ is the density of energy states, $Z$ is the partition function, and $\beta = 1/k_B T$. Of course, we are not looking for this distribution as we need a new dis-

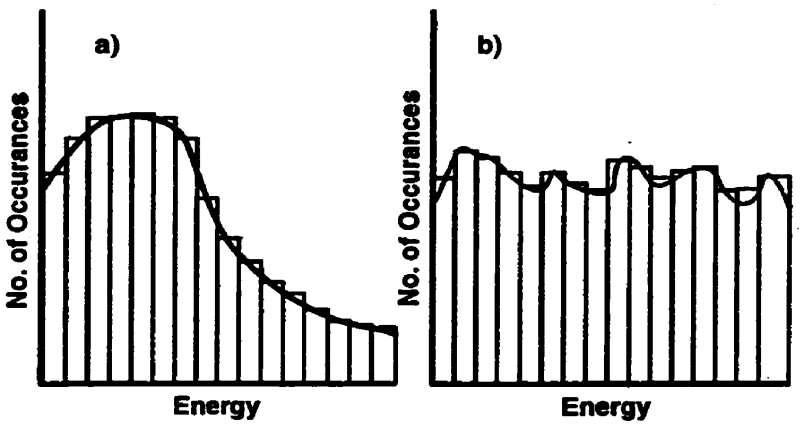


Figure 2.2: a) A representation of a histogram of a typical Boltzmann weighted simulation. b) A respesentation of the same histogram preformed with a multicanonical weight. The histogram is approximately uniform.

tribution that is constant (Fig. 2.2b) over all energies. This distribution is created by writing the inverse temperatures as a function of the energy and by adding a function $\alpha$ to the Boltzmann distribution in the following way,

$$P'(E) = constant = \frac{n(E)e^{-\beta'(E)E-\alpha'(E)}}{Z'} \tag{2.7}$$

where $\beta'(E)$ and $\alpha'(E)$ are functions of energy. In order to derive a method by which $P'(E)$ can be determined for the specific density of states of our system, it is written as a function of $P(E)$ by examining the ratio $P'(E)/P(E)$. This leads to the form [51],

$$P'(E) = \frac{P(E)e^{-(\beta'(E)-\beta(E))E-(\alpha'(E)-\alpha(E))}}{\tilde{Z}} \tag{2.8}$$

$$\tilde{Z} = \sum P(E)e^{-(\beta'(E)-\beta(E))E-(\alpha'(E)-\alpha(E))} \tag{2.9}$$

where $\tilde{Z}$ is the new partition function. The functions $\beta(E)$ and $\alpha(E)$ are chosen to be those for the Boltzmann weight function, where $\beta(E)$ is a constant function of energy, $\beta(E) = \beta$, and $\alpha(E)$ is zero for all energies. Thus, if $P(E)$ is the known Boltzmann probability distribution, a probability distribution $P'(E)$ that is flat can be created; however, $P(E)$ is not known *a priori* because the density of states is not known for the system. If the density of states was known, all quantities could be calculated directly and the problem would be trivial. We settle for a rough estimate for $P(E)$ and reply on a recursive method for determining the proper forms of the functions $\beta'(E)$ and $\alpha'(E)$. This recursion relation can be constructed by grouping the energies of the system into bins as demonstrated in Fig. 2.2. If we create a histogram to estimate $P(E)$, and introduce the condition that the new histogram, $P'(E)$, has adjacent bins of equal magnitude, the new functions of $\alpha'$ and $\beta'$ can be calculated. This condition implies,

$$P'(E_{i+1}) - P'(E_i) = 0 \ , \tag{2.10}$$

where $E_i$ are the energies of the $i^{th}$ bin. This yields the necessary equations to generate the $\alpha'$ and $\beta'$ functions needed.

Solving Eq. 2.10 using Eq. 2.8 generates the method for obtaining P'(E) derived by Hansmann & Okamoto[33].

1)We first perform a simulation under a simple canonical distribution, at a temperature where the system is well above any transition. The energy scale is discretized across the energy range of interest.

2)We estimate the probability density by a histogram, $P(E)$. For all P(E) that are above an arbitrary cutoff, generate a function S(E) with,

$$S(E_i) = \ln(P(E_i)) \ . \tag{2.11}$$

3) From the function S(E) we generate the new multicanonical parameters $\alpha'(E)$ and $\beta'(E)$, according to the following equations:

$$\beta'(E_i) = \begin{cases} \beta_0 & E_i \geq E_{\max} \\ \beta_0 + \frac{S(E_{i+1}) - S(E_i)}{E_{i+1} - E_i} & E_{\min} \leq E_i < E_{i+1} < E_{\max} \\ \beta'(E_{\min}) & E_i < E_{\min} \end{cases} \tag{2.12}$$

and,

$$\alpha'(E_i) = \begin{cases} 0 & E_i \geq E_{\max} \\ \alpha'(E_{i+1}) + [\beta'(E_{i+1}) - \beta'(E_i)]E_{i+1} & E_i < E_{\max} \end{cases} \tag{2.13}$$

Here $E_{\max}$ and $E_{\min}$ respectively signify the upper and lower bounds of the energy range which is to be re-weighted to a flat histogram. $E_{\max}$ is arbitrarily set to a convenient value which is typically the energy at which the maximum in $P(E)$ occurs. This is chosen as all energies above this would be properly sampled by a normal Monte Carlo simulation conducted at a temperature $\beta_0$. $E_{\min}$ is usually chosen to be the ground state energy of the system, but this is not a requirement.

4) Begin a new simulation with the Metropolis weighting function of

$$w'(E) = e^{-\beta'(E)E-\alpha'(E)} . \tag{2.14}$$

There is a significant problem in this method. If $P'(E)$ could be accurately determined using this method, the simulation could have been performed using $P(E)$. Normally, $P(E)$ is not determined accurately enough in a single attempt to yield the desired histogram; thus, a recursion method is used to increase the accuracy of the histogram. If the weight function from step 4, $w'(E)$, is used to conduct a new simulation, the same four steps can be used to construct a better weight function w"(E); however, $\beta''$ and $\alpha''$ will be a function of $P'(E)$, $\beta'$, and $\alpha'$.

$$\beta''(E_i) = \beta'(E_i) + \frac{S'(E_{i+1}) - S'(E_i)}{E_{i+1} - E_i} \tag{2.15}$$

$$\alpha''(E_i) = \alpha''(E_{i+1}) - [\alpha'(E_{i+1}) - \alpha'(E_i)] \tag{2.16}$$

$$+[\beta''(E_{i+1}) - \beta'(E_{i+1}) - \beta''(E_i) - \beta'(E_i)]E_{i+1} \tag{2.17}$$

From Eq. 2.12 and Eq. 2.13 $\beta''$ and $\alpha''$ can be written in terms of the original Boltzmann variables.

$$\beta''(E_i) = \beta_0 + \frac{S'(E_{i+1}) - S'(E_i) + S(E_{i+1}) - S(E_i)}{E_{i+1} - E_i} \tag{2.18}$$

$$\alpha''(E_i) = \alpha''(E_{i+1}) + [\beta''(E_{i+1}) - \beta''(E_i)]E_{i+1} \tag{2.19}$$

Therefore by keeping a running tally of $S(E)$ such that Eq. 2.11 is now

$$S(E) \Leftarrow S(E) + \ln(H(E)) \ , \tag{2.20}$$

where $H(E)$ is the histogram of the current probability distribution, then steps 2 through 4 can be repeated until a sufficiently flat histogram is achieved for all energy bins of interest.

A single production run is then made with the final calculated weight function. The averages collected from this simulation run are not weighted according to the Boltzmann integral, but according to the following from,

$$\frac{\int dE \ A \ w'(E)}{\int dE \ w'(E)} \ . \tag{2.21}$$

Therefore, the production run values need to be re-weighted back to the original distribution. The averages under the Boltzmann distribution are calculated from the multicanonical values with the following formula,

$$\langle A \rangle = \frac{\int dE \ A \ w(E)/w'(E)}{\int dE \ w(E)/w'(E)} \ , \tag{2.22}$$

where $w(E)$ is the Boltzmann weight function, and $w'(E)$ is the multicanonical weight function used in the production run. Eq. 2.22 shows how any temperature can be examined, as the Boltzmann temperature is merely an adjustable parameter in $w(E)$. This allows the entire temperature range to be examined in a single production run, as a Boltzmann temperature is not explicitly used in the production

run.

As has been shown above, this method has advantages for the calculation of quantities at low temperatures over regular Monte Carlo techniques. This method does have some limitations. It should be noted that the calculation of the new weighting function could take considerable CPU time, which in the end does not lead to significant advantages over typical Monte Carlo procedures. Also, in systems with glassy low temperature regimes, it can be difficult to obtain convergence towards a uniform distribution when determining the proper weighting function. Despite these disadvantages, the multicanonical technique is well suited for systems with first order phase transitions, as the technique can be used to accurately determine the location of the transitions and properties of the specific heat and susceptibility curves for finite size scaling analysis. In this respect, the multicanonical technique surpasses typical Monte Carlo.

# Chapter 3

# Minimal Helix Model

Helices are the most commonly occurring secondary structures in proteins. For this reason they have been the focus of many studies for the past 40 years, as researchers attempt to understand the intricacies of protein folding. Understanding these secondary structures is a significant advance toward a complete understanding of proteins in general.

In this chapter, a minimal off-lattice model of a helical segment is constructed using an anisotropic potential. The model is examined in detail for the structural states occurring, the scaling behavior of the different structural regions, and the effect the anisotropy of the potential has on the results. The construction and analysis of this model is the main focus of this thesis.

The traditional view of the coil-helix transition has been to treat the system as a quasi one-dimensional Ising system[27, 28]. Using this idea, many mean-field theories describing the nature of the transition have been put forth, with the most

well known being that of Zimm and Bragg[27] which is discussed in Chapter 1. The understanding of these structures is that the system can be treated as a single first order like transition of a coil to a helix, which, due to the one dimensional nature of the problem, is not a true phase transition in the thermodynamic limit. This leads to a maximum helical length that is inversely proportional to $\sqrt{\sigma_z}$.

## 3.1 Recent Work

Some recent theoretical works have shown that the coil-helix transition can behave as a phase transition when external media are considered. Carri and Muthukumar[52] have suggested that the coil-helix transition in the presence of an absorbing membrane will become a true second-order phase transition. Park and Sung have studied transmembrane helix formation, with accompanied first-order adsorption transition[53]. Buhot and Halperin[54] have examined helical brushes, which at a critical density undergo a first order phase transition. On the experimental side, much of the interest in the coil-helix transition has been in examining the propensity of various amino acids to form helical structures[38]. Other interesting studies include those by Kumar and Manju[55] who have recently examined the characteristics of $\alpha$-helices in globular proteins and characterized them as kinked, rigid, or curved, with the largest fraction falling in the curved category. In Samulski's[56] study of the coil-helix transition, collapsed configurations were observed to occur at a higher temperature than the coil-helix transition.

In recent years, it has become more popular to study these structures with the

use of computer simulations in order to obtain a better physical picture of the system. This has proven to be fruitful as these simulations have shown that the coil-helix transition may involve multiple transitions[57], and that the system could be characterized by critical exponents[58, 59]. It has also allowed helical structures to be examined closely on a molecular scale through statistical analysis[33] and dynamic simulations[34, 35].

To study helix formation via a computer simulation one must create a model that has a helical ground state at low temperatures. This has typically been done by attempting to mimic a protein's interactions, and then choosing a protein sequence that is known to be a helix former. This approach is referred to as the all-atomic method which is discussed in detail in chapter 1. The problem with this method is that a protein's potential energy function is extremely complex. It is not only a function of the amino acids in the sequence, but also a function of the type of solvent environment. Although helical ground states have been successfully simulated using these models[33], they tend to require large amounts of computer resources. It is therefore desirable to attempt to understand helix formation on a fundamental level, and to determine the necessary interactions to create these helical structures. As well, what is the effect of altering these interactions on the observed transitions?

Minimal models have become very popular in recent years for deciphering complex systems. An example is the case of protein systems where minimal models have proven very useful in yielding insights into the folding mystery[1]. While most of the current research is still based on minimal lattice models, there have been numerous off-lattice models developed as well[60, 37]. In this chapter, we propose a

new model similar to those studied by several groups[61, 62, 63, 64] that examines specifically the dipole type interactions in polymer chains. This model could also be viewed as an experimental version of a Zimm-Bragg type model. An in depth study is also conducted of this new simplified polymer model, which has a helical ground state created from a limited number of adjustable parameters. We will also demonstrate how the alteration of the fundamental nature of the potential affects various aspects of the helix-coil transitions and the implication to the helix-coil transitions in helical protein segments.

## 3.2 New Minimal Helix Model

Before constructing a simplified model, the forces that are of interest must first be determined. In protein secondary structure, most of the stability of the structure is generated from interactions not restricted to interact with their local monomers. Therefore, these are the interactions that are captured in this model. The basic backbone of the protein also needs to be simplified; therefore, the model is made generic enough to describe a wide range of systems. The final step in creating the model is designing a potential that has the desired system behavior. In this model, a helical ground state structure is the requirement that must be implemented.

### 3.2.1 The Backbone

To examine the effects of directional binding on the entire structure of a polymer, we first model each residue as a single monomer that might interact with another

monomer through the excluded volume interaction, which we define as having a radius $d$, in addition to the attraction. Thus, the amino acids in Fig. 1.1 are treated as single spheres, which is a valid approximation for polymer systems. We would like to stress that in making such an approximation, we do not restrict ourselves to protein systems as we are concerned about examining the fundamental aspects of helical formation in all types of polymer chains. Therefore, in addition to helical forming protein structures, this model would apply to other polymer systems that undergo a coil-helix transition. Materials such as poly(*gamma*-benzyl-L-glutamate), poly($\beta$-benzyl-L-aspartate)[65], and poly(ethylene oxide)[66] are examples of synthetic molecules that undergo a coil-helix transition.

We must now consider the aspects of a polymer system that are most important to the formation of helical structures. First, the covalent bonds between monomers in this polymer system are approximated as unbreakable, as these bonds will be much stronger than the non-local interaction of interest. To further simplify the model, the bond lengths are fixed to a length of $a = 1$, as these fluctuations would be much smaller than those of the rest of the system. In addition to the fixed bond length constraint, a persistence length is added to this backbone chain because helical structures occur on a length scale of approximately the persistence length of the polymer backbone; this effect should be included explicitly in the model. The effect of persistence can be added to the system in many different ways. One of the options is to fix the azimthutal bond angle between the adjacent bonds so that the backbone of the polymer forms a worm-like chain(see Fig. 3.1). This creates a relatively rigid structure not unlike a polypeptide chain which has the backbone

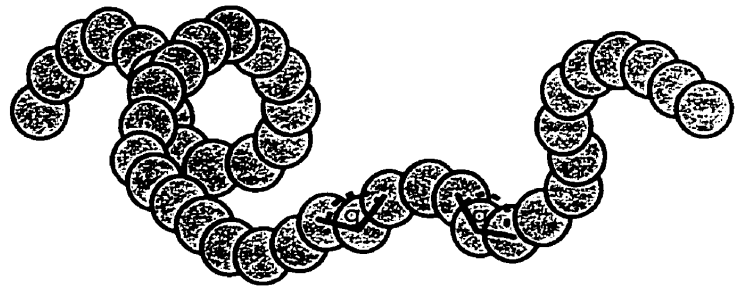

Figure 3.1: A depiction of a worm-like chain molecule. The azimuthal bond angle is fixed to give the chain a presistence length.

bond angles maintained by a potential an order of magnitude stronger than the non-local dipole forces that are to be examined. These constraints create the overall backbone of the system on which the helical model can be built.

### 3.2.2 The Potential

Clearly, an attraction now needs to be introduced between monomers in order to obtain a stable helical structure at low temperatures. One could attempt to construct such a potential by addressing the following questions: a) If an isotropic attraction is introduced, what is the ground state? b) What kind of attractive forces are needed to produce a helical ground state?, and c) Is there a unique potential that produces helical states in such a system? These are intriguing questions to a polymer physicist.

With some thought, it is clear that an isotropic potential interaction is not sufficient to produce helical ground states. As we will discuss later in this arti-

cle, an isotropic attraction binds monomers indiscriminately in any direction and produces a collapsed ground state at low temperatures with no specific crystalline structure[67, 68]. The question is then, what is needed to produce the desired helical ground state? This can be answered by examining the type of interactions in real protein systems. One of the most dominant forces in protein folding is the hydrophobic interaction[5, 69]. Although this is an important force in the protein's overall structure, it is not responsible for stabilizing or generating the helical secondary structures in proteins. Simulations of all atomic models of homoalanine have shown that helical states can be formed in a vacuum[33], which suggests that the hydrophobic interactions can be disregarded when creating a minimal potential. The remaining interactions left to stabilize the helix can be divided into two categories, local and non-local. Local interactions are those between nearest neighbor amino-acids such as torsional interactions, which are usually represented on a Ramachandran plot. In such a plot, there is typically not one unique energy minimum, meaning that several protein structures are available to the atom. It is possible to create a minimal model of a helix using only these forces, as Potthast has done[37]; however, these interactions are not generally considered the dominant interaction in the stabilization of the helical structures[33]. For this reason, we have chosen to exclude these types of interactions in order to consider only the dominant stabilizing force.

The remaining forces are non-local, as these interactions can create bonds with any molecule (not only with their immediate neighbors) along the chain. These interactions are comprised of electrostatic forces, dipole interactions, and hydrogen

bonds. Hydrogen bonding is generally considered to have the largest contribution. All of these forces contain a couple of common features. First, these forces are considered to interact over short distances, as the electrostatic forces are screened over long distances by the solvent molecules. Thus, in our model, we characterize this range through a parameter, $\sigma$. Second, due to the planar structure of an amino acid, the non-local interactions also contain an element of anisotropy in the overall nature of the interaction. It is this directional binding that leads to the stable formation of unique secondary structures ($\alpha$-helix or $\beta$-sheet) in proteins.

Therefore, the only ingredient that we need in order to make up a helical ground state is an anisotropic potential. The orientation of this anisotropy will produce different secondary structures; thus, it is directed perpendicular to the bond angle plane to produce a helix structure. To define the orientation of the $i^{\text{th}}$ residue, we consider the nearest neighbor and define a bond orientation unit vector $\hat{u}_i$.

$$\hat{u}_i = (\vec{r}_i - \vec{r}_{i-1}) \times (\vec{r}_{i+1} - \vec{r}_i)/\sin\theta \ , \tag{3.1}$$

where $\vec{r}_i$ is the position of the $i^{\text{th}}$ monomer, and $\theta$ is the fixed bond angle. This vector is pictorially represented in Fig. 3.2.

The preference for various alignments of these bond orientation vectors generates the anisotropic potential. In a helical state, these unit vectors should point in the same direction. In the same spirit as the square well potential for other polymeric

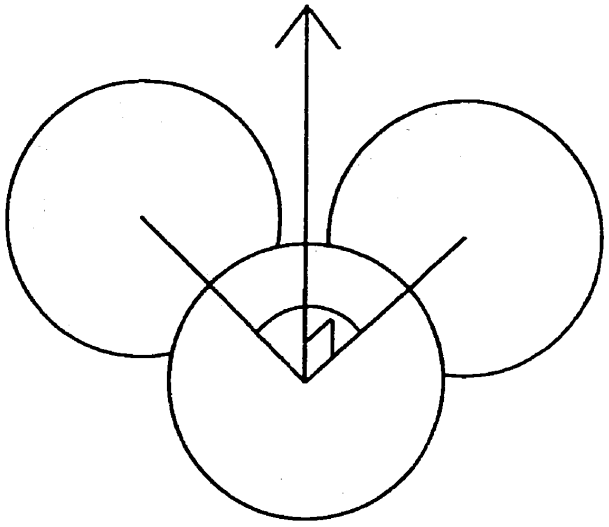

Figure 3.2: Pictorial representation of the bond direction vector.

systems, we choose the binary potential to have the form

$$V_{ij} = \begin{cases} 0 & for\ \sigma \leq r_{ij} \\ -\epsilon[\hat{u}_i \cdot \hat{r}_{ij}]^m - \epsilon[\hat{u}_j \cdot \hat{r}_{ij}]^m & for\ d \leq r_{ij} < \sigma \\ \infty & for\ 0 \leq r_{ij} < d \end{cases} \tag{3.2}$$

where $d$ is the excluded volume radius, $\sigma$ is the attraction radius, $\hat{r}_{ij} = (\vec{r}_i - \vec{r}_j)/|\vec{r}_i - \vec{r}_j|$ is the unit vector defining the relative positions between monomers, and $\epsilon$ represents the strength of the attraction. Later, the parameter $\epsilon$ is rescaled into the temperature of the system and we consider the reduced temperature $\tilde{T} = k_B T/\epsilon$.

### 3.2.3 Simulation Details

This final form for the model contains only a few adjustable parameters. The most significant one is the exponent $m$, which controls the strength of the anisotropy, with $m = 0$ producing an isotropic attractive field. Note that the $m = 0$ case is merely the square-well potential that has attracted considerable attention recently. In the low temperature regime, the possible states of a polymer and the nature of the transitions in these systems are of particular interest in this case. There is some flexibility in value $m$, and in this study we consider several different values. There are some considerations which should be kept in mind when deciding on an appropriate choice. For this work, a symmetric potential is desired, and thus, $m$ is restricted to even values $(2, 4, 6, 8$,etc). When a sufficiently small value for the fixed azimuthal bond angle is used, all of these choices create a helical ground state; however, it is found that as the bond angle is increased to allow more monomers per helical loop, the values 2 and 4 no longer produce a helical ground state. This occurs because the potential is insufficiently anisotropic to favor a helical state within the geometry of the system. The values 6 and 8 produce helical states, but the value 8 requires more computational effort and should not produce significantly different results from the value 6. Thus, $m = 6$ is chosen to represent our helical polymer state.

The other parameters in the model, varying in a reasonable physical range, have less profound effects on the nature of the transition. For this study, the diameter of the monomers is chosen to have a value $d = (3/2)a$, where $a$ is the bond length between monomers so that we have a somewhat smooth monomer-

monomer surface contour. The range of the attractive force is chosen to be $\sqrt{45/8}a$, which is in line with the conventional choice of the attraction for a square well to coincide with a Lennard Jones potential[67, 70]. The fixed bond angle also adds an additional parameter to the model. For the data presented here, we have used an angle of $\pi/3$. In doing so, we keep to the typical bond angles observed along the backbone of a protein molecule. With these choices of parameters, we consider the thermodynamics of the minimal model as a function of $N, m, \tilde{T}$.

In order to sample the conformations of the worm-like chain, a Monte Carlo pivot algorithm is used. Introduced by Madras and Sokal[42], this non-local algorithm is an effective way to sample the configurational space of polymer chains. In the case of a worm-like chain with both fixed bond length and azimuthal angle, it is perhaps the most efficient as local moves are difficult to calculate because of the fixed bonded angle restrictions. A detailed description of this algorithm is given in chapter 2.

In addition to the pivot algorithm, the multicanonical technique was utilized to obtain a clearer picture of the temperature dependence of the system. Once the appropriate weight function was determined, a production run was conducted for $5 \times 10^8$ Monte Carlo steps. Data points were collected every 100 steps for a total of $5 \times 10^6$ data points for each production run. These points were then averaged with the appropriate Boltzmann weight function in order to calculate the quantities of interest.

## 3.3 Definition of Measured Quantities

Among numerous quantities that can be measured in this type of simulation, the most common is the the specific heat per molecule defined as

$$C_v = \frac{\langle E^2 \rangle - \langle E \rangle^2}{N\tilde{T}^2} \ . \tag{3.3}$$

The specific heat is an ideal quantity to examine because it provides direct information about the free energy of the system. Any significant structural changes are reflected by anomalous behavior in the specific heat. In a phase transition, this anomaly in $C_v$ would increase with system size, while a simple crossover between regions would manifest as a smooth hump that does not increase with system size. In our model, we expect to see various types of behavior due to fluctuations of the positional arrangement of molecules, and the ordering of the bond directional vectors specified in Eq. 3.1.

To assess the overall arrangement of the monomers, the squared radius of gyration is also used. This is an ideal measure of the size of the polymer because it can be directly measured through light scattering experiments. This quantity will reveal the major structural change that occurs, and is calculated using the following equation,

$$\langle R_g^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle (\vec{r}_i - \vec{r}_{cm})^2 \rangle \ , \tag{3.4}$$

where $\vec{r}_{cm}$ is the vector defining the center of mass. One way to identify structural changes within the polymer is to examine the scaling of the radius of gyration. At

high temperatures, when a polymer's structural characteristics are dictated by the repulsive forces of the excluded volume interaction, it is well known that $R_g$ should scale with system size with an exponent of $\nu \approx 3/5$. At cooler temperatures, when attractive forces dominate over the repulsion, the polymer acts as a molten liquid or globule and $R_g$ scales with an exponent $\nu = 1/3$. In extended or rod like phases, $R_g$ would scale with the length and with an exponent $\nu = 1$.

It is also worth dissecting the radius of gyration further to examine the structural changes, especially in the helix regime. We can define an overall helical axis for the entire polymer as the vector which is the sum of all the bond direction vectors. By breaking the radius of gyration into components both parallel and perpendicular to this vector; we can define the following quantities,

$$\hat{U} = \sum_{l=2}^{N-1} \vec{u}_l \Big/ \left| \sum_{l=2}^{N-1} \vec{u}_l \right| \tag{3.5}$$

$$\langle R_{g\parallel}^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \left[ (\vec{r}_i - \vec{r}_{cm}) \cdot \hat{U} \right]^2 , \tag{3.6}$$

$$\langle R_{g\perp}^2 \rangle = \langle R_g^2 \rangle - \langle R_{g\parallel}^2 \rangle . \tag{3.7}$$

During helix formation, we would expect these two parameters to diverge from each other as the parallel component increases and the perpendicular component decreases to approach the square diameter of the helix.

To describe the orientation fluctuations of the bond directional vector, we need to define an appropriate orientational order parameter. The question is, what is the single parameter which correctly describes these types of fluctuations? In real

protein models, the orientational order of an amino acid within the molecule is described through its Ramachandran angles. For a specific type of structure, a range of values is specified to describe the amino acid's conformation. As our model is not designed with the Ramachandran angles, we attempt to classify the orientational fluctuations by several more quantitative methods. To do this, we introduce a set of order parameters, each of which describes the different correlated fluctuations of the system. The first parameter, $H_1$, we create is similar to the Ramachandran $\phi$ and $\psi$ parameters used to describe a helix in a protein molecule. Here the configuration of a group of atoms is examined to determine whether they are in a helical configuration using a predetermined criterion. To determine if a cluster is in a helical configuration, the distance between the $i$ and $i+3$ monomer is calculated. If this distance falls within the length criterion for a helix, then the unit is counted as helical. We attempted several definitions of unit helicity that showed little difference in the qualitative behavior of this order parameter.

The above order parameter relies on the arbitrary definition of a helical unit, and for this reason, is somewhat unsatisfactory. It is worth defining order parameters that describe the ordering and do not use a predetermined criteria for defining helicity. Thus, the other order parameters defined are based on the vector properties which define the relative orientation of the bond directional vectors. As mentioned earlier, these vectors should all approximately align when a helical state is approached. Thus, we define,

$$H_2 = \sum_{i=2}^{N-2} \hat{u}_i \cdot \hat{u}_{i+1} \tag{3.8}$$

$$H_3 = \sum_{i=2}^{N-1} \hat{u}_i \cdot \hat{u}_{\text{mid}} \tag{3.9}$$

$$H_4 = \left( \sum_{i=2}^{N-1} u_i \right)^2 , \tag{3.10}$$

where $\hat{u}_{\text{mid}}$ is the bond vector of the $\left(\frac{N}{2}\right)^{th}$ monomer (in practice, $\left(\frac{N}{2}\right)$ is rounded down if not an integer). The order parameter $H_2$ describes the local correlations of the orientational order in the helix, and indicates the onset to helix formation. However, this parameter will not yield any information on long range order, and thus, it will not be useful in describing the nature of the transition. The parameter $H_3$ is similar to $H_2$ in that it examines the correlation of bonds, but in this case, a central bond vector is correlated with all the other bond vectors along the chain. We choose a central monomer because the end monomers are subject to large fluctuations due to fewer constraints on the monomer's conformation. A potential drawback of using this parameter is the fact that the central monomer that is chosen may be located at a fracture region while the rest of the chain is helical, a situtation that would lead to an improper description of the chain. The final parameter, $H_4$, is merely the sum of all the bond vectors. This order parameter is perhaps the best for examining the nature of the coil-helix transition because it should accurately describe the net helical growth within the molecule while accounting for fractured regions that reduce the value of the order parameter.

We can calculate susceptibilities for the radius of gyration and for all of the

orientational order parameters with the following equation,

$$\chi_A = \frac{1}{\bar{T} N_A} \left[ \langle A^2 \rangle - \langle A \rangle^2 \right] , \tag{3.11}$$

where $N_A$ is the number of units of the quantity $A$. These quantities will reveal when fluctuations of the order parameters are significant. In the case of phase transitions, all parameters should show anomalies that increase with system size. Most importantly, the parameters will demonstrate the type of ordering that occurs at various temperatures.

The most popular method for examining the coil-helix transition is through the use of the Zimm-Bragg parameters, $s_z$ and $\sigma_z$. These parameters are typically obtained from the equations[33],

$$\langle n_h \rangle / N_h = \frac{1}{2} - \frac{1 - s_z}{2\sqrt{(1 - s_z)^2 + 4 s_z \sigma_z}} \tag{3.12}$$

$$\langle l_h \rangle = 1 + \frac{2 s_z}{1 - s_z + \sqrt{(1 - s_z)^2 + 4 s_z \sigma_z}} \tag{3.13}$$

where $\langle n_h \rangle$ is the average number of helical units, $N_h$ is the total possible number of helical units, and $\langle l_h \rangle$ is the average length of a helical segment. In order to obtain these parameters, the only quantities needed are the average helicity and the average helical segment length. Obtaining these values is not straightforward. Although any of the four helical order parameters could be used to calculate the average helicity, the parameter $H_1$ is the most suitable for calculating the segment length. The other parameters require the introduction of an arbitrary condition to

determine the length of a helical segment. This would leave the average helicity as a constant and the average segment length a varying parameter depending on the condition used to determine the break points in the helix. Therefore, $H_1$ is left as the best candidate to measure these parameters because the condition used to define a segment as helical is the same one used to define helical length. However, depending on the condition set for determining helicity, the values of the Zimm-Bragg parameters could also be changed. For example, using two valid conditions to determine the parameter $H_1$ generates a 35% difference in the Zimm-Bragg values at the transition temperature. Thus, the Zimm-Bragg parameters in this model are subject to inconsistencies from the particular specification of a unit being helical in our model. Therefore, it is not instructive to calculate exact values for these parameters, but rather to assess the temperature trends of these functions, which are essentially unchanged by different conditions for helicity.

## 3.4 Observed Helical States

We now examine the behavior of our model as a function of temperature. First, we describe the case in which the parameter, $m$, in the potential is 6, although we can produce helical states with the other studied values of non-zero $m$. We initially examine the heat capacity of our system. In Fig. 3.3, the heat capacity per molecule for chains of $N = 13$, 19, 26, 33, 39, and 51, are displayed with symbols on every tenth temperature unit. The figure presents several interesting features, as the system appears to show three separate structural transitions. There are strong

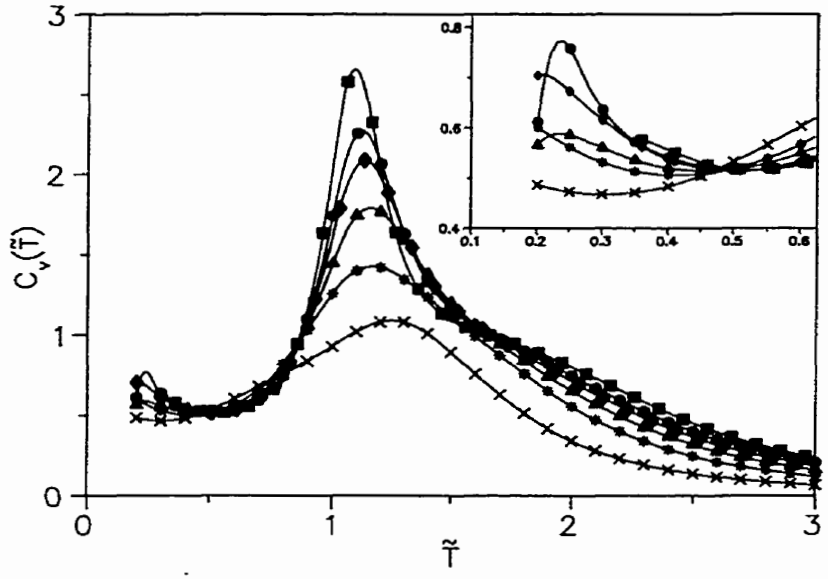


Figure 3.3: Specific heat for $m = 6$ for polymer sizes 13 (×), 19 (∗), 26 (▲), 33 (♦), 39 (•), and 51 (■). The low temperature data for the $N = 51$ is missing in the inset as we were unable to collect sufficiently accurate values at these temperatures.

peaks near temperatures $\tilde{T} \approx 1$ and $\tilde{T} \approx 0.3$. Around $\tilde{T} \approx 2$, we also observe a weak shoulder that appears to be a third transition. The peak at $\tilde{T} \approx 0.3$ represents a change between two helical regions, the peak at $\tilde{T} \approx 1$ represents the coil-helix transition, and the shoulder represents a collapsing transition. These regions are further studied using the other quantities mentioned above. We have depicted structures believed to be representative of the different regions in Fig. 3.4 to assist in visualizing the types of transitions in the system.

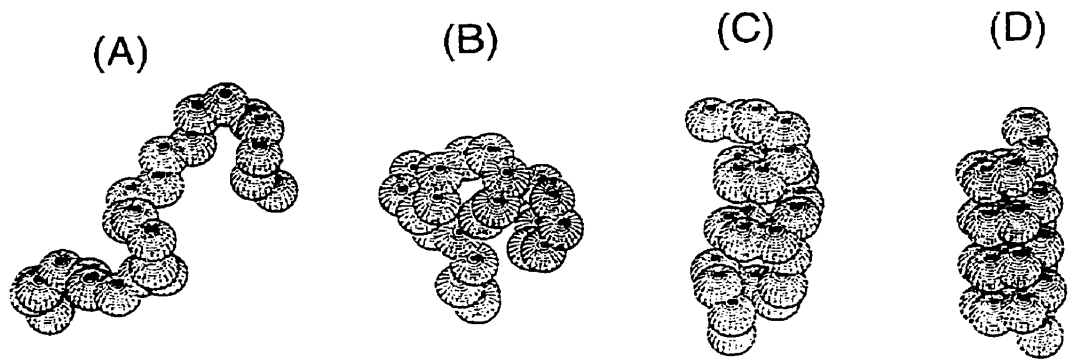


Figure 3.4: Snapshots of the configurations of a 26mer at various temperatures: (a) $k_BT/\epsilon \approx 4$ (coil), (b) $k_BT/\epsilon \approx 1.3$ (globular) (c) $k_BT/\epsilon \approx 0.8$ (helix I) and (d) $k_BT/\epsilon \approx 0.2$ (helix II). The size of the beads represents the actual hard-core interaction between non-adjacent monomers, and the attraction force range is $\sqrt{45/8}a$, where $a$ is the bond length. A highly directionalized potential with $m = 6$ (see Eq. (3.2)) is used.

### 3.4.1 Coil Region

In the high temperature regime, the polymer is in a random coil state (see Fig. 3.4a), where the repulsion of the polymer dominates causing swollen conformations. In this region, the polymer's size is expected to scale with system size and to have an exponent of $\nu = 0.589$[71] in the long chain limit. Calculating the radius of gyration for polymers at a temperature of $\tilde{T} = 10$ (see Table 3.1), we obtain a scaling exponent of $\nu = 0.64(5)$ by fitting the data corresponding to the four largest polymers to a power-law. The squares in Fig. 3.5 show the scaling of these values vs. the polymer size on a log-log plot. The calculated exponent is larger than expected, which can easily be accounted for as we are not in the asymptotic limit of large polymers where the predicted exponent should be valid. Although we do

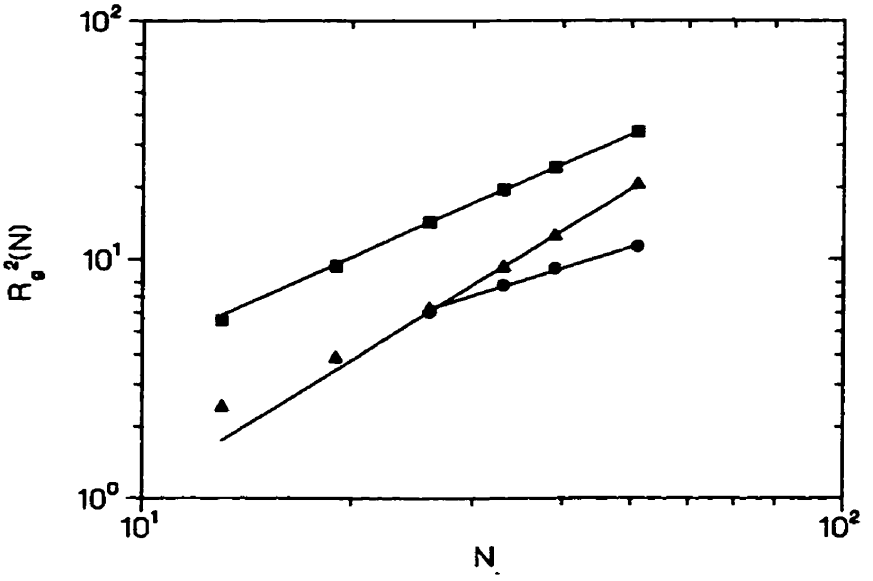


Figure 3.5: Scaling of the squared radius of gyration for $m = 6$ in the coil (■), globular (●), and helical (▲) regimes.

not obtain precise agreement with the expected exponent, the relative values of the exponents in the various temperature regions are important.

### 3.4.2 Molten Globular Region

Upon cooling to a temperature of approximately $\tilde{T} \approx 2$, there is a significant change in the size of the polymer (see Fig. 3.4b). This is signified by the shoulder in the peak at $\tilde{T} \approx 1$ of the heat capacity curve. In Fig. 3.6, the squared radius of gyration demonstrates this collapse by showing a dramatic reduction in the polymers size. In the larger sized polymers, we see the size increase again at the coil-helix transition. This is caused by the rearrangement of atoms into a helical form that is elongated

Table 3.1: The value of the squared radius of gyration at various temperatures for the $m = 6$ case. The values for $R_g$ at $\tilde{T} \approx 1.3$ and $\tilde{T} \approx 0.6$ are based on the values when $R_g$ reaches it minimum size and when $R_g$ is maximum (when in a helical configuration) respectively.

| $N$ | $R_g^2(T = 10)$ | $R_g^2(T \approx 1.3)$ | $R_g^2(T \approx 0.6)$ | $R_g^2(T = 0.25)$ |
|---|---|---|---|---|
| 13 | 5.5573(2) | - | - | 2.40618(6) |
| 19 | 9.365(1) | - | - | 3.8636(1) |
| 26 | 14.256(2) | 6.0334(5) | 6.3567(2) | 6.2080(2) |
| 33 | 19.479(3) | 7.8235(8) | 9.5400(3) | 9.2503(2) |
| 39 | 24.184(4) | 9.174(1) | 12.8463(4) | 12.4080(2) |
| 51 | 34.083(5) | 11.321(2) | 21.0162(5) | 20.4131(3) |

in one direction and has a larger overall size than the collapsed globular. In the smaller polymers, this collapse is not observed because the final helical states are comparable in size to that of the collapsed state. This type of collapse has been observed in real protein systems. Samulski's experimental studies of the coil-helix transition in polypeptide chains show this type of globular collapse prior to the coil-helix transition[56]. Also, Pitard et al.[62] analytically considered anisotropic potential forms in flexible polymers, and have shown that this type of polymer undergoes a collapse before orientation ordering; however, the structures studied in their work were not of a helical nature.

To determine whether the polymer is truly making a transition to a globular state, we consider only the larger polymers and examine the scaling of the size of a polymer when in its most compact state (defined by the minimum in the radius of gyration). The results are displayed in Table 3.1, and by fitting this data to

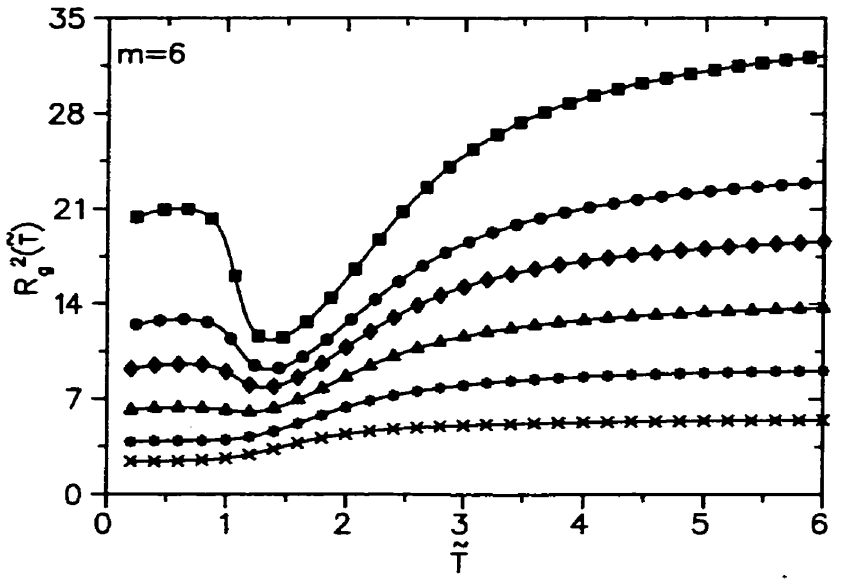


Figure 3.6: Squared radius of gyration for $m = 6$. Shown are the polymer lengths 13 (×), 19 (∗), 26 (▲), 33 (♦), 39 (•), and 51 (■).

a power-law, we obtain a value of $\nu = 0.45(15)$ for the scaling exponent. This value of $\nu$ is well below that of the exponent for the coil regime, but is still above the value predicted for the globular phase ($\nu = 1/3$[72]). The circles in Fig. 3.5 demonstrate this scaling, and we observe that the exponent decreases for larger system sizes. This is reasonable as we are still in the comparatively small size limit for the polymer. To further examine this transition, we examine the parallel and perpendicular squared radii of gyration. These are illustrated for the $N = 39$ case in Fig. 3.7. Both of these quantities show a smooth decrease in the temperature range of interest. This confirms that the polymer is collapsing isotropically. In Fig. 3.8, the four order parameters are displayed for the $N = 39$ chain. All four show no significant increase about the temperature of the collapsing transition,

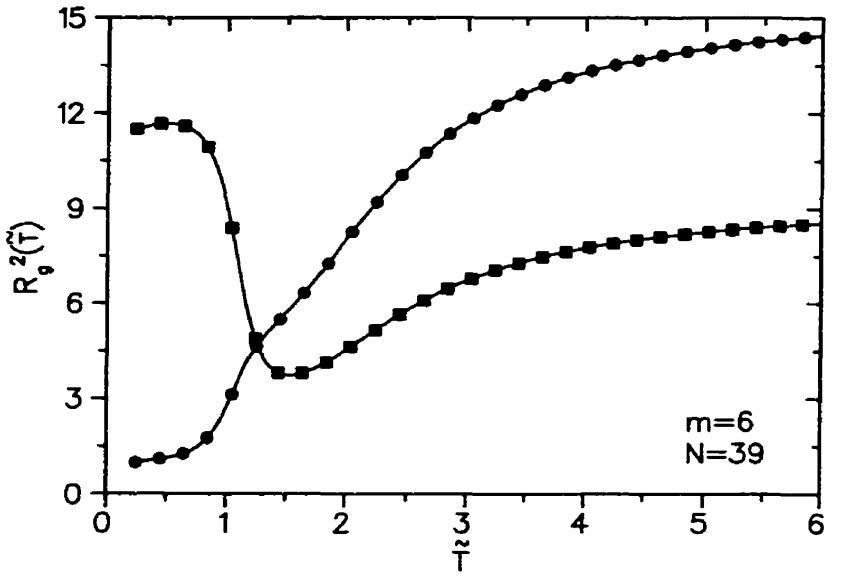


Figure 3.7: Perpendicular (•) and parallel (■) components of the squared radius of gyration for the $N = 39$ polymer and potential $m = 6$.

indicating that there is no orientational ordering developing and that the collapse is strictly a spatial phenomenon. We believe that for large polymers, the molecule enters a true globular phase. In this phase, the monomers would be closely packed and uniformly distributed, just like a liquid state for a collection of small molecules.

For an alternative perspective of the collapsing transition, we can examine the susceptibility for the radius of gyration. Fig. 3.9 illustrates these curves for all of the polymer sizes considered. The figure shows a broad peak near $\tilde{T} = 2.5$ that increases with system size, and the heights and locations of each of the peaks are listed in Table 3.2. The location of the peak roughly represents the $\Theta$ temperature of the molecule, which is the point where repulsive and attractive forces within the polymer balance. This data reveals that the collapsing transition occurs prior to

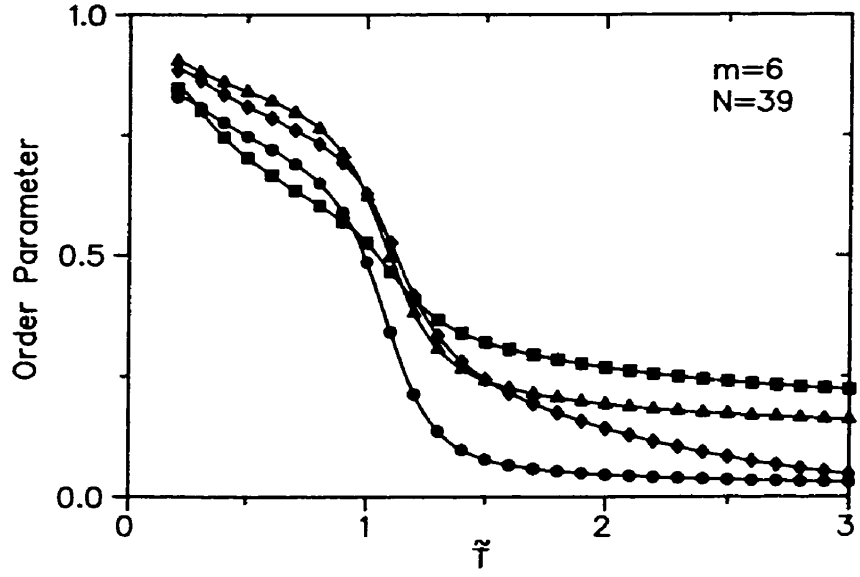


Figure 3.8: Order parameter $H_1$ (■), $H_2$ (♦), $H_3$ (•), and $H_4$ (▲), for $m = 6$ and $N = 39$.

any orientational ordering.

### 3.4.3 Helix I Region

The strongest peak on the $C_v$ plot corresponds to a transition from an isotropic collapse state (Fig. 3.4b) to a helical state (Fig. 3.4c) observed near $\tilde{T} \approx 1$. The exact locations of the peaks and their maximum values are given in Table 3.2. The data shows that the height of the specific heat curves increase systematically with polymer size. The scaling of these peaks will be examined later in this work when the nature of the transitions is studied.

Accompanying the strong peak in the $C_v$ curve, we have observed drastic changes in the orientational order parameters. In Fig. 3.8, the four helical order parameters,

Table 3.2: Data on the transitions in the $m = 6$ case for various sizes. Shown is the globular-helix transition temperature, $T_c$, the height of the $C_v$ at $T_c$, and the width of $C_v$ when $b = 0.9$, $\Gamma(C_v)$. Also shown is the $\Theta$ temperature, $T_\Theta$, the susceptibility of $R_g$ at $T_\Theta$ and the foldability parameter, $\sigma$.

| $N$ | $T_c$ | $C_v(T_c)$ | $\Gamma_{C_v}$ | $T_\Theta$ | $\chi_{R_g}(\Theta)$ | $\sigma$ |
|---|---|---|---|---|---|---|
| 13 | 1.25(1) | 1.09(1) | 0.37(2) | 1.71(1) | 0.0666(5) | 0.27(1) |
| 19 | 1.17(1) | 1.42(3) | 0.35(2) | 2.13(1) | 0.0936(7) | 0.45(1) |
| 26 | 1.16(1) | 1.79(5) | 0.25(2) | 2.35(1) | 0.136(1) | 0.51(1) |
| 33 | 1.13(1) | 2.08(7) | 0.18(2) | 2.46(1) | 0.184(2) | 0.54(1) |
| 39 | 1.12(1) | 2.29(9) | 0.17(2) | 2.54(1) | 0.226(2) | 0.56(1) |
| 51 | 1.09(1) | 2.69(13) | 0.13(2) | 2.64(1) | 0.318(2) | 0.59(1) |

$H_1$, $H_2$, $H_3$, and $H_4$ for the system size of $N = 39$, all show significant increases in their values crossing this transitions. Susceptibility curves for these order parameters are illustrated in Fig. 3.10, with each clearly showing a distinct peak at the transition. The maximum heights of each of the peaks are also listed in Table 3.3.

In addition, the squared radius of gyration also shows an increase at this temperature as the polymer becomes extended along the helical axis. The extension is further demonstrated by the sudden increase in $R_{g\parallel}$ and decrease in $R_{g\perp}$. For larger polymers, we observe a turnover in the radius of gyration as $R_{g\perp}$ begins to shrink in size causing the polymer to take on a more compact helical form. If we scale the radius of gyration in the region where the turnover in the radius of gyration occurs, this corresponds to scaling the helices of the helix I region. The values for the square radius of gyration at the turnover are listed in Table 3.1, and by fitting these values to a power-law, we obtain an exponent of $\nu = 0.90(9)$. This

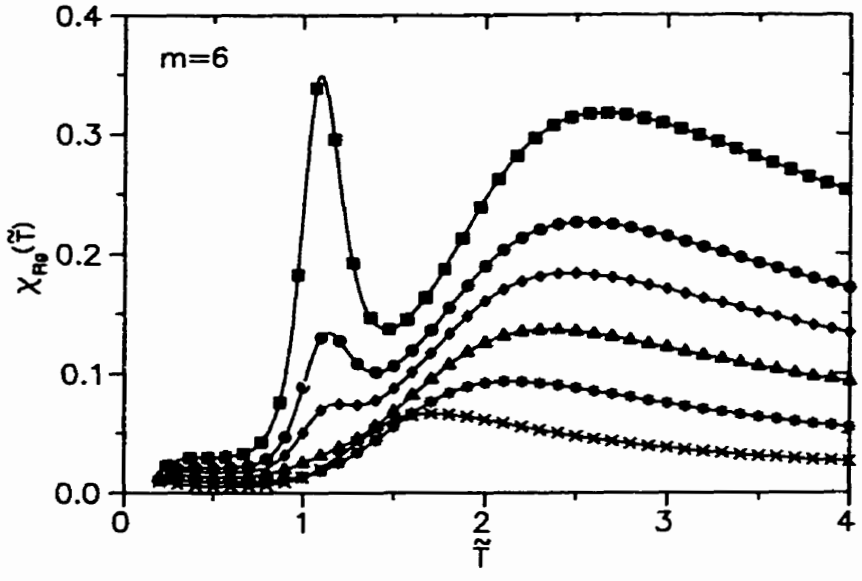


Figure 3.9: Susceptibilities of the radius of gyration for sizes 13 (×), 19 (∗), 26 (▲), 33 (♦), 39 (•), and 51 (■).

is clearly larger than the exponent for the coil regime and agrees fairly well with the predicted value of $\nu = 1$.

### 3.4.4 Helix II Region

The final point of interest in Fig. 3.3 is the anomalous peak at $\tilde{T} \approx 0.3$, which is enlarged in the inset. This peak in the heat capacity curve is due to a solid-solid type transition between two helical states. The transition is accompanied by a second change in the four order parameters as reflected by their susceptibilities (Fig. 3.10) for $N = 39$. The susceptibility of the order parameter $H_1$ has a profound second transition. This parameter is set to designate a group of monomers as helical,

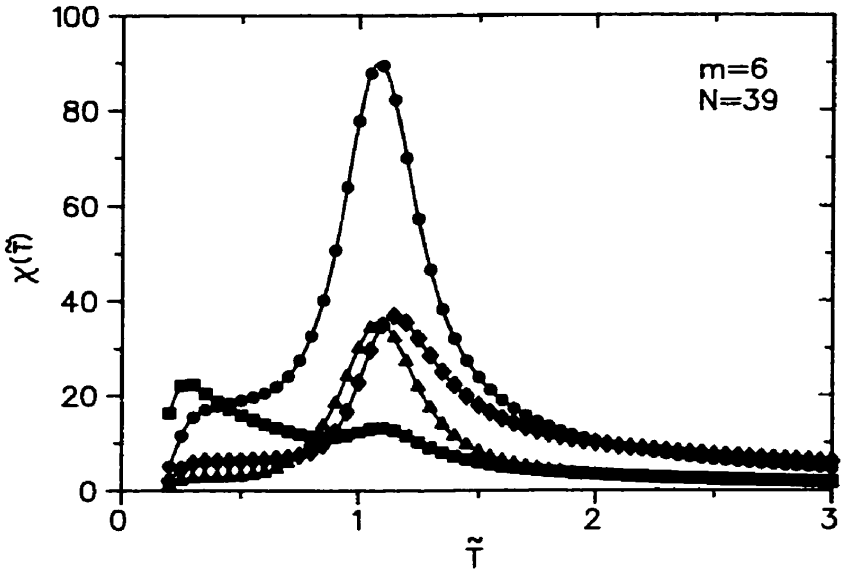


Figure 3.10: Susceptibilities $\chi_{H_1}$ (■), $\chi_{H_2}$ (♦), $\chi_{H_3}$ (•), and $\chi_{H_4}$ (▲), for $m = 6$ and $N = 39$.

and can be modified to be sensitive to the more perfectly formed helical states. The increased fluctuations in this parameter demonstrate that the activity around this transition temperature is probably related to a crystallization of the entire polymer chain into a helical state. This solid-solid transition is known to occur in related polymer systems, as Zhou et al.[19, 67] have demonstrated multiple-low temperature transitions in polymer and protein-like systems. The structures of the two states differ as the higher temperature helix I state has a high degree of helical ordering yet is loosely packed, possibly with dangling end segments. The radius of gyration does not display a significant change here as the polymer does not undergo a major structural change. We calculate the Zimm-Bragg parameters (Fig. 3.11) using the order parameter $H_1$. $H_1$ is adjusted to be sensitive to perfectly formed

Table 3.3: The maximum heights of the susceptibility, for $m = 6$. The blank entries indicate no peak at the location of the helix-coil transition.

| $N$ | $\chi_{H_1}$ | $\chi_{H_2}$ | $\chi_{H_3}$ | $\chi_{H_4}$ |
|---|---|---|---|---|
| 13 | - | - | - | 0.232(7) |
| 19 | - | 0.561(5) | 0.798(6) | 0.353(1) |
| 26 | - | 0.695(8) | 1.299(9) | 0.553(1) |
| 33 | 0.329(7) | 0.88(1) | 1.88(1) | 0.775(2) |
| 39 | 0.366(12) | 1.05(2) | 2.36(2) | 0.955(2) |
| 51 | 0.454(14) | 1.39(2) | 3.36(2) | 1.331(3) |

helical states. The typical behavior for the Zimm-Bragg parameters is observed in the figure, with a deviation in the standard shape of these curves due to the second transition. In the parameter $\sigma_z$ there is an initial decrease at the coil-helix transition, as would normally be observed, followed by another sharp drop near the second peak in the specific heat, confirming that there is a significant change in the average number of breaking points of the helices in the two regions.

We can best describe the structural differences between the two helical states by considering the very long chain limit. In this limit, we would expect the helix I polymer to behave like a rescaled worm-like chain that would contain bendable helix units, while the helix II polymer would behave like a rigid rod of perfectly ordered helical units. Experimentally these two types of helices have been observed. In a study of numerous protein helical segments by Kumar and Bansal[55], they classify helices into three groups: kinked, curved, and rod. If we disregard the kinked molecules, the curved and rod helices certainly correspond to our two helical

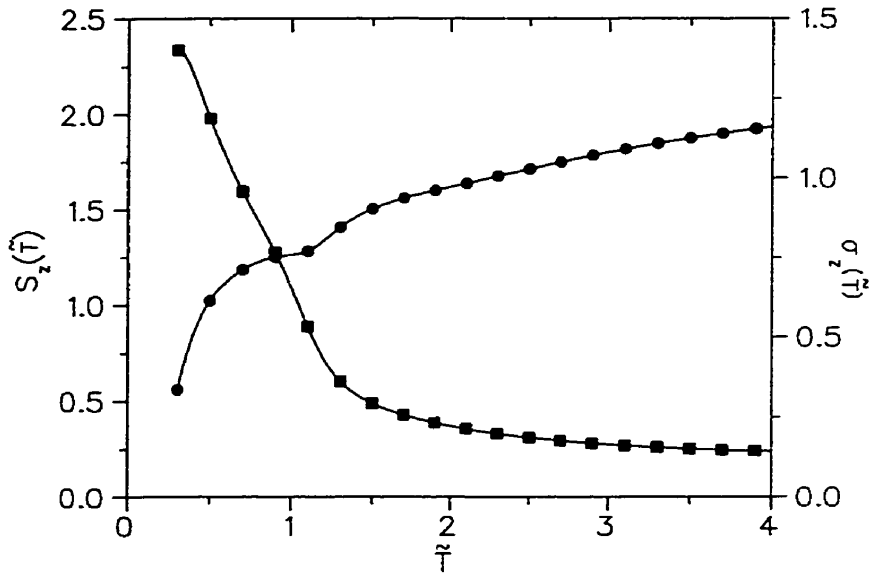


Figure 3.11: Zimm-Bragg parameters $s_z$ (■) and $\sigma_z$ (•) as a function of temperature for $N = 39$.

regions. In our model, we examine the rescaled temperature that can be interpreted in two ways. We may consider the system to have a constant interaction strength, $\epsilon$, while we adjust the temperature, $T$, or we may consider a fixed $T$ while we allow $\epsilon$ to vary. When comparing heteropolymers, the later is the more appropriate view since experiments are usually conducted at constant temperature. Hence, we can also consider experimentally observed helices as homopolymers with a constant interaction parameter based on the average interaction between monomers. This would mean that the curved helices observed experimentally correspond to the helix I region, and have on average a weak interaction between monomers. Similarly, the rods would correspond to the helix II region and would have a much stronger average interaction. In Kumar and Bansal's work, the largest fraction of observed helices are

curved, which corresponds with helices just below the helix-coil transition. This suggests a consistency between the two systems, as we would expect the higher temperature helix I region to be more populated.

If we examine the parallel and perpendicular components of the squared radius of gyration in the low temperature region (Fig. 3.7), we observe a continuous decrease in the perpendicular component and a turnover in the parallel component. This is suggestive of a change in the structural behavior, as we would expect the parallel component to increase to a constant value. We can also examine the scaling of the radius of gyration at very low temperatures, which corresponds to the helix II region. The square radius of gyration is listed in Table 3.1 for the temperature $\tilde{T} = 0.25$. The triangles in Fig. 3.5 show the scaling of the data with $N$; fitting the data corresponding to the four largest polymers to a power-law yields an exponent of $\nu = 0.9(1)$. The exponent is the same as that of the helix I region; however, we would expect to see some deviation if we could simulate longer helices because the size of the helices in the helix I region would fluctuate more due to their flexibility. We speculate that the change between the two states is generated by a change in the allowed magnitude of energy fluctuations. At high temperatures, the helix can flex in the interior of the chain, but this causes fluctuations, as the monomers move collectively to bend the helix, which are of the order of $6\Delta E$ (where $\Delta E$ is the energy fluctuation from a single monomer). When the chain is cooled, these collective fluctuations can be frozen out leaving a rigid rod with end monomers that are only able to fluctuate slightly.

## 3.5 Examination of Finite Size Scaling

### 3.5.1 Scaling of the Coil-Helix Transition

Although we are working on rather small system sizes, it is worth attempting some finite size analysis of the data in order to examine the nature of the transitions. We can examine several quantities to determine the critical exponents of the system. We define a temperature gap for the transition $\Gamma = T_2 - T_1$ (where $T_1 < T_c < T_2$), such that $C(T_1) = bC(T_c) = C(T_2)$, where $b$ is some fraction less than one. This gap should scale as,

$$\Gamma \sim N^{-\frac{1}{d\nu}} , \tag{3.14}$$

where $d$ is the effective dimension, and $\nu$ is the critical exponent for a correlation length. The effective dimension of the system is not known thus, it is always used in conjunction with the exponent $\nu$. From this point on $d\nu$ is treated as a single parameter. The maximum in the specific heat will then scale as

$$C_{max} \sim N^{\frac{\alpha}{d\nu}} , \tag{3.15}$$

and the maximum in the susceptibility of the proper order parameter for the system will scale as

$$\chi_{max} \sim N^{\frac{\gamma}{d\nu}} , \tag{3.16}$$

where $\alpha$ and $\gamma$ have their usual meaning in critical phenomena[74].

The helix-coil transition is typically not considered to be a true phase transition

in the thermodynamic limit of $N \rightarrow \infty$; however, for small system sizes, the system should still behave as a first order phase transition due to system cooperativity[72, 73]. It is not realistic to ask whether our system exhibits a phase transition because with this method, exceptionally large system sizes need to be studied. Despite this, we examine the system to determine its scaling behavior, as this system should still be characterizable by a set of effective critical exponents[58, 59].

From the data in Table 3.2, we have calculated the critical exponents for the globular to helix transition as $d\nu = 1.04(9)$, and $\alpha = 0.70(15)$. To obtain the exponent $\gamma$, we use the susceptibility data of the $H_4$ order parameter in Table 3.3. As mentioned earlier, we suspect that this order parameter will best describe the fluctuations of the bond direction vectors as they align during the transition. From this data, a value of $\gamma = 1.3(2)$ was obtained. In a typical first order phase transition, these exponents are expected to follow the relation $d\nu = \alpha = \gamma = 1$[75]. The exponent $d\nu$ appears to be in agreement with the notion that the system exhibits a first order phase transition; however, the exponents $\alpha$ and $\gamma$ appear to differ significantly. The discrepancies in this system are likely due to the small system size and large errors associated with the heights of the peaks. It appears for the specific heat that any errors are likely to underestimate the peak height, which is more significant within larger polymer sizes. It is also unclear what effect the coil-globular transition has on the peak heights measured. Although the values differ from those expected, the model does show some agreement with the scaling results of the all-atom simulations of protein molecules (for details see Ref.[59]). This suggests that the system is within this size range where the coil-helix transition can be treated

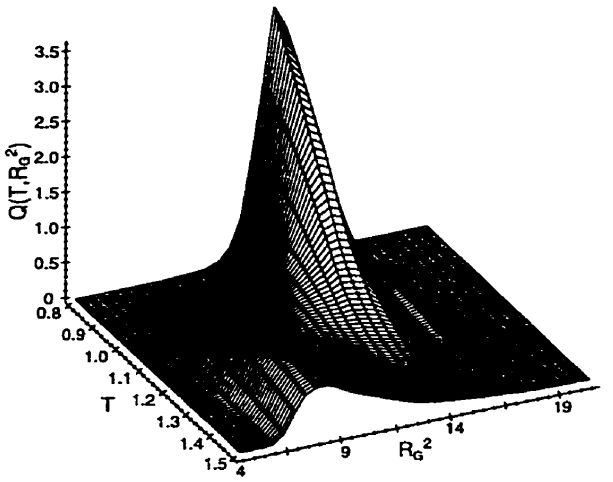


Figure 3.12: Frequency of a particular value of the radius of gyration as a function of temperature for $N = 39$.

as a first order phenomenon.

For a closer look at the transition, we re-examine the squared radius of gyration as a function of temperature through the parameter $Q(\tilde{T}, r_g^2)$, which examines the distribution of polymer sizes as a function of temperature. A diagram such as this should reveal whether or not the polymer is making a discontinuous jump to the helical state, which is predicted to occur for cooperative first order systems. A histogram of the size of the polymer at a specific temperature is defined in the following way,

$$Q(r_g^2, T) = \left\langle \sum_{R_g^2(T)} \delta \left[ R_g^2(T) - r_g^2(T) \right] \right\rangle \ , \tag{3.17}$$

where $r_g$ is a specific value of the radius of gyration. Figure 3.12 illustrates a three-dimensional plot of this parameter in the temperature range of interest for

the $N = 39$ case. Cooling the polymer from high temperatures shows that the polymer's structure localizes to a compact form prior to the helix-coil transition. At the transition temperature, it appears that a double peak develops as the polymer enters into a helical structure. This behavior would suggest a first order phase transition because the system appears to make a discontinuous change in size at the transition. This is supported by the increasing height of the radius of gyration peak at a temperature $\tilde{T} = 1$, in Fig. 3.9. We see that the three largest polymer sizes have a well defined peak that is growing rapidly in size; however, there is not enough data to attempt a meaningful finite size scaling analysis. Fig. 3.12 also demonstrates that at cooler temperatures, the polymer merely localizes into a defined helical structure with a very narrow size distribution.

### 3.5.2 Scaling of Other Transitions

We now turn our attention to the nature of the coil-globular transition. The collapsing transition in stiff polymer systems is typically believed to be first order[72], due to a discontinuous change in the polymer's size at the $\Theta$ temperature. From a figure similar to Fig. 3.12 we see only a smooth crossover in system sizes. There are two reasons why we may not observe this discontinuous change in our simulation. First, the polymers that are modeled are very small, which has the tendency to broaden the transition due to surface effects[72]. We would expect to see some sort of first order behavior if we were in the range of $N >> 100$, therefore, there are significant deviations due to the small polymer sizes. The second reason for this discrepancy is that the polymer may not be fully collapsed prior to the helix

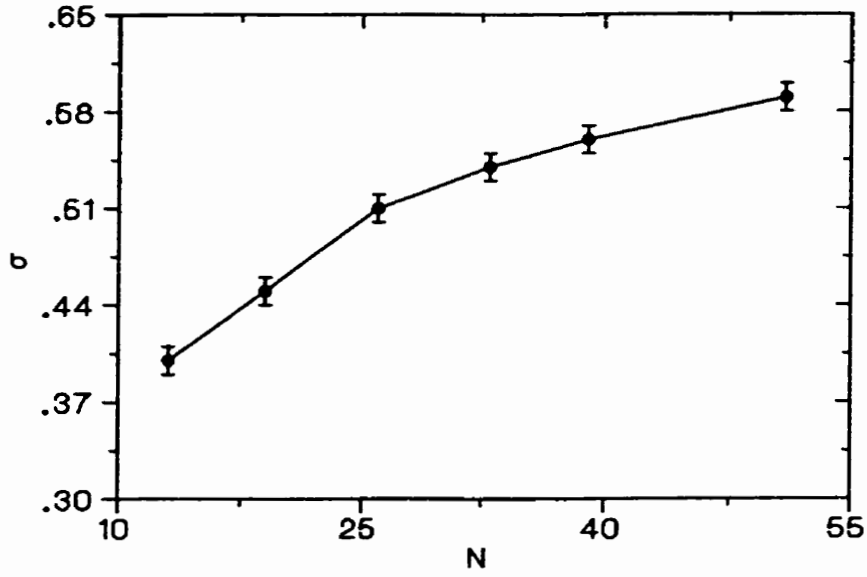


Figure 3.13: Foldability of the polymer as a function of the polymer's length. The figure shows that the foldability parameter is increasing with $N$, suggesting the polymer is becoming more difficult to fold.

transition because the coil-globular and globular-helix transitions occur so closely together. Therefore, this effect may mask any discontinuous changes in the radius of gyration. If we re-examine the susceptibility of the radius of gyration, we indeed see that the peak height increases with polymer size; however, this peak is not well defined. Although we do not observe the expected behavior in this case, we do know that the polymer undergoes a change between scaling regimes, supporting the argument for a phase transition.

An interesting result that arises from examining the location of the collapse

transition is the calculation of the foldability condition[76],

$$\sigma = \frac{T_\theta - T_c}{T_\theta} , \tag{3.18}$$

where $T_\theta$ is the collapsing transition temperature and $T_c$ is the globular to helix transition temperature. This parameter is typically used as a measure of folding properties of protein-like molecules. The faster the molecule folds, the smaller the value of $\sigma$. We have calculated this parameter as a function of system size for our model (Table 3.2). As the system size is increased, this parameter increases as well, suggesting that dynamically, it becomes more difficult to fold these structures as the system size is increased. As the foldability parameter is related to the cooperativity of the system[76], and in Zimm-Bragg's helix-coil transition theory the cooperativity is related to the nucleation parameter $\sigma_z$[73], the increase in the foldability parameter implies that the cooperativity of the system has dropped. This means that $\sigma_z$ increases as this drop occurs. Although it appears (see Fig. 3.13) that the rate of increase in $\sigma$ is slowing at large $N$, it is not clear whether it will reach a finite asymptotic value. This increase in the foldability parameter may be a manifestation of the instability of the system due to the one-dimensional nature of the model.

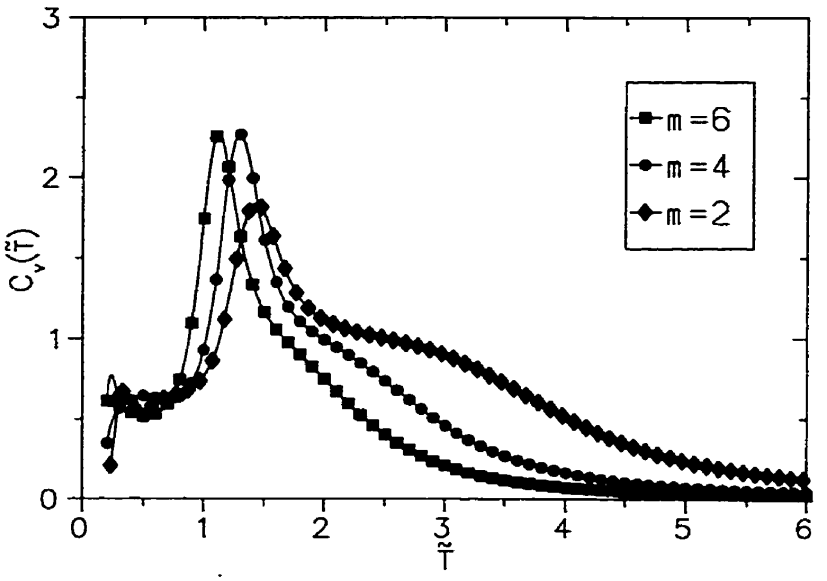


Figure 3.14: Specific heat for $N = 39$ with varying values of $m$.

## 3.6 Importance of Anisotropy

### 3.6.1 Cases m = 2,4

It is of interest to try to understand the effect that the power of $m$ has on the heat capacity curves. Clearly, the smaller the value of $m$, the more isotropic the interaction. When using these smaller powers of $m$, the other parameter values used in the model affects the ground state. For example, if a slightly larger fixed bond angle is used, the ground state can change from helical to a helical-like state that has a high helical content but is not a properly formed helix. One will also notice a larger number of compact globular structures are favored, making the sampling of low energy states more difficult.

Table 3.4: Temperature of the coil-helix transition, $T_c$, the specific heat at $T_c$, $C_v$, temperature of collapsing transition, $T_\Theta$, maximum of the susceptibilities for the radius of gyration at $T_\Theta$, and foldability parameter. In the $m = 0$ case there is no globular-helix transition. Therefore, the transition temperature, $T_c$, is the location of the transition that follows the globular transition as the temperature is lowered.

| $N$ | $m$ | $T_c$ | $C_v(T_c)$ | $T_\Theta$ | $\chi_{R_g}(T_\Theta)$ | $\sigma$ |
|---|---|---|---|---|---|---|
| 19 | 4 | 1.39(1) | 1.39(3) | 2.63(1) | 0.0762(7) | 0.47(1) |
| 39 | 4 | 1.29(1) | 2.27(9) | 3.19(1) | 0.182(1) | 0.60(1) |
| 19 | 2 | 1.72(1) | 1.32(5) | 3.56(1) | 0.0569(6) | 0.52(1) |
| 39 | 2 | 1.42(1) | 1.83(17) | 4.46(1) | 0.133(1) | 0.68(1) |
| 19 | 0 | 0.72(1) | 1.57(14) | 6.64(1) | 0.0304(4) | 0.89(2) |
| 26 | 0 | 0.82(1) | 1.36(17) | 7.84(1) | 0.0424(6) | 0.89(2) |
| 33 | 0 | 0.62(1) | 1.3(2) | 8.62(1) | 0.0552(8) | 0.92(2) |
| 39 | 0 | 1.30(2) | 1.1(3) | 9.12(2) | 0.0669(9) | 0.85(2) |

Simulating our model with a weaker anisotropic interaction in the potential invokes a change in the coil to globular transition. An increase in the isotropic nature of the potential creates a greater propensity for the molecule to exist in the globular state, hence the transition occurs at a higher temperature as seen in Fig. 3.14 for $N = 39$. The cusp in the heat capacity due to the globular transition is enlarged for the lower $m$ powers. We calculate the location and heights of these peaks for $m = 2, 4$ for the polymer sizes 19 and 39, which are displayed in Table 3.4. The locations of these peaks are found through the susceptibility of the radius of gyration, as was done for the $m = 6$ case. We also see that this isotropic nature of the transition has little effect on the second helix transition (near $\tilde{T} = 0.3$) because the net isotropic nature has little influence on the helical conformations.

One of the interesting debates in the field of helix folding is whether the col-

lapsing transition would actually occur in protein helices. A direct transition is assumed from the coil state to the helix state as suggested by the very name of the transition[27]. As well, all-atomic simulations of protein models such as the ones by Okamoto and Hansmann[33] do not show the globular collapse clearly, which suggests that in these simulations, the nature of the amino acid interactions is highly directionalized with few interactions not directed along the helical axis. The data from Table 3.4 is plotted in Fig. 3.15 and shows that the coil-globular and globular-helix transitions move closer together as the off-helical axis component is reduced. In those particular all-atomic simulations, the solvent effects have been ignored. Thus, our results suggest that in a predominately hydrophobic helix, one would observe a collapsing transition prior to the helical transition.

The influence of the anisotropy potentially has a profound effect on the foldability of the protein. As the value of $\sigma$ becomes smaller, the foldability of the sequence increases, as this parameter is believed to be closely linked to the cooperativity of the transition. Using the data in Tables 3.2 and 3.4, the parameter $\sigma$ has been calculated. The results suggest that as the value of $m$ is decreased, the foldability of the helix is also decreased. Hence, there should be marked improvement in the folding times of helical structures whose overall potential has a large degree of anisotropy. Fig. 3.16 shows the behavior of the parameter $\sigma$ as a function of the parameter $m$. The figure demonstrates that $\sigma$ is a decreasing function of $m$, and that an asymptote is approached suggesting a limiting value of the cooperativity for the transition.

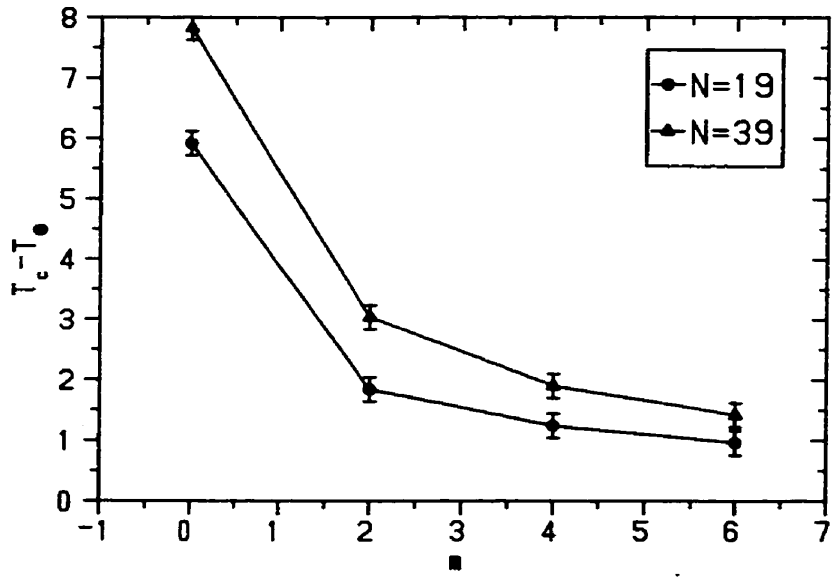


Figure 3.15: Separation of the coil-globular and globular-helix transitions as a function of the anisotropy. We see that the two transitions approach each other as the anisotropy is increased. In the $m = 0$ case there is no globular-helix transition. Therefore the separation is calculated as the difference between the high temperature globular transition and the second transition occurring at a lower temperature.

### 3.6.2 Case m=0: An Isotropic Potential

The other extreme case is an isotropic potential with $m = 0$. Here we are merely dealing with the square well potential. This type of polymer has been studied extensively in the past, most recently by Zhou et al.[67] and Irbäck et al.[68]. Zhou et al. modeled a flexible, off-lattice polymer with a square well potential. They show that the polymer first undergoes a collapsing transition followed by a first order phase transition. In a similar manner, Irbäck et al. studied flexible and semi-flexible, off-lattice polymers with a Lennard Jones potential. Here the results confirm the same collapsing transition, but no evidence of a first order transition

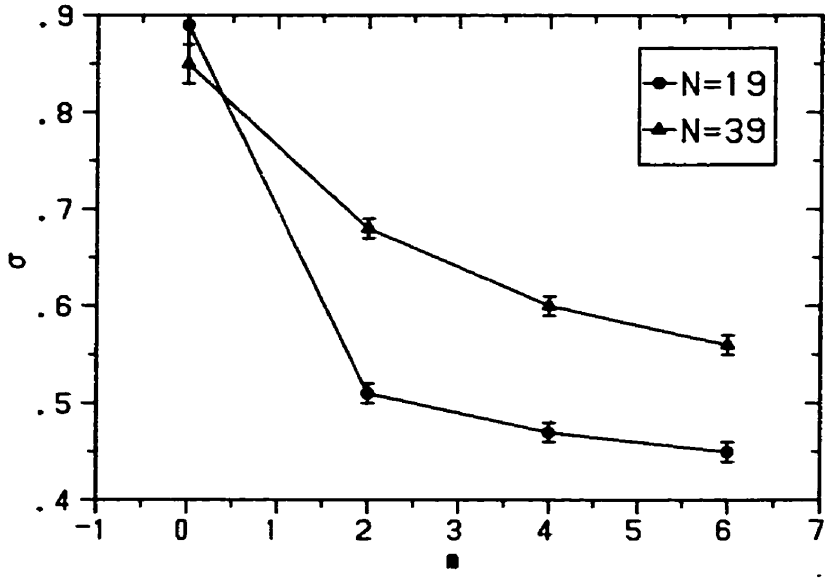


Figure 3.16: Foldability as a function of the anisotropy of the potential. The function appears to reach a limiting value as the anisotropy is increased. In the $m = 0$ case there is no globular-helix transition. Therefore the foldability is calculated from the difference between the high temperature globular transition and the second transition occurring at a lower temperature.

following the collapse. Our $m = 0$ case is equivalent to these two models because we simulate a stiff polymer off-lattice with a square well potential. Our results should yield some insight into these conflicting results.

For this type of system, we are not able to determine what is the native state since the low energy states can have significant structural differences. This makes sampling the low energy configurations difficult even with the multicanonical technique, as the number of low energy configurations is large and each configuration is somewhat disconnected from the others on the energy landscape. This leads to simulations which need to run longer in order to collect reasonable low temperature

information. The multicanonical technique relies on the simulations entering and leaving the low temperature states a repeated number of times. If the low temperature states are glassy, the simulation should technically enter and leave each of the glassy states several times in order to obtain accurate results. Therefore, a system with a large number of glassy states is not suited for the multicanonical technique used here, as the production run needs to be run a very long time; however, due to the fact that the model used here is so easily altered, to examine this case we attempt to study the system.

The specific heat curves for all polymer sizes studied are shown in Fig. 3.17. This figure illustrates two distinct features for the polymer sizes of interest. .The first smaller hump in the specific heat is a collapsing transition, while the second has been postulated to be a transition to a crystal-like state. We observe from the figure that the collapsing transition now occurs at a much higher temperature (Table 3.4). The location of the collapsing transition is calculated through the susceptibility of the radius of gyration. This transition occurs at a higher temperature in this model as the net potential on a monomer is greater. This translates into an increase in our parameter $\epsilon$ in Eq. 3.2.

The low temperature behavior is very interesting. In the study by Zhou et al. of a square well potential with a flexible chain, a first order phase transition to a crystalline structure is observed. Irbäck et al.'s simulations of a system with a Lennard Jones potential suggest a transition, but the peaks in the specific heat do not appear to be size dependent, thus implying no first order transition. Our model of a stiff polymer with isotropic square well attraction demonstrates a low

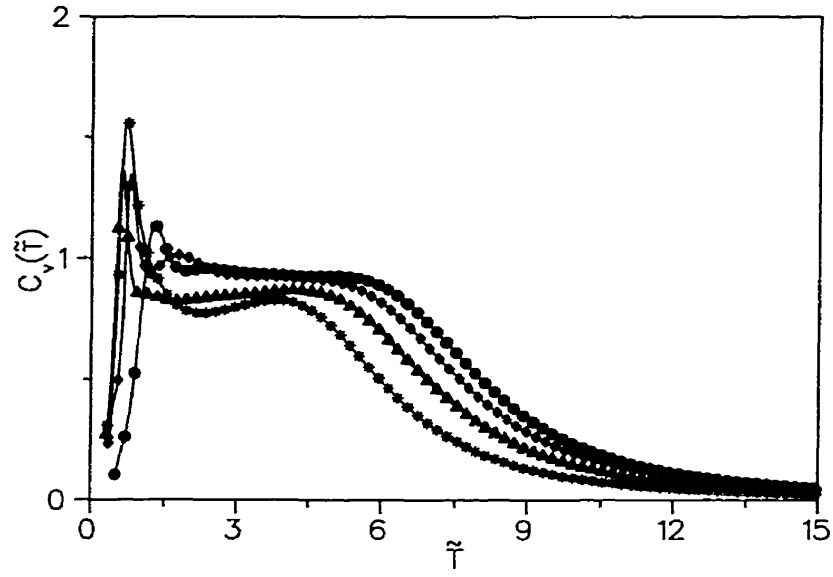


Figure 3.17: The specific heat for $m = 0$ for sizes 19 (∗), 26 (▲), 33 (♦), and 39 (•).

temperature transition that does not appear to have a size dependent anomaly in the specific heat. Due to the lack of a well-defined ground state, we are unable in these simulations to accurately determine the peak heights and locations for finite size scaling. The curves, however, do not show any significant difference in peak height for the low temperature range, and the overall shape of the curve demonstrates the same behavior as Irbäck et al.. Although these results are not conclusive, they lend support to the conclusion that this system does not have a first order phase transition.

## 3.7 Summary

We have presented a minimal model for examining the coil-helix transition in helical forming molecules. The potential is based on the non-local anisotropic interactions found in protein molecules, and it is versatile enough to allow one to study the various effects that it has on the coil-helix transition. We have successfully simulated helical states using potentials with the parameter $m = 2, 4, 6$, which has enabled us to obtain a systematic understanding of the structures generated by anisotropic potentials.

We have concentrated the bulk of our analysis on the $m = 6$ case. This has shown that the coil-helix transition involves four states: 1) coil, 2) molten-globular 3) helix I, and 4) helix II. The coil state is the typical repulsion-dominated state of polymer molecules. The globular state is a collapsed state formed prior to the transition to the helix. This state appears to be coupled to the helix transition through the content of anisotropy in the potential. As the anisotropy of the potential is increased the coil state is more likely to make a direct transition to the helical state. The helix I state is a poorly formed helix that is highly flexible, while the helix II state is a crystalline form of the helix. The final helix-helix transition that is observed appears to be due to a crystallization process in the helix, whereby the helix takes on a rigid rod-like conformation.

We have examined the nature of the coil-globular and globular-helix transitions within the model. We have attempted to show that the coil-globular transition is a first order phase transition, as expected for a semi-flexible polymer; however, the

small size of the polymers used have obscured the results. The finite size scaling of the globular-helix transition has demonstrated that this transition behaves like a first order phase transition. We also see that the polymer undergoes a discontinuous change in its size at the transition temperature, indicative of a first order transition. In the case where we edit our potential to create an isotropic interaction, we do not observe any phase transition-like behavior at low temperatures. Although the data is not conclusive, there is no significant increase in the low temperature heat capacity peaks to indicate a phase transition. These results lend support to the recent conclusion about such systems.

One of the most interesting results is the relationship between the relative temperature differences of the coil-globular and globular-helix transitions. These differences can be altered by changing the content of anisotropic interactions in the potential. Also, this content appears to control the cooperativity of the formation of helices, and as the potential is made more anisotropic, the foldability of the helix is increased. This suggests that the anisotropy in protein structures plays a key role in the ability of the molecule to fold quickly to a specific ground state. Thus, further study of the dynamic behavior and folding of the model is needed.

# Chapter 4

# Dynamic Helix Model

Attempting to understand the complex dynamic nature of proteins is one of the most challenging problems in molecular biology. As is discussed in chapter 1, there have been numerous minimal models created to attempt to understand the complicated funnel-like landscapes of proteins; however, most of the work regarding the study of secondary structure dynamics has been done with all-atomic models. There have been no systematic simulation studies of secondary structure dynamics; therefore, it is advantageous to create a minimal dynamic model to provide a method for performing such studies.

The minimal model created in chapter 3 has provided a method for studying the statistical properties of a helical segment; however, we have not been able to determine any dynamic features of the system. Extending the helix model to simulate the dynamics presents a good opportunity for examining the features of actual protein type objects through a minimal model. This would provide a systematic

method by which these structures could be studied.

## 4.1 Introduction

The lattice protein models have provided a great deal of insight into the dynamic behavior of proteins; however, they are limited in their comparison to a real protein system. It would be more advantageous to study minimal models in an off-lattice setting, where the system adopts more realistic protein structures. Perhaps the most popular model for off-lattice folding studies of proteins is the Gō model[17, 18]. Here a heteropolymer with isotropic interactions is used, and the interaction matrix is chosen such that the desired native state is the structure of interest. Although this method can be applied to a wide range of problems[17, 18, 19, 20, 77], it too has limitations. In this model, structures are created out of heterogeneity, which plays a dominant role in structure selection, but secondary structures also occur in homopolymers. This suggests that additional interactions should be included in order to capture properties not dependent on the heterogeneity of the sequence.

In this chapter, the model of the previous chapter is extended to examine the dynamics of helical structures. The dynamics of the model are simulated with a dynamic Monte Carlo algorithm and the potential is modified to accommodate for the local movement of monomers. The potential is still based on a non-local anisotropic interaction that selects a helical ground state structure. The generic and robust nature of this model allows for the systematic study of the folding properties of helices, and the effect of different perturbations on the basic model. In addition,

simple modifications, such as the addition of heterogeneity, could be made to the potential to create alternative protein structures. For example, $\beta$-sheets and helix-barrel structures would be possible candidates, as this model would not only be limited to the simulation of helices.

## 4.2 The Dynamic Model

To create the dynamic version of the model from chapter 3, the same steps are implemented to devise the overall features of the model. In this model, we are still interested in the non-local interactions within a helical segment; thus, to study effects associated with these interactions the three basic considerations of chapter 3 must be included in a basic polymer model: 1) persistence, 2) excluded volume interactions, and 3) an anisotropic potential directed perpendicular to the curvature of the helix.

To create the persistence effect, we use the same worm-like polymer chain as the backbone polymer. The sub-unit monomers, or amino acids, are represented as spheres of radius, $d$, and the monomers are connected together using rigid bonds of length $a$. Fixing the azimuthal bond angle created the worm-like chain previously used. This is a very restrictive constraint, as any small local movement of the polymer would involve the rearrangement of numerous monomers to maintain the constraint. To avoid this problem, we allow the bond angles to fluctuate slightly

under the energy constraint,

$$U_{\text{Angle}} = \sum_{i=2}^{N-1} \frac{K}{2}(\cos\theta_i - \cos\theta_0)^2 \ , \tag{4.1}$$

where $K$ represents the strength of the bond restoring force, and $\theta_0$ represents the average bond angle. This energy condition maintains the properties of the worm-like chain backbone while allowing for small local movements, as the potential is harmonic about the $cos\theta_0$ rather than $\theta_0$. The potential could be constructed such that the potential is harmonic about $\theta_0$; however, the backbone would not have the properties of a worm-like chain. Also, the bond angle will only fluctuate a small amount, so both possibilities for the bond constraint would be nearly equivalent.

To create the last two effects for helix formation, that of an excluded volume interaction and the anisotropic potential, we devise a modified Lennard Jones type bond interaction. This will be beneficial for the dynamics over the square well potential used earlier, as there will be no large discontinuities in the potential. As stated above, this potential must be anisotropic. To include this effect in the potential, a bond orientation vector, $\hat{e}$, is defined from the vector $\hat{u}$, which is a unit vector perpendicular to the bond angle plane (see Fig. 3.2). The vector $\hat{u}$ can also be used to define the orientation of the potential as it was in the previous model; however, both right and left handed helices are equally favored using this definition. There are two reasons for wishing to break the symmetry in the helices in this model. First, from a realistic perspective, left and right hand helices typically do not form in the same protein helical segment. Therefore, breaking this symmetry

is not an unreasonable addition. Second, the dynamics of folding are significantly different if both hands are allowed to form in the same segment, as domains with very long lifetimes would be created, thus constructing a rather unphysical picture when examining protein folding. These domains are equivalent to the domains in an Ising model that form at low temperatures with one up and one down. They are never resolved because the probability that one of the domains could be flipped is unlikely. To break the symmetry between the two helices the vector $\hat{u}$ is tilted to correct for the pitch of the helix so that a new vector $\hat{e}$ points directly along the helical axis when a helical structure is adopted. The vectors are defined as follows,

$$\hat{u}_i = |(\vec{r}_{i+1} - \vec{r}_i) \times (\vec{r}_i - \vec{r}_{i-1})| \tag{4.2}$$

$$\hat{c}_i = |(\vec{r}_{i+1} - \vec{r}_{i-1})| \tag{4.3}$$

$$\hat{e}_i = \sqrt{1 - d_s^2}\hat{u}_i + d_s\hat{c}_i \tag{4.4}$$

where $d_s$ is the separation of sub-units along the helical axis, such that the pitch is $p = nd_s$ given that there are $n$ monomers per loop. The form of the potential is then specified as,

$$U_{ij}(r) = \begin{cases} 0 & for\ \sigma \leq r_{ij} < \infty \\ V(r)\left[(\hat{e}_i \cdot \hat{r}_{ij}) \cdot (\hat{e}_j \cdot \hat{r}_{ij})\right]^m & for\ d \leq r_{ij} < \sigma \\ \infty & for\ 0 \leq r_{ij} < d \end{cases} \tag{4.5}$$

where $\hat{r}_{ij} = (\vec{r}_i - \vec{r}_j)/|\vec{r}_i - \vec{r}_j|$. The potential is truncated when r is beyond $\sigma$, the attraction radius, and a hard wall potential is used when r is below $d$, the excluded

volume radius. These conditions are implemented to minimize computational time. The function $V(r)$ is a Lennard Jones potential of the form,

$$V(r) = -\epsilon\left(\frac{a_0}{r^{12}} - \frac{a_1}{r^6}\right) \ . \tag{4.6}$$

where the parameters $a_0$ and $a_1$ are chosen to yield a continuous potential form between the limits $d$ and $\sigma$. The size of each monomer is selected to give a somewhat smooth contour to the shape of the worm-like chain, and has a value of $d = (3/2)a$, where $a$ is the bond length between monomers. The radius of the attractive interaction can have a range of valid values, and we select a value of $\sigma = \sqrt{45/8}a$. The parameter $m$ controls the anisotropy of the potential and is set to $m = 6$ as discussed in the previous chapter[57, 78]. The last two variables are $\epsilon$ and $K$. Typically, the interactions governing the bond angle fluctuations tend to be an order of magnitude larger than non-local interactions[22]. Therefore, $K$ is selected as $K = 10\epsilon$. The final adjustable parameter is $\epsilon$ which is scaled into the temperature to produce the reduced temperature unit $\tilde{T} = k_B T/\epsilon$.

To simulate the dynamic motion of particles a Monte Carlo simulation with local moves is used. In the simulation, a monomer is selected at random and rotated a small amount around the $\hat{c}_i$ unit vector as is discussed in chapter 2. This condition does not allow the bonds between adjacent monomers to fluctuate, thus keeping the bond separation between monomers constant. This is a reasonable assumption as covalent bond interactions are typically in the order of 100 times larger than the non-local bonding effect of interest in this model[22]. The distance the monomer is

pushed is randomly selected between $-\delta$ and $\delta$, where $\delta$ is chosen such that the rate of acceptance of moves is approximately 50% and has a value of 0.2. A single Monte Carlo time step consists of $N$ attempted moves of randomly selected monomers.

This model presents an ideal way to examine the folding behavior of a helix. By setting the temperature of our system below the coil-helix transition, we are able to confirm that the native state of this model is a helix. We use a reduced temperature of $\bar{T} = 0.6$ to obtain a stable helical state. Knowing this, the dynamic behavior of the model can be examined. Typically, in folding studies of minimal models of proteins, an understanding of the folding properties can be obtained from a study of the mean first passage times (MFPTs)[1, 11, 79]. The MFPT is defined as the time required for a molecule to first enter the native state when it is started in an arbitrary configuration. This time will be dependent on the temperature of the system[80], but this is not of interest in this study. The same temperature as above is used for all the simulations because helices will be stable at this temperature.

## 4.3 Helical Order Parameters

In addition to setting the temperature, an order parameter that characterizes the native state needs to be defined. Defining an order parameter that accurately describes the helicity in our model is not difficult. The problem lies in determining when the molecule is in the native helical state. To explain, the helical native state is not a single state because fluctuations allow the helical segment to flex. It is reasonable to assume that as long as the helical segment is not broken, then it can

be defined as the native state although the energy is not completely minimized. The only question that remains is, what defines a broken helical segment?

In a protein segment, helicity is defined by the Ramachandran angles $\phi$ and $\psi$. If all the angles fall within a specified range, the segment is considered to be in a helical form. The minimal model constructed here does not contain equivalent angles to the Ramachandran angles. This leaves us to define an alternate order parameter for helicity. For this model, we define two order parameters and use both as indicators that a helical segment has been reached. The first order parameter measures the local correlation of the bond orientation vectors and is defined as follows,

$$H_1 = \frac{1}{N_s - 1} \sum_{i=1}^{N-1} \hat{e}_i \cdot \hat{e}_{i+1} \ , \tag{4.7}$$

where $N_s$ is the number of helix forming monomers in the system. This parameter ensures that adjacent monomers, including the ends, are nearly helical. The second order parameter examines the net behavior of the bond orientation vectors and is defined as,

$$H_2 = \frac{1}{N_s} \left| \sum_{i=1}^{N} \hat{e}_i \right| \ . \tag{4.8}$$

This vector determines whether or not the segment is fractured near the middle. A break mid-segment may not produce a significant drop in the first order parameter, but it could produce a large variation in the second.

As both of these order parameters are continuous, we define an arbitrary value for each order parameter to define a segment as helical. When the values of both parameters reach a value greater than 0.95, the segment is considered helical. Spec-

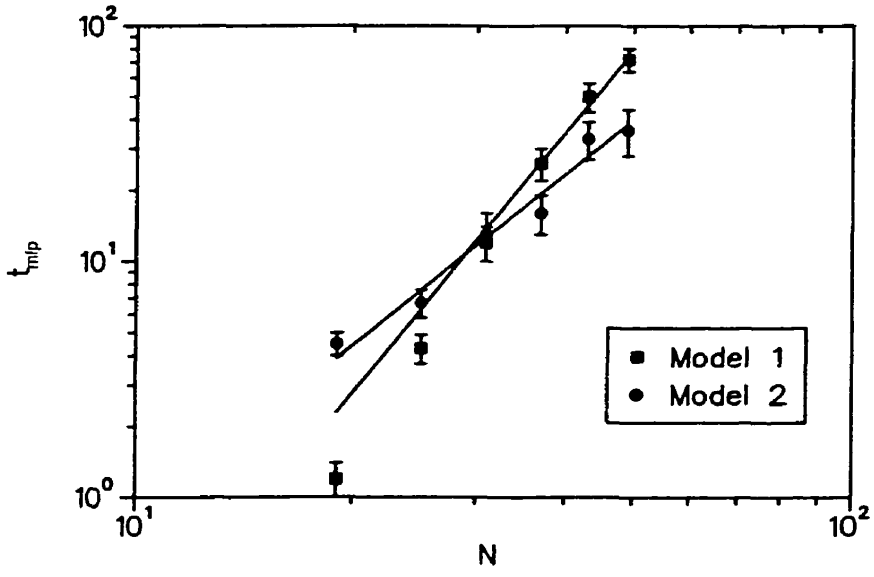


Figure 4.1: Scaling of average folding time vs. polymer length. Helical segment without tethered segments(■), and helical segment with two tethered segments (•).

ifying this value insures that the segments are near perfect helices, and should not affect the general dynamic trends we have set out to observe. A range of values near 0.95 should be acceptable because a value that is set too high to be representative of the allowable fluctuation should not affect the results significantly. This can be explained as follows. When a segment enters the native well, it should quickly find the bottom of the well on a time scale much shorter than the first passage time. It will not remain in the bottom for a extended period of time, as the structure will fluctuate within the well. Setting a high cutoff value should unsure accurate first passage time results, as it will ensure that the structure is in its native energy well with only marginal error to the MFPT.

Table 4.1: The folding data for a helical segment. $N$ is the total segment length, $N_s$ is the number of helix forming segments, $t_{\text{mfp}}$ is the mean first passage time, $t_{\text{max}}$ is the maximum folding time allowed, and % DNF, is the percentage that did not fold.

| $N$ | $N_s$ | $t_{\text{mfp}}(\times 10^3)$ | $t_{\text{max}}(\times 10^6)$ | % DNF |
|---|---|---|---|---|
| 19 | 19 | 1.2(2) | 10 | 33 |
| 25 | 25 | 4.3(6) | 20 | 10 |
| 31 | 31 | 12.(2) | 50 | 10 |
| 37 | 37 | 26.(4) | 100 | 10 |
| 43 | 43 | 50.(7) | 150 | 20 |
| 49 | 49 | 72.(8) | 200 | 15 |

## 4.4 Nucleation and Folding

### 4.4.1 Model I

Using this condition, we study the MFPTs of helical segments. To do this, the first passage times for 40 segments for each length of $N_s = 19$, 25, 31, 37, 43, and 49 were calculated. The MFPT was then determined and the results are shown in Table 4.1. Not all the segments reached a helical state in the maximum number of steps allowed for the simulation, but the percentage of non-folding segments is low when compared to other models[1]. In minimal lattice models, a segment which only folds 40% of the time is still a good folding protein. By examining the percentage of non-folding segments, we can conclude that our model is a good folder.

Work by Gutin et al.[79] showed that good folding protein sequences have MF-

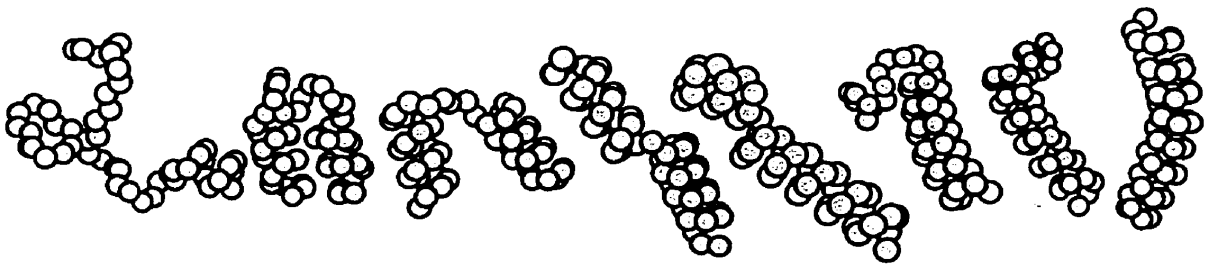

Figure 4.2: Illustration of a typical folding scenario in Model I for $N_s = 49$.

PTs that obey a power-law behavior when scaled with system size,

$$t_{mfp} \sim N^{\lambda} , \tag{4.9}$$

where $t_{mfp}$ is the mean first passage time, and $\lambda$ is the characteristic exponent. The exponent $\lambda$ varied depending on how the sequences were designed. For example, a randomly designed chain scaled with an exponent of $\lambda_{ran} \approx 6$, while a sequence designed from a Miyazawa and Jernigan[6] potential scaled with an exponent of $\lambda_{MJ} = 4.5$[79]. This showed that sequences designed from protein-like potentials were better folders when the sequences being folded were longer. Using the data from Table 4.1, we can determine the exponent associated with the helical model by plotting the data on a log-log plot as in Fig. 4.1. By fitting the data with a least squares method to Eq. 4.9, a value of $\lambda_{M1} = 3.7(2)$ is obtained. This exponent suggests that the model demonstrates characteristics of a well-designed protein sequence.

Knowing that the model dynamics are characteristic of a good folding protein, the typical folding process of a helix can be examined. In Fig. 4.2, a time lapse im-

age of the helical segment is shown as it folds. The figure clearly shows a nucleation propagation process at work, which is the expected mechanism for a cooperative system. In this model, the nucleation occurs at the ends of the chain and not in the middle. This has the obvious explanation that the mobility of the end molecules is significantly higher due to reduced confinement restrictions on movement of the end monomers. Once the nucleation of the end of the segment occurs, the helices propagate inwards. At some point, the two parts of the helical segment meet with a discontinuity. The initial helical formation is only a small fraction of the net folding time, while the resolution of the discontinuity requires the majority of the time. This happens as propagation of the helical segment occurs along the contour of the backbone. There is more freedom of movement in this direction, as the helical segment moves through longitudinal fluctuations. In some cases, the formed helical segment rotates in conjunction with the longitudinal motion facilitating the helical propagation. This contour propagation is also responsible for generating the discontinuities in the helical segment, as propagation is usually halted by sharp bends in the chain contour. Resolving the discontinuity requires transverse fluctuations, which are limited by the helical confining geometry. Therefore, this step in the folding process requires the largest fraction of the total folding time. This can be further demonstrated by Fig. 4.3, which shows the behavior of the two order parameters for a typical folding event. The local helicity order parameter, $H_1$, increases rapidly at the beginning showing that the helical content is rising significantly. The global order parameter, $H_2$, also increases but then undergoes large variations as transverse fluctuations of the molecule occur.

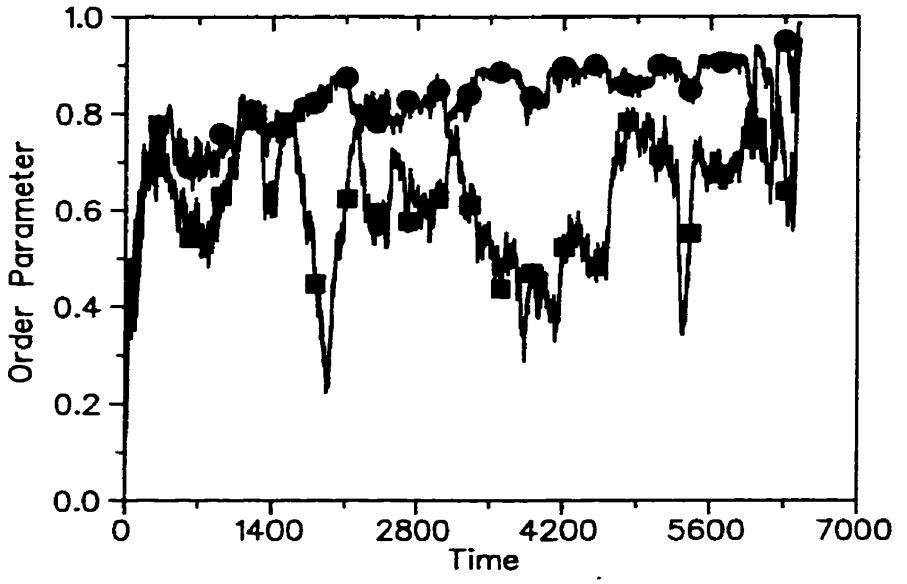


Figure 4.3: Typical development of the order parameters with time for $N = 49$. The local parameter $H_1$ (•) and the global parameter $H_2$ (■) are shown on the graph. The fluctuation in the global parameter demonstrates the resolution of a discontinuity.

### 4.4.2 Model II

This leads to the question, can the folding times of the helical segments be improved by altering the nucleation properties of the chain? More importantly, is the scaling behavior of the chain altered by such changes? These questions are answered by examining a slightly modified version of the helical model presented. In this model, two segments of six monomers are attached to the ends of the helical segment, but the segments only interact through an excluded volume interaction. This has two effects on the system. First, the non-attractive monomers increase the weight of the system, thereby creating a system with slower characteristic fluctuation times.

This means an overall slower dynamic behavior. Second, the end monomers of the attractive segment will no longer have increased mobility and will have nearly the same likelihood of nucleation as the interior monomers. For this model, the chain is now considered helical if the attractive monomers, not the added segments, meet the requirements stated above for helicity.

It would appear that adding the two non-attractive segments would have a net effect of slowing the dynamics of the segment; however, this is not the case. By adding the additional segments, the MFPTs are decreased. Table 4.2 shows the MFPTs and it is clear that for the larger segments, the folding times are nearly 50% shorter. This dramatic decrease in the folding times is accounted for in the following way. The reduction of the ability of the end segments to nucleate the helix causes a more uniform distribution of nucleation sites. Also, the net probability of creating a nucleation site is now much lower. This means that the initial nucleation is longer, but that a nucleation site that already exists has a much longer time to propagate through the entire segment before a second nucleation site occurs. Thus, there is a significant reduction in the folding times as a discontinuity in the segment does not have to be resolved. Fig. 4.4 illustrates this folding process where only a single nucleation site is formed and the segment propagates throughout the segment. Note that the nucleation site is located in the middle of the segment. It should be mentioned that multiple nucleation sites can still occur. This results in a dislocation that needs to be resolved, and hence, an increase in the folding times; however, these events occur less frequently. In the first model, a discontinuity was formed in every simulation, while in the second model, a discontinuity only formed

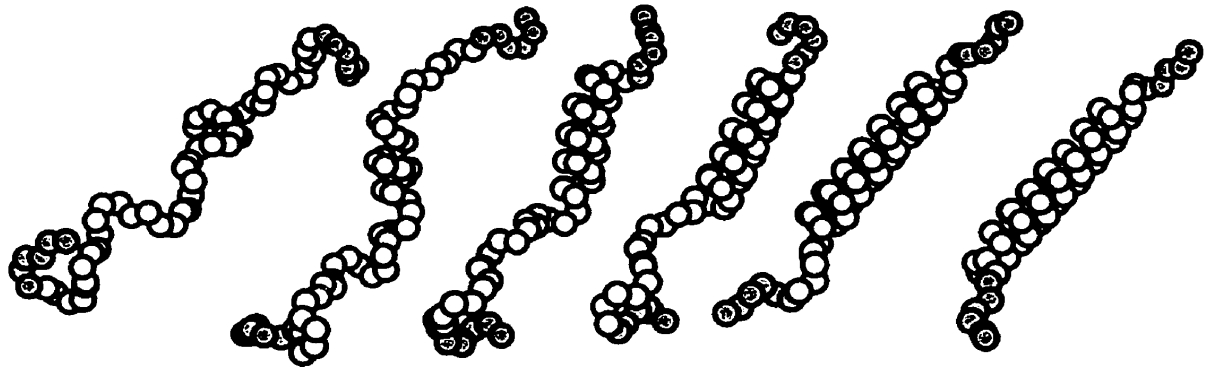

Figure 4.4: An illustration of a folding event for $N_s = 49$ in Model II where only a single nucleation site is formed. Note the central location of the nucleation site. The dark monomers are those with no attractive potential.

in approximately 50% of the simulations. This is determined by visual inspection.

Also, the added segments have the unexpected effect of assisting the resolution of discontinuities, as the end segments decrease the longitudinal motion and increase the transverse motion. In rare events, a complete helical segment is formed with the opposite hand. This keeps both order parameters small enough that a good measure of the MFPT cannot be obtained, but this only occurs in a limited number of simulations and these segments are therefore considered non-folders. These two effects account for the observed folding times, and it is clear that the reduction in the number of nucleation sites plays a very significant role in the folding times.

Using the data in Table 4.2, a log-log plot of the data has been made in Fig. 4.1. The same power-law behavior that was observed for the previous model remains in this model. The exponent, which is calculated from a least squares fit to the data, has a value of $\lambda_{M2} = 2.4(3)$. This is significantly lower than the exponent for the

Table 4.2: The folding data for a helical segment with non-helix forming segments tethered to both ends.

| $N$ | $N_s$ | $t_{\mathrm{mfp}}(\times 10^3)$ | $t_{\max}(\times 10^6)$ | % DNF |
|---|---|---|---|---|
| 31 | 19 | 4.5(5) | 25 | 33 |
| 37 | 25 | 6.7(9) | 50 | 20 |
| 43 | 31 | 13.(3) | 100 | 25 |
| 49 | 37 | 16.(3) | 150 | 23 |
| 55 | 43 | 33.(6) | 200 | 15 |
| 61 | 49 | 36.(8) | 250 | 13 |

previous model, and demonstrates that the folding is fundamentally different for this model. It shows that longer helices can be formed if the probability of seeding a segment is relatively small, and if there is only one nucleation site.

### 4.4.3 Model III

To fully understand the folding behavior of this type of model, we examine yet a third similar model. In this model, we attach only a single non-attractive segment to one end of the segment. In this case, we might expect to see a third set of dynamics. We measure the helicity of a segment in the same way as the second model by only considering the attractive segments. The folding times for this model are shown in Table 4.3. The data in this table is calculated based on twenty simulations and is not sufficiently accurate to obtain a reliable value for the exponent; however, the exponent is estimated from this data and we believe it should lie between the values of the other two models. The exponent calculated from a least squares fit has a value of $\lambda_{M3} = 3.5(1.0)$.

Table 4.3: The folding data for a helical segment with a non-helix forming segment tethered to one end.

| $N$ | $N_s$ | $t_{\mathrm{mfp}}(\times 10^3)$ | $t_{\max}(\times 10^6)$ | % DNF |
|---|---|---|---|---|
| 25 | 19 | 3.0(5) | 25 | 20 |
| 31 | 25 | 7.0(8) | 50 | 35 |
| 37 | 31 | 10.(2) | 100 | 35 |
| 43 | 37 | 39.(5) | 150 | 40 |
| 49 | 43 | 28.(5) | 200 | 15 |
| 55 | 49 | 73.(14) | 250 | 10 |

The dynamics of this model are a combination of the two already discussed. The nucleation of the helical segment occurs at the free end of the segment, as in the first model. This is likely to be the only nucleation site thus reducing the probability of having to resolve a discontinuity; however, the propagation of the helical segment occurs through longitudinal fluctuations along the chain contour and can be retarded by sharp changes in the chain contour. This slowing of the propagation often provides an opportunity to generate a second nucleation site in the remaining segment which gives rise to a discontinuity that retards the dynamics. Only approximately 15% of the segments now fold with a single nucleation site. A comparison of the folding times to the first and second models shows that this model is relatively similar to the first model, but a slightly smaller exponent for this model would be projected because 15% can fold with a single nucleation site. Therefore, it appears that the exponent is far more sensitive to the probability of multiple nucleation sites, as more than one nucleation site is likely to decrease the foldability of the segment.

These results also demonstrate the significance of "hot" sites, or conserved residues[12, 13], within a protein. These sites are the key nucleation regions of the folding process and form first in order to seed the creation of the native state. The "hot" sites are also responsible for the stability of the structure as their mutation has a large effect on foldability. The results above are in agreement with these observations, but we can add an additional conclusion. It is not important to have a dominating nucleation site as long as propagation can proceed throughout the entire segment without an alternate nucleation site forming and conflicting with the folding; however, the simplest way that this can be attained is to have a region that has a higher probability of seeding the segment, which would increase the time for propagation through the structure.

## 4.5 Potential Anisotropy and Folding

As discussed in chapter 3, the anisotropy of the potential was projected to play a significant role in the foldability of the helical segment. The results of Klimov and Thirumalai[76] projected that the foldability of a protein is related to the relative separations of the coil-globular and globular-folded transition through the parameter $\sigma$(see Eq. 3.18). In chapter 3, we showed that decreasing the anisotropy increases the value of $\sigma$, which is projected to increase the folding times. This dynamic model presents an ideal method for studying this proposal.

To conduct this simulation, will use the second model of a helical segment with two non-interacting segments attached. This model is used as it demonstrates

the best folding characteristics. In the above simulations, the potential used a parameter of $m = 6$. For this simulation, a value of $m = 2$ is used. The MFPTs were then collected for the same helical segment lengths as in Model II, using twenty simulations in the average. The data collected is shown in Table 4.4. Further, the exponent for this model was determined to be $\lambda_{m=2} = 3.1(5)$.

Comparing the data from Tables 4.2 and 4.4, as well as the corresponding exponents, shows the projected reduction in the folding times. The folding times for the $m = 2$ model are much larger, and the exponent characterizing the scaling behavior of the system is also larger. In this model, the folding dynamics should be the same as those observed in the $m = 6$ model; however, a much slower folding and different scaling behavior is observed. What is different about the dynamics when the anisotropy is reduced? As the results from the previous chapter suggest, the stability of the globular state has been changed. This increased stability brought on by the reduction in anisotropy should reduce the probability of nucleation and decrease the rate of helical propagation. This would account for the observed increase in the exponent and why the long helical segments require considerably more time to fold.

## 4.6 Summary

In this chapter, a dynamic minimal model of a helix forming polymer segment was presented. The model utilizes a non-local interaction to create a segment with a helical ground state, similar to non-local interactions in proteins such as dipoles and

Table 4.4: The folding data for a helical segment using a value of $m = 2$ for the potential with non-helix forming segments tethered to both ends.

| $N$ | $N_s$ | $t_{\text{mfp}}(\times 10^3)$ | $t_{\text{max}}(\times 10^6)$ | % DNF |
|---|---|---|---|---|
| 31 | 19 | 2.4(4) | 25 | 40 |
| 37 | 25 | 6.6(2) | 50 | 15 |
| 43 | 31 | 14.(6) | 100 | 30 |
| 49 | 37 | 29.(10) | 150 | 40 |
| 55 | 43 | 59.(16) | 200 | 40 |
| 61 | 49 | 69.(18) | 250 | 35 |

hydrogen bonding. The model creates the helical ground state by using a persistent polymer backbone and directionalized potential directed along the helical axis. This model is sufficiently general that other protein structures could also be simulated using this model.

By studying the dynamics of this model, we have shown that the MFPTs from the coil state to the helical state scale as a power-law with system size, as is expected for a protein-like system. The calculated power-law exponents were shown to be consistent with other toy protein models. In addition, MFPTs are sensitive to the nucleation probability of the segment, as is expected for a protein system; however, not only are the times altered, but the scaling exponents for the system are altered as well. It is also clear that nucleation is not the only important factor in folding, and that the relationship between propagation and nucleation is the dominating factor in creating segments with fast folding characteristics.

In addition to studying the effects of nucleation on helix formation, the effect that anisotropy in the potential has on folding was also examined. The projection

from the previous chapter of a increase in folding times due to a decrease in the anisotropy of the potential was confirmed. Not only were folding times longer, but the scaling exponent was also altered by the changes to the potential. These results support the conclusion that fast-folding proteins prefer to fold in an "all or nothing" type of process.

# Chapter 5

# Lattice Prions

Prions are perhaps the greatest challenge to theorists attempting to understand proteins. As has already been discussed, proteins have the characteristic feature of a funnel-like landscape that directs the folding of a sequence to a single native state. Prions appear to be the exception to this rule, as these mysterious proteins apparently have two native states. This presents a very perplexing problem: what is different about the energy landscape of a prion that gives it this dual native state property? Not only are there two native states, but one state is favored during folding in such a way that it behaves similarly to a typical fast-folding protein. This feature has probably given prions the chance to develop biological functionality.

## 5.1 Introduction to Prions

Prions are protein structures that have just recently come to the forefront of scientific interest. They have been implicated as the cause of some rare diseases, such as Bovine Spongiform Encephalopathy ("Mad Cow Disease"), Creutzfeldt-Jakob Disease, Kuru, and Scrapies. These diseases are caused by one of the two native states of the prion. Most of the time, a prion will adopt one structure, but in odd situations, the prion will take on the other form. For convenience, the state to which a prion typically folds will be referred to as the native state. The other state to which a prion folds on rare occasions, will be referred to as the prion state. It is the prion state that causes the diseases mentioned above.

There is also a rather novel transmission mechanism that accompanies these diseases. Stanley Prusiner[81] has proposed the protein only hypothesis for which he won a Nobel Prize in 1997. Although this model is not completely proven, there is mounting evidence to its validity[82]. In this model for the transmission of an infectious agent, there is no need for nucleic acids such RNA and DNA, as the infectious agent is the protein in the prion state. The term prion is used to denote a proteinaceous infectious particle. The idea that proteins themselves can transmit diseases is a novel and fascinating proposition. In the prion diseases mentioned above, the protein in the prion state acts as a template to convert the already present proteins, which are in the native state, to proteins in the prion state. There are several possible scenarios for the conversion process, but none have been proven correct[83]. In each of the conversion mechanisms, the process

needs to be seeded by the infectious prion. There are several possible theories for the origin of the infectious prion. One theory is that the infectious protein forms are consumed and accumulated in the body because the infectious form of the prion is protease resistant[83]. A second theory is that the infectious particles originate spontaneously within the body through sufficient protein mutation causing the protein to favor the infectious form over its regular structure. Another possibility is the occurance of a rare kinetic event that causes the structure to adopt the infectious form. All of these are possible scenarios for the origin of the infectious prion particles.

In all of these diseases, there is a common prion protein $PrP^C$, which is only slightly genetically different in each disease. In each case, the prion gene sequence PrP has two different protein conformations, $PrP^C$ and $PrP^{Sc}$. $PrP^{Sc}$ is the infectious form of the PrP protein. These two proteins have no detectable sequence differences and are formed from the same gene[84]. The only difference lies in the conformation adopted. The $PrP^C$ state is predominately $\alpha$-helix with little $\beta$-sheet structure, while the $PrP^{Sc}$ state contains a large fraction of $\beta$-sheet.

There are also some interesting properties to these proteins that are related to their ability to transmit diseases. For example, the conversion process can occur across species (i.e. "Mad Cow Disease" prions in cows can cause Creutzfeldt-Jakob disease in humans); however, infection is more efficient if the genetic sequences of the PrP genes of the two species differ only by a small amount. This implies that the disease can only be transmitted between species in some special cases. For example, humans would not be able to contract Creutzefeldt-Jakob disease from

the $PrP^{Sc}$ proteins in mice (Scrapies). Also, $PrP^{Sc}$ can be denatured using GdnSCN into an unfolded state, but the process is not reversible[85]; however, under some conditions such as altered salinity and acidity, the $PrP^{Sc}$ infectious state can be recovered[86].

Understanding the properties and folding mechanisms of prions has been of great interest, especially in the last ten years, as these proteins have such unusual properties. Lattice models have also been used to attempt to understand the novel properties of these proteins, and work by Abkevich et al.[39] has shown that a protein's energy landscape can be kinetically partitioned. This simply means that two states of similar energy could have different folding times such that one state is more kinetically accessible than the other. This partitioning is believed to originate from the differences in the number of local and non-local contacts in the two conformations[87]. In this chapter, we use lattice models to further examine the landscape of prion molecules. In contrast to the approach of Abkevich et al., who designed their prion molecules to have specific structural features, we attempt to identify prion sequences in a simple 27-mer model based strictly on the sequence's energy spectrum and folding properties.

## 5.2 Model

For this study, the 27-mer model discussed in chapter 1 is used in conjunction with the dynamic moves for a cubic lattice discussed in chapter 2. This model is the same as those used by Šali et al.[1]. An interaction matrix representing the

forces between the different amino acids in the chain is generated. The interactions are chosen randomly from a Gaussian distribution centered at -2.0 and standard deviation 1.0. Each matrix that is generated is a particular realization of a sequence. The protein-like structures for these sequences are considered arrangements of the sequence in a cubic structure, as cubes are the most compact chain arrangements and contain the maximum number of contacts. The configuration that will be adopted by the sequence as the native structure is the lowest energy cube. This state is deemed the native state because it is probably the lowest possible energy configuration, which does not have to be a cubic configuration. For most matrices, the lowest energy cube is the lowest energy state, but this is not always the case. The simulations of each sequence are conducted at different temperatures. The temperature is set such that the following criterion is met[11],

$$\sum_{i=\mathrm{CSA}} \left( \frac{e^{-\frac{E_i}{k_B T}}}{\sum_{i=\mathrm{CSA}} e^{-\frac{E_i}{k_B T}}} \right)^2 = 0.2 \ , \tag{5.1}$$

where CSA are all the compact self-avoiding states or cubic arrangements. This condition for determining the temperature has been suggested by Šali et al. to be a good estimate for the ideal temperature at which the simulations should be conducted[1].

Šali et al. define a random sequence as protein-like as one that can reach its native state 40% of the time in a simulation five times longer than the mean first passage time (MFPT) for a typical fast-folding sequence. We use this criterion to look for prion-like sequences in this 27-mer model. In order to examine the

properties of prions, we first must determine the sequences that exhibit the behavior of a prion. The criteria used to label a sequence as prion-like are as follows:

1) The two lowest energy compact states must have similar energies, and are referred to as the native states of the sequence

2) The two native states must be significantly separated from the energy spectrum of the other compact states by a large gap.

3) One of the native states must be a good folder.

4) The two native states must have considerably different structures.

5) A large energy barrier must kinetically separate the native states.

These conditions have the following meanings. Condition 1 originates from experimental evidence that suggests the two conformations need only be marginally different in stability[83]. The second condition arises from the work by Šali et al. that shows protein-like sequences have large energy gaps separating the native state from the spectrum of compact states. We require in condition 3 that at least one sequence be a good-folding sequence. This implies that the two native states should be separated by a significant energy gap from the spectrum of compact states. Condition 4 comes from experimental evidence which shows that prion structures have major structural differences. The final condition implies that if these two sequences are to be kinetically partitioned, then the two native states must be separated by a large energy barrier.

These rules are implemented in the simulation as follows. First, random sequences are generated and the energies of all the compact configurations are determined. Sequences with an energy spectrum similar to the one shown in Fig. 5.1a

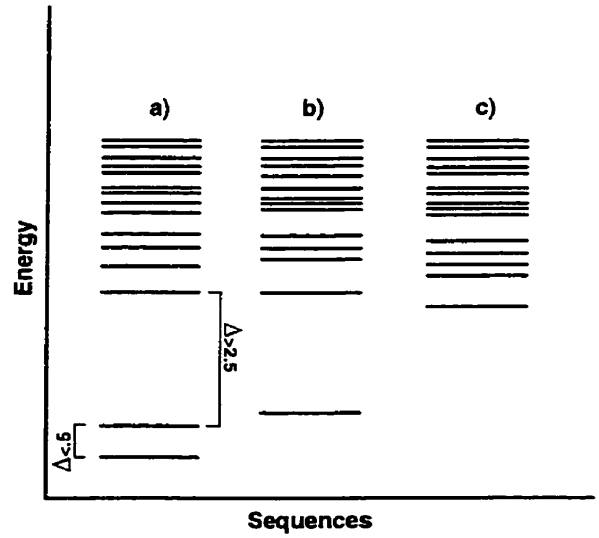


Figure 5.1: a) The energy spectrum of a prion-like sequence. b) The energy spectrum of a good folding sequence. Note the large gap between the lowest and second lowest energy states. c) The energy spectrum of a sequence with glassy characteristics.

are deemed possible prion candidates. The spectrum is based on conditions 1 and 2, and is constructed by accepting sequences that have two native states separated by less than 0.5, and a gap of greater than 2.5 between the second and third lowest energy cubes. The sequences that meet this criterion then have their native state structures examined for similarities. These structural differences are accessed by determining the number of monomers situated in the same positions of the cube. To identify these monomers, the cubic structures are written out with the same orientation. If structures have less than three monomers in the same position, the structures are considered sufficiently different. This method is only one possible choice for determining whether two structures have different conformations

that would give them a higher probability of being kinetically partitioned. This method was chosen as it is computationally simple and does not constrain the type of bonding that could occur, as bonding is believed to be important for kinetically partitioning the two native states. About 0.03% of the sequences sampled meet these energy and structural criteria.

The sequences that are obtained from the above method are then folded 50 times for $1 \times 10^9$ steps. The results of the simulation are studied to determine if conditions 3 and 5 are met. Those sequences that meet these requirements are called prion-like. To facilitate the discussion, the lowest energy native state will be referred to as $N_1$ and the other native state will be $N_2$. In the folding results, we look for sequences that meet the good folding condition. If either the $N_1$ or $N_2$ state is reached in 40% of the simulations, then the sequence is considered a good folder. We are only interested in sequences that exhibit good folding behavior to one of the two native states, as $PrP^C$ is know to be a very fast-folding molecule[88]. We also look for the kinetic accessibility of the other native state in the folding process. In other words, is the other state typically part of the folding pathway? We are only interested in those sequences that are kinetically partitioned by a large energy barrier. Therefore, the unpreferred native state should rarely be visited in folding events when compared with the preferred native state. By keeping track of the compact states visited during each folding event, we are able to obtain an idea of the kinetic partitioning that occurs.

## 5.3 Results

For this study, 273 different sequences were examined that had the required constraints on the energy spectrum and structural differences. Most of the sequences fell into one of two categories. The majority of the sequences were good-folding sequences, but also had the secondary native state visited during a large fraction of the folding events. This implied that the secondary state was not very kinetically partitioned from the primary native state, and would only be a transient state in the folding pathway. We also found that in most of these cases, the primary native state is the $N_1$ state. In the other cases, the sequences did not appear to be good folding sequences, as neither the $N_1$ nor the $N_2$ states were visited repeatedly.

Among these sequences, we found several sequences that demonstrated unusual properties suggesting that these sequences are prion-like. We examined two of these sequences in detail. Both sequences where given considerably more computational effort and were folded 2000 times each. The first sequence (labeled Sequence 54) demonstrates the characteristic properties of a prion that would be observed experimentally. In real prion systems, the $PrP^C$ state is believed to be marginally less stable than the infectious $PrP^{Sc}$[83] state, and under typical biological conditions, the $PrP^C$ is the favored state in the folding pathway. Sequence 54 shows the same behavior as these proteins because 72.1% of the time the native state, $N_2$ (the higher energy state), is reached without passing through the native state, $N_1$. In addition, in 6.8% of the folding simulations, the native state, $N_1$, is reached without passing though the $N_2$ state. In approximately 1.7% of the simulations, the sequence

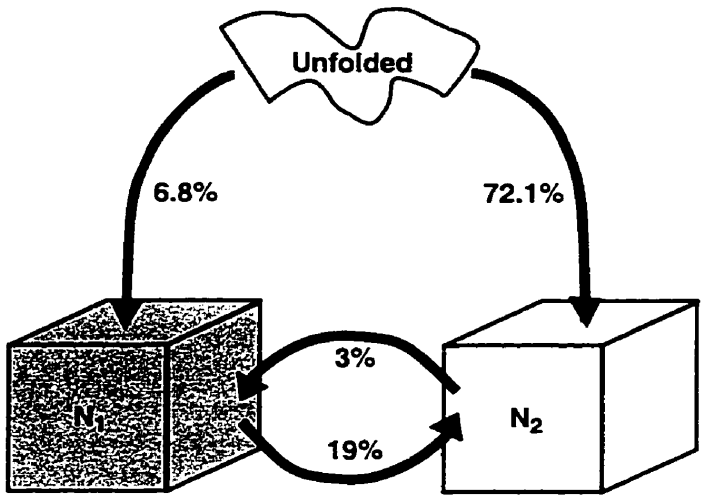


Figure 5.2: Diagram displaying the successful folding percentages between the various states for Sequence 54. The remainder of the folding events not shown are unsuccessful folding events where the target configurations were not reached. The results beginning in the unfolded states are based on 2000 simulations, while the results between the two native states are obtained from 100 simulations each. Each simulation is $1 \times 10^9$ Monte Carlo steps.

passed through the $N_2$ state before entering the $N_1$ state. In the remainder of the simulations, the lowest energy structures reached were a variety of higher energy cubes in which we have little interest. To further understand the folding behavior observed, 100 simulations of the same length were conducted, where the simulation started from the $N_2$ state and was folded to the $N_1$ state. Another 100 simulations were conducted in the reverse direction, starting from the $N_1$ state and folded to the $N_2$ state. In the simulation from $N_2$ to $N_1$, only 3% of the simulations reached the $N_1$ state. In the reverse situation, 19% of the simulations reached the $N_2$ state from the $N_1$ state. These results are depicted in a diagram in Fig. 5.2.

The above results for Sequence 54 demonstrate prion-like behavior even in this very simplified toy system of proteins. The native state, $N_2$, is a fast-folding structure. The alternate native state, $N_1$, also appears to be kinetically partitioned from the general folding landscape as is suggested by the small percentage of crossovers from one native state to the other. This data further suggests that there are two separate folding funnel pathways which are depicted in Fig. 5.3; this idea was first presented by Abkevich et al.[39]. We infer this from the fact that the percentage of crossover events is not symmetric, and that there is a higher probability of moving from $N_1$ to $N_2$. Also, in an intermediate state model where a protein folds first to a stable intermediate and then to its ground state, we would expect to observe the percentages in the reverse order. This senario is highly unlikely as it has been shown both in lattice models[10] and in experiments[89] that fast-folding proteins do not fold with the aid of intermediates. The temperatures at which the simulations are conducted do not exclude the possibility of the structure unfolding and refolding. Therefore, once the structure has unfolded, there is a higher probability to favor the $N_2$ funnel over the $N_1$ funnel. This explains the asymmetry in the observed crossover folding results. To further study this behavior, the temperature of the system was lowered by 5%. At the reduced temperature, folding to the two native states was still achieved with approximately the same success rates, while the crossover between the native states was nearly eliminated. This further supports the idea of two seperate folding funnels.

Using the idea that the landscape of a prion is separated into separate funnels, we estimate the MFPTs of the direct folding events to each of the native states.

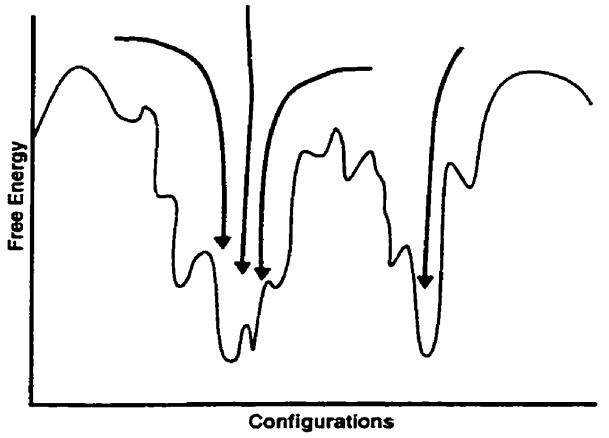


Figure 5.3: This is a representation of the landscape of a prion-like sequence. In this landscape there are two deep energy minima separated by a significant energy barrier.

The data for these times is shown in Table 5.1. From the data, it is clear that folding to the native state, $N_2$, occurs quickly, and that the folding times to the $N_1$ state are of similar magnitude. These times are calculated based on the fact that the other native state is not encountered during the simulation. This is not in contradiction with previous results which suggest that the folding times for one of the two native states should be much longer than the other in kinetically partitioned landscapes[39]. Our results are based on the structure not encountering the kinetic trap of the other native state. If we included the time spent in the other energy minimum, then the folding times to the $N_1$ state would be much longer. The folding events in which the sequence becomes kinetically trapped in the other energy minimum for some period of time are excluded from these results because a real protein system would have significantly deeper energy minima. A very deep energy

minimum would make an unassisted, spontaneous structure change unlikely. Thus, the favoring of one funnel over the next is related to the availability of transition states to the structure[39]. The folding times for the crossover events are presented for only those that were able to make the transition in the time allotted, which is about ten times longer than the average MFPTs for a typical fast-folding protein sequence. These times, in theory, should vary greatly depending on the length of time spent in the respective energy minima, and the probability of encountering the alternate transition states to the other native state. Thus, the results are not representative of the mean folding times between energy minima as we end the simulations at $1 \times 10^9$ steps; however, comparing these results demonstrates the kinetic partitioning of the two states because the mean crossover times for the successful simulations are twice as long as the MFPTs. As these crossover times are only representative of those simulations that fold the fastest, we expect the actual crossover times to be much longer. Another point of interest in regards to the MFPTs is that they are approximately three times longer than those sequences that do not exhibit any kinetic partitioning. This suggests that the transition states are quite structurally different, and that the added competition to direct the folding slows the overall dynamics of these molecules.

The structures that this sequence adopts are not designed to have any particular properties; this provides an opportunity to study the structural features for clues as to the nature of the partitioning. The two states for the sequence are shown in Fig. 5.4, and the energies for the $N_1$ and $N_2$ are -78.596 and -78.413 respectively. This is an energy difference of 0.183, while the energy gap between the $N_2$ state

Table 5.1: The folding times for Sequence 54. Shown is the number of events for each average, followed by the mean first passage time.

| Path | No. Events | Time |
|---|---|---|
| $U \rightarrow N_1$ | 136 | $3.0(3) \times 10^8$ |
| $U \rightarrow N_2$ | 1442 | $3.24(7) \times 10^8$ |
| $N_1 \rightarrow N_2$ | 19 | $6.0(6) \times 10^8$ |
| $N_2 \rightarrow N_1$ | 3 | $6.2(7) \times 10^8$ |

and the next lowest energy cube is 2.541. The analysis of direct structural features such as preferred geometry is difficult to deduce because the system size is so small; however, some speculation on the nature of the bonding occurring can be done.

To analyze the nature of the bonding it is constructive to break the types of bonds down in the following way: 1) local vs. non-local, and 2) similar vs. different. Local bonds are defined as bonds that form between the $i^{th}$ and $i^{th}+3$ monomers, and all the rest are considered non-local. This type of distinction has been considered very important in understanding the folding nature of proteins[87]. Similar bonds are those bonds that occur in both native structures. Distinguishing between the similar and different bonds will be important in discussions of prion structures because the different bonds will probably play a role in determining the folding behavior. In Fig. 5.4, the similar and different-local bonds are shown to assist in a discussion of the folding process.

Table 5.2 presents a breakdown of the bonds into the various classes. The number of bonds, the average bond strength, and the standard deviations in the average, are all given. In this sequence, both the similar-local and similar-non-local

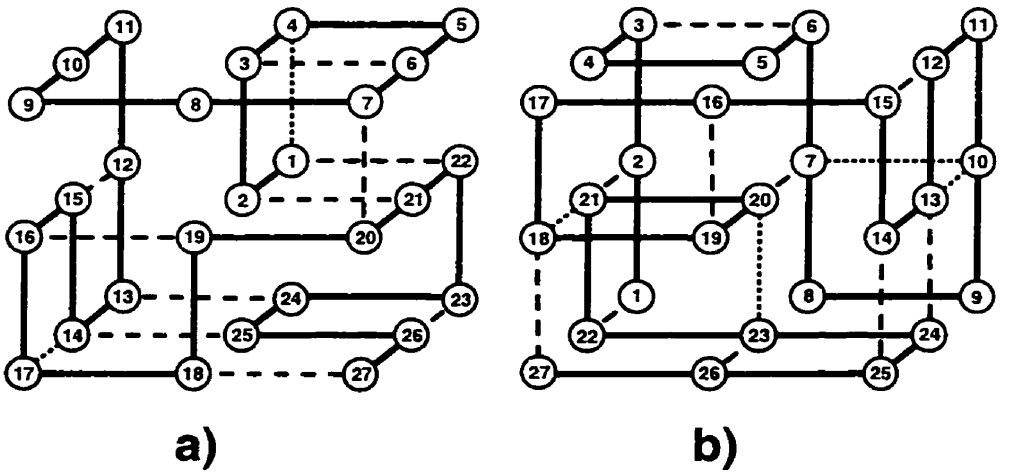


Figure 5.4: The two native states of Sequence 54, where a) is the $N_1$ native state and b) is the $N_2$ native state. Shown are the similar bonds (long dashes) and the different local bonds (short dashes).

bonds have average interaction strengths that are much stronger than the average, which is preset by the mean of the Gaussian to $-2.0$. These similar bonds probably play a significant role in the observed good folding properties of the sequence, as Abkevich et al.[87] have shown that dominant local and non-local contacts produce faster folding sequences. It also appears that the similar-non-local bonds are much stronger than the different-non-local contacts.

The different contacts should play a leading role in the kinetic partitioning of the sequence. The different-non-local contacts appear only to be slightly stronger than the average and are similar in value between the two native states, but the $N_2$ state has a much wider variation in the bond strengths as is shown by the large standard deviation. This variation in the different-non-local bonds probably has some role in the kinetic partitioning, but it is not the only necessary condition.

Table 5.2: Bonding breakdown for Sequence 54. Shown are the similar-local(SL), similar-non-local(SNL), different-local(DL), and different-non-local(DNL), for the $N_1$ and $N_2$ native states.

| | SL | SNL | DL ($N_1$) | DL ($N_2$) | DNL ($N_1$) | DNL ($N_2$) |
|---|---|---|---|---|---|---|
| No. Bonds | 4 | 6 | 2 | 4 | 16 | 14 |
| Average Energy | -3.11 | -3.57 | -1.65 | -2.69 | -2.59 | -2.41 |
| Stand. Dev. | 0.70 | 0.97 | 0.55 | 0.52 | 0.53 | 1.08 |

The different-local contacts present a much different picture, as the variation in the average energy is large. In the good folding $N_2$ native state, the bond strengths are above average and there are more of them in this structure. This is fully consistent with the conclusion that dominant local contacts generate the kinetic partitioning in prions[39]. An interesting observation is that a large number of different-local contacts are not needed, and that these contacts do not have to be excessively strong. What role the different-local bonds have on partitioning is unclear, but these bonds probably play some role in folding.

In studies of real prion systems, it is believed that the $PrP^{Sc}$ state must be marginally more stable than the $PrP^{C}$ state in order to obtain the observed disease propagation behavior[83]. A second sequence, labeled Sequence 49, exhibits the kinetic partitioning similar to Sequence 54, but with the difference that the $N_1$ native state is the favored funnel over the $N_2$ state. For this sequence, out of the 2000 simulations conducted, 63.4% of the time the native state is reached without encountering the secondary native state, $N_2$. In 4.65% of the simulations the $N_2$ state is reached without encountering the state $N_1$. For a small fraction of 0.4%,

$N_2$ is encountered while finally ending up in the primary native state of $N_1$. In the rest of the simulations, folding is not achieved. Further examination of the kinetic partitioning was performed by conducting 100 simulations to observe the number of times the alternate native state could be obtained in $1 \times 10^9$ steps. Starting in the native state, $N_1$, the state $N_2$ was reached 5% of the time, while in the reverse situation, 15% of the simulations were able to obtain the $N_1$ state. These results are represented in a diagram in Fig. 5.5. For this sequence, the temperature was also lowered by 5% with near elimination of crossover between the two native states. Similar folding precentages to the various native states were still obtained at this temperature.

These results demonstrate a similar kinetic partitioning of the two native states and suggest that the landscape is similar to the landscape of Sequence 54 discussed above. The MFPTs for the various pathways are displayed in Table 5.3. These results demonstrate similar folding times to Sequence 54, leading to the conclusion that these two systems only differ in the fact that the preference for a particular native state is reversed. This has the implication that in real prion systems, the alternate native state (the $PrP^{Sc}$ state) need not be more stable in order to observe the conversion to the alternative conformation. If a catalytic process could alter the accessiblity of the transition states to the various folding funnels, then it is not a necessary requirement that the primary state be of higher energy. These two structures provide an ideal opportunity for studying whether the transition state accessibility could be altered to switch the observed behavior; however, this idea is not explored in this work.

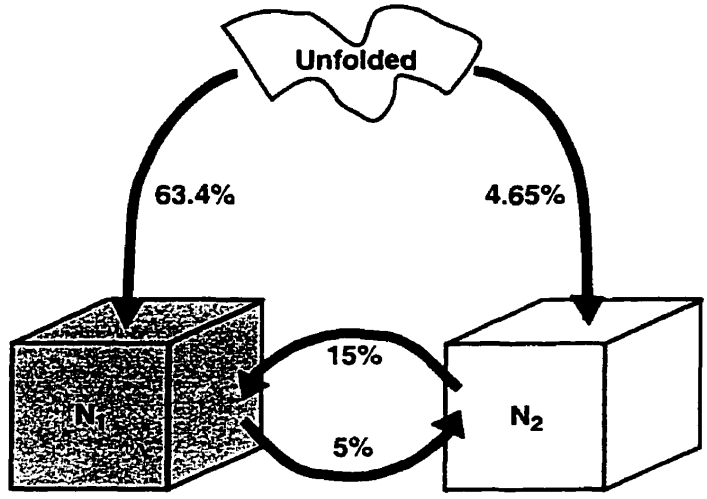


Figure 5.5: Diagram displaying the folding percentages between the various states for Sequence 49. The remainder of the folding events not shown are unsuccessful folding events where the target configurations were not reached. The results beginning in the unfolded states are based on 2000 simulations each, while the results between the two native states are obtained from 100 simulations each. Each simulation is $1 \times 10^9$ Monte Carlo steps.

The structural similarities and differences between the two native structures of Sequence 49 can be observed in Fig. 5.6. In this case, the energies of the native states, $N_1$ and $N_2$, are -78.705 and -78.677 respectively, which is a difference of 0.028. The energy gap between the $N_2$ state and the next lowest energy level is 2.504. The observed energy between the two states is much smaller in this case; however, this probably has little effect on folding kinetics, as the energy barrier between the two states is what is important. The smaller split in energy may provide a slightly larger barrier in the transition from $N_2$ to $N_1$.

If the bonding structure of this sequence is examined, a similar pattern to that

Table 5.3: The folding times for Sequence 49. Shown is the number of events for each average, followed by the mean first passage time.

| Path | No. Events | Time |
|---|---|---|
| $U \rightarrow N_1$ | 1268 | $3.80(7) \times 10^8$ |
| $U \rightarrow N_2$ | 93 | $3.6(3) \times 10^8$ |
| $N_1 \rightarrow N_2$ | 5 | $4.6(10) \times 10^8$ |
| $N_2 \rightarrow N_1$ | 15 | $6.46(7) \times 10^8$ |

of Sequence 54 is found. The similar contacts are above average in strength with the similar-local contacts only slightly lower in strength than those of Sequence 54. The slightly lower similar-local bond strengths may account for the reduced percentage of folding structures to the primary folding state. In terms of different-non-local contacts, the same average energies are observed; these energies are lower than the energies of the similar non-local contacts. The standard deviations show a different result than Sequence 54, as the larger variation still lies with the $N_2$ state not with the primary folding state. In Sequence 49, the $N_1$ state is the favored folding state; however, this does not discount the idea that the variation in the different-non-local bonds plays an important role in partitioning because the standard deviations in both structures' different-non-local bonds are large. The different-local bonds present a more interesting result. In these contacts, there are more contacts in the primary folding state ($N_1$) than in the secondary folding state, but the energies of the different-local contacts are much weaker than the energies of the contacts in the secondary state. This is in contrast with the conclusion that local contacts play a significant role in partitioning, which may mean that different-local bonding has

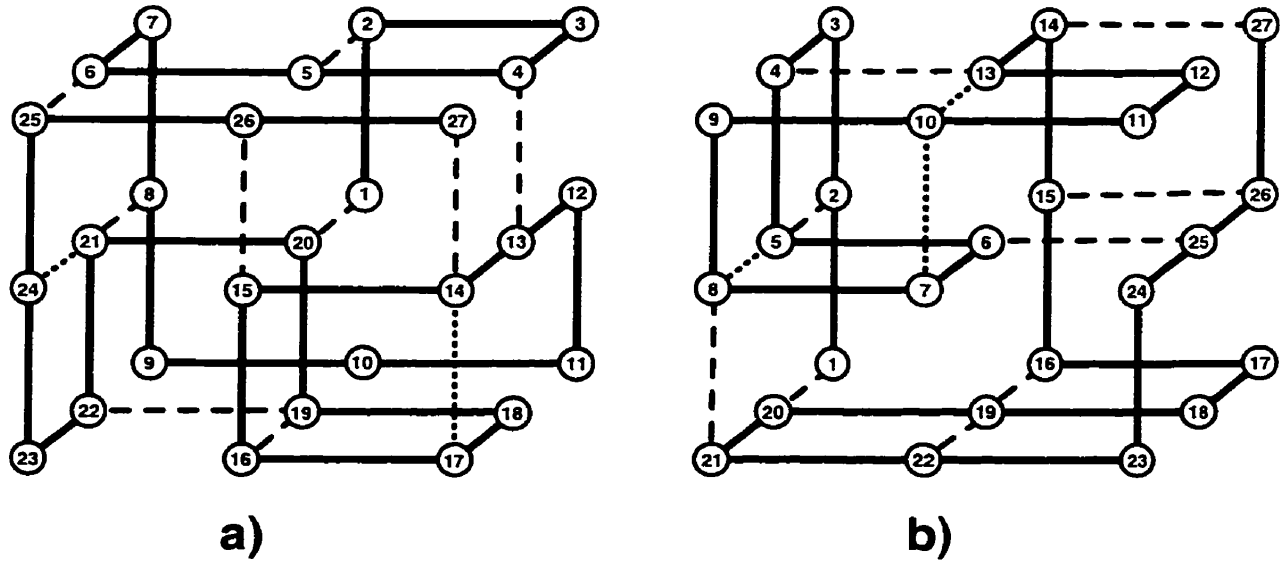


Figure 5.6: The two native states of Sequence 49, where a) is the $N_1$ native state and b) is the $N_2$ native state. Shown are the similar bonds (long dashes) and the different local bonds (short dashes).

little to do with partitioning.

We have found several other sequences that demonstrate possible prion-like behavior. The results of simulations with these structures are outlined in Table 5.5. These sequences all show a preference toward one of the two native states, with most of the sequences favoring the $N_1$ native state. What makes these possible prion-like candidates is the very small number of sequences that fold first to the $N_2$ state and then to the $N_1$ state. For sequences similar to Sequence 49, we need this number of crossover events to be small, because there is probably less than a 20% chance that the sequence can reach the $N_1$ state from the $N_2$ state if all $1 \times 10^9$ steps are used to make this jump. Sequences that are similar to Sequence 54 should have even

Table 5.4: Bonding breakdown for Sequence 49. Shown are the similar-local(SL), similar-non-local(SNL), different-local(DL), and different-non-local(DNL), for the $N_1$ and $N_2$ native states.

| | SL | SNL | DL ($N_1$) | DL ($N_2$) | DNL ($N_1$) | DNL ($N_2$) |
|---|---|---|---|---|---|---|
| No. Bonds | 3 | 7 | 3 | 2 | 15 | 16 |
| Average Energy | -2.82 | -3.07 | -1.48 | -2.62 | -2.68 | -2.47 |
| Stand. Dev. | 0.79 | 0.80 | 0.40 | 0.57 | 0.73 | 0.82 |

fewer crossover events; therefore, the number of these crossover events provides a good screening method for picking out sequences with good kinetic partitioning.

Examining the bond breakdown in these sequences yields few common features, which makes determining what is responsible for kinetic partitioning difficult. There appear to be no common features in the different-local and similar-local bonds, although the kinetic partitioning has a tendency to favor the structure with the stronger different-local bonds. Another point worth mentioning is the relatively small number of local bonds, and, in particular, the small number of different-local bonds, in comparison to the number of different-non-local bonds. This suggests that local bonding only plays a small role in the folding kinetics of these structures. The only striking feature is that all the similar-non-local bonds appear to be much stronger than the average, while the different-non-local bonds appear to be only marginally stronger. These results suggest that the combination of strong similar-non-local and marginally weaker different-non-local bonds are necessary in effective kinetic partitioning; however, we find that this is not a sufficient condition for observing prion-like behavior.

In addition to the sequences discussed above, two sequences were found that appear to exhibit two kinetically partitioned states, as there was little crossing between the two native states; however, both native states were approximately equal in the number of times they were visited, and both would be classed as good-folders. In these sequences, the bonding analysis showed a large number of similar bonds. The features discussed above were still present in these sequences, but the failure to prefer a single native state is why they were not considered prion-like. These sequences do demonstrate that the number of different bonds does affect the accessibility of the transition states. This further implies that the bonding in the native states is not necessarily of great importance in the overall dynamics, but that it is the nature of the transition states that controls the overall behavior. Therefore, the bonding that occurs in the transition states is important, but the formation of transient bonds in this stage of folding also probably plays a key role.

## 5.4 Summary

Using the simplified 27-mer lattice model, we are able to sort through the possible sequences and determine which of them exhibit prion-like behavior. Even at this most basic level of complexity, we are able to observe sequences that mimic the properties of real prions with remarkable similarity. Two sequences have been studied in detail for their folding properties. Each sequence has a preferred native state that is kinetically partitioned from its other native state. The direct MFPTs to each state are approximately equal. An observed asymmetry in the ability to

Table 5.5: Folding data and bond analysis for several prion-like sequences. Shown are the percentages of folding events that are successful to a particular native state, and the percentage of events that fold first to $N_1$ then to $N_2$. A bonding breakdown is also given for the similar-local(SL), similar-non-local(SNL), different-local(DL), and different-non-local(DNL) bonds for the $N_1$ and $N_2$ native states.

| | 151 | 171 | 204 | 218 | 220 |
|---|---|---|---|---|---|
| $\%N_1$ | 86 | 88 | 10 | 72 | 71 |
| $\%N_2$ | 3 | 4 | 49 | 13 | 6 |
| $\%N_2 \Rightarrow N_1$ | 2 | 2 | 1 | 3 | 5 |
| No. SL/SNL | 1/4 | 2/8 | 0/9 | 3/8 | 1/9 |
| SL | -4.0 | -3.9(3) | - | -3.1(2) | -2.0 |
| SNL | -3.8(6) | -3.1(10) | -3.1(6) | -3.3(6) | 3.0(7) |
| No. $N_1/N_2$ | 6/2 | 3/4 | 4/6 | 3/2 | 3/1 |
| DL($N_1$) | -2.6(9) | -2.6(6) | -2.5(6) | -3.3(7) | -2.6(11) |
| DL($N_2$) | -1.9(6) | -2.2(3) | -2.4(4) | -2.3(1) | -1.5 |
| No. $N_1/N_2$ | 17/21 | 15/14 | 15/13 | 14/15 | 15/17 |
| DNL($N_1$) | -2.6(8) | -2.7(7) | -2.5(9) | -2.5(7) | -2.6(7) |
| DNL($N_2$) | -2.6(8) | -2.8(9) | -2.5(10) | -2.7(10) | -2.6(8) |

cross between the two native states suggests that the landscape of these sequences is broken into two separate funnels.

In addition to the analysis of the folding, a breakdown of the bonding within the native states was conducted to find the features of those sequences that generate prion-like behavior. The bonds were divided into similar vs. different and local vs. non-local bonds. The only commonality found among the sequences were the strong similar-non-local bonds with marginally weaker different-non-local bonds. This feature was present in the structures; however, it was not a sufficient condition to produce a prion-like sequence. Based on these analyses, we speculate

that the nature of the native states plays a small role in creating the prion-like characteristics, and that the nature of the bonding in the transition states of these sequences is where the controlling behavior exists.

# Conclusions

Throughout this thesis, minimal models have been used to study the features of biological systems. The simplicity of the models has allowed certain fundamental features to be understood without the inclusion of the atomic complexity. The focus of the thesis has been to create minimal models and to use them to study secondary helical structures in proteins. As well, a specific class of proteins called prions were studied with a well defined minimal model.

In chapter 3, a minimal model of a helical forming segment was created with a focus on anisotropic interactions that were not confined to bond with local neighbors. The model uses a simple anisotropic potential in which the anisotropy can be systematically varied. Using this model, we have studied the statistical properties of the helical forming segment and have demonstrated several interesting features: **1)** We have shown that the model contains four states: 1) coil, 2) globular, 3) helix I, and 4) helix II. The transitions between these various states were studied in detail, and the potential anisotropy was varied to observe the effect on the transition. **2)** These results show that the relative locations of the coil-globular transition and globular-helix transition can be altered as the anisotropy is varied. A strong

anisotropy creates an "all-or-nothing" type transition as the coil-globular and globular-helix transitions occur at nearly the same temperature. A weak anisotropy causes these transitions to occur at significantly different temperatures.

**3)** We also predict that the anisotropy should have an effect on the folding dynamics to a helical state. The foldability of the segment should be significantly altered by changing the anisotropy, and a strong anisotropy leads to a faster folding helical segment.

**4)** Finally by studying our model with an isotropic potential, we lend support to the conclusions of other groups[68] that this system should not have a low temperature first order phase transition.

To continue with the study of these helical segments, we modify our static helix model in chapter 4 to create a dynamic model of a helix. The model is similar to the model of chapter 3; however, it can be used to study a wider variety of problems. Using this model, the mean first passage times of folding were calculated, and the dynamic folding process was examined. Here we observed the interplay between nucleation and propagation in the dynamics of folding.

**5)** We demonstrated that the folding times for the helical segments obeyed a power-law behavior when scaled with system size.

**6)** When the nucleation properties were modified by adding small non-helical forming segments to the ends of the helix, this power-law behavior was altered. It also demonstrated the importance of a single nucleation site in the folding process.

**7)** In addition to the nucleation properties, the effect of anisotropy on the folding times was explored. The results confirmed the conclusions of chapter 3, which

projected better folding behavior with increased anisotropy.

Finally, in chapter 5, a well established minimal model is used to attempt to study prions. These unique proteins have an energy landscape unlike other proteins, as two kinetically stable native states exist. Using a minimal model that has been used extensively to study proteins, we attempt to determine if prion-like sequences exist at this simplified level.

**8)** We successfully determined several structures that exhibit similar behaviors to a prion, and analyze the bond composition to determine what features separate a prion sequence from a protein sequence. The results show no simple pattern in structures, except that the bonds that are similar between the two native states are stronger than the bonds that are different in these structures. At present, we are unable to identify the criteria that will select out a prion sequence from those generated randomly; however, finding sequences that exhibit prion-like behavior provides a method for further studying these molecules.

These models have demonstrated the types of issues that can be addressed through a simplified system. The creation of a model that contains some essential features, yet can be systematically varied, provides a useful tool for studying protein systems. The models in this work have yielded results that could not easily be obtained with more complex all-atomic models.

# Bibliography


[1] A. Šali, E. Shakhnovich, M. Karplus, Nature **369**, 248 (1994)

[2] K.F. Lau and K.A. Dill, Macromolecules **22**, 3986 (1989)

[3] C.B. Anfinsen, Science **181**, 223 (1973) and references therein.

[4] C. Levinthal, J. Chim. Phys. **65**, 44 (1968)

[5] H. Li, C. Tang, N.S. Wingreen, Phys. Rev. Lett. **79**, 765(1997)

[6] S. Miyazawa and R.L. Jernigan, Macromolecules **18**, 534 (1985)

[7] H. Li, C. Tang, N.S. Wingreen, Science **273**, 666 (1996)

[8] K. Yue and K.A. Dill, Proc. Nat. Acad. Sci. USA **89**, 4163 (1992)

[9] R. Mélin, H. Li, N.S. Wingreen, C. Tang J. Chem Phys. **110**, 1252 (1999)

[10] E.I. Shakhnovich, Curr. Opin. Struct. Biol. **7**, 29 (1997)

[11] A. Šali, E. Shakhnovich, M. Karplus, J. Mol. Biol. **235**, 1614 (1994)

[12] E.I. Shakhnovich, V. Abkevich, O. Ptitsyn, Nature **379**, 96 (1996);

[13] M. Skoroboagatiy and G. Tiana, Phys. Rev. E **58**, 3572 (1998)

[14] A.M. Gutin, V.I Abkevich, E.I. Shakhnovich, Biochemistry **34**, 3066 (1995)

[15] L.A. Mirny, V.I Abkevich, E.I. Shakhnovich, Fold. Des. **1**, 103 (1996)

[16] A.R. Fersht, L.S. Itzhaki, N.F. ElMasry, J.M. Matthews, D.E. Otzen, Proc. Natl. Acad. Sci. USA **91**, 10426 (1994)

[17] H. Taketomi, Y. Ueda, N. Gō, Int. J. Peptide Protein Res. **7**, 445 (1975);

[18] N. Gō and H. Abe, Biopolymers **20**, 991 (1981); N. Gō and H. Abe, Biopolymers **20**, 1013 (1981)

[19] Y. Zhou, M. Karplus, Proc. Nat. Acad. Sci. USA, **94**, 14429(1997)

[20] N.V. Dokholyan, S.V. Buldyrev, H. E. Stanley, E.I. Shakhnovich, cond-mat/9812284 (1998)

[21] N.V. Dokholyan, S.V. Buldyrev, H. E. Stanley, E.I. Shakhnovich, cond-mat/0012288 (2000)

[22] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, J. Comp. Chem. **4**, 187(1983)

[23] P. A. Kollman and S.J. Weiner, J. Comp. Chem. **2**, 287 (1981); S.J Weiner, P.A. Kollman, D.A. Case, U. Chandra Singh, C. Ghio, G. Alagona, Jr. S. Profeta, J. Am. Chem. Soc. **106**, 765 (1984)

[24] H. Kawai, Y. Okamoto, M. Fukugita, T. Nakazawa, and T. Kikuchi, Chem. Lett. **1991** 213 (1991); Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai, Protein Engineering **4**, 639 (1991)

[25] R.E Dickerson and I. Geis, *The structure and action of proteins*, Harper and Row, New York, 1969.

[26] F. Eisenmenger, U.H.E. Hansmann, Chem. Phys. Lett. **268**, 86 (1997)

[27] B.H. Zimm, J.K. Bragg, J. Chem. Phys., **31**, 526(1959)

[28] S. Lifson and A. Roig, J. Chem. Phys. **34**, 1963(1961)

[29] H. Qian, J.A. Schellman, J. Phys. Chem., **96**, 3987(1992)

[30] Z.L. Schulten, B.E. Ramirez, P.G. Wolynes, J. Phys. Chem., **99**, 2177(1995)

[31] J.A. McCammon, S.H. Northrup, Biopolymers, **19**, 2033(1980)

[32] J.G. Saven, P.G. Wolynes, J. Mol. Biol., **257**, 199(1996)

[33] Y. Okamoto, U.H.E. Hansmann, J. Phys. Chem., **99**, 11276(1995)

[34] V. Daggett, P.A. Kollman, I.D. Kuntz, Biopolymers **31**, 1115(1991)

[35] M. Takano, T. Takahashi, K. Nagayama, Phys. Rev. Lett., **80**, 5691(1998)

[36] G. Hummer, A.E. Gracía, S. Garde, Phys. Rev. Lett. **85**, 2637 (2000)

[37] F. Potthast, cond-mat/9806217 (1998)

[38] A. Chakrabartty, J.A. Schellman, R.L. Baldwin, Nature, **351**, 586(1991)

[39] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Proteins **31**, 225 (1998)

[40] N. Metropolis, A.W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, J. Chem. Phys. **21** 1087 (1953)

[41] M.P. Allen, D.J. Tildesley, *Computer Simulations of Liquids*, Clarendon Press, Oxford, 1994

[42] N. Madras and A.D. Sokal, J. Stat. Phys. **50** 109 (1988)

[43] J.P Ryckaert and A. Bellemans, J. Comput. Phys. **23**, 327 (1977)

[44] H.J. Hilorst and J.M. Deutch, J. Chem. Phys. **63** 5153 (1975)

[45] B.A. Berg, Int. J. Mod. Phys. C **4**, 249(1993)

[46] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451(1992)

[47] J. Lee, M.Y. Choi, Phys. Rev. E, **50**, R651(1994)

[48] B.A. Berg, T. Neuhaus, Phys. Rev. Lett., **68**, 9(1992)

[49] B.A. Berg, T. Neuhaus, Physics Letters B, **267**, 249(1991)

[50] U.H.E. Hansmann, Y. Okamoto, Physica A, **212**, 415(1994)

[51] A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett., **61**, 2635(1988)

[52] G. Carri and M. Muthukumar, Phys. Rev. Lett. **82**, 5405 (1999)

[53] P.J. Park and W. Sung, Phys. Rev. Lett. **80**, 5687(1998)

[54] A. Buhot and A. Halperin, Europhys. Lett. **50**, 756 (2000)

[55] S. Kumar and M. Bansal, Biophysical Journal **75**, 1935 (1998); S. Kumar and M. Bansal, Biophysical Journal **71**, 1574 (1996)

[56] E.T. Samulski, J. Polymer Sci. **46**, 335 (1974)

[57] J.P. Kemp and Z.Y. Chen, Phys. Rev. Lett. **81**, 3880 (1998)

[58] N.A. Alves, U.H.E. Hansmann, Phys. Rev. Lett. **84**, 1836 (2000)

[59] J.P. Kemp, U.H.E. Hansmann, Z.Y. Chen, EPJB **15**, 371 (2000)

[60] T.B. Liverpool, R. Golestanian, K. Kremer, Phys. Rev. Lett., **80**, 405(1998)

[61] M. Muthukumar, J. Chem. Phys., **104**, 691(1996)

[62] E. Pitard, T. Garel, H. Orland, J. Phys. I (France) **7**, 1201 (1997); T. Garel and H. Orland, cond-mat/9902147

[63] P.J. Camp, J.C. Shelley, G.N. Patey, Phys. Rev. Lett. **84**, 115(2000)

[64] T. Yoshizaki and H. Yamakawa, J. Chem. Phys. **98**, 4207(1993)

[65] H. Pivcová, D. Doskočilová, E.M. Bradbury, Polymer **20**, 139 (1979)

[66] R. Yang, X.R. Yang, D.F. Evans, W.A. Hendrickson, J. Baker, J. Phys. Chem. **94** 6123 (1990)

[67] Y. Zhou, C.K. Hall, M. Karplus, Phys. Rev. Lett., **77**, 2822(1996)

[68] A. Irbäck, E. Sandelin, J. Chem. Phys. **110**, 12256 (1999)

[69] Z. Wang and H.C. Lee, Phys. Rev. Lett. **84**, 574(2000)

[70] A.E. Sherwooda and J.M. Prausnitz, J. Chem. Phys. **41**, 429 (1964)

[71] M. Muthukumar and B. Nickel, J. Chem. Phys. **80**, 5839(1984)

[72] A.Y. Grosberg, A.R. Khokhlov, *Statistical Physics of Macromolecules*, AIP Press, New York, 1994.

[73] R. Majumdar, R.K. Pathria, JMS-REV Macromol. Chem. Phys., **C25(2)**, 191(1985)

[74] K. Binder, D.W. Heermann, *Monte Carlo simulation in statistical physics*, Springer-Verlag, Berlin 1988.

[75] M. Fukugita, H. Mino, M. Okawa, A. Ukawa, J. Stat. Phys. **59**,1397 (1990)

[76] D.K. Kilmov, D. Thirumalai, Folding and Design **3**, 127 (1998)

[77] N.V. Dokholyan, S.V. Buldyrev, H. E. Stanley, E.I. Shakhnovich, cond-mat/9812291 (1998);

[78] J.P. Kemp and Z.Y Chen, Biomacromolecules, (accepted)

[79] A.M. Gutin, V.I. Abkevich, E.I. Shakhnovich, Phys. Rev. Lett. **77**, 5433 (1996)

[80] M. Cieplak, T.X. Hoang, M.S. Li, cond-mat/9907269 (1999)

[81] S.B. Prusiner, Science **252**, 1515 (1991)

[82] G.C. Telling, P. Parchi, S.J. DeArmond, P. Cortelli, P. Montagna, R. Gabizon, J. Mastrianni, E. Lugaresi, P. Gambetti, S.B. Prusiner, Science **274**, 2079 (1996)

[83] F.E. Cohen, S.B. Prusiner, Annu. Rev. Biochem. **67**, 793(1998)

[84] N. Stahl, M.A. Baldwin, D.B. Teplow, L. Hood, B.W. Gibson, Biochemistry **32**, 1991 (1993)

[85] S.B. Prusiner, D. Groth, A. Serban, N. Stahl, R. Gabizon, Proc. Natl. Acad. Sci. USA **90**, 2793 (1993)

[86] G.S. Jackson, L.L.P. Hosszu, A. Power, A.F. Hill, J. Kenny, H. Saibil, C.J. Craven, J.P. Waltho, A.R. Clarke, J. Collinge, Science **283**, 1935 (1999)

[87] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, J. Mol. Biol. **252**, 460 (1995)

[88] G. Wildegger, S. Liemann, S. Glockshuber, Nature Struct. Biol. **6**, 550 (1999)

[89] L.L.P. Hosszu , N.J. Baxter, G.S. Jackson, A. Power, A.R. Clarke, J.P Waltho, C.J. Craven, J. Collinge, Nature Struct. Biol. **6**,740 (1999)