Minimal model for the secondary structures and conformational conversions in proteins

by

Hideo Imamura

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, 2005

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Better understanding of protein folding process can provide physical insights on the function of proteins and makes it possible to benefit from genetic information accumulated so far. Protein folding process normally takes place in less than seconds but even seconds are beyond reach of current computational power for simulations on a system of all-atom detail. Hence, to model and explore protein folding process it is crucial to construct a proper model that can adequately describe the physical process and mechanism for the relevant time scale. We discuss the reduced off-lattice model that can express α-helix and β-hairpin conformations defined solely by a given sequence in order to investigate a protein folding mechanism of conformations such as a β-hairpin and also to investigate conformational conversions in proteins. The first two chapters introduce and review essential concepts in protein folding modelling — physical interaction in proteins, various simple models, and also review computational methods, in particular, the Metropolis Monte Carlo method, its dynamic interpretation and thermodynamic Monte Carlo algorithms. Chapter 3 describes the minimalist model that represents both α-helix and β-sheet conformations using simple potentials. The native conformation can be specified by the sequence without particular conformational biases to a reference state. In Chapter 4, the model is used to investigate the folding mechanism of β-hairpins exhaustively using the dynamic Monte Carlo and a thermodynamic Monte Carlo method — an efficient combination of the multicanonical Monte Carlo and the weighted histogram analysis method. We show that the major folding pathways and folding rate depend on the location of a hydrophobic. The conformational conversions between α-helix and β-sheet conformations are examined in Chapter 5 and 6. First, the conformational conversion due to mutation in a non-hydrophobic system and then the conformational conversion due to mutation with a hydrophobic pair at a different position at various temperatures are examined.

Acknowledgements

I would like to express my gratitude for many people whom I have encountered during my Ph.D. program and for many others who have influenced me directly and indirectly. In particular, I would like to thank Dr. Chen for supervising my projects and various help during my Ph.D program. I like to thank my committee members Dr. Gingras, Dr. Ha, Dr. Meiering, the external examiner Dr. Linhananta and my defence committee members Dr. Duhamel and Dr. Forrest. In addition I express my gratitude to Dr. Duhamel, Dr. Linhananta and Dr. Ha for their helpful comments on my thesis.

Further, I would like to thank Dr. Lemak who had provided numerous valuable information about protein folding and computational methods and Andy Inglot for many things. I would like to thank Dr. Jun, Dr. Okabe, Dr. Lai, Dr. Chan, Dr. Huang, Dr. Wallin, Dr. Stuber and Dr. Hamza for their various advice. I thank Dr. Licinio for his careful guidance in Venice and Trieste and for various insightful comments on my projects. I like to thank late Dr. Lepock who introduced me to the field of biophysics. While I cannot thank my family enough, I thank them for their support, nevertheless.

Teaching assistantship from the Department of Physics, University of Waterloo, and research assistantships from NSERC are greatly acknowledged. Computational time was generously provided by the Hydra computer cluster of Dr. Gingras and by SHARCNET.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Proteins

What are proteins?

Proteins are linear polymers consisting of twenty amino acids [1–10]. The combinations of these amino acids make life possible and flourish. Proteins tirelessly help an organism to sustain life in the self-regulation — the coordinated regulation of all physical, chemical and biological process inside the organism — and in the self-replication [2]. To describe the relationship between the primary sequence of a protein and its function, it is customary to use an analogy, comparing a protein molecule with a text written in twenty letters of amino acids, alphabets with a side chain. Morse code, using only four symbols, can deliver an instruction sufficiently hence a sequence consisting twenty amino acids can also deliver a more complicated instruction. Written in 20 letter code, short protein chains can be compared with short books. Then how can we interpret the text, which the nature seems to accomplish with ease? Is the text definite and identical in all the copies of a given book, that is, are all the molecules of a given protein identical conformation or a heterogeneous mixture of different conformations? How should we interpret the exact meaning of the text, that is, how does a given primary sequence determine the biological function of the protein? What is the role of misprint, that is, what

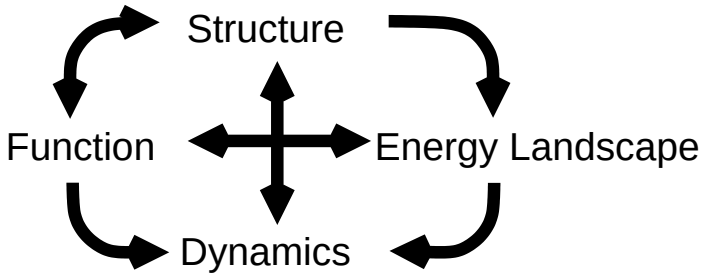

Figure 1.1: A biological system has four components.

is the role of mutation [2]? The task of deciphering the genetic code became manageable recently but relating the genetic code with its structure and function is not simple. Evolution has made it possible over the long period of time for proteins to perform a specific function efficiently but evolution did not make it easy for us to understand the relationship between a primary sequence and its corresponding native conformation.

One can relate a primary sequence with a text but a protein is more than just a text. Proteins have diverse functions and can act as the structural elements such as bricks, plasters and girders of organisms and also provide the energy necessary for life processes. Enzymes are proteins that catalyze biochemical reactions such as familiar food digestion to ensure the smooth operation of a complex organism. Transport proteins like hemoglobin facilitate the movement of molecular oxygen and other essential compounds. Antibodies are proteins that neutralize foreign materials that can be harmful to an organism. In a sense proteins can be regarded as vital molecular machines [1, 2, 8]. It is now possible to synthesize proteins by genetic engineering [8], and there is hope of making new proteins that work for specific function like a new enzyme. Producing such a protein requires: the ability to predict the most stable conformation of a given sequence, the ability to design new conformation, the ability to predict the kinetic accessibility of the conformation, the ability to design the precise features for a specific binding in the fold and the ability to design the precise orientation of the groups in the protein for an efficient catalytic function [8]. Unfortunately we are not fully in command of any of the prerequisites at this point.

What factor distinguishes a simple biological system from a simple system encountered, say, in physics? In biology there is a factor that physicists have rarely encountered in traditional physics. For example, there are three fundamental properties of the systems in physics; structure, energy level, and dynamics. These elements are crucial to understand biological systems. In addition to these key concepts, there is another crucial property in biology that makes life possible; that is function (Fig. 1.1). Furthermore an energy level is better to be replaced by an energy landscape. These four concepts are closely interconnected with each other but it is difficult to establish the basic physical principles connecting each element [7]. Before understanding the function of biological systems and incorporating the function with the other elements, we must first understand the relationship between structure, energy landscape and dynamics.

Protein folding

The first step to understand proteins, for physical scientists in particular, is to find the native structure for a given primary sequence. It seems reasonable and feasible to attempt to find the ground state of a given sequence because in principle it is simply a global minimization problem or a global optimization. This seemingly simple question is not so simple. A similar type of a well-known problem, the traveling salesman's problem — finding the optimal path to visit several cities in the shortest time — is notoriously difficult and many methods such as nonlinear programing, neural network, and generic algorithm are applied to solve this problem [11]. These mathematical algorithms are also used to find the native conformation of a protein by mathematicians but these methods will not illustrate the mechanism of protein folding [11].

Finding the ground state for its corresponding primary sequence became a well-defined physical problem after Haber and Anfinsen demonstrated that ribonuclease can spontaneously regain its full function in vitro [12]. They observed that ribonuclease can fold and refold without any extra unknown biological agents in vivo and therefore concluded that the information encoded in the primary sequence of ribonuclease contains not only the covalent structure but also the secondary and tertiary structure of the molecule [12]. Anfinsen and Scheraga [13] later stated that

"It is currently believed that the three dimensional structure of the native protein in a given environment (solvent, pH, ionic strength, presence of other components, temperature, etc.) is the one in which the Gibbs free energy of the *whole* system is a minimum with respect to all degrees of freedom, i.e., that the native conformation is determined by the various interatomic interaction and hence by the amino acid sequence, in a *given environment*. Whether or not this local minimum in the free energy hyperspace is the global one is at present an unsettled question."

The basic question of the protein folding problem can be stated as: Given a specific sequence of amino acids, how do physical and chemical forces determine its myriad properties, especially the essential unique folded structure of a globular protein in a given environment? This problem is often regarded as solving the "second genetic code" and solving the folding problem would advance biotechnology, in principle permitting the design of a totally new protein from scratch. Theoretical studies of protein folding have focused on a number of issues [5]. First, what are the sequence requirements for proteins to fold rapidly and be stable in their native conformations? Second, what are the thermodynamic mechanism(s) of protein stabilization and the kinetic mechanism(s) of folding? Third, are there special native structures (structural motifs) that are more likely to correspond to the native structures of foldable proteins? Fourth, what is the best approximation for protein-folding energetics (potentials)? These are the questions often asked by theorists in this field.

Since proteins suffer the multiple-minima problem, we can apply the ideas already developed to solve the multiple-minima problem in other fields of physical science. One of the relevant ideas to avoid the multiple-minima trap is known as the principle of minimal frustration [14, 15]. The principle of minimal frustration states that there is a smooth slope to the free energy landscape because a protein evolved to equip harmonious cooperativity. In this formalism proteins are regraded as minimally frustrated heteropolymers. This means that the energy landscape of real protein folding is not globally flat with random fluctuations like a random heteropolymer. The energy landscape is rugged but with a preferred direction toward a unique native conformation. The folding landscape of proteins is necessarily rugged because polypeptide chains can sample many conformations

during their folding process and have the possibility of making inappropriate contacts between residues. Roughness arises from interactions between improper pairs and misoriendted structures in a protein. The landscape of protein folding must have some ruggedness to reflect this diversity of energy interactions. This results in a competition between the tendency toward the folded native state and trapping because of the ruggedness of the funnel. The landscape cannot be flat as assumed in the Levinthal's paradox, which states that there are too many configurations for a protein to find its ground state randomly [16]. He suggested that if the search is completely random through the entire space of possible conformations, a protein of normal size needs billions of years to search its ground state through all of the available configurations. This led him to postulate that a protein must follow a specific path that guides it to the native conformation and therefore folding must be under kinetic control. According to him, "If the final folded state turned out to be the one of lowest conformational energy, it would be a consequence of biological evolution and not of physical chemistry [17]." The search, however, is far from random as Levinthal had proposed. The search, is guided by the free energy landscape that leads to the ground state or, at least, eliminates the vast majority conformations as irreverent states so that protein can reach the ground state within the time scale of cell life without ever visiting astronomically large physically irrelevant conformations [18].

The function of a protein is closely related to its three dimensional structure, thus we like to predict the three dimensional conformations from the primary sequence of amino acids. Experimental determination of a protein structure is still a slow process and it usually requires either the X-ray crystallography or NMR methods for relatively small proteins. For proper X-ray diffraction analysis, a well-ordered protein crystal of size about 0.5 mm is required to produce a clear diffraction pattern and the crystallization of proteins is still difficult [8]. Until recently the correct prediction of protein conformations can be achieved better by homology modeling [19], fold recognition methods [20] than through models governed by potentials derived from the first physical principles [21]. The group led by Scheraga, however, has been developing the physics-based united residue (UNRES) force field, and they manage to predict a large fragment of proteins irrespective of

the types of structures (α, $\alpha + \beta$ or β) with sufficient accuracy, better than some homology methods and their results were presented at the Critical Assessment of Techniques for Protein Structure Prediction (CASP5 and CASP6) [22]. Their initial trial conformation still relies on a generic algorithm that uses knowledge based information and they are trying to improve UNRES such that the UNRES can predict a conformation from a scratch only based on a given sequence without the help of knowledge based information, that is, data previously measured in experiments. Models based on physical principles are ultimately needed to understand how a protein works and to address questions of the existence of the apparent degenerative native state in proteins such as the prion proteins, to simulate protein folding pathways, to understand the mechanisms of protein folding, and to study the interactions of proteins with other biomolecules.

Misfolding

Misfolding refers to the failure of a protein to its native conformation efficiently or the failure to maintain the conformation due to its reduced stability as a result of either environmental change or mutation [23]. Recently a new class of diseases were recognized and often called as conformational diseases including one of the most puzzling medical problems such as the prion diseases and Alzheimer's disease. There is a clear difference between most other genetic diseases and the conformational diseases. Most genetic diseases are due to an initial misfolding of the newly synthesized protein. On the other hand, in the conformational diseases, at least some of the protein are correctly folded and released in their normal form — in the beginning they are functional. For some unknown and known reasons proteins undergo conformational conversion and transform into non-functional or weakly functional forms, which eventually leads to the aggregation and deposition of the proteins.

Well known examples of conformational diseases are serpins and prions. Conformational disorders involving serpins — serine protease inhibitors — involve some conformational conversion. The biologically active serpin is not at its thermodynamically most stable conformation; over time, active inhibitor slowly converts to a more stable inactive form [24]. This is one of the examples demonstrating that

some protein folding reactions can be governed by kinetics rather than thermodynamics alone [23]. Prions — proteinacious infectious particles that lack nucleic acid — are now believed to cause the spongiform encephalopathies, the diseases characterized by porous structures like a sponge in a postmortem examination on the brain of a patient or animal affected by these diseases. Affected individuals show the deterioration of health, loss of balance, mental capacity associated with the loss of the basic functions of various parts of the brain [25]. The diseases include Creutzfeld-Jacob Disease (CJD), new variant Creutzfeld-Jacob Disease (vCJD), Gerstmann-Straussler-Scheinker syndrome (GSS) and Kuru (Kuru means shivering in a Papua New Guinean language). While the structural details are not known, the proposed structural change involves a transition from primarily helical structures to primarily β-sheet structures. This conformational conversion in prion proteins can arise spontaneously as a result of a genetic mutation in the prion. In more disturbing cases, there is evidence that the transition is also propagated in normal prions by introduction of abnormal form of prions [26]. This conformational infection can take place either between individuals of same species or individuals of different species as is suspected with the new variant of Creutzfeld-Jacob disease. The abnormal form of prions is believed to act as template for the transformation of the normal form to the abnormal one. Hence the understanding of the various process involved in prion diseases at the molecular level is required to tackle prion diseases.

1.2 Modeling protein structures

1.2.1 The structure of proteins

Amino acids and primary structure

Twenty natural amino acids that build up proteins can take either the L or the D steric form where they are mirror image of each other. In natural proteins only the L form shown in Fig. 1.2 exists. If a cup of milk consisted of D form of amino acid proteins then its proteins have no nutritional value except as diet milk because

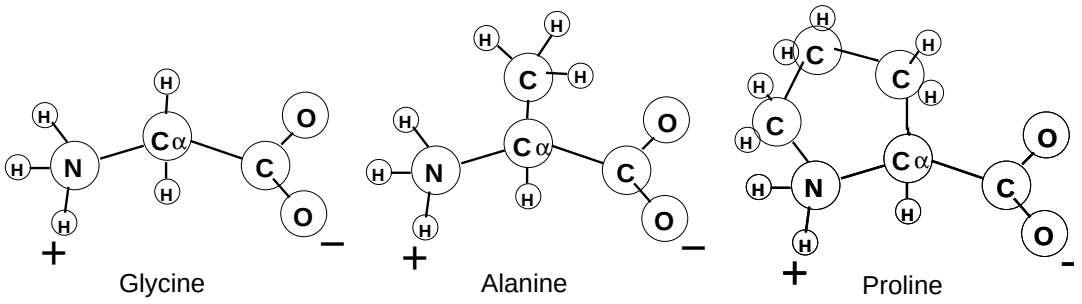

Figure 1.2: Amino acids: Glycine and alanine have a hydrogen and a CH_3 as a side chain, respectively and proline has a unique side chain geometry. All amino acids are shown in the charged zwitterion forms which are the most common molecule forms at pH 7.

biological systems distinguish these chiral difference and only use the L form of the proteins. The origin of the handedness of the amino acids is not known and there seems no apparent reason why the L form is the dominant form of amino acids [2, 4, 6]. The overall amino acid compositions of most proteins are similar. The 20 amino acids arise at certain frequencies that are similar in most proteins. On the other hand, membrane proteins tend to contain more hydrophobic residues than soluble proteins and the fibrous structural proteins tend to contain repetitive amino acid patterns [6].

Each protein consists of a polypeptide chain made of amino acids linked by peptide bonds. The twenty amino acids differs only in their side chains or R-groups. The amino acids can be divided into three classes by the chemical nature of the side chain: hydrophobic side chains, alanine (Ala), valine (Val), leucine (Leu), isolucine (Ile), phenylalanine (Phe), proline (Pro) and methionine (Met); charged ones aspartic acid (Asp), glutamic acid (Glu), lysine (Lys) and arginine (Arg); polar ones, serine (Ser), theronine (Thr), cystein (Cys), asparagine (Asn), glutamine (Gln), histidine (His), tyrosine (Tyr), and tryptophan (Trp). Glycine (Gly) has only a hydrogen atom as a side chain and can be classified either as the fourth class by itself or as the first one. The backbone of a protein conformation is connected by covalent bonds. Covalent bond can be regarded as a pair of electrons that the two connected atoms share. This sharing of the electrons results in a strong

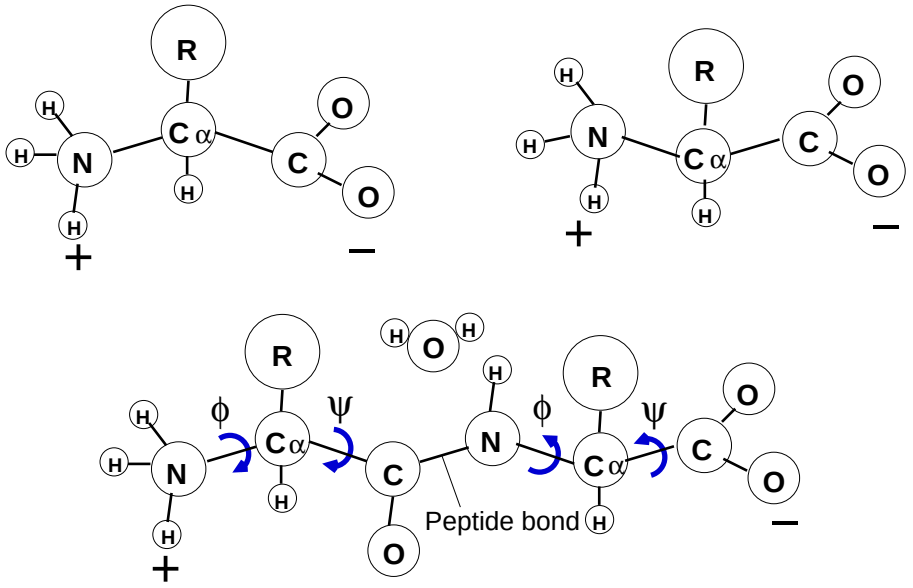

Figure 1.3: Peptide chain: Two generic amino acids (top) form a peptide bond and release a water molecule (bottom). R represents a side chain, + and − indicate partial charge.

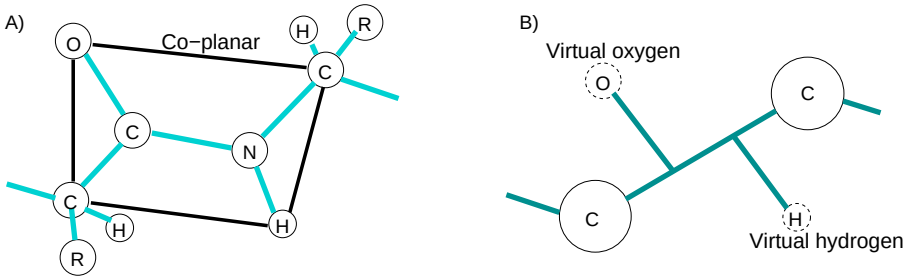

Figure 1.4: Co-planar atoms: A) Six atoms lie on the same plane. B) These configuration can be approximated by two backbone carbon beads and virtual oxygen and hydrogen atoms.

attraction between the atoms. Covalent bonds can be regarded as permanent on the time scale of a protein's life time. These bonds hold the protein backbone and sidechains together. In particular two amino acids are connected by a peptide bond, where a hydrogen atom from the NH_2 amino group and the OH unit from the terminal COOH are removed to form a bond between N and C atoms (Fig. 1.3). The peptide bond is planar which is caused by a considerable delocalization of the lone pair of electrons of the nitrogen onto the carbonyl oxygen (Fig. 1.4). The C-N bond is consequently shortened about 10% and it has double-bond character. Hence the twisting of the peptide is rare except for proline residues [7, 8].

The primary sequence of proteins is an unbranched polypeptide chain consisting of L-amino acids linked by peptide (amide) bonds between the carboxyl of one reside and the amino group of the next. A primary sequence must be unbranched because the information of the primary sequence is stored in the one dimensional DNA sequence and a DNA sequence chain cannot store the information of a branched sequence [1]. A primary sequence can be cross bridged by a disulfide bond (-S-S-) in which two cystein residues are linked by their thiol groups [8].

Secondary structure

The polypeptide chains of proteins are often organized into hydrogen-bonded structures. There are two major types of such conformations.One is the α-helix conformation (Fig. 1.5) and another is the β-sheet conformation (Fig. 1.6). The α-helix is stabilized mainly by hydrogen bonds between the carboxyl group C=O on the ith monomer and the amino N-H group on the $(i + 4)$th monomer. The C=O groups are parallel to the axis of the helix and point almost straight at the NH groups to which they are hydrogen-bonded. The length of the α-helix varies from five to forty residues and the average length is about ten residues [9]. One side of helices often is hydrophobic character and another side is hydrophilic character. This amphipathic nature of the helices plays a critical role in the interaction with other structures. When helices are bundled together, the hydrophobic side of the helix faces inside to form the hydrophobic core. An helix also has a net dipole moment since the dipoles of the polypeptide backbone align nearly parallel to the axis of the helix.

Unlike the curled helical structure, an extended polypeptide chain can also form complementary hydrogen bonds with a parallel extended chain. This parallel structure binded by the hydrogen bonds can match up with another extended chain to build up a β-sheet structure. Two arrangements are possible: one is a parallel β-sheet in which all the chain are aligned in the same direction, and the other is an anti-parallel β-sheet in which the directions of chains are alternating. Parallel β-sheets are found in globular proteins where two separated β-strands, which are far away from each other along the sequence, can be aligned in parallel direction. Secondary structures are often characterized by their Ramachandran diagram [27] defined by the range of the preferable dihedral angles ϕ and ψ (Fig. 1.3). In the model we discuss in the thesis, the backbone heavy atom are simplified and the angles ϕ and ψ are not defined.

The secondary structural elements further pack into a compact, native conformation. This three dimensional structure is called the tertiary structure of a single protein. Larger proteins with more than about two hundreds residues may contain organized structural units called domains. Some proteins consist of more than one peptide chain forming a quaternary structure such as hemoglobin — four individual proteins, or sub-units comprise the functional form of hemoglobin. The thesis will focus mainly on the simplified secondary structure units.

1.3 Physical interactions for protein folding

To reach the accuracy required for a realistic analysis of the conformation and dynamics of macromolecules, it is necessary to use a relatively complex form for the empirical potential function and must be relatively simple to facilitate efficient computations. Typical force fields for high resolution models such as all atom models incorporate a harmonic restoring force between bonded nearest neighbors, a relatively stiff angle potential between, a dihedral torsional potential to allow hindered rotation about a bond, and non-bonded interactions between the separated atoms. Reduced models tend to use more specialized potentials. For example hydrophobic and hydrogen potentials must be explicitly incorporated because the potentials depending on the detail interactions at the atomic level must be approx-

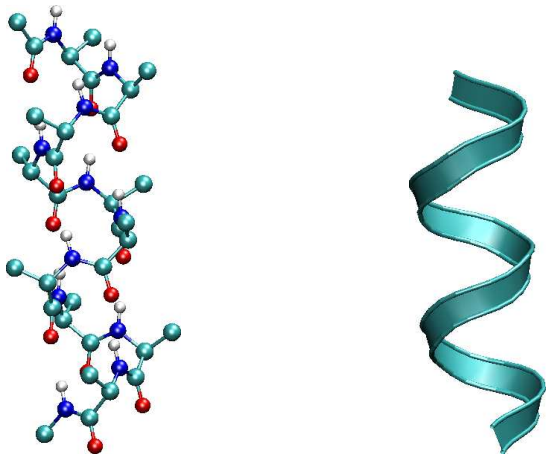

Figure 1.5: Polyalanine α-helix: The atomic presentation left, and ribbon presentation right.

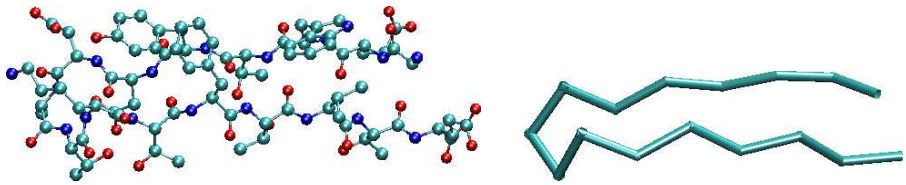

Figure 1.6: β-hairpin of GB1: The atomic presentation left, and back bone presentation right.

imated by computationally tractable alternative potentials. Often other potentials like a chirality factor for a conformation to reduce the degeneracies of the native conformation are used. We will discuss simplified models in detail in the next section.

1.3.1 Physical interactions

A typical set of key potentials of a high resolution model includes the following potentials: the bond length, bond angle, Lennard-Jones, electrostatic and dihedral potential terms.

$$V_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angle}} K_\theta(\theta - \theta_{eq})^2 + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{D R_{ij}} \right] + \sum_{\text{dihedral}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \tag{1.1}$$

where each summation for the potentials covers all the non-redundant terms in each potential and all the parameter are given in reduced units.

The term $\sum_{\text{bonds}} K_r(r - r_{eq})^2$ represents the stretching of a covalent bond around the equilibrium length r_{eq} where K_r is the constant of proportionality and first proposed by Pauling as the one-two pair interaction [28]. This term is sometime set to a constant in simple models for simplicity. In the bond stretching, its time scale is short and the deviation from r_{eq} is comparatively small, this term is often averaged out to be constant for the time scale of the rotational motion of the residues in simple models. Strictly, the true bond stretching is not harmonic. An alternative, more realistic bond stretching potential is the Morse potential [29]: $V_{\text{Morse}} = D_e[1 - e^{-a(r - r_{eq})}]^2$ for each bond length where D_e is the depth of the potential energy function and a controls the width. This potential is not particularly computationally attractive and rarely used in simple models for protein folding. To approximate the Morse function, cubic and higher terms can be included with properly avoiding the error associated with higher terms [35].

The term $\sum_{\text{angle}} K_\theta(\theta - \theta_{eq})^2$ represents the bending of the bonds from the equilibrium angle θ_{eq} and the deviation from the equilibrium angle is small hence

K_θ is normally set to be large. The higher order terms in this potential are also added to improve the accuracy in some detail calculations. For the bond angle potential that particularly involves hydrogen atoms, the so-called Urey and Bradley interaction, a shifted harmonic potential defined by the distance between the first and third atoms, can be used [30, 31]. Simple models, however, tend to use the simple harmonic bending potential energy or set each bond angle to a fixed value.

The Lennard-Jones potentials are expressed as $\sum_{i<j}[A_{ij}/R_{ij}^{12} - B_{ij}/R_{ij}^{6}]$ where R_{ij} is a distance between the ith and jth residues and A_{ij} and B_{ij} are constants specific to the atoms i and j. The attractive part of the potential arises from induced mutually fluctuating dipoles. In a simple model these potentials are replaced by a simple excluded volume term such as a hard square wall or a shifted L-J type potential.

The term $\sum_{i<j}[q_i q_j/DR_{ij}]$ measures the electrostatic interactions where D is the dielectric constant and for the vacuum $D = 1.0$. If all of the atoms and charges in the system are explicitly represented and atomic polarization is included, the dielectric constant should be just unity. If the solvent is not explicitly included in the systems like many simplified models, then the dielectric constant of water is about 80 and that of the interior of a protein ranges between 2-4 in a protein [8]. This large difference in the dielectric constants makes it difficult to select a proper dielectric constant. The electrostatic interactions can be calculated straightforwardly by Coullumb's law to all the partial charges but the calculation is time consuming because of the long range of electrostatic interactions. Fortunately the calculation can be simplified since no monopoles, or free charges are observed inside a protein (salt bridge are ion pairs hence dipoles), all partial charges form di- or multipoles. Therefore the actual range of electrostatic interactions is rather short. Further in β-sheet conformations, adjacent dipoles are antiparallel so that electrostatic interactions cancel each other at longer distance. In α-helices, the dipoles form lines cancelling each other except at both ends and consequently antiparallel helix bundles are more favorable than parallel ones [3]. In practice, the last two potential interactions, in simple models, are applied to residues separated by three or four residues because the residues closer than this distance are interacting through the bonding potentials like the bond length, bond angle and dihedral potential discussed

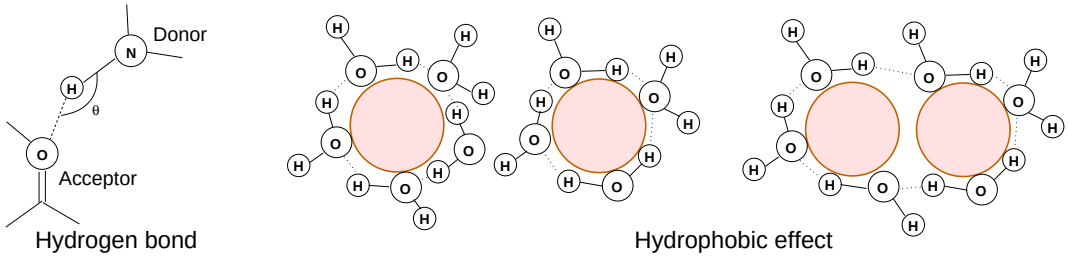

Figure 1.7: Hydrogen bonding and hydrophobic effect: Left the non-polar residues encaged by water molecules feel the hydrophobic effect.

below.

The term $\sum_{\text{dihedral}}(V_n/2)[1+\cos(n\phi-\gamma)]$ characterizes the twisting of the dihedral angle defined by the consecutive four residues where ϕ, V_n, n and δ are a dihedral angle, its force co bnstant, multiplicity, and phase, respectively. The dihedral potential of this form for ethylene and ethane was used since 1930's [32]. These periodic energy dependence arises from two sources: the varying extent of interaction of the H's on the carbon with the H's on the other carbon and the orbital-orbital interactions of the electrons centered about the two carbons [33]. These are based on the quantum mechanical effect. To create potential field in classical limit, it was found that the hindered rotation about a single or a partial double bonds cannot be represented by other potentials such as the bond length, bond angle, Lennard-Jones, and electrostatic potentials with sufficient accuracy [34]. If non-bounded interactions (Lennard-Jones and electrostatic potentials) are made stronger enough to ensure the proper barrier to the torsional motion, then they are no longer a good representation of the intermolecular interactions. Hence it is common practice to use the dihedral angle term presented above.

In addition to these potentials, other types of phenomenological potentials, which can be represented by the potentials listed above in principle, are also defined in various simplified models. For instance, hydrogen bonding and hydrophobic interaction terms are widely used in simple models. In some force fields, electrostatic potentials are assumed to be able to represent partial charge interactions for the hydrogen bonding, and the proper solvent particles can encompass the hydrophobic

effect. Hydrogen bonding is the most common element in organic compounds and particularly important in biological systems [7, 8]. In a peptide chain, hydrogen atoms are covalently bounded to strongly electronegative elements such as N and O (Fig. 1.7). The hydrogen arises when two electronegative atoms bind to the same proton. The bonds are asymmetric; the proton is at its normal covalent bond distance from the atom to which it is formally bonded, and at a distance from the other somewhat shorter than the usual van der Waals contact distance [8]. The hydrogen bonding interaction has been represented in various forms [35], and can be described as the simple Lennard-Jones $6-12$ or a Lennard-Jones $10-12$ interactions between the donor hydrogen atom and the acceptor atom. In our model we use the shifted Lennard-Jones $6-12$ between the virtual hydrogen and oxygen atoms whose direction is defined by the adjacent residues. A more complicated hydrogen bonding function is sometimes used which incorporates the directionality of hydrogen bonding $V_{\text{hydrogen}} = \sum_{i<j}[(A_{ij}/R_{ij})^6 - (B_{ij}/R_{ij})^4]\cos^4\theta$ where θ is the angle spanned by the acceptor, hydrogen and donor shown in Fig. 1.7 [36]. The importance of hydrogen bonds was initially emphasized by Mirsky and Pauling in 1936 [37], and they stated that the importance of hydrogen bonds in protein structure can hardly be overemphasized. Later Kauzmann, however, suggested that hydrophobic interaction not hydrogen bond interactions is the dominant force in proteins [38]. The first strong evidence that hydrogen bonds contribute favorably to protein stability was shown in studies of the interaction of tRNA synthetase [39]. Since then many other experimental studies have reached similar conclusions [40]. In particular the hydrogen bonding plays crucial roles to define the secondary structures [6, 8, 10].

The hydrophobic effect is induced by the interaction between polar solvent molecules like water molecules and this effect can be integrated into the system often by a temperature independent 6-12 Lennard-Jones potential in a reduced model [41–43]. The detail molecular mechanism of the hydrophobic effect continues to be the subject of some debate but qualitatively the hydrophobic effect can be explained as follows. Water is made up with a dynamic, loose network of hydrogen bonds. The presence of a nonpolar compound imposes a local rearrangement in this network. The water molecules like to preserve the number of hydrogen bonds and they line up around the nonpolar molecule. The hydrophobic solute does not

cause large enthalpy changes in the polar solvent but decreases the entropy of the system due to the increase in local order. Consequently a hydrophobic molecule is force to coalesce into the hydrophobic region of a protein by the regaining of entropy of water [8] (see Fig. 1.7). It has an entropic component hence the hydrophobic effect is known to become weaker at lower temperature. This peculiar nature of the hydrophobic effect results in cold denaturation of proteins [6, 10]. At higher temperatures, the entropy term becomes dominant, hence the hydrophobic effect become stronger. In simple models the hydrophobic potential is set to be temperature independent for simplicity [41–43].

Including the hydrodynamic effect in simple protein models is difficult [45, 46]. This is partly because the hydrodynamic effect is a force not a potential, hence it is difficult to approximate the hydrodynamic effect with a simple tractable form. In the frame work of a Monte Carlo algorithm, the hydrodynamic interaction can, in principle, be incorporated by using Oseen tensor [44], but in practice this approach has not been usually taken [45]. In one recent study, the hydrodynamic effect is examined in the study of homopolymer collapse and protein folding [46]. It was shown that the hydrodynamic effect speeds up homopolymer collapse in a simple model [46]. In their model the polymer dynamics was simulated by MD and the dynamics of the solvent is computed by special stochastic calculation method. They found, however, that there is no significant impact on the overall dynamics of protein folding of three systems, an α-helix and β-hairpin of 21 residues and a helix-bundle of 65 residues in a simple, native contact biased Gō model, which is explained in the next section. They argued that the hydrodynamic effect will influence the dynamics of the folding process of a polymer chain if the solvent motion, which is generating the hydrodynamic effect, is sufficiently correlated. In a homopolymer collapse, the solvent motion become sufficiently correlated to induce the collective flow that accelerate the chain collapse. On the other hand, in the protein folding process, the interaction in a protein system is heterogeneous and consequently the solvent will not collectively flow in a specific direction and the hydrodynamic effect does not particularly accelerate the folding process in the studied systems [46].

1.4 Models for protein folding

1.4.1 Molecular dynamics simulations

Molecular dynamics (MD) methods have provided atomistic details on biomolcular systems on equilibrium properties such as free energies in explicit or implicit solvent environment and it can characterize detail atomistic movements for a short time scale [31, 47]. All-atom approach for the elucidation of proteins is fundamentally hampered by the limitation on the integration time-step of MD simulations, which is around a femtosecond (10^{-15}) for typically more than 10000 particles. In order to encompass the timescale of tens of microseconds in a conventional MD method even for a small protein, 10^{10} integration steps are necessary, and this seems to be beyond reach of contemporary computational power [48]. One of the longest molecular dynamics simulation for complete folding of a protein is performed by Duan and Kollman who have managed to simulate a complete folding process of a villin headpiece subdomain (36-residues) for a micro second range [49]. Statistically reliable sampling of folding trajectories, however, requires just more than a few trajectories. In order to overcome the limitation of the folding timescale, several tricks have been invented such as smoothing protocols to accelerate dynamics, overcoming the energy barriers by using umbrella potentials, simulating along a reaction path without explicit reference to time, optimization protocols instead of physical ones, high temperature method, Langevin dynamics, massively distributing computing and so on [48]. The last three methods can properly estimate the time scale of the dynamics. The high temperature MD is used to study the conformational conversion between α- and β-conformations [50]. The massively distributing computing scheme assumes that a single timescale dominates the process and computes a large number of independent trajectories, then once they observe a reactive transition, they will restart the massively distributing computing from the conformation obtained from the reactive transition, and repeat this process until a sufficient sampling has been obtained. From this calculation, the rate of the folding process can be calculated and this method was applied to β-hairpin folding [51]. Another approach to overcome the difficulty of timescale is to use Gō-potentials, in which the native contacts in the system are given higher priority [52, 53] and a Gō-model is

used for β-hairpin folding study [54].

A proper empirical force field, similar to the one described in the previous section, is essential for an all-atom simulation, and it is best to use a force field that has previously been shown to be the most effective for a particular system. Many force fields for biomolecular systems have been developed. Widely used empirical force fields are CHARMM [55] and AMBER [56], which are developed to represent the various forces among biomolecules derived from the experimental data sets in the frame work of classical mechanics. There are a large number of other force fields specialized for particular physical systems; for example, BMS for nucleic acids [57], ECEPP for proteins [58], GROMOS for biomolecules [59], OPLS for biomolecules [60], and PFF for proteins [61]. One can find more detail information on the force fields in the original papers cited above or books [35, 62] and which provides a good starting point on molecular forces and computational chemistry.

1.4.2 Simple Models

Simple on- and off-lattice models

The first exact enumeration of short chains on lattices was performed by Orr [63] who enumerated all the possible self-avoiding conformations on lattice for chains less than 10 monomers in length. Longer chains were explored in the polymer theory including scaling laws in end-to-end distances and renormalization-group methods [64]. At first, it may not be clear that simple lattice models can be applied to investigate the general features of protein folding behavior. Single domain proteins do not exceed much beyond 200 residues. Specific interactions that lead to the unique architecture of native conformation cannot be fully represented by lattice models. The dynamics simulated on the lattice with limited move sets by the Monte Carlo method seems not so ideal for the dynamics of a protein in aqueous solution. Despite these concerns, a large number of protein lattice models have indeed demonstrated a number of predictions many of which have been affirmed in experiments on the general protein folding [18, 65, 66] and on the chaperone protein GroEL [67]. In the context of protein folding, lattice model models were first

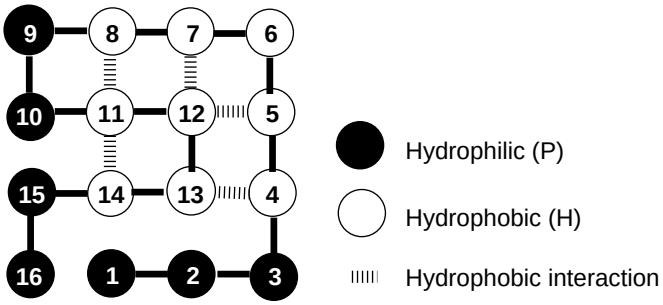

Figure 1.8: Two dimensional HP lattice model:

introduced by Gō and coworkers [52], then Dill, Chan and coworkers have popularized the HP model — using only two types of residues — for protein folding and applied it to diverse topics [18, 65]. They used exact enumeration for thermodynamic characteristics and the Metropolis Monte Carlo method to measure folding kinetics [68].

In the early development of a simple off-lattice model for protein folding, Levitt and Warshel pioneered the attempt to simplify the model of a globular protein and used a simplified model using a bead per a residue for pancreatic trypsin inhibitor. Every backbone residue is a sphere with stiff rigid side chains and successive residues are linked by the virtual bonds. The model is governed by torsional, van der Waals, hydrogen bonding, hydrophobic type, and non-bonded local interactions. In addition, the model has non physical pushing and pulling potentials to overcome local energy barriers, which were later criticized [69]. A simple off-lattice model became popular only after the computational power became sufficiently fast in the early 90's, while increasing sophisticated on-lattice models have been developed by Kolinski and coworkers since late 80's [69].

HP models

The HP model reduces 20 amino acids to only the two types, hydrophobic (H) and polar (P) [65, 70]. For example, in Fig. 1.8 the native conformation of 2D-lattice model of sixteen residues is shown. There are five attractive hydrophobic HH interactions in the native state. All the other contacts like HP and PP have

zero energy contribution to the system.

The model is one of the simplest models possible and its simplicity enables to perform calculation that is not possible in more complicated models. The HP model and its variants have been used as a convenient toy model to experiment ideas and they became valuable tools. The strengths of the HP model are as follows [65]. 1) All the conformations can be enumerated. 2) Exact nature of a model is valuable for precise analysis. This means that models have two components: the physical model itself and the mathematical approximations for further analysis. Keeping the number of parameters minimum makes it possible to understand the consequences of a model, rather than the consequences of the choices of parameters. 3) Models with the least microscopic detail and the greater extraction of principle can provide broad information about how protein-like behavior is encodable in other types of chain molecules than proteins. Even a failure of the HP model provides some insights on the requirement for proteins. 4) Simple lattice models that explicitly account for specific monomer sequences, chain connectivity, and excluded volume are useful for testing analytical theories such as mean-field treatments of heteropolymer collapse and spin glass models. They found that homopolymers do not collapse to unique states and heteropolymers collapse to very few structures hence heteropolymers are suitable for protein folding models, and larger code alphabets promote uniqueness of the conformation [65]. The HP models are widely used as on-lattice models, but they are also used as off-lattice models in order to discuss the fitness of sequences against mutations in evolution (long time scale) [71,72] and general folding properties depending on HP sequences in shorter chains [73–75] and longer chains [76,77] and adsorption on patterned surfaces including a many-body interaction version of the HP potentials [78]. The simple secondary structures are also discussed in the HP model [65].

Gō-like models — native contact biased models

The Gō-model was initially developed by Gō and coworkers who investigated the dynamics and thermodynamics on the square lattice models as idealized protein systems [52, 53]. A simple Gō-model contains only native contact potential. For example, the HP model shown in Fig. 1.8 can be transformed into a Gō-

model if the hydrophobic interactions are restricted to the native contacts, that is, H_4H_{13}, H_5H_{12}, H_7H_{12}, H_8H_{11} and $H_{11}H_{14}$. While in the original HP model, an H monomer interacts with any other H monomers, but in the Gō-version HP model, a H monomer only interact with its corresponding native pair partners. Hence the native biased potentials can reduce the energy frustration in the system. Gō et al. originally constructed a set of more complex potentials [52, 53]. The original Gō-model included three types of interactions: short range local conformational propensities, long range native biased potentials and hydrophobic potentials. The trajectories are sampled by the Metropolis Monte Carlo method. They found long range interactions increase the folding cooperativity but the local native residue propensities decrease the folding cooperativity — tendency for residues to change their physical properties as a group [33], in this case, tendency for residues to change conformations as a group — but speed up the folding process and stabilize the native conformation.

Gō-models have been enjoying renewed interest after some experimental and theoretical analysis suggested that the topology of the native conformation might be a determinant factor for the folding rate [79, 80]. Various types of Gō-models currently exist; one dimensional residue-based phenomenological Ising like models [81–83], lattice models [84, 85], and off-lattice coarse grained [88, 90–93], or all atom models [54, 94, 95]. Ising like models simplify the systems in such a way that only native or non-native state exists. Some model further simplifies the system by allowing only continuous sequential native state [82] in order to reduce the parameter and formulate a problem in analytical form. Lattice and off-lattice Gō-models can be constructed that folds into a preferred native conformation, and used to explore the biologically pertinent questions that may be difficult to formulate and be tested in conventional non-Gō-models. One of the most attractive features of Gō-models is that the folding is faster in Gō-models than the models with non-native interactions. Hence statistically sufficient sampling is possible even for detailed all atom models. One of the weakness of Gō-models is that Gō-models can not be used for a structural prediction from a given sequence, and that Gō-models ignored non-native interaction with little physical justifications. Even though the all interaction is attractive in Gō-models and there is small local energetic frustration, but

topological constrains can still cause the frustration in the systems. The solvent averaged free energy of the chain configuration E tends to decrease because interactions are mostly favorable, but the formation of the native conformation results in more specific order in the systems hence the loss of chain entropy S. At temperature T, the free energy $F = E - TS$ has a barrier along the reaction coordinate due to the loss of chain entropy that is not compensated by the gain in E [96].

Torsional potential models

The pioneering works on the simple off-lattice models — presenting the residue with a single bead — with a set of minimal number of potentials were developed by Thirumalai's group. The models are governed by local interaction potentials such as bond angle potentials, conformation specific dihedral angle potentials to β- or α- conformations and for a flexible turn residues and by long range interaction potentials such as hydrophobic interactions and excluded volume interactions. The models are used for a β-bundle [41] and an α-helix bundle [86]. In these models there is no side chain or the hydrogen bonding interaction. Later these models were improved by the additional hydrogen bonding between virtual hydrogen and oxygen atoms — which do not have any volume — that interact only with the native contact pair [87, 88]. The models are later applied to study a combination of α- and β-conformations [90]. The β-bundle models with Gō potential applied to its hydrophobic interactions have been also studied to examine mutational effect of the systems [91, 92].

Conformational independent models

It is desirable to design a model that can form an α-helix and a β-hairpin depending only on a given sequence. This seemingly simple task is probably impossible and has never been accomplished with any model that presents one particular residue as only a single bead, which can be classified as a one-bead model [89]. One bead model can not define a conformation without reference dependent potentials like the native biased dihedral potentials discussed above. In order for a model to be independent of a reference conformation, some additional factors must be implemented in the model. This can be done by adding an extra residue for a side chain or a virtual

hydrogen donor and acceptor which improve the specificity of local interactions, and the model become closer to a real peptide chain. Our models presented in this thesis are one of the first generations of such models and of the simplest models that can accommodate the helical and β-sheet conformations [97–100]. In other models, additional virtual hydrogen donors and acceptors have also improved the ability to have particular conformations, but their hydrogen bond interaction kept a native bias contact and its dihedral angle are β- and α-conformation specific and their models remained the reference dependent ones [87, 88]. There are two bead models for protein systems but they are specialized for the system but they are not designed to investigate general protein folding [89].

Four or six bead models are used in various physical systems to investigate protein folding and notably conformational conversion in proteins which we discuss in Chapter 5 and 6. In four-bead models developed by the group led by Hall, the side chain is presented by a single bead and three heavy atoms on the backbone N, C_α and C (Fig. 1.3) are also presented in the models and can describe the hydrogen bonding explicitly. The model is used to examine simple α-helices [101] and α-helix bundles [102]. They investigated the conformational conversion between α-helix and β-hairpin structures in polyalanine chains [103], and also intermolecular aggregation of β-sheet structures [104]. The model is simplified compared to all atom models but it is still too demanding for straight forward molecular dynamics to cover the long time scale of full trajectories and obtain the sufficient sampling, therefore they used a discontinuous molecular dynamics method which uses simplified potentials with a set of rules defined to facilitate faster calculation of trajectories [106].

A six-bead model contains the 3 backbone heavy atoms N, C_α and C as well as the atom H attached to N, the atom O attached to C and the sidechain in six beads of different size. In these models the hydrogen bonds are naturally defined between the atom H and O and the hydrophobic interaction is between sidechains. This approach was developed by Takada and coworkers and Irback's group. Takada and coworkers [107–109] investigated general aspects of protein folding while Irback's group have focused on the α-helix bundles [110–112] and the secondary structures [113] and an extended version of this six-bead model has also been investigated by others [114]. These models have not been explicitly used to investigate

conformational conversion.

A totally different approach to the modeling of protein structures has been developed by considering two simple topological parameters; the thickness of the tube r and the range of the attractive interaction R [115]. When the sizes of these two parameters are comparable or the ratio of the two is around unity, helical and sheet conformations arise. When the tube is slightly thinner, that is, r/R is slightly less than one, the tube forms a helix. When the tube is slightly thicker, that is, r/R is slightly greater than one, the tube forms a hairpin [115]. The model can accommodate the various conformations such as a β-hairpin and a β-sheet, a helix and others but the model can not be meaningfully used to investigate protein folding. Later the model has been improved and the new model contains the directionally biased hydrogen potential normal to the plane defined by three consecutive monomers, whose definition is similar to those in [116, 117] and that used in this thesis, the hydrophobic interaction, local and non-local radius of curvature constrains. The improved model can tackle biologically relevant problems like complicated phase diagram containing α-helix and β-sheet structures [118]. Note that the conformations accommodated in this model are remarkably similar to our model and this is due to the emphasize on the directionality of the residue orientation in the two approaches.

Miscellaneous models

Various lattice models at higher resolutions have been investigated by Kolinski and coworkers since the 1980's [69, 119, 124]. The lattice model has an advantage over the off-lattice model for the computation efficiency because the number of the possible interaction energies can be calculated beforehand and be saved as a look-up table. Hence the computational time is shorter than that of off-lattice models. The most serious weakness of the simple lattice models is the geometrical constrains. Monomer cannot move freely like off-lattice continuous models. High resolution lattice models, however, can partly overcome this weakness and the capability of accommodating the complicated structures has improved over the years. Side chains can be added to lattice models and can be installed on-lattice as well as off-lattice [119]. Some lattice models were improved by implementing the side chain

as a simple representation of the hydrogen bonding for β-conformations [120, 121] and for α- and β-conformations [122].

There are many other simplified models for protein folding and we have only discussed the models pertinent to the thesis. Those who are interested in these models can check many excellent reviews such as the ones on HP models [21,65,123] and the ones on various other simple models [69, 89, 119, 124–126].

1.5 Outline of the next Chapters

In the coming chapters, we will discuss an off-lattice reduced model which is simple and efficiently sample conformations, but at the same time it can present the essential secondary structures of proteins defined by a given primary sequence. In Chapter 2, we will discuss main computational techniques, Monte Carlo algorithms and particularly the Multicanonical algorithm and the weighted histogram analysis method for efficient thermodynamic calculations, used throughout the thesis. In Chapter 3, we will present a model which is the basis for the models in later chapters. We show that the model can present both α-helix and β-sheet conformations using simple potentials. The native conformation can be specified by the sequence without particular conformational biases to a reference state. In particular, we show that we can transform an α-helix to a β-hairpin by inserting a neutral monomer, whose the effective strength of hydrogen bonds, into the middle of the α-helix. We also show that many other complicated conformations composed of α-helices and β-sheets are possible. In Chapter 4 we discuss the folding mechanism of β-hairpins. A β-hairpin is a simplest secondary structure unit, but its folding and thermodynamic properties have not been sufficiently examined experimentally. This experimental difficulty of β-hairpins are caused by their strong tendency to aggregate and unfold in solution with sufficient concentration for experimental measurements. After the experimental benchmark on the kinetics of β-hairpin folding had been set, many computational studies have been performed but no exhaustively study on the folding mechanism of has been done. Introducing a hydrophobic interaction into the model, we can exhaustively investigate the folding mechanism of β-hairpins in the presence of a hydrophobic pair by using the dynamic Monte Carlo and a thermody-

namic Monte Carlo method. We show major folding pathways are found to depend on the location of a hydrophobic pair and we observed all the folding pathways previously described by others. In Chapter 5 and 6, the conformational conversions between α-helix and β-sheet conformations are examined. These discussions are one of the first to study the conformational conversions in computational models. We introduce idealized mutational effects on the hydrogen bonds in the middle of the chain and illustrate the conformational conversions by using the reduced free energy with respect to α and β-conformations in Chapter 5. Then the conformational conversions due to mutation in the presence of a hydrophobic pair at different positions are examined in Chapter 6. The phase diagrams are used to describe the conformational change influenced by the effective strength of the hydrogen bonds in the middle of the chain, scaled temperature and a hydrophobic pair at various positions. We suggest possible implications to biological systems that may display similar conformational conversions. In Chapter 7 we summarize the results and discuss further applications of the model.

Bibliography

[1] Dickerson RE, Geis I. The structure and action of proteins. New York : Harper Row, 1969, p120.

[2] Vol'kenshtein MV. Molecules and life New York : Plenum Press, 1973, p513.

[3] Schulz GE, Schirmer RH. Principles of Protein Structure. New York : Springer-Verlag, 1979, p314.

[4] Cantor CR. Biophysical Chemistry : Part I: The Conformation of Biological Macromolecules San Francisco : W.H. Freeman, 1980, p341.

[5] Creighton TE. Protein folding New York : W.H. Freeman, 1992. p547.

[6] Creighton TE. Proteins. New York: W H Freeman; 1993, p507.

[7] Flyvbjerg H, Hertz J, Jensen MH, Sneppen K. Physics of biological systems: from molecules to species. New York : Springer, 1997, p366.

[8] Fersht A. Structure and mechanism in protein science. New York : WH Freeman, 1999, p631.

[9] Branden C, Tooze J. Introduction to Protein Structure. New York : Garland, 1999, p410.

[10] Finkelstein AV, Ptitsyn OB. Protein physics. San Diego : Academic Press, 2002, p354.

[11] Merz KM. Jr. The Protein Folding Problem and Tertiary Structure Prediction. Boston : Birkhauser, 1994, p581.

[12] Haber E, Anfinsen CB. Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. J Biol Chem 1961;236:422-424.

[13] Anfinsen CB, Scheraga HA. Experimental and theoretical aspects of protein. folding. Adv Prot Chem 1975;29:205-300.

[14] Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 1987;84:7524-8.

[15] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins. 1995;21:167-95.

[16] Levinthal C. in Mossbauer Spectroscopy in Biological Systems, eds. Debrunner P, Tsibris JCM, Munck E. (Univ. of Illinois Press, Urbana), 1969;22-24.

[17] Levinthal C. Are there pathways for protein folding? J Chim Phys Phys-Chim Biol 1968;65:44-45.

[18] Dill KA, Chan HS. From Levinthal to pathways to funnels Nat Struct Biol 1997;4:10-19.

[19] Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM. A "FRankenstein's monster" approach to comparative modeling: merging the finest fragments of Fold-Recognition models and iterative model refinement aided by 3D structure evaluation. Proteins 2003;53 Suppl 6:369-379. Samudrala R, Xia Y, Huang E, Levitt M. Ab initio protein structure prediction using a combined hierarchical approach. Proteins 1999;Suppl 3:194-8.

[20] Sippl MJ. Knowledge-based potentials for proteins. Curr Opin Struct Biol. 1995;5:229-235. Fischer D, Rice D, Bowie JU, Eisenberg D. Assigning amino acid sequences to 3-dimensional protein folds. FASEB 1996;10:126-136.

[21] Chan HS, Kaya H, Shimizu S. Computational Methods for Protein Folding: Scaling a Hierarchy of Complexities, in Current Topics in Computational Biology T. Jiang, Smith T, Yu Y, Zhang MQ, eds., Boston : MIT Press, Boston, pp. 403-447.

[22] Oldziej S, Czaplewski C, Liwo A, Chinchio M, Nanias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kazmierkiewicz R, Ripoll DR, Pillardy J, Saunders JA, Kang YK, Gibson KD, Scheraga HA. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests. Proc Natl Acad Sci USA 2005;102:7547-7552.

[23] Pain RH ed. Mechanisms of protein foldin. New York : Oxford University Press 2000, p433.

[24] Carrell RW, Lomas DA. Conformational disease. Lancet 1997;350:134-138.

[25] Telling GC. Prions and prion diseases. Norfolk : Horizon Bioscience, 2004, p307.

[26] Huang ZW, Prusiner SB, Cohen FE. Scrapie prions: A three-dimensional model of an infectious fragment. Fold Des 1996;1:13-19.

[27] Ramachandran GN. Sassiekharan V. Conformation of polypeptides and proteins. Adv Prot Chem 1968;28:283-437.

[28] Pauling L. The Nature of the Chemical Bond. Ithaca : Cornell University Press, 1940, p450.

[29] Morse PM. Diatomic molecules according to the wave mechanics. II. Vibrational levels. Phys Rev 1929;34:57-64.

[30] Urey HC, Bradley CA. The vibrations of pentatonic tetrahedral molecules. Phys Rev 1931;38:1969-1978.

[31] Brooks CL, Karplus M, Pettitt BM. Proteins : a theoretical perspective of dynamics, structure, and thermodynamics. New York : J. Wiley, 1988, p259.

[32] Nielsen HH. The torsion oscillator-rotator in the quantum mechanics. 1932;40:445-456.

[33] Hopfinger AJ. Conformational properties of macromolecules. New York : Academic Press, 1973, p339.

[34] Scheraga HA. Calculations of conformations of polypeptides. Adv Phys Org Chem 1968;6:103-184.

[35] Leach AR. Molecular modelling 2nd ed. New York : Prentice Hall, 2001, p744.

[36] Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 1985;28:849-57.

[37] Mirsky AE, Pauling L. On the structure of native, denatured, and coagulated proteins. Proc Natl Acad Sci USA 1936;22:439-447.

[38] Kauzmann WW. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1-63.

[39] Fersht AR. The hydrogen bonding on molecular-recognition. Trends Biochem Sci 1987;12:301-304.

[40] Myers JK, Pace CN. Hydrogen bonding stabilizes globular proteins. Biophys J 1996;71:2033-2039.

[41] Honeycutt JD, Thirumalai D. The nature of folded states of globular proteins. Biopolymer 1992;32:695-709.

[42] Favrin G, Irbäck A, Wallin S. Folding of a small helical protein using hydrogen bonds and hydrophobicity forces. Proteins 2002;47:99-105.

[43] Nguyen HD, Marchut AJ, Hall CK. Solvent effects on the conformational transition of a model polyalanine peptide.Protein Sci 2004;13:2909-2924.

[44] Doi M, Edwards SF. The theory of polymer dynamics. New York : Oxford University Press, 1986, p391.

[45] Binder K. Applications of the Monte Carlo method in statistical physics. Berlin : Springer-Verlag, 1984, p311.

[46] Kikuchi N, Ryder JF, Pooley CM, Yeomans JM. Kinetics of the polymer collapse transition: The role of hydrodynamics. Phys Rev E 2005;71:061804.

[47] McCammon JA, Harvey SC. Dynamics of proteins and nucleic acids. Cambridge : Cambridge University Press, 1987, p234.

[48] Elber R. Long-timescale simulation methods. Curr Opin Stru Bio 2005;15:151-156.

[49] Duan Y, Kollman PA. Abstract Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science. 1998;282:740-744.

[50] Levy Y, Jortner J, Becker OM. Solvent effects on the energy landscapes and folding kinetics of polyalanine. Proc Natl Acad Sci USA 2001;98:2188-2193.

[51] Zagrobic B, Sorin EJ, Pande V. β-hairpin folding simulations in atomistic detail using an implicit solvent mode. J Mol Biol 2001;313:151-169.

[52] Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulations. Int J Peptide Protein Res 1975;7:445-459.

[53] Gō N. Theoretical studies of protein folding. Ann Rev Biophy BioEng 1983;12:183-210.

[54] Zhou Y, Linhananta A. Role of hydrophilic and hydrophobic contacts in folding of the second β-hairpin fragment of protein G: molecular dynamics simulation studies of an all-atom model. Proteins 2002;47:154-162.

[55] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins J Phys Chem B 1998;102:3586-3616.

[56] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force

field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179-5197.

[57] Langley DR. Molecular dynamic simulations of environment and sequence dependent DNA conformations: the development of the BMS nucleic acid force field and comparison with experimental results. J Biomol Struct Dyn 1998;16:487-509.

[58] Zimmerman SS, Scheraga HA. Stability of cis, trans, and nonplanar peptide groups. Macromolecules. 1976;9:408-16. Nemethy G, Pottle M, Scheraga HA, Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. J Phys Chem 1983;87:1883-1887. Kang YK, No KT, Scheraga HA. Intrinsic torsional potential parameters for conformational analysis of peptides and proteins. J Phys Chem 1996;100:15588-15598.

[59] Schuler LD, Daura X, Van Gunsteren WF. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. J Comp Chem 2001;22:1205-1218.

[60] Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensenet WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides J Phys Chem B 2001;105:6474-6487.

[61] Kaminski GA, Stern HA, Berne BJ, Friesner RA, Cao YX, Murphy RB, Zhou R, Halgren TA. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. J Comput Chem 2002;23:1515-1531.

[62] Cramer CJ. Essentials of computational chemistry Hoboken, N.J : J. Wiley, p596, 2004.

[63] Orr WJC. Statistical treatment of polymer solutions at infinite dilution. Trans Faraday Soc. 1947;43:12-27.

[64] Domb C. Self-avoiding walks on lattices. Adv Chem Phys 1996;15:229-259.

[65] Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding - A perspective from simple exact models. Protein Sci 1995;4:561-602.

[66] Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 2000;25:331-339.

[67] Betancourt MR, Thirumalai D. Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. J Mol Bio 1999;287:627-644.

[68] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087-1092.

[69] Skolnick J, Kolinski A. Computer simulations of globular protein folding and tertiary structure. Annu Rev Phys Chem 1989;40:207-35.

[70] Socci ND, Onuchic JN. Kinetic and thermodynamic analysis of protein-like heteropolymers: Monte Carlo histogram technique. J Chem Phys 1995;103:4732-4744.

[71] Nelson ED, Onuchic JN. Proposed mechanism for stability of proteins to evolutionary mutations. Proc Natl Acad Sci USA 1998;95:10682-10686.

[72] Nelson ED, Teneyck F, Onuchic JN Symmetry and kinetic optimization of proteinlike heteropolymers Phys Rev Lett 1997;79:3534-3537.

[73] Irbäck A, Peterson C, Potthast F. Evidence for nonrandom hydrophobicity structures in protein chains. Proc. Natl Acad Sci USA 1996;93:9533-9538

[74] Irbäck A, Peterson C, Potthast F. Identification of amino acid sequences with good folding properties in an off-lattice model. Phys Rev E 1997;55:860-867.

[75] Irbäck A, Sandelin E. Monte Carlo study of the phase structure of compact polymer chains. J Chem Phys 1999;110:12256-12262.

[76] Khokhlov AR, Khalatur PG. Protein-like copolymers: computer simulation. Physica A 1998249:253-261.

[77] Khokhlov AR, Khalatur PG. Conformation-dependent sequence design (engineering) of AB copolymers. Phys Rev Lett 1999;82:3456-3459.

[78] Lee NK, Vilgis TA. Preferential adsorption of hydrophobic-polar model proteins on patterned surfaces. Phys Rev E 2003;67:050901.

[79] Baker D. A surprising simplicity to protein folding. Nature 2000;405:39-42.

[80] Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. Biochemistry. 2000;39:11177-11183. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985-994.

[81] Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci USA 1999;96:11305-11310.

[82] Muñoz V, Henry ER, Hofrichter J, Eaton WA. A statistical mechanical model for β-hairpin kinetics. Proc Natl Acad Sci USA 1998;95:5872-5879.

[83] Muñoz V. What can we learn about protein folding from Ising-like models? Curr Opin Struct Biol. 2001;11:212-216.

[84] Ozkan SB, Bahar I, Dill KA. Transition states and the meaning of Phi-values in protein folding kinetics. Nat Stru Bio 2001;8:765-769.

[85] Faisca PFN, da Gama MMT. Native geometry and the dynamics of protein folding. Biophys Chem 2005;115:169-175.

[86] Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a de novo designed four helix bundle protein. J Mol Biol 1996;263:323-343.

[87] Klimov DK, Betancourt MR, Thirumalai D. Virtual atom representation of hydrogen bonds in minimal off-lattice models of alpha helices: effect on stability, cooperativity and kinetics. Fold Des 1998;3:481-496.

[88] Klimov DK, Thirumalai D. Mechanisms and kinetics of β-hairpin formation. Proc Natl Acad Sci USA. 2000;97:2544-2549.

[89] Tozzini V. Coarse-grained models for proteins. Curr Opin Struct Biol 2005;15:144-150.

[90] Sorenson JM, Head-Gordon T. Matching simulation and experiment: A new simplified model for simulating protein folding. J Comp Biol 2000;7:469-481.

[91] Shea JE, Nochomovitz YD, Guo ZY, Brooks CL. Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist beta-barrel model. J Chem Phys 1998;109:2895-2903.

[92] Nymeyer H, García AE, Onuchic JN. Folding funnels and frustration in off-lattice minimalist protein landscapes. Proc Natl Acad Sci USA 1998;95:5921-5928.

[93] Cieplak M, Hoang TX, Robbins MO. Folding and stretching in a Gō-like model of titin. Proteins 2002;49:114-124.

[94] Zhou Y, Karplus M. Interpreting the folding kinetics of helical proteins. Nature 1999;401:400-403.

[95] Shimada J, Kussell EL, Shakhnovich EI. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. J Mol Biol 2001;308:79-95.

[96] Gō-ing for the prediction of protein folding mechanisms. Proc Natl Acad Sci USA 1999;96:11698-11700.

[97] Chen JZY, Imamura H. Universal model for α-helix and β-sheet structures in protein. Physica A 2003;321:181-188.

[98] Imamura H, Chen JZY. Conformational conversion of proteins due to mutation. Europhys Lett 2004;67:491-497.

[99] Imamura H, Chen JZY. Dependence of folding dynamics and structural stability on the location of a hydrophobic pair in β-hairpins. (Submitted).

[100] Imamura H, Chen JZY. Conformational conversion due to mutation and a hydrophobic pair. (To be submitted)

[101] Smith AV, Hall CK. α-helix formation: discontinuous molecular dynamics on an intermediate-resolution protein model. Proteins 2001;44:344-360.

[102] Smith AV, Hall CK. Assembly of a tetrameric α-helical bundle: computer simulations on an intermediate-resolution protein model. Proteins 2001;44:376-391.

[103] Nguyen HD, Marchut AJ, Hall CK. Solvent effects on the conformational transition of a polyalanine peptide. Protein Sci 2004;13:2909-2924.

[104] Nguyen HD, Hall CK. Phase diagrams describing fibrillization by polyalanine peptides. Biophys J 2004;87:4122-4134.

[105] Nguyen HD, Hall CK. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. Proc Natl Acad Sci USA 2004;101:16180-16185.

[106] Alder BJ, Wainwright TE. Studies in molecular dynamics. I. General method. J Chem Phys 1959; 31: 459-466.Rapaport DC, Molecular dynamics simulation of polymer chains with excluded volume. J Phys A Math Gen 1978;11: L213-L217. Bellemans A, Orban J, Van Belle D. Molecular dynamics of rigid and non-rigid necklaces of hard discs. Mol Phys 1980;39:781-782.

[107] Takada S, Luthey-Schulten Z, Wolynes PG. Folding dynamics with nonadditive forces: a simulation study of a designed helical protein and random heteropolymer. J Chem Phys 1999;110:11616-11628.

[108] Takada S. Protein folding simulation with solvent-induced force field: folding pathway ensemble of three-helix-bundle proteins. Proteins 2001;42:85-98.

[109] Fujisuka Y, Takada S, Luthey-Schulten ZA, Wolynes PG. Optimizing physical energy functions for protein folding. Proteins 2004;54:88-103.

[110] Irbäck A, Sjunnesson S, Wallin S. Three-helix-bundle protein in a Ramachandran model. Proc Natl Acad Sci USA 2000;97:13614-13618.

[111] Irbäck A, Sjunnesson S, Wallin S. Hydrogen bonds, hydrophobicity forces and the character of the collapse transition. J Biol Phys 2001;27:169-179.

[112] Favrin G, Irbäck A, Wallin S. Folding of a small helical protein using hydrogen bonds and hydrophobicity forces. Proteins 2002;47:99-105.

[113] Irbäck A, Samuelsson B, Sjunnesson F, Wallin S, Thermodynamics of β and α structure formation in proteins. Biophys J 2003;85:1466-1473.

[114] Knott M, Chan HS. Exploring the Effects of Hydrogen Bonding and Hydrophobic Interactions on the Foldability and Cooperativity of Helical Proteins Using a Simplifed Atomic Model. Chem Phys 2004;307:187-199.

[115] Banavar JR, Maritan A. Geometrical approach to protein folding: a tube picture. Rev Mod Phys 2003;75:23-34.

[116] Kemp JP, Chen JZY. Formation of helical states in wormlike polymer chains. Phys Rev Lett 1998;81:3880-3883.

[117] Trovato A, Ferkinghoff-Borg J, Jensen MH. Compact phases of polymers with hydrogen bonding. Phys Rev E 2003;67:021805.

[118] Hoang TX, Trovato A, Seno F, Banavar JR, Maritan A. Geometry and symmetry presculpt the free-energy landscape of proteins. Proc Natl Acad Sci USA 2004;101:7960-7964. Hoang TX, Trovato A, Seno F, Banavar JR, Maritan A. Geometrical model for the native-state folds of proteins. Biophys Chem 2005;115:289-294.

[119] Kolinski A, Skolnick J. Reduced models of proteins and their applications. Polymer 2004;45:511-524.

[120] Zhdanov VP. Folding time of ideal β sheets vs. chain length. Europhys Lett 1998;42:577-58.

[121] Dimitrievski K, Kasemo B, Zhdanov VP. Chain length scaling of protein folding time: β sheet structures. J Chem Phys 2000;113:883-890.

[122] Borg J, Jensen MH, Sneppen K, Tiana G. Hydrogen bonds in polymer folding. Phys Rev Lett 2001;86:1031-1033.

[123] Dill KA. Polymer principles and protein folding. Protein Sci 1999;8:1166-1180.

[124] Skolnick J, Kolinski A. Monte Carlo approaches to the protein folding problem. Adv Chem Phys 1999;105:203-242.

[125] Thirumalai D, Klimov DK, Dima RI. Insights into specific problems in protein folding using simple concepts. Adv Chem Phys 2002;120:35-76.

[126] Head-Gordon T, Brown S. Minimalist models for protein folding and design. Curr Opin Str Bio 2003;13:160-167.

Chapter 2

Simulation methods

2.1 Introduction

The Metropolis Monte Carlo (MC) was originally developed for calculating equilibrium properties of physical systems [1–3]. The method can be generalized to describe stochastic processes of a non-equilibrated system if a physically realistic move in the system, which satisfies the detailed balance condition [2], is used [2–4]. The dynamic interpretation of the MC algorithm for the protein folding process has been widely used in many studies [4–7]. MC has solidified the theoretical basis of protein folding based on statistical mechanics and made it possible to perform exhaustive calculations in various aspects of protein folding and the role of MC in this field cannot be overemphasized [8–11].

2.2 The Metropolis Monte Carlo method

In the Metropolis algorithm, configurations are generated from a previous state using the transition probability which depends on the energy difference between the initial and final states and a simulating temperature [12]. This trajectory is called a Markov chain. The Markov chain is developed by Markov who tried to establish the limiting laws of probability in their most general form and later

established a Markov chain: a sequence of random variable in which the future variable is determined by the present variable but is independent of the way in which the present state arose from its predecessors — the formal definition can be found in reference [13]. The sequence of conformations generated by a Markov process follows a time ordered path. Time steps are measured by Monte Carlo steps — the number of each move attempt divided by the number of the residues that can be moved. The time dependent behavior can be described by a master equation

$$\frac{\partial P_n(t)}{\partial t} = -\sum_{n \neq m} [P_n(t) W_{n \to m} - P_m(t) W_{m \to n}], \tag{2.1}$$

where $P_n(t)$ is the probability of the system being in the conformation n at time t, and $W_{n \to m}$ is the transition rate for $n \to m$. In equilibrium $\partial P_n(t)/\partial t = 0$ and the two terms on the right-hand side of 2.1 should be equal. This leads to the detailed balance relationship

$$P_n(t) W_{n \to m} = P_m(t) W_{m \to n}. \tag{2.2}$$

The probability of the nth state in a classical system is given by

$$P_n(t) = e^{-E_n/k_B T}/Z \tag{2.3}$$

where Z is the partition function. The partition function is usually unknown at this point but one can eliminate the partition function by the use of a Markov chain — each new conformation is generated directly from the preceding one. If we obtain the nth conformation from the mth conformation, the relative probability is the ratio of the individual probabilities and therefore the denominator can be canceled. Hence only the energy difference in the two conformations is required for the computation;

$$\Delta E = E_n - E_m \tag{2.4}$$

Any transition rate which satisfies detailed balance is acceptable and, in the Metropolis method, the transition rate is defined by

$$W_{n \to m} = \tau_0^{-1} e^{-\Delta E/k_B T}/Z \qquad \Delta E > 0 \tag{2.5}$$

$$= \tau_0^{-1} \qquad \Delta E < 0 \tag{2.6}$$

where τ_0 is the time required to attempt to move a single residue. τ_0 depends on the type of the move set and the number of the movable residues in a given peptide system.

A typical Metropolis Monte Carlo calculation for a peptide chain, including a simple single local move, consists of the following steps.

1. Choose an initial state.

2. Choose a residue i.

3. Rotate the residue around the axis defined by the neighboring residues within the range of a constant constrain.

4. Calculate the energy change ΔE after the move. Typically only the energies associated with the moved residue i need to be calculated to save computational time because the rest of energies among residues that did not moved are unchanged. This may be obvious but many beginners will not notice this.

5. Generate random number r of range $0 < r < 1$.

6. If $r < exp(-\Delta E/k_B T)$ accept the move.

7. Go to the next residue and go to step 3.

Using the steps described above, one simulates a folding dynamics from a well-thermalized initial conformation to the native state and obtains the statistics for the folding time and other folding properties and also can obtain the equilibrium properties of a system after sufficiently long simulations. The sampling by the Metropolis algorithm becomes inefficient at a lower temperature because the system tends to be trapped in and can not escape from local minima and consequently fails to efficiently sample a proper distribution.

2.2.1 Move set in Monte Carlo algorithm in polymeric systems

The calculation of equilibrium systems can take advantages of various efficient algorithms to speed up the calculation — simulated tempering, generalized ensemble

techniques, and many specialized move sets such as cluster algorithms [2]. The dynamical calculation, however, mostly relies on the basic Metropolis algorithms with a basic local move set that should be dynamically correlated such that time steps as well as dynamical trajectories can be interpreted as dynamical quantities. Only recently histogram methods have been applied to calculate the nonequilibrium properties of the systems but their use is currently limited to small systems [14].

Residues in an off-lattice model are not confined on the lattice grids, therefore the movements of residues are more flexible than those of on-lattice models. Consequently more realistic motion and geometrical presentation are possible [2]. In a simple cubic lattice model, in particular long chain systems, one must ensure that a move set should satisfy the ergodicity, that is, all the conformations should be accessible given sufficient time, even though some inaccessible conformations may not significantly contribute to the overall distribution of the conformations [2].

In order to implement dynamics using Monte Carlo algorithms only local moves can be legitimately used. Often local kink-jump method (Fig. 2.1) is sufficient for a simple small system. In larger systems, some other global move sets may be required for an efficient computation. Global moves such as pivot, wave-like and reptational-like moves (Fig. 2.2) are often used but these moves must be used with care and the ratio of the use of the various move sets must be properly designed to avoid distortion of dynamics by a particular move set type [15]. The pivot movement is often used to accommodate the diffusional movement of a secondary structure unit. In our study physically appropriate local rotational moves known as kink-jump moves [2] are used in all simulations, starting with the selection of a monomer at random. If the selected monomer is not a terminal, the monomer is rotated around the axis connecting the two adjacent monomers; if the selected monomer is a terminal, which has only one neighboring monomer, the monomer is rotated around the axis defined by the nearest and second nearest monomers.

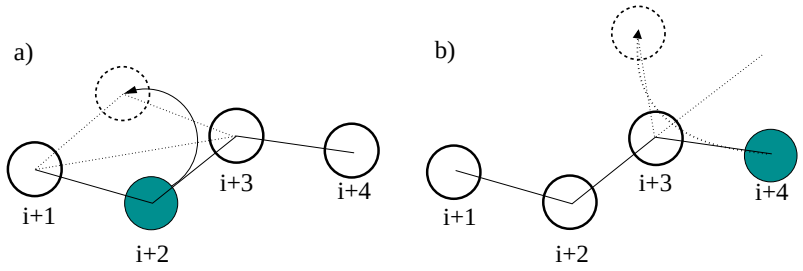

Figure 2.1: Local move: Kink-jump

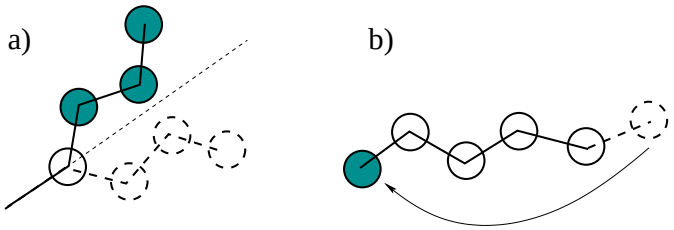

Figure 2.2: Global move: a) pivot, b) reptation.

2.2.2 Dynamic interpretation of the Metropolis Monte Carlo method

A trajectory generated through the Metropolis Monte Carlo method is correlated since each new conformation is generated directly from the preceding conformation [3]. This correlation can be regarded as a dynamic correlation of a well defined model with stochastic dynamics; hence the base of the dynamic interpretation of the MC method. It is possible to associate the scale of time t of subsequent conformations. The time is usually normalized such that N single residue transitions are performed within a unit time. The dynamic evolution of the probability $P_n(t)W_{n\to m}$ is governed by the master equation

$$\frac{\partial P_n(t)}{\partial t} = -\sum_{n\neq m}[P_n(t)W_{n\to m} - P_m(t)W_{m\to n}], \tag{2.7}$$

and it represents a transition probability per unit time. Hence the Monte Carlo dynamics can be considered as reasonable microscopic dynamics — the numerical

realization of a master equation in discrete time.

This stochastic dynamics can not be expected to reproduce dynamical processes on the time scale of C-C bond length or bond angle vibrations ($10^{-14}-10^{-13}s$) [3,16]. Conformational and structural relaxation on the longer time scales, however, arise by rotational motions around a torsional angle along the chain ($\approx 10^{-11}s$). These rotational motions are activated processes that arise with a waiting activation time distribution depending on the Newtonian dynamics on the shorter time scale [3]. For relaxation processes consisting of many of such jumps, the dynamics can be approximated by a stochastic process, and on these time scales the Monte Carlo dynamics can be expected to be a representation of the real physical dynamics of the system [3]. For a shorter time scale, molecular dynamics on an all-Atom model is suitable to calculate near equilibrium properties [16, 17].

Rey and Kolinski [7] demonstrated that the MC dynamics in a high resolution lattice model can reproduce similar folding pathways of a two-helix bundles obtained by Brownian dynamics on the off-lattice version of the system and hence they demonstrated the MC dynamics is suitable and efficient for the examination of the dynamical aspect of protein folding for the long time scale. Kikuchi et al. [18] have directly shown by using a simple system that the MC dynamics can be used to solve the Fokker-Planck equation, an approximation to the the master equation for Markov processes. Cieplak et al. [19] demonstrated that the MC calculation using a simple lattice protein folding model matches the time evolution of the native conformation with the time evolution based on a Master equation formalism. Chubykalo et al. [20] have shown that MC time steps can be converted to real time steps in a magnetic system.

2.3 Efficient thermodynamic Monte Carlo algorithms

The appropriate simulation techniques is required to efficiently solve the multiple-minima problem in order to adequately characterize thermodynamic pictures of the systems. To overcome the multiple-minima problem, we adapt a multicanonical

WHAM Monte Carlo algorithm, which is a hybrid of the standard multicanonical Monte Carlo method and weighted histogram analysis method (WHAM). In this method, the standard multicanonical process that determines a multicanonical weight factor is regarded as a set of trajectories that is performed under a different umbrella potential to obtain proper sampling in the potential energy space. Then the statistics obtained by these trajectories with non-Boltzmann weight factor is reweighed and normalized into the properly weighted statistics by using the WHAM algorithm.

The merit of this method is that it does not require to find a multicanonical weight factor that yields a sufficiently flat probability distribution histogram, which can be quite tedious and problematic for complex systems, thus computational time is greatly reduced. We briefly review the multicanonical algorithm, which is the basis of our thermodynamic calculations, and the weighted histogram analysis method.

2.4 Multicanonical Method

The efficiency of canonical simulations of complex systems such as spin glasses and biopolymers at fixed temperature is severely limited by the multiple-minima problem. At lower temperatures the systems tend to get trapped in one of many local energy minima, thus it is very difficult to obtain an accurate canonical distribution at lower temperatures by conventional canonical Monte Carlo methods.

It is desirable to construct the simulation algorithm which can compute accurate thermodynamic pictures of complex physical systems with multiple local minima over a wide range of temperatures, in particular at lower temperature where the multiple minima problems become increasingly amplified and the transition between local minimum states becomes so slowed down that the sufficient sampling becomes practically impossible [30]. One way to overcome this multiple-minima problem is to simulate a system in a generalized ensemble where each state is weighted by non-Boltzmann probability weight factor so that potential energy space is properly sampled in the range where one tries to investigate. This non-Boltzmann weight

enhances the probability that those unlikely states between local maxima occur more frequently and hence increasing the sampling range, accuracy and efficiency. Finding such a non-Boltzmann weight is not trivial and requires multiple trials and errors and many algorithms have been developed for this purpose in the past.

One of the most popular non-Boltzmann, generalized-ensemble methods is a multicanonical algorithm. Similar algorithms which are based on the similar mathematical principle are developed under various name such as entropic sampling first applied on Potts model [21], adaptive umbrella sampling of the potential energy on the alanine dipeptide [22], random walk algorithm on the threonine and met-Enkephalin [23], and density of states Monte Carlo on Ising and Potts models [27]. We will briefly discussed some of the similarities and differences among these methods after the outline of the basic of the multicanonical algorithm below.

We follow the basic implementation of the multicanonical MC algorithm developed by Berg [32] and Hansmann and Okamoto [33]. The multicanonical MC can be performed in the following steps [33].

1. Perform a canonical Monte Carlo simulation at a sufficiently high temperature T_0 where the system is well above any transition. The weight factor is given by the $w(E) = e^{-\beta_0 E}$ with $\beta_0 = 1/k_B T_0$.

2. Initialize and set each bin of $S(E)$ to zero, where E is discretized with bin width δE.

3. Collect the energy distribution obtained in the previous multicanonical loop — the simulation computed with a particular multicanonical weight factor — as a histogram $H(E)$ which has an equally partitioned bin width as $S(E)$.

4. After a single multicanonical loop, update the array $S(E)$ defined by

$$S_{k+1}(E) = S_k(E) + \ln H(E). \tag{2.8}$$

   where the subscript k indicates the entropy is accumulated to approximate the weight factor of the next simulation loop.

5. Construct the continuous function $S'(E)$ from the discrete $S(E)$ between $E_{\min}$ and $E_{\max}$. In our calculation, $E_{\min}$ is set to slightly smaller than the ground state energy and $E_{\max}$ is around $E = 0$, below which meaningful conformations arise. In the standard construction of the multicanonical parameter, $\alpha(E)$ and $\beta(E)$ are used to approximate the $S(E)$ with

$$S(E) = \beta(E)E + \alpha(E). \tag{2.9}$$

   This determination procedure is not unique and various alternative methods are possible. We approximate the function $S'(E)$ by using Lagrange Interpolation which will be described below, and therefore the $\beta(E)$ factor that is identified as an inverse of temperature is not used.

6. Begin a new loop with the multicanonical weight factor defined by

$$w_{\mathrm{mu}}(E) = e^{-S'(E)}. \tag{2.10}$$

7. Iterate step 3 through 6 until the obtained distribution becomes reasonably flat with acceptable deviation of an order of magnitude.

Once the multicanonical weight factor is estimated then for the production run of the multicanonical MC, one proceeds with the following steps [33].

8_{MU}. With this obtained multicanonical weight factor, perform a multicanonical simulation at one given temperature slightly higher than the characteristic transition temperature of the system.

9_{MU}. During this simulation, collect statistics for the physical variables of interest such as the energy, dihedral angles for the conformations, radius of gyration, structural order parameters and so on. These variables will be weighted according to the multicanonical distribution.

10_{MU}. From this single simulation, the canonical distribution at any temperature for a wide range of temperatures using a reweighting technique.

Once the production run is finished then we can obtain proper canonical statistics of the physical quantities by reweighting. Remember that, in the canonical ensemble, distributions are weighted with Boltzmann factor $w_{\mathrm{B}}(E) = e^{-\beta E}$ and the resulting probability distribution can be written as

$$p_{\mathrm{B}}(E,\beta) = \frac{g(E) w_{\mathrm{B}}(E,\beta)}{Z_{\mathrm{B}}} \tag{2.11}$$

where $g(E)$ is the density of states and $Z_{\mathrm{B}} = \sum_i e^{-\beta E_i}$ the partition function. In the multicanonical ensemble, the probability distribution is defined by

$$p_{\mathrm{mu}}(E) = \frac{g(E) w_{\mathrm{mu}}(E)}{Z_{\mathrm{mu}}} = \text{constant} \tag{2.12}$$

where $Z_{\mathrm{mu}} = \sum_i e^{-S(E_i)}$ is the multicanonical partition function. This relation tells that all the energies have equal weight. This feature enables the system to sample configurations that may be suppressed in canonical distribution.

The relevant reweighting technique widely used today was proposed in a multiple histogram method [37, 38]. Despite its advantage over the canonical MC for more efficient sampling, the reweighting can be done simply by choosing the inverse temperature β and solving for the Boltzmann probability distribution using the following expression

$$p_{\mathrm{B}}(E,\beta) = \frac{p_{\mathrm{mu}}(E) w_{\mathrm{B}}(E,\beta)/w_{\mathrm{mu}}(E)}{Z_{\mathrm{B}}/Z_{\mathrm{mu}}}. \tag{2.13}$$

Using this expression, we can now obtain an estimate for a given physical quantity in the canonical distribution. The average value of some quantity A, obtained through the canonical distribution, is given as

$$\langle A \rangle_B = \sum_i A_i p_{\mathrm{B}}(E_i,\beta) \tag{2.14}$$

where i is a sum over the physical variables obtained using Boltzmann statistics. The average value of A which is obtained from multicanonical simulation scheme is weighted by the multicanonical distribution and it can be properly reweighted into canonical Boltzmann statistics by the relation

$$\langle A \rangle_B = \sum_i \frac{A_i w_{\mathrm{B}}(E_i,\beta)/w_{\mathrm{mu}}}{Z_{\mathrm{B}}/Z_{\mathrm{mu}}}. \tag{2.15}$$

The statistical properties at an arbitrary temperature can be obtained from this expression.

2.4.1 Multicanonical and entropic sampling algorithms

The name of multicanonical method stems from its use of the inverse temperature $\beta(E)$ factor to approximate the entropy $S(E)$ of the system accumulated in the form of a histogram expressed by Equation 2.9. However this linear approximation of $S(E)$ by using the $\beta(E)$ and $\alpha(E)$ factor is arbitrary and calling $\beta(E)$ as an inverse of temperature is called to be artificial by Scheraga and Hao [29]. In our simulation we have approximated $S(E)$ by $a(E)E^2 + b(E)E + c(E)$ except for the bins corresponding to E_{min} and E_{max} as described in the next subsection. The primary physical significance lies on the value of the entropy $S(E)$ and not the value of individual coefficients $a(E)$, $b(E)$ and $c(E)$ and the weighting function can be conveniently written as $e^{-S(E)}$ without introducing any additional arbitrary parameters in our systems. The name *multicanonical* is used here mainly because the computational procedure described above are given by Hansmann and Okamoto who named the method multicanonical algorithm [33]. A similar computational iterative procedure was also described in less details and it was named *entropic sampling* [21]. While we will discuss only the MC method in this thesis, the generalized-ensemble method can also be applied to molecular dynamics methods.

2.4.2 Estimating smooth continuous entropy $S'(E)$ from a histogram

Suppose that we have a histogram $H(E)$ and that we would like to estimate a line between the middle points of E_{i+2} and E_{i+3} in the situation depicted in Figure 2.3. We first obtain the two curves $L1$ which passes the middle point of E_{i+1}, E_{i+2} and E_{i+3} and $L2$ which passes the middle point of E_{i+2}, E_{i+3} and E_{i+4} by Lagrange interpolation and then the line we like to define is expressed by $(L1+L2)/2$ between points E_{i+2} and E_{i+3}. If no neighboring point exists such as the energy bins of $E_{\min}$

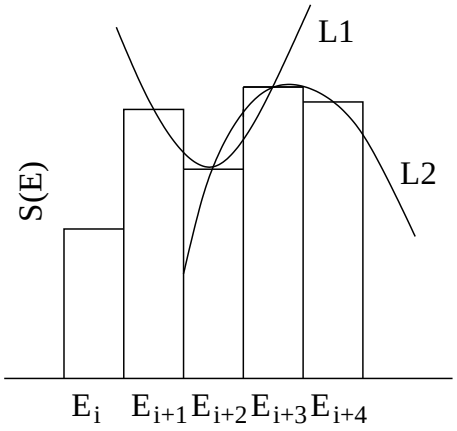

Figure 2.3: Constructing the continuous function S(E) from the histogram S(E).

and $E_{\max}$, then we simply approximate these lines by a straight line passing through $E_{\min}$ and $E_{\min+1}$, and $E_{\max-1}$ and $E_{\max}$.

This approach connects the points located in the middle of bins, that is $(E_{i+1} + E_{i+2})/2$ and $(E_{i+2} + E_{i+3})/2$ instead of the left or the right edges these energy bins as has been done in [33]. This makes the continuous curve closer to the real distribution of states. The advantage of this curve fit estimation over the simple linear fit estimation is observed to be insignificant in the computations in simple systems [34]. We have not measured quantitative differences in two methods but it was noticed that the selection of the left or the right edges in the linear approximation made differences in calculations [35]. Hence, the curve fit estimation using the middle points should provide better computational results even though the improvement might be small.

2.5 Reweighting trajectories computed with multicanonical weight factor

The multicanonical weight factor $w_{\mathrm{mu}}(E)$ is not known a priori. Obtaining the proper weight factor that can sample evenly over the wide range of energies is

not trivial and tedious for complex systems such as biopolymers and protein systems [30]. The obtained weight after many trials and errors still often fails to sample sufficient conformations, and the system is often stuck at the lower energy conformations. In practice, it is impossible to obtain the ideal multicanonical factor with a completely flat potential energy distribution, and it is considered to be a sufficient weight factor as long as one gets a random walk in potential energy space with a tolerance of an order of magnitude deviation [30].

How can we improve this computational procedure? One way to do this is to invent more efficient estimation method of the multicanonical weight factor $w_{\mathrm{mu}}(E)$. Another way is to interpret the multicanonical Monte Carlo procedure in different perspective. We can make use of the fact that the proper energy sampling is already achieved during the preparation process of the multicanonical Monte Carlo simulations, during which one tries to estimate the proper multicanonical weight factor $w_{\mathrm{mu}}(E)$. Hence we can treat each simulation of a multicanonical Monte Carlo loop as an individual simulation performed with a different umbrella weight $S(E)$. From the kth loop we can obtain the estimate of $g_k(E)$ by using the single histogram method developed by Ferrenberg and Swendsen [36]. After many loops, we can combine the results of the data calculated with the multicanonical weight at each loop. Then we obtain the optimal estimate for the density of states to obtain a wide range of parameter values in the form of continuous functions by the weighted histogram analysis method (WHAM) [37, 38]. The weighted histogram analysis method can reweigh and normalize the trajectories computed with non-Boltzmann weight to the trajectories with proper statistical weight. This multicanonical-WHAM can provide similar statistical information of the trajectory without a long multicanonical production run. Single histogram method, a algorithm much simpler but similar to the multicanonical method described above, can calculate the thermodynamic properties at temperatures other than the simulated temperature if a sufficiently accurate density of states is estimated in the relevant energy range. Hence thermodynamic quantities near the simulated temperature can be sufficiently obtained. The multiple histogram method optimally combines simulated data sets obtained at many different temperatures and can provide the thermodynamic quantities at a wider range of temperature. The weighted histogram analysis method estimates

the density of states from data computed with umbrella potential, which can be applied to various quantities such as temperature, distance, entropy and energy and others.

After we computed sufficient trajectories with the multicanonical-type non-Boltzmann weights, we needed to estimate the probability of the system in a particular bin in an ensemble defined by the potential E and temperature T. The probability of obtaining the variable A can be written as

$$P(A_j) = Z_0^{-1} \sum_E g(E) e^{-E/k_B T}. \tag{2.16}$$

An optimal estimate of the density of states can be obtained by combining the results of all M simulations and then minimizing the statistical error of the density of states $\delta^2 g(E)$ [37, 38], and is determined by the set of self-consistent equations known as WHAM equation, given in our case by

$$g(E) = \frac{\sum_{i=1}^{M} \sum_E N_i(E)}{\sum_{i=1}^{M} n_i z_i^{-1} e^{-S_i(E)}}, \tag{2.17}$$

$$z_i = \sum_E g(E) e^{-S_i(E)},\ i = 1,\ ...,\ M. \tag{2.18}$$

where $g(E)$ is the density of states and $N_i(E)$ is the number of observations of a state with a particular E collected from the trajectory saved during the ith simulation and n_i is the number of snap shots taken during the same period. A self-consistent solution $g(E)$ of the coupled equations 2.17 and 2.18 can be iteratively determined. $g(E)$ is then used to define the following weight function

$$\Gamma(E) = \frac{\sum_{i=1}^{M} \sum_E N_i(E)}{g(E)}. \tag{2.19}$$

Hence, to determine $\Gamma(E)$ we need to collect statistics on variables E. The probability of variable A_j can be now calculated through the saved trajectories by weighing

each configuration by a corresponding factor given in the following expressions

$$P(A_j) = Z_0^{-1} \sum_{i=1}^{M} \sum_{t_i=1}^{n_i} N(A_j(t_i)) e^{-E(t_i)/k_B T} \cdot \Gamma(E(t_i)) \tag{2.20}$$

where

$$Z_0 = \sum_{i=1}^{M} \sum_{t_i=1}^{n_i} e^{-E(t_i)/k_B T} \cdot \Gamma(E(t_i)) \tag{2.21}$$

and $N(A_j(t_i))$ is the number of observations of A_j in the ith simulation. The average value of a variable A_j can be then calculated by using the relationship

$$\begin{aligned} \langle A_j \rangle &= \sum A_j P(A_j) \\ &= \sum A_j Z_0^{-1} \sum_{i=1}^{M} \sum_{t_i=1}^{n_i} N(A_j(t_i)) e^{-E(t_i)/k_B T} \cdot \Gamma(E(t_i)). \end{aligned} \tag{2.22}$$

We have discussed, in this chapters, background information on main computational methods required in the thesis. In the next four chapters we introduce the reduced protein model and apply it to the protein secondary structures, β-hairpin folding and conformational conversions in proteins.

Bibliography

[1] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087-1092.

[2] Binder K. Monte Carlo simulation in statistical physics. Berlin, New York: Springer-Verlag; 1988. p127. Newman MEJ, Barkema GT. Monte Carlo methods in statistical physics. New York: Oxford University Press; 1999. p475.

[3] Binder K. Monte Carlo and molecular dynamics simulations in polymer sciences. Oxford : Oxford University Press, 1995, p587.

[4] Chan HS, Dill KA. Transition states and folding dynamics of proteins and heteropolymers. J Chem Phys 1994;100:9238-9257.

[5] Kolinski A, Skolnick J. Reduced models of proteins and their applications. Polymer 2004;45:511-524.

[6] Sali A, Shakhnovich E, Karplus M. Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. J Mol Biol 1994;235:1614-1636.

[7] Rey A, Skolnick J. Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins. Chem Phys 1991;158:199-219.

[8] Karplus M, Shakhnovich E. Protein folding: Theoretical studies of thermodynamics and dynamics. In: Protein Folding. Creighton TE. (ed.). New York: W. H. Freeman 1992:127-196.

[9] Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding - A perspective from simple exact models. Protein Sci 1995;4:561-602.

[10] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167-95.

[11] Chan HS, Dill KA. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics Proteins 1998;30:2-33.

[12] Landau DP, Binder K. A guide to Monte Carlo simulations in statistical physics. New York : Cambridge University Press, 2000, p384.

[13] Kalos MH. Monte Carlo methods. New York : J. Wiley Sons, 1986.

[14] Lee HK, Okabe Y. Reweighting for nonequilibrium Markov processes using sequential importance sampling methods. Phys Rev E 2005;71:15102. Dickman R. Reweighting in nonequilibrium simulations. Phys Rev E 1999;60:R2441-R2444.

[15] Skolnick J, Kolinski A. Computer simulations of globular protein folding and tertiary structure. Annu Rev Phys Chem. 1989;40:207-35.

[16] McCammon JA, Harvey SC. Dynamics of proteins and nucleic acids. Cambridge : Cambridge University Press, 1987, p234.

[17] Brooks CL, Karplus M, Pettitt BM. Proteins : a theoretical perspective of dynamics, structure, and thermodynamics. New York : J. Wiley, 1988, p259.

[18] Kikuchi K, Yoshida M, Maekawa T, Watanabe H. Metropolis Monte Carlo method as a numerical technique to solve the Fokker-Planck equation. Chem Phys Lett 1991;185:335-338.

[19] Cieplak M, Henkel M, Karbowski J, Banavar JR. Master equation approach to protein folding and kinetic traps. Phys Rev Lett 1998;80:3654-3657.

[20] Chubykalo O, Nowak U, Smirnov-Rueda R, Wongsam MA, Chantrell RW, Gonzalez JM. Monte Carlo technique with a quantified time step: Application to the motion of magnetic moments. Phys Rev B 2003;67:64422.

[21] Lee J. New Monte Carlo algorithm: entropic sampling. Phys Rev Lett 1993;71:211-214.

[22] Mezei M. Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias. J Comp Phys 1987;68:237-248.

[23] Bartels C, Karplus M. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. J Comp Chem 1997;18:1450-1462.

[24] Souaille M, Roux B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Comp Phys Comm 2001;135:40-57.

[25] Bouzida D, Rejto PA, Verkhivker GM. Monte Carlo study of ligand-protein binding energy landscapes with the weighted histogram analysis method. Int J Quan Chem 1999;73:113-121.

[26] Roux B The calculation of the potential of mean force using computer simulations. Comput Phys Commun 1995;91:275-282

[27] Wang F, Landau DP. Efficient, multiple-range random walk algorithm to calculate the density of states. Phys Rev Lett 2001;86:2050-2053. Wang F, Landau DP. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. Phys. Rev. E 2001;64:056101.

[28] Jeffreys H, Jeffreys BS. Methods of Mathematical Physics. 3rd ed. Cambridge : Cambridge University Press, 1988, p260.

[29] Scheraga HA, Hao MH. Entropy sampling Monte Carlo for polypeptides and proteins. Adv Chem Phys 1999;105:243-272.

[30] Mitsutake A, Sugita Y, Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. Biopolymers 2001;60:96-123.

[31] Lemak AS, Gunn JR. Rotamer-specific potentials of mean force for residue pair interactions. J Phys Chem B 2000;104:1097-1107.

[32] Berg BA, Neuhaus T. Multicanonical ensemble: a new approach to simulate first-order phase transitions. Phys Rev Lett 1992;68:9-12.

[33] Hansmann UHE, Okamoto Y. Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. Physica A 1994;212:415-437.

[34] Zhan L. Personal communication.

[35] Kemp KP. Personal communication.

[36] Ferrenberg AM, Swendsen RH. New Monte-Carlo technique for studying phase-transitions. Phys Rev Lett 1988;61:2635-2638.

[37] Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. Phys Rev Lett 1989;63:1195-1198.

[38] Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg J. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comp Chem 1992;13:1011-1021.

[39] Leach AR. Molecular modelling. 2nd ed. New York : Prentice Hall, 2001. p744.

Chapter 3

Universal model for α-helix and β-hairpin structures in protein

We introduce a unified potential-energy model that successfully reproduces the directionally dependent hydrogen bonding effect in proteins. The model can be used to represent well defined secondary structures, both α-helices and β-hairpins, whose conformations are solely determined by the sequence of the modelled protein. The model provides structural insight into the physical mechanism of such problems as structural conversion due to mutation and double native conformations with different α-helix and β-sheet contents.

3.1 Introduction

The current understanding of the characteristics of protein folding is widely based on statistical-physics models of polymers that capture the essential interactions in real protein systems (see, e.g., [1–3] and also the first chapter of the thesis). The reduction of the degrees of freedom of the coordinates in such a model, compared to those of all-atom models, allows the accumulation of adequate statistics in computer simulations. These models, commonly called "minimal models", have been

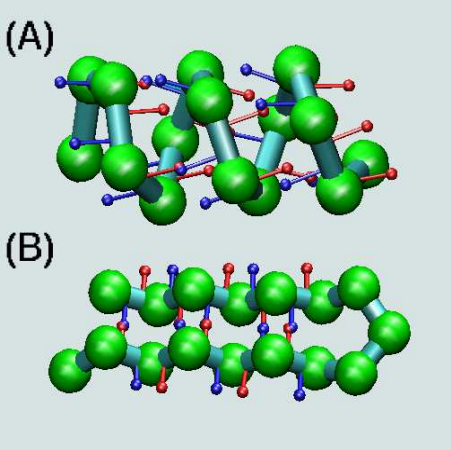

Figure 3.1: Sketch of an α-helix (A) and a β-hairpin (B). In the model, we have defined a virtual-oxygen atom (the red spheres) and a virtual-hydrogen (the blue spheres) atom for each polymer bond that are capable of making an OH bonding based on the vector relationship (see text). The size of a volume-excluding monomer (green) has been reduced in this plot for visual ease.

successfully used to explore the underlying physical mechanism of structural formation, folding dynamics and protein-protein interaction. A more challenging task is to relate minimal models to real protein conformations in which α-helix (Fig. 3.1A) and β-hairpin (Fig. 3.1B) are two essential types of secondary structures. Residue-based minimal models that separately deal with each of these two types of structures have been extensively studied in recent years [3–10]. In most cases, individual residue must be given a dihedral-angle potential energy that is specifically biased to the α-helix or β-sheet structures.

Until recently there has been no off-lattice minimal model that can describe the protein secondary structures solely based on its sequence without a predefined reference conformation. Many important structural-biology problems, such as the effects of mutation of protein sequence on the native structures [11] and the double native states in a prion protein [12], rely on a deeper understanding of the roles played by basic potential interactions, such as hydrogen bonding and hydrophobic interactions.

Hence, it is indispensable to develop a reduced potential-energy model that can universally represent the physical mechanism of both α-helix and β-sheet formation. From the protein-folding perspective, the three-dimensional native conformation of a given sequence of a heteropolymer, is entirely determined by global energy minimization and may contain different secondary structure contents for a different sequence. For this purpose, we present in this chapter a simple novel model that captures the intriguing α-helix and β-sheet duality of structural information, which can be used to reproduce both α-helix and β-sheet, mixed α-helix and β-sheet, and α-helix bundles according to residue sequence in the model.

3.2 The model

We consider a simple chain system which can fold into a β-hairpin, an α-helix or a helix-turn-helix conformation depending on the presence of neutral monomers in the middle, whose effective strength of their hydrogen bonds are set to be zero, and the length of the chain and possible native conformations are shown in Fig. 3.1.

A chain consists of various numbers of monomers connected linearly by bonds of fixed length l. Monomers are allowed to rotate about the axis connecting the adjacent monomers, subject to a stiff harmonic potential energy [3] that prefers a bond angle of $\theta = 1.833\,rad$ (105°) among three successive beads and it can be written as

$$V_\theta = \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{3.1}$$

where $k_\theta = 20\epsilon_h/(rad)^2$ where ϵ_h is the strength of the hydrogen bond in unit of energy in this model. Other reduced models, based on residue-level approximations, contain an additional potential energy that controls the preferred torsional angle formed by three consecutive bonds, relating four consecutive residues; such an energy is usually defined to drive the local bond arrangement to an α- or a β-configuration [3, 4, 8, 10]. This type of torsional potential is not used in our study.

In our model the preferred torsional angles are directly determined by a combination of a directionally biased interaction [15], as described below. The formation

of the β-hairpin and α-helix, for example, can be formed by hydrogen bonding between non-local monomers that are far apart along the chain. The excluded volume occupied by a monomer has a diameter of $6l/5$ and two nearest-neighbor residues are not subject to excluded-volume interaction. This diameter is somewhat greater than the bond length and reflects the geometry of amino acids in proteins. The hydrogen bonding is effectively described by the interaction between any pair of virtual oxygen O and hydrogen H; in particular, for the ith monomer that is described by the position vector $\vec{r}_i$, an O_i and a H_i were placed at the tips of the vectors $\vec{r}_i^{(O)}$ and $\vec{r}_i^{(H)}$ respectively, directed away from $\vec{r}_i$:

$$\vec{r}_i^{(O)} = \frac{1}{3}(\vec{r}_{i+1} - \vec{r}_i) + 3l/5\vec{n}_i \tag{3.2}$$

$$\vec{r}_i^{(H)} = \frac{2}{3}(\vec{r}_{i+1} - \vec{r}_i) - 3l/5\vec{n}_i \tag{3.3}$$

where $\vec{n}_i$ is the normal direction of the plane defined by the $(i-1)$th, ith, and $(i+1)$th monomers and expressed as $\vec{n}_i = (\vec{r}_i - \vec{r}_{i-1}) \times (\vec{r}_{i+1} - \vec{r}_i)/|(\vec{r}_i - \vec{r}_{i-1}) \times (\vec{r}_{i+1} - \vec{r}_i)|$. Except for the first and second nearest neighbors, any pair of O and H interact with each other through a shifted Lennard-Jones potential,

$$V_{OH} = 4\epsilon \left[\left(\frac{l}{r + r_0} \right)^{12} - \left(\frac{l}{r + r_0} \right)^{6} \right] \tag{3.4}$$

where $r_0 = 2^{1/6} \times 5l/6$ is the optimal bonding distance, and r is the OH distance. The strength of an effective hydrogen bond among a non-perturbed OH pair yields a bonding energy $-\epsilon$ for $r = 0$. The strength of an effective hydrogen bond associated with a neutral monomer is set to be zero. This hydrogen bonding of the 12-6 Lennard-Jones potential (Eq. 3.4) can be replaced by a 12-10 Lennard-Jones potential without significantly affecting the essential physical mechanism of structural formation. The main focus of this chapter is to describe conformations generated by hydrogen bonding, but we can incorporate hydrophobicity by introducing an hydrophobic potential energy between hydrophobic residues; non-hydrophobic monomers do not participate in the interaction. Again, we use a simple Lennard-Jones form,

$$V_H = 4\epsilon_H[(6l/5r)^{12} - (6l/5r)^6] \tag{3.5}$$

where $\epsilon_H = 0.7\epsilon$ represents the effective strength of a hydrophobic potential and r is the distance between hydrophobic residues. Irbäck and coworkers [5] investigated the effect of relative strength of hydrophobic and hydrogen attractions in similar reduced models; they have used a similar directionally biased hydrogen bond potential designed for α-helical bundles where $\epsilon_H = 0.78\epsilon$. We have examined systems corresponding to various values of ϵ_H, ranging from 0.3ϵ to 1.2ϵ, by introducing a pair of cross-strand hydrophobic residues in our model and in this chapter we use $\epsilon_H = 0.7\epsilon$.

3.3 Results and discussion

We have generated a large number of different types of native structures displayed in Fig. 3.2 with the model. Figure 3.2A shows the native conformation corresponding to the global energy minimum, obtained from multicanonical-annealing Monte Carlo simulations [14].

Homopolymer: No symmetric native conformation is observed in the systems shorter than $N = 5$ and the α-helix structure becomes stabilized starting at $N = 6$. On average, each OH bonding formed in a right-handed helix as displayed in Fig. 3.2A yields an energy -0.82ϵ. We have successfully generated these single helical structures from $N = 6$ to $N = 35$. Though a single helix is in principle the ground energy state for $N > 35$, we have not verified this due to the inefficiency in current computational power. We have also observed in our simulations that a higher local energy minimum is possible, corresponding to a rather stable left-handed helix (not shown) which contains one hydrogen bonding less than the right handed helix does. The key factor to make the right-handed helix more energetically fovorable, as in real proteins, is the introduction of the $1/3l$ bond shift used in defining vectors that give rise to virtual O and H positions. In an earlier, preliminary version of the model [15], both left- and right-handed helices are energetically equal.

Next, we consider a series of heteropolymers in which up to three neutral monomers that have no OH binding are introduced in the middle of the polymer chain. Practically, the virtual-O and H atoms associated with any bonds connected

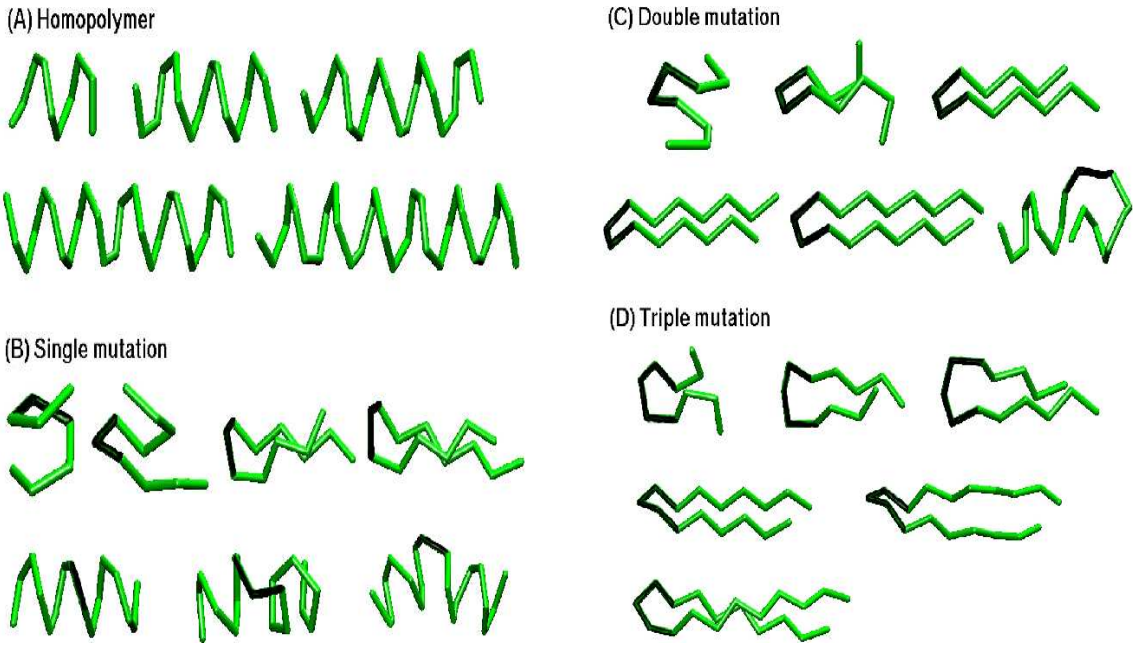

Figure 3.2: Native structures of homopolymer (A) and heteropolymer (B-D). In heteropolymers the hydrogen bonding has been removed from the black bonds. In plot (A) $N = 10$, 16, 20, 26, 30 polymers are shown as the representatives of the α-helix structure. In plot (B) one neutral monomer is introduced in the middle. No identifiable structures are shown in the first two plots for $N = 8$, 10 polymers. $N = 12$, 14 polymers show β-hairpin conformation and $N = 16$, 18, 20 polymers show α-helix conformation. In plot (C) two neutral monomers are introduced in the middle. $N = 9$ polymer has no clearly identifiable structure and $N = 11,\ 13,\ 15,\ 17$ polymers display a β-hairpin turn. Starting from $N = 19$, polymers from helical structures with a kink in the middle. In plot (D) three neutral monomers are introduced in the middle. Polymers with number of monomers less than $N = 10$ display no particular structure. $N = 12,\ 14,\ 16,\ 18,\ 20$ polymers are shown to display β-hairpin conformation.

to the neutral monomers are removed. All other physical conditions, including bond-angle and excluded-volume constraints, remain the same as above.

Heteropolymer with a neutral monomer: The neutral residue disrupts the OH bonding in the middle. While no particular structures can be identified for global energy minima of chains with $N < 10$, polymers of lengths $N = 12$ to $N = 14$ display β-hairpin native conformations, as displayed in Fig. 3.2B. The first OH bond near the β-turn is weak because of the angular constraint of the middle neutral bonds. On average, the OH bonding energy is -0.87ϵ in a β-hairpin conformation. To shorten the bonded OH distance, the two arms of the polymers twist themselves and the β-hairpin is not completely flat, as can be seen in Fig. 3.2B. $N = 16$ presents the critical polymer length beyond which an α-helical conformation becomes the native conformation. This can be understood by a simple examination of the number of OH-bonds. Six OH bonds can form in a $N = 16$ β-hairpin with a total energy -5.2ϵ. Although this β-hairpin conformation can be shown to correspond to a local energy minimum, a single α-helix would produce seven OH bonds with an energy approximately -5.6ϵ, corresponding to the global energy minimum. Beyond $N = 16$, the global energy minimum corresponds to a structure that contains two helices connected to a kink in the middle. The introduction of every new monomer yields a new OH bonding in helix, whereas two new monomers are needed to form an OH bonding in a two-stranded β-hairpin conformation. We have successfully produced such helical conformations up to $N = 32$.

Heteropolymer with two neutral monomers: No particular structures can be identified less than $N = 11$. Polymers with total monomer length $N = 11$ to 17 display a flat β-hairpin conformation corresponding to the ground-state energy minimum. The three bonds connected to the two neutral monomers can make a much better hairpin turn with the given bond angle constraints. As a result, each OH bonding attains an average energy of -0.98ϵ. $N = 19$ represents the critical polymer length beyond which α-helices can be stabilized in a ground state. These helical structures contain two-α helices connected to the two middle neutral monomers and each separate arm forms an almost perfect right-handed helix. For $N = 18$, the α-helix structure has the total OH bonding energy -7.2ϵ, which can be

compared to a local energy minimum of -6.9ϵ corresponding to a double stranded β-hairpin.

Heteropolymer with two neutral monomers: More configurational freedom exists in the middle section. The polymer can form a much better hairpin turn. For polymer length from $N = 12$ to $N = 20$, the ground-state energy minimum corresponds to a β-hairpin conformation. An average OH bonding energy of -1.05ϵ can be achieved, implying an almost perfect OH bonding which would yield $-\epsilon$ per bond without the angle bond penalties. The lower value of the average bonding energy less than minus one is attained through a weaker interaction between nearby OH pairs that are not directly bonded.

A critical polymer length of $N = 20$ exists, beyond which α-helices become stable structures. Notice that the conformational preference to β- and α-conformations of the $N = 20$ chain differs from the previous two critical systems. Here, the β-hairpin structure corresponds to a global minimum and the kinked α-helical structure to a local minimum, yielding energies -7.4ϵ and -7.1ϵ respectively. The α-helix structure, though, is more kinetically accessible than that of the lower-energy β-hairpin, due to the amount of non-local OH bonding needed to form a β-hairpin. This observation is similar to the one obtained by Borg and coworkers who introduced directional hydrogen bonding into a well known lattice model to enhance the secondary structures. They found that a metastable α-helix is easily attainable while a groundstate β-hairpin exists [16]. The upper limits of the strand length in antiparallel double stranded β-hairpins in peptides [17] have recently been examined using NMR by Stanger et al. [18]. They have shown that beyond a certain number of residues, two-stranded β-hairpin will destabilize, in agreement with our modeling result here.

So far, we have described the native conformations associated with the ground-state energies for both homopolymers and heteropolymers. A detailed free-energy study at finite temperatures will provide us more information on the structural stability and folding pathways, as will be presented elsewhere [19]. In the reminder of this chapter, we discuss a number of possible applications of this model in protein folding.

The structural conversion due to mutation in a protein sequence is an area of

crucial biological importance [20]. Despite numerous models developed for protein folding problems, transforming α-helices into β-hairpins or vice versa by a simple mutation of the sequence has been rarely reported in molecular simulation studies [21, 22]. One exception is the recent study by Nelson and Onuchic, who have investigated mutational effects and developed a quantifying method for mutation using a simple off-lattice alpha carbon model, however, with no specific secondary structures identifiable. The stability of α-helices and β-hairpins can be affected by introducing mutants in the sequence placed at strategic locations. Recent experiments performed by Cordes and coworkers have suggested that mutations adjacent to the antiparallel β-sheet of the Arc repressor are sufficient to change the local secondary structure to a right-handed helix [11]. From the mutation perspective, we may view the heteropolymer structures described above as the mutated counterpart of the homopolymer of the same length or vice versa. With the mutation of the middle monomers in a homopolymer model, we see that an α-helix (Fig. 3.2A) undergoes a structural conversion to a β-hairpin (Fig. 3.2B, 3.2C and 3.2D).

Protein structure design is another area where a good minimal model is highly desirable. We stress that only OH bonding is effectively considered in the simple model above. This does not preclude us from introducing other types of interactions, such as the hydrophobicity, into the model. We can show that the introduction of additional *isotropic* attractions between monomers stabilizes more complex native structures with mixed α-helices and β-sheets [19]. For example, Fig. 3.3 is the ground-state-energy conformation of a $N = 32$ heteropolymer chain based on our model presented above, in which neutral-monomer insertions have been placed at strategic positions, shown in black. This motif is representative of the zinc finger domain of the Cys_2His_2 type that is the most abundant DNA binding motif in eukaryotic transcription factors [23]. In another example, by placing neutral monomers at the $(16-18)$th and $(34-36)$th positions of a $N = 51$ homopolymer and introducing a weak hydrophobic attraction between monomers, whose strength is 0.3ϵ, we have reproduced a bundled structure with three helices (Fig. 3.4). The interplay of the hydrophobic and hydrogen-bonding interactions has been the recent subject of study [5]. Irbäck and coworkers, for example, studied the three α-helix bundle structure in detail and examined the significance of the magnitude of hy-

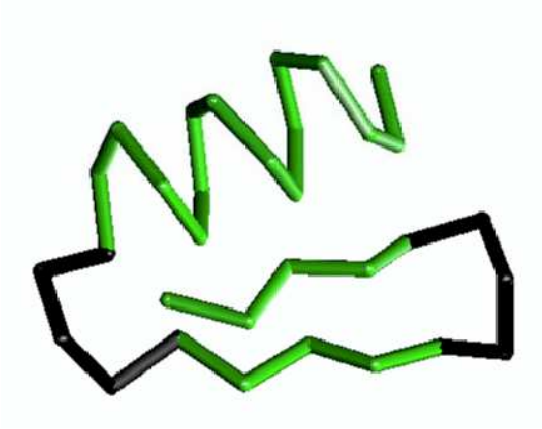

Figure 3.3: The α-β motif obtained with the help of additional hydrophobic attraction that stabilizes the tertiary structures. The bonds in black are neutral. Hydrophobic atoms are placed at monomer $i = 2, 3, 7, 8, 12, 13, 28, 30$, and 32, counted from the beginning terminal of the α helix.

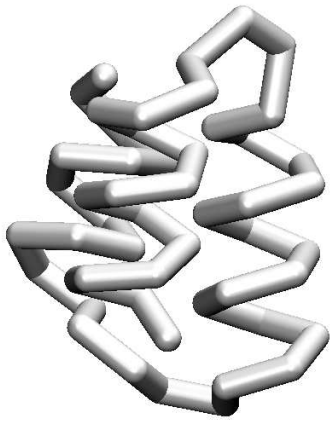

Figure 3.4: Three bundle α-helices: Hydrophobic atoms are placed at monomer $i = 4, 5, 9, 10, 14, 15, 22, 23, 27, 28, 32, 33, 40, 41, 45, 46, 50$, and 51, and nutural residues are placed at $i = 16, 17, 18, 34, 35$ and 36. counted from the upper left terminal end of the α-helix.

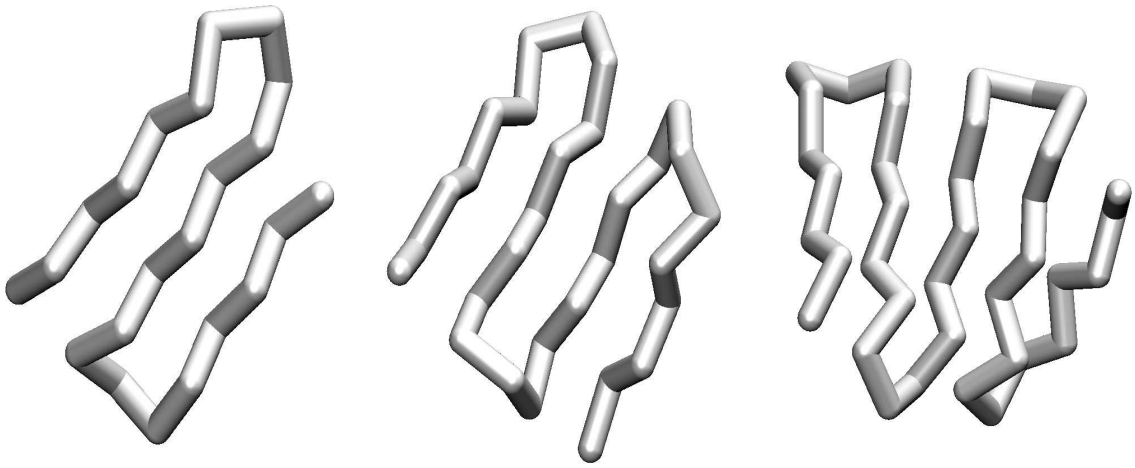

Figure 3.5: Multiple beta stranded β-sheets: Each β-turn contains three neutral residues and each strand contains five residues.

drogen bonding in the system. Similar structures have been extensively studied by a number of other authors [3, 6, 24]. Multiple stranded β-sheet conformations are also possible by inserting neutral residues at proper β-turn sites. The β-sheets with three, four and five strands, which were computed by the canonical MC method, are shown in Fig. 3.5. These conformations shown do not contain hydrophobic interactions. Similar types of multiple β-hairpin conformations are studied in experiments [27–29] and computations [30].

Finally, our model can be used to study proteins with dual native conformations that form different α-helix and β-sheet contents. For example, prion proteins are known to display completely different native structures of similar energies [12]. An α-helix dominated normal structure is kinetically more accessible, and a β-sheet rich disease structure is energetically more stable. The two native conformations are partitioned in the energy landscape [25]. Chen et al. [26] have recently extended the model presented here to simulate the energy landscape and folding kinetics of a minimal prion-like homopolymer model. They have shown that the competition between the hydrogen bonding and hydrophobic interactions yields two energetically favorable secondary structures, an α-helix and a two stranded β-hairpin (see also

Ref. [16]). The numerical evidence also shows that the folding from a denatured random conformation prefers to yield an α-helix, although the β-hairpin structure has a lower energy.

3.4 Conclusion

We have carried out numerical simulations of a new minimal model of proteins which sufficiently addresses the ability of directionally biased hydrogen bonding that can define the secondary structures in proteins. Both α-helix and β-hairpin conformations as well as more complicated conformations such as an α/β-compound, helix bundles and multi-stranded β-sheet structures can be stabilized in hetero-polymers following the information encoded in the sequence without *a priori* biased potentials to a specific native conformation. This simple model, which is based on a residue-level approximation with an explicit introduction of virtual hydrogen and oxygen atoms interacting through a shifted Lennard-Jones potential, gives rise to a large variety of protein-like conformations and can be used as the base for further extension to other models.

In the following chapters we will apply this model β-hairpin folding in details and to the interconversion between β-hairpin and α-helix conformations.

Bibliography

[1] Chan HS and Dill KA. The protein folding problem. Phys Today 1993;46:24-32.

[2] Šali A, Shakhnovich EI, Karplus M. How does a protein fold. Nature 1994;369:248-251. Shakhnovich EI. Proteins with selected sequences fold into unique native conformation. Phys Rev Lett 1994;72:3907-3910.

[3] Honeycutt JD, Thirumalai D, The nature of folded states of globular proteins. Biopolymers 1992;32:695-709.

[4] Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a de novo designed four-helix bundle protein. J Mol Biol 1996;263:323-343.

[5] Irbäck A, Sjunnesson F, Wallin S J. Biol. Phys Hydrogen bonds, hydrophobicity forces and the character of the collapse transition 2001;27:169-179.

[6] Takada S, Luthey-Schulten Z, Wolynes PG. Folding dynamics with nonadditive forces: a simulation study of a designed helical protein and random heteropolymer. J Chem Phys 1999;110:11616-11628.

[7] Hardin C, Luthey-Schulten Z, Wolynes PG. Backbone dynamics, fast folding, and secondary structure formation in helical proteins and peptides. Proteins 1999;34:281-294.

[8] Shea JE, Nochomovitz YD, Guo ZY, Brooks CL III. Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist beta-barrel model J Chem Phys 1998;109:2895-2903.

[9] Shea JE, Brook III CL, From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 2001;52:499-535.

[10] Sorenson JM, Head-Gordon T. Matching simulation and experiment: A new simplified model for simulating protein folding. J Comp Biol 2000;7:469-481.

[11] Cordes MHJ, Walsh NP, McKinght CJ, Sauer RT. Evolution of a protein fold in vitro. Science 1999;284:325-327.

[12] Prusiner SB. Molecular biology of prion diseases. Science 1991;252:1515-1522.

[13] The choice of this particular angle follows Ref. [3]. A variation of this angle does not affect the main conclusion.

[14] Hansmann UHE, Okamoto Y. Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. Physica A 1994;99:415-437.

[15] Kemp JP, Chen ZY. Formation of helical states in wormlike polymer chains. Phys Rev Lett 1998;81:3880-3883.

[16] Borg J, Jensen MH, Sneppen K, Tiana G. Hydrogen bonds in polymer folding. Phys Rev Lett 2001;86:1031-1033.

[17] Fisk JD, Powell DR, Gellman SH. Control of hairpin formation via proline configuration in parallel β-sheet model systems. J Am Chem Soc 2000;122:5443-5447.

[18] Stanger HE, Syud FA, Espinosa JF, Giriat I, Muir T, Gellman SH, Length-dependent stability and strand length limits in antiparallel β-sheet secondary structure. Proc Nat Aced Sci 2001;98:12015-12020.

[19] Imamura H, Lemak A, Chen JZY. to be published.

[20] Bussemaker HJ, Thirumalai D, Bhattacharjee JK. Thermodynamic stability of folded proteins against mutations. Phys Rev Lett 1997;79:3530-3533.

[21] Skorobogatiy M, Tiana G. Mapping of mutation-sensitive sites in proteinlike chains. Phys Rev E 1998;58:3572-3577.

[22] Nelson ED, Onuchic JN. Proposed mechanism for stability of proteins to evolutionary mutations. Proc Natl Acad Sci USA 1998;95:10682-10686.

[23] Berg JM, Shi Y. The galvanization of biology: a growing appreciation for the roles of zinc. Science. 1996;271:1081-1085.

[24] Zhou Y, Karplus M. Folding thermodynamics of a model three-helix-bundle protein Proc Natl Acd Sci USA 1997;94:14429-14432. Interpreting the folding kinetics of helical proteins. Nature 1999;401:400-403.

[25] Abkevich VI, Gutin AM, Shakhnovich EI. Theory of kinetic partitioning in protein folding with possible applications to prions. Proteins 1998;31:335-344.

[26] Chen JZY, Lemak A, Lepock JR, Kemp JP. Minimal model for studying prion-like folding pathways. Proteins 2003;51:283-288.

[27] Deechongkit S, Nguyen H, Powers ET, Dawson PE, Gruebele M, Kelly JW. Context-dependent contributions of backbone hydrogen bonding to β-sheet folding energetics. Nature 2004;430:101-105.

[28] Syud FA, Stanger HE, Mortell HS, Espinosa JF, Fisk JD, Fry CG, Gellman SH. Influence of strand number on antiparallel β-sheet stability in designed three- and four-stranded β-sheets. J Mol Bio 2003;326;553-568.

[29] Ferguson N, Johnson CM, Macias M, Oschkinat H, Fersht A. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. Proc Nat Acad Sci USA 2001;98:13002-13007.

[30] Brown S, Fawzi NJ, Head-Gordon T. Coarse-grained sequences for protein folding and design. Proc Nat Acad Sci USA 2003;100:10712-10717.

Chapter 4

Dependence of folding kinetics and structural stability on the location of a hydrophobic pair in β-hairpins

We study the dependence of folding time, nucleation site and stability of a model β-hairpin on the location of a cross-strand hydrophobic pair, using a coarse-grained off-lattice model with the aid of Monte Carlo simulations. 6500 independent folding trajectories have been produced dynamically by our simulations, forming the basis for extensive statistical analysis. Four folding pathways, zipping-out, middle-out, zipping-in, and reptation, have been closely monitored and discussed in all 7 sequences studied. A hydrophobic pair placed near the β-turn or in the middle section effectively speed up folding; a hydrophobic pair placed close to the terminal ends or next to the β-turn encourages stability of the entire chain.

4.1 Introduction

4.1.1 General

A deeper understanding of the physical mechanisms behind the unique structural and folding properties of proteins has recently emerged from both experimental and modelling approaches. The remarkable diversity and simplicity of structures are mainly controlled by a balance between a few essential types of non-covalent interactions: van der Waals, hydrogen-bonding, hydrophobic and electrostatic interactions [1]. These interactions, together with the conformational entropy (reflecting the freedom in bond fluctuations), are also responsible for the overwhelming efficiency of the dynamical properties such as folding in proteins.

In this chapter we are particularly concerned about the key factors that influence the folding dynamics and structural stability of a simple β-hairpin (see Fig. 4.1), consisting of two antiparallel strands. The physical mechanism for the formation of the α-helix has been a more traditional topic and extensively explored both experimentally [2] and theoretically [3]. In contrast, much less attention has been paid to the β-sheet, until Muñoz and Eaton et al. [4] presented their seminal work on the folding timescale and mechanism of a β-hairpin — the C-terminal of protein GB1. Unlike the α-helix in which local interactions involving neighboring residues are mainly responsible for the structural stability [5], the location of the cross-strand interactions of a β-hairpin sequence, which are mostly non-local, has been identified as one of the major factors influencing the structural properties [6–9]. The effects of the cross-strand interactions on the β-hairpin stability and folding efficiency, dependent on the peptide sequence, have been the main focus of some recent experimental and modelling studies [4, 6–36]. These studies have provided a clear evidence of the importance of cross-strand interactions; the conclusions from these studies, however, are not always consistent.

Recent computer simulations have been performed to address issues related to β-hairpin folding based on two categories of models: all-atom [23–31] and minimal models [4, 22, 32–39]. All-atom models have been used to directly generate the microscopic folding trajectories of a β-hairpin — mostly for the C-terminal of protein

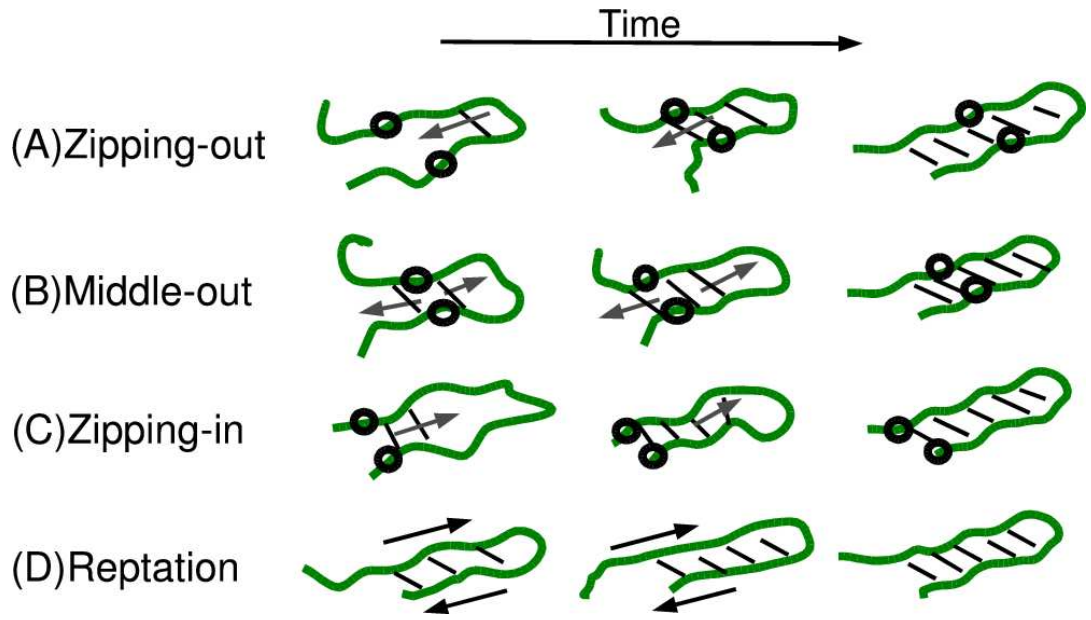

Figure 4.1: Summary of major folding pathways seen in our computer simulations: (A) zipping-out, (B) middle-out, (C) zipping-in, and (D) reptation. Zipping-out was observed in all 7 sequences and is more common in sequences containing no hydrophobic pair or a hydrophobic pair close to the β-turn. Middle-out was observed in sequences containing a hydrophobic pair located in the middle or near the terminal ends. Zipping-in was more often observed in sequences containing a hydrophobic pair located near the terminal ends. In reptational folding the chain makes a displaced β-hairpin first and then shifts the strands to finally reach the native β-hairpin, step by step; this pathway has observed for all 7 sequences.

GB1 in previous studies, in connection to experimental observations. These models offer a unique perspective on the detailed physical properties of protein folding under a given environment; such details and idealized environment cannot always be achieved by current experimental techniques. However, the complexity of all-atom models is also computationally too demanding for thorough characterization of the thermodynamics and dynamics in a systematic way; a comprehensive investigation of all-atom models for the β-hairpin, which requires the consideration of a class of sequences and multiple simulations of folding trajectories for a given sequence, is beyond the reach of current computational capacity. Therefore, one must make a proper choice between the faithful representation of a real system and the necessary simplification without sacrificing the essential physics. A simple minimal model, constructed from a careful parameterization of essential physical features, can be used to characterize the generic properties of β-hairpin systems. The advantages of minimal models lie in their ability to rapidly accumulate meaningful statistics on folding pathways, dynamics and thermodynamics, and in the ease of controlling most relevant parameters to the systems [40, 41].

In this chapter, we use an off-lattice minimal model to systematically explore the effects of cross-strand hydrophobic interaction of a β-hairpin on the folding time and structure stability. The model contains monomers that can interact with reach other through effective hydrogen bonding and hydrophobic interactions. Our treatment does not invoke the simplification of native contact-energy approximation, known as the Gō-approximation, used in some other models [42]. The Gō-approximation, however, does not reflect the actual dynamics in which non-native-contact residues can interact with each other. The current model is an extension to the previous work on characterizing the folding pathways for a prion-like model [38, 39] and the structural conversion between the α-helix and β-hairpin due to mutation in a peptide sequence [39]. In protein folding modeling, one of the major challenges is to find a single universal set of potential parameters for any protein conformations that can properly fold into a native conformation depending on only information encoded in the sequence [43] and our model meets this criterion.

Most analysis, based on computations performed to simulate the β-sheet formation, examined the free energy landscape, which is a thermodynamic perspective.

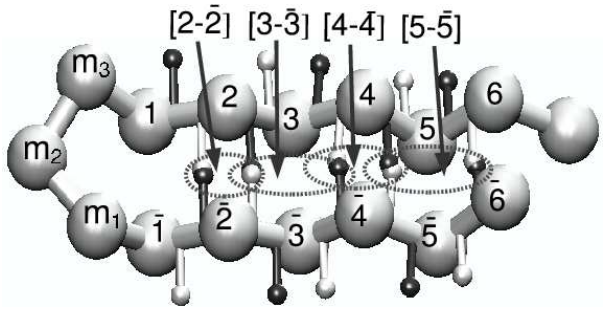

Figure 4.2: The native conformation and native hydrogen pair arrangement. Large gray monomers stand for α-carbon, small white spheres virtual Oxygen residues, and small black spheres virtual Hydrogen residues. One non-hydrophobic and six hydrophobic sequences have been studied in this chapter; for the latter, a cross-strand hydrophobic pair is selectively placed at symmetric residues. The native hydrogen bonding pairs are identified by the notation $[2-\bar{2}]$, $[3-\bar{3}]$, $[4-\bar{4}]$ and $[5-\bar{5}]$.

Only a limited number of computer simulations have directly been used to reproduce the entire folding trajectory. In this chapter, we examine both thermodynamic properties and dynamic folding events; the latter is based on 6500 independent folding simulations starting from a random conformation of the modelled chain, for the 7 sequences considered at various temperatures. Both multicanonical and canonical Monte Carlo techniques have been used in this chapter: the former has been used to produce the thermodynamical properties and the latter the dynamics. These folding simulations have been carefully analyzed in relationship to previously reported folding pathways of a β-hairpin, as summarized in Fig. 4.1.

4.1.2 Experimental and theoretical works on β-hairpins

The β-hairpin is structurally simple, but experimental determination of the structural properties has been challenging, partially because most natural β-hairpins are unstable in water and often form aggregates [8]. Recent improvements in the NMR technique as well as in selecting and creating a stable β-hairpin have made it possible to characterize the stability of the β-hairpin, in particular the effects caused by

the position of a hydrophobic pair and other long range interactions. However, it is too early to draw a conclusive physical picture based on the experimental findings alone.

Some groups have found that a cross-strand interaction placed near the turn or the middle part enhanced the stability of β-hairpins and one placed near the terminal ends only weakly enhanced the stability. For example, Espinosa et al. [11] reached this conclusion, using NMR analysis to examine the positional effect of a hydrophobic pair on the stability of 3 isomeric, β-hairpin forming peptides; these peptides contain the same residues but differ only by the separation between the D-Pro-Gly β-turn and the hydrophobic pair. In a different study, Santiveri et al. [12] examined a series of sequences that only differ by the location of two residues and have made a similar conclusion.

On the other hand, others have stressed the importance of cross-strand interactions far away from the β-turn. Kiehna and Waters [13] demonstrated that the substitution of a hydrophobic Phe-Phe pair, placed close to the turn or at the end terminals, would in each case increase the folded β-hairpin population in the samples. Griffiths-Jones et al. [14] found that a salt bridge at the terminal ends, between the cationic side-chain of Lys and the C-terminal carboxylate group, contributes to stability of a 16-residue β-hairpin consisting of natural amino acids using NMR analysis. Drawing a similar conclusion, Ciani et al. [15] found that a salt bridge near either the β-turn or the terminals can enhance stability of a β-hairpin; Fesinmeyer et al. found that the salt bridge between two lysine residues and a glutamic acid in mutant trpzip4 β-hairpins can enhance the stability [16].

The first experimental determination of the folding rate of a β-hairpin, GB1, was reported by Muñoz et al. [4] using infrared spectroscopy coupled with laser-induced temperature-jump technique. The temperature dependency of the native β-hairpin population of GB1 was also measured and the native hairpin population is estimated around 80% at 273K and 45% at 298K. Using a similar laser temperature-jump technique, Xu et al. [17] found that a synthesized β-hairpin of fifteen residues has a broad conformational transition but folds fast, only a few time slower than an α-helix. Dyer et al. [18] analyzed the formation rate of various β-hairpin and argued that the sequence whose hydrophobic pair is placed closer to the turn folds

faster than the sequence whose hydrophobic pair is placed further away from the turn, which is predicted by the theoretical model of Muñoz and Eaton et al. [4, 22] This acceleration was attributed to a direct reduction of the entropy in a smaller loop. The issue, however, has not yet been settled because some contradictory observation [19] and prediction [33, 34] exist.

Computational modelling enables us to gain microscopic insight into the folding mechanism of β-hairpins, in complementary to experimental investigations. Muñoz and coworkers [4, 22] have developed an Ising-like model, which was used to calculate the thermodynamic quantities analytically, to further investigate the physical mechanism in conjunction to their experimental observation [4] They suggested that the folding starts from the nucleation of the native structure near the turn and then would follow by the propagation of other hydrogen bonding. This folding scenario, called "zipping-out" in Fig. 4.1, was also supported by other numerical evidences based on various models. For example, Kolinski et al. [32] studied a high-resolution lattice model using Monte Carlo dynamics and entropy-sampling Monte Carlo, and obtained a similar physical picture. Klimov and Thirumalai [33, 34] constructed a simplified off-lattice model containing side groups of 16-residues to represent the C-terminal fragment of protein GB1; they used the multiple histogram technique to study the thermodynamics and Langevin method to study the folding dynamics. They suggested that a hydrophobic cluster closer to the β-turn reduces the folding cooperativity and stability and leads to slightly longer folding times; they also suggested that the hydrophobic interactions closer to the hairpin ends enhance cooperativity and stability and speed up folding rates. Note that their finding is different from the suggestion made by Muñoz and Eaton et al. [4, 22] and the experimental observations [17, 18] that the folding time is shorter in a system where a hydrophobic pair is placed closer to the turn. These two groups [32–34] also detected a small number of the hydrophobic core driving the folding process in their folding trajectories. For comparison, in this study, we found that "zipping-out" is one of the most frequently observed pathways (probably mixed with other pathways) in all types of sequences, no matter where the hydrophobic pair is placed. A hydrophobic pair placed near the β-turn or in the middle section effectively speeds up folding; a hydrophobic pair placed close to the terminal ends or next to the

β-turn encourages stability of the entire chain.

Dinner et al. [23] have stressed a different folding mechanism of the C-terminal of protein GB1 — hydrophobically driven folding pathway; the nucleation mainly starts from the hydrophobic cluster and then the entire native structure forms next to nucleation. Their study was based on the free energy analysis obtained by the multicanonical Monte Carlo method on an all-atom model with an implicit solvent treatment at 300K. This hydrophobic driven folding pathway has been supported by other all-atom model studies, including the molecular dynamics simulations performed by Pande and Rokhsar [24], the molecular dynamics simulation performed by Zagrovic et al. [25] using a distributed-computing/ensemble-dynamics approach, the free-energy analysis made by García et al. [26] using the replica exchange method, and the discontinuous molecular dynamics simulation performed by Zhou and Linhananta [27] based on an all-atom model with the Gō-approximation. For comparison, in this study, we have also found that the location, where the hydrophobic hair is placed, has a high probability to become the nucleation site. Depending on where the location is, the hydrophobic driven pathway could correspond to one of the "zipping-in", "middle-out" and "zipping-out" pathways shown in Fig. 4.1. For a given sequence, other folding pathways may coexist with the hydrophobic driven pathway (also see Zhou et al. [28]).

In addition to these folding pathways, Wei et al. [29–31], based on an energy minimization technique [45], have suggested yet another possibility — reptational folding (Fig. 4.1D). Previous models based on native contact energies would have missed the reptational folding because these models do not incorporate the non-native interactions necessary for reptational folding. Wei at al. observed the existence of displaced β-hairpins, which may be related to reptational folding (Fig. 4.1D). In simulations [23, 25, 32, 35] these conformations are often regarded as off-pathway conformational traps rather than molten globular conformations leading to the native states. The presence of displaced β-hairpins are, however, rarely reported in experiments. For comparison, in this study, we found an adequate evidence that non-native bonding, in particular corresponding to displaced β-hairpins, is significant during the folding. We have also found a clear evidence, through dynamic folding simulations, of a dynamic stage that produces the final move in the

reptational folding.

Furthermore, a computational and modelling approach that does not necessarily incorporates the basic concepts of Monte Carlo and molecular dynamics has also been used to study relevant systems [36]. In a Gaussian-potential treatment where the potential between residues is written in a quadratic form — which is accurate only for small deviations from the native state — Guo et al. [36] noted that the formation of a β-hairpin starts from the turn in the system of weak hydrophobic interaction and it starts from the hydrophobic core in the system of stronger hydrophobic interaction.

4.2 The model and method

4.2.1 The minimal model

In recent years, a number of minimal models have been developed for computational studies of a β-hairpin, [4, 22, 32–38] and β-barrels that contain more exotic structures [47, 48]. The minimal model used for the current chapter is an extension of a previous model proposed by our group [37–39]. In particular, we consider a chain consisting of 16 monomers connected linearly by bonds of fixed length l, which, when folded into a β-hairpin, has a native conformation shown in Fig. 4.2. Monomers are allowed to rotate about the axis connecting the adjacent monomers, subject to a stiff harmonic potential energy that prefers a bond angle of 105°. Most other minimal models, based on residue-level approximations, contain an additional potential energy that deals with a preferred torsional angle formed by the three consecutive bonds between four consecutive residues; such an energy is usually proposed to drive the local bond arrangement to form an α or β configuration. This type of torsional energy, however, is not considered here. In our model the preferred torsional angles directly follow from a combination of directionally biased interactions, [49] as described below. The formation of the β-configuration, for example, could be driven by the hydrogen bonding between non-local monomers that are far apart along the chain [37, 39].

The excluded volume occupied by a monomer has a diameter of $6l/5$ and two nearest-neighbor residues are not subject to excluded-volume interaction. This diameter is somewhat greater than the bond length and reflects the geometry of the resulting α helix, in matching real systems [37]. The hydrogen bonding is effectively described by the interaction between any pair of virtual oxygen O and hydrogen H; in particular, for the ith monomer that is described by the position vector $\vec{r}_i$, an O_i and a H_i were placed at the tips of the vectors $\vec{r}_i^{(O)}$ and $\vec{r}_i^{(H)}$ respectively, directed away from $\vec{r}_i$:

$$\vec{r}_i^{(O)} = \frac{1}{3}(\vec{r}_{i+1} - \vec{r}_i) + 3l/5\vec{n}_i \tag{4.1}$$

$$\vec{r}_i^{(H)} = \frac{2}{3}(\vec{r}_{i+1} - \vec{r}_i) - 3l/5\vec{n}_i \tag{4.2}$$

where $\vec{n}_i$ is the normal direction of the plane defined by the $(i-1)$th, ith, and $(i+1)$th monomers. Except for the first and second nearest neighbors, any pair of O and H interact with each other through a shifted Lennard-Jones potential,

$$V_{OH} = 4\epsilon\left[\left(\frac{l}{r+r_0}\right)^{12} - \left(\frac{l}{r+r_0}\right)^{6}\right] \tag{4.3}$$

where $r_0 = 2^{1/6}\times 6l/5$ is the optimal bonding distance, and r is the OH distance. An effective hydrogen bonding is reached when $r = 0$ that allows the system to acquire a bonding energy $-\epsilon$. According to this model, we have previously demonstrated that α-helices and β-hairpins can be stabilized as the chain length, temperature, and sequence vary [37–39].

In order to extend this previous model to include hydrophobicity, we consider here an additional hydrophobic potential energy between pairs of hydrophobic residues; non-hydrophobic monomers do not participate in the interaction. Again, we use a simple Lennard-Jones form,

$$V_H = 4\epsilon_H[(6l/5r)^{12} - (6l/5r)^6] \tag{4.4}$$

where $\epsilon_H = 0.7\epsilon$ represents the effective strength of a hydrophobic potential and r is the distance between two hydrophobic residues. Irbäck and coworkers [50] investigated the effect of relative strength of hydrophobic and hydrogen attractions

in similar minimal models; they have used a similar directionally biased hydrogen bond potential designed for α-helical bundles and considered $\epsilon_H = 0.78\epsilon$. In this chapter, we use $\epsilon_H = 0.7\epsilon$ to investigate of the impact of a cross-stranded hydrophobic pair on β-hairpin formation.

Physically, the stabilization of a β-hairpin is critically dependent on the residue type near the β-turn. The basic strength of a backbone hydrogen bonding connecting a residue pair does not widely vary among different pairs [51]. Some amino acids, however, have significantly different side-chain conformation than others and they can promote or disrupt the formation of hydrogen bonding in the secondary structures. For example, glycine has a proton on the side chain hence is flexible, and proline has a unique ring structure hence the backbone hydrogen can not contributes to β-sheet formation [52]; these residues often disturb and break the hydrogen-bond network of an α-helix and β-strands, and they frequently comprise a reverse turn of the β-sheet where no intra-strand backbone hydrogen bonding is formed [1]. In addition, a non-standard amino acid, α-amino isobutyric acid (Aib), which contributes weaker backbone hydrogen bonding, was designed for enforcing a β-turn as well [53]. This can be reflected in our model by effectively erasing $V_{\rm OH}$ for these residues. Schematically, the residues at the β-turn, labelled m_1, m_2 and m_3, are shown in Fig. 4.2 without the associated virtual O and H.

In this chapter, we focus on the effects caused by the hydrophobic interaction, as the sequence of the 16mer model varies. We systematically investigate the physical properties of 7 separate sequences, including a chain containing no hydrophobic interaction (labelled by S0) and 6 chains, each containing a hydrophobic pair located at $(1,\bar{1})$, $(2,\bar{2})$, $(3,\bar{3})$, $(4,\bar{4})$, $(5,\bar{5})$ and $(6,\bar{6})$ (labelled by S1, S2, S3, S4, S5, and S6, respectively); All these seven sequences display a native conformation of a β-hairpin, found by an energy minimization procedure, with the specific interaction parameters mentioned above.

In comparison to all-atom models, we have made several approximations to remove atomic details. By keeping the essential features of proteins in the model, we are able to examine the most important structural properties. This simplification allows us to investigate the entire folding process with sufficient statistics, which is currently not possible in more elaborate all-atom models.

4.2.2 Computational methods

We used the Metropolis Monte Carlo (MC) method [55] to perform the computer simulations required for both equilibrium and folding studies. MC was originally developed for calculating the equilibrium properties of physical systems; [55, 58] its generalization to describing a stochastic process of a nonequilibrate system is also possible if a physically tractable move and the Boltzmann weights are used [58–60]. The thermodynamic and dynamic interpretation of such a MC algorithm for protein folding has been the basis of recent theoretical studies [41, 61, 63–68]. Rey and Kolinski [61], for example, have shown that the MC dynamics in a fine lattice model can reproduce folding pathways of a three-helix bundle similar to those obtained by the Brownian dynamics. In another example, Cieplak et al. demonstrated that the MC computation matches the time evolution of the native conformation obtained through the Master equation formalism for a simple lattice model [62]. To be suitable for simulating the dynamical trajectories presented in Sects. 4.3.2 to 4.3.4, only local MC moves have been designed and the Boltzmann weight have been used with a selected T. If a monomer is not terminal along the chain, the monomer was rotated around the axis connecting the two adjacent monomers; if a monomer is terminal, with only one neighboring monomer, the monomer was rotated around the axis defined by the nearest and second nearest monomers. The acceptance of the move is then determined by a MC criterion. The rotational angle is a random number selected from the range $[-\pi/2, \pi/2]$, which leads to an acceptance rate of about $20 - 30\%$ [69]. A Monte Carlo step (MCS) in the simulation amounts to 16 attempted moves as the system contains 16 monomers. Whenever the concept of time is used in this chapter, we use one MCS as the basic dimensionless time unit. The relationship between the time measured here and the real physical time can be estimated up to an undetermined constant factor. In all 6500 folding simulations conducted in the current chapter, *all* reached the native conformation as the final structure — some may take a long computational time.

To obtain the thermodynamic properties presented in Sects. 4.3.1 and Sect. 4.3.6 with sufficient conformational samples, we computed simulation trajectories according to a biased weighing factor (the so-called multicanonical scheme) that enables a system to sample the energy space uniformly without being trapped

in local minima [56]. The bias in the weighing scheme is then removed by the weighted histogram analysis method (WHAM), [57] which translates the physical observables measured with the biased non-Boltzmann weight into those normalized by the Boltzmann weight [57].

In the rest of this section, we describe the technical procedure used to find the physical quantities characterizing the system; these include the structural overlap parameter, native state, β-hairpin, displaced β-hairpin, nucleation site, various characteristic times, and the free energy landscape.

4.2.3 The structural-overlap parameter x

The structural overlap parameter x is a measurement of the microscopic conformational similarity between the configuration under examination and a reference configuration. Computationally, x is defined as [70]

$$x = \frac{1}{N'} \sum_{i}^{N-3} \sum_{j>i+3} \Theta(D_x - |r_{ij} - r_{ij}^0|), \tag{4.5}$$

where N' is a normalization constant, $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ the distance between the ith and jth residues of the examined configuration, and r_{ij}^0 is the same pairwise distance but based on the reference configuration. $\Theta(\zeta)$ is the step function,

$$\Theta(\zeta) = \begin{cases} 1, & \zeta \geq 0 \\ 0, & \zeta < 0 \end{cases}$$

and in the current chapter, D_x is set at $0.2l$. This value $D_x = 0.2l$ is same as that used in [70]. This value allows to eliminate possible ambiguity in the conformational measurement that may arise for larger D_x values but at the same time this value allows some conformational fluctuation. We consider all non-redundant, inter-residue square distances of a given configuration except those between the adjacent residues — the latter is always fixed. This definition ensures that x becomes unity when the two configurations are almost identical. The use of x is computationally more efficient in comparison with computing the root mean square deviation (RMSD) that requires minimization in its calculation. This is particularly true when the evaluation of x needs to be performed frequently.

4.2.4 Native hydrogen bond and native state

Computationally, the configuration of the native state is obtained from energy minimization of the total potential energy, and corresponds to the global energy minimum. For all 7 sequences considered in this chapter, the native states have the configurations displayed in Fig. 4.2. We define that the formation of β-like hydrogen bonds (BHB) is associated with not one single pair, but two consecutive OH pairs, each pair reaching a distance within $0.4l$; this definition ensures that proper anti-parallel hydrogen bonds are actually formed. For example, the notation $[2-\bar{2}]$ refers to native BHB between monomers 2 and $\bar{2}$, when two OH pairs circled in Fig. 4.2 form. For the current system, a native β-hairpin contains five pairs of OH *bonds* but the number of BHB is four. In the computation, the chain is considered to have reached a native state when two conditions are met: all 4 native BHBs formed and the structural-overlap parameter x is greater than 2/3. Though the final native configurations, obtained from global energy minimization, differ slightly from each other in the 7 sequences considered here, we have found that the structural-overlap parameter x efficiently registers all these conformations as equivalent conformations ($x = 1$).

4.2.5 Nucleation site

During folding, a nucleation site with well defined local structure would form first and the rest follows. Computationally we identify a nucleation site when a native BHB forms as the precursor to the final approach to the native state. Note that starting from a random configuration, a particular BHB may form and then break during the entire folding process; the initially formed BHB, if it is short-lived, might not be the actual nucleation site. Technically, after the native conformation is reached, we trace back trajectories in time to find the configuration where the first native pair has permanently formed; this site is then considered as the nucleation site. The transition time from the formation of the nucleation site to the formation of the native state is denoted by $\tau_{\rm zip}$.

4.2.6 Folding time, unfolding time, equilibrium simulations and other characteristic time measures

Starting from a random conformation equilibrated at a high temperature $\tilde{T} = k_B T/\epsilon \gg 1$, we measure the first-passage time that the system takes to reach the native state and define it as the folding time. For each given sequence, the average folding time $\tau_{\rm fold}$ discussed below is based on 400 independent simulations of folding events for $\tilde{T} = 0.12$ and 0.11, 300 independent simulations for $\tilde{T} = 0.13$ and 0.14, 200 independent simulations for $\tilde{T} = 0.10$ and 0.15, and 100 independent simulations for $\tilde{T} = 0.09$. The unfolding time $\tau_{\rm unfold}$ is measured by the number of MC steps the system takes to undergo the transition from the native β-hairpin to an unfolded state where $x \leq 1/3$; the average unfolding time of the system is obtained by averaging over 400 events. The relative errors of the folding and unfolding times can be estimated to be approximately four, six, eight and ten percent for folding times based on 400, 300, 200 and 100 simulations, respectively. For equilibrium properties discussed below, we have performed eight independent simulations for every sequence, each consisting of 7×10^7 MCS after 10% of the computation spent on initial equilibration. The measurements of the physical properties were performed at every other 150 MC steps.

4.2.7 Free energy

The reduced free energy as a function of order parameters is obtained from the statistics accumulated in folding by dynamical MC method. We measure the number of times, histogram $\Omega(x, R/\ell)$, that a particular state — specified within a grid system that is spanned by R and x — is visited by the system during the simulation. The reduced free energy is then calculated from

$$F(x, R/\ell)/k_{\rm B}T = A - \ln \Omega(x, R/\ell) \tag{4.6}$$

where A is a normalization factor to set the base line of the reduced free energy zero. This dynamic free energy is not strictly thermodynamic free energy. For equilibrium systems, the reduced free energy is calculated by a Multicanonical-WHAM algorithm.

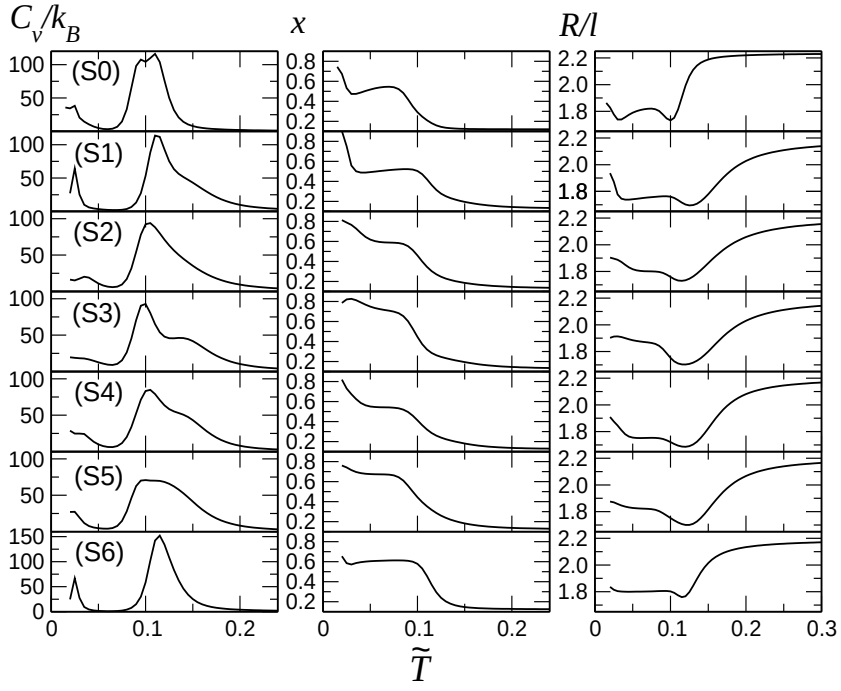

Figure 4.3: The reduced heat capacity C_v/k_B, structural overlap parameter x and the reduced root mean square radius of gyration, R/l, as functions of the reduced temperature $\tilde{T} = k_B T/\epsilon$. Each row of the plots belongs to the same sequence and each column contains the plots for the 7 sequences considered in this chapter.

4.3 Results and discussion

4.3.1 Thermodynamics

The characteristics of the folding properties are dependent on the selected temperature. To start with, we examine the thermodynamic properties of the sequences as a function of temperature, in particular, the heat capacity $C_v/k_B = (\langle (E/\epsilon)^2 \rangle - \langle (E/\epsilon) \rangle^2)/\tilde{T}^2$ calculated from the fluctuations in energy, root mean square radius of gyration R, and structural overlap parameter x defined above. The computation was carried out by the use of the multicanonical technique and reweighted by WHAM, [56, 57] which allows us to examine these thermodynamic

quantities (see Fig. 4.3) as functions of the reduced temperature, in one single long production run for each sequence.

The most distinctive feature of the C_v/k_B plots is that all 7 sequences display strong peaks somewhat above $\tilde{T} = 0.1$, which is a signature for structural transition from a coil state, stable at high temperatures, to the native state, stable below the transition temperatures. The width of the peak, however, varies from sequence to sequence. For the non-hydrophobic case S0, for example, the peak starts at a lower temperature than those of other systems (Fig. 4.3). When the temperature is lowered from a large value, the heat capacity remains small until approximately $\tilde{T} = 0.15$ and reaches a maximum approximately located at $\tilde{T} = 0.11$. This rapid increase is accompanied by a reduction of R, which decreases from 2.2ℓ, symbolizing the approximate size of a denatured chain, to approximately $0.18l$ above $\tilde{T} = 0.10$ as the chain collapses into a more compact conformation. The final folding in the native state, which can be characterized by a relatively large x, would not occur until $\tilde{T} = 0.08$; at this temperature, the chain attains an elongated conformation, hence corresponds to a slight increase in R.

This physical picture is significantly altered for sequences containing a hydrophobic pair (S1-S6). As the temperature is lowered, the chain collapses at higher temperatures, both reflected by the earlier reduction in R, and by the formation of a smaller secondary elevation on the right side of the peak of heat capacity plot. In addition, their folding temperatures, which can be defined by equating x with, say, 0.4, are higher than that of S0, which can be seen in the $x - \tilde{T}$ plots. The extra hydrophobic interaction in these sequences promotes the intra-chain attraction, which in turn encourages earlier formation of compact structures as the temperature is lowered.

Note that in about a 20% range of the reduced temperature, x varies from $x \approx 0$ (reflecting the random-coil state) to $x \approx 0.8$ (the β-sheet state). The magnitude of the temperature range is comparable to that resulting from a two-state model, presented by Muñoz et al. [4,22], where only two states, random coil and β-hairpin, are believed to be sufficient for characterizing the system. The issue of whether or not other, relatively stable intermediate states exist in a simple β-hairpin have been debated recently [20, 21, 23]. We see no clear thermodynamic indication of a stable

intermediate state from our simulations. The dynamic stages of folding, described in the next few subsections, are presented for the convenience of analyzing the process; most folding events display the cooperative behavior. The only exception to this is the reptation folding pathway, discussed in Sect. 4.3.8.

4.3.2 Folding time

We have conducted a large number of folding simulations for the 7 sequences at various temperatures. Each simulation event displays a complete folding process starting from a denatured configuration to the final formation of the native state. The amount of Monte Carlo steps, taken to accomplish this, is considered as the folding time. The average folding time, τ_{fold}, is listed in Table 4.1.

The selection of the temperature for the folding simulations has a profound impact on the simulation efficiency hence the folding time. Consider any given sequence at various simulation temperatures. At a low temperature (lower than those corresponding to the peak in Fig. 4.3) the system is most likely to be trapped in local minima, showing the glassy behavior [68]. On the other hand at a temperature higher than the folding temperature, the thermal fluctuations prevent the system from forming a native bonding and therefore the folding time also become increasingly longer. Within the numerical errors, the shortest folding time can be related to a temperature at which the heat capacity attains a maximum, i.e., $\tilde{T} = 0.12$.

At a reduced temperature above $\tilde{T} = 0.12$, the folding times are shortened as a hydrophobic pair is introduced into the chain. The folding time is drastically shortened when the hydrophobic pair is located near the turn and becomes higher as the location of the hydrophobic pair moves further away from the β-turn. The dependence of the folding time on the location of the hydrophobic pair agrees well with the dependence suggested by Muñoz and Eaton et al [4, 22] and Kolinski et al. [32], despite the difference between the modeling techniques. As well, this dependence agrees with experimental results of of Xu et al. [17] and Dyer et al. [18] On the other hand, a recent experimental [19] and theoretical [33, 34] studies have demonstrated a folding-time dependence that is very different from that described in Table 4.1.

At a reduced temperature below $\tilde{T} = 0.11$, however, the folding time becomes longer in some hydrophobic sequences. Because of the isotropic nature of the hydrophobic interaction, the hydrophobic pair can attract each other more easily in folding; once the monomers reach a distance shorter than the force range of the attraction, the two monomers are bounded in an energy trap which cannot be overcome at a low temperature. An interaction that would otherwise stabilize the native conformation may actually create a kinetic trap to frustrate the energy landscape; Viguera has recently described such an observation in a small β-sheet protein Bergerac-SH3 [71]. Hence, the change in the folding time after the introduction of the hydrophobic pair is dependent on both the location and the simulation temperature.

Practically the optimal choice of the folding temperature that can be used in computational study is approximately the location of the peak in the heat capacity. In our study we also face another dilemma; that is, we wish to use the same simulation temperature for all 7 sequences, in order to make a comparison. For this reason, we used $\tilde{T} = 0.12$ to perform the folding simulations for all 7 sequences. The native structure nucleation, described in the next section, was examined at other adjacent temperatures as well and we found that the nucleation probability is more greatly affected by the location of the hydrophobic pair than by $\tilde{T}$. Our selection of the simulation temperature for dynamics is similar to the approaches taken in other simulation studies [26, 28, 32, 33, 36, 72].

It is interesting to note that this selection of temperature is also consistent with that used in experimental determination of the folding dynamics. For example, to determine the folding time of the C terminal of protein GB1 (a β-hairpin) [4] and other synthesized β-hairpin systems, [17, 18] the experiments have been performed at approximately 300K. This temperature is slightly higher than the peak of the heat capacity curve measured in the cooling experiments of the C terminal of the protein GB1 by differential scanning calorimetry [73].

Average time ($10^4 MCS$)	Bonding site	$\tilde{T}$	Sequence S0	S1	S2	S3	S4	S5	S6
τ_{fold}	all	0.09	732.86	827.85	894.85	564.98	963.43	1037.01	1017.68
		0.10	179.26	213.73	183.07	201.46	215.77	277.83	302.23
		0.11	93.96	64.61	77.34	77.26	97.56	131.54	140.40
		0.12	86.38	38.60	47.65	48.41	55.71	75.41	78.77
		0.13	123.08	36.96	36.90	38.45	45.51	58.88	75.71
		0.14	245.85	49.69	49.13	42.90	47.54	75.58	98.06
		0.15	619.93	105.54	69.58	70.36	72.08	97.86	160.49
τ_{zip}	any	0.12	2.38	1.94	1.49	1.41	1.31	1.55	1.60
	$[2-\bar{2}]$		2.81	2.05	1.63	1.70	1.40	1.68	1.70
	$[3-\bar{3}]$		1.78	1.24	1.24	0.88	1.41	1.74	1.68
	$[4-\bar{4}]$		0.85	0.38	0.35	0.50	0.78	0.99	1.41
	$[5-\bar{5}]$		2.85	0.84	0.37	0.80	0.48	1.31	0.94
$\tau_{\text{first}}^{\text{HP}}$	–	0.12	–	0.39	1.85	2.96	2.54	2.29	1.08
τ_{first}	any	0.12	7.57	4.60	4.65	5.65	4.50	4.30	4.04
	$[2-\bar{2}]$		16.11	7.53	7.32	12.45	16.73	19.39	17.09
	$[3-\bar{3}]$		22.28	15.82	11.32	12.79	13.67	20.61	18.50
	$[4-\bar{4}]$		17.43	12.68	14.26	10.55	8.09	9.85	11.99
	$[5-\bar{5}]$		10.58	8.43	9.73	9.34	5.47	5.22	4.61
τ_{lifetime}	$[2-\bar{2}]$	0.12	0.311	0.765	0.314	0.255	0.111	0.188	0.167
	$[3-\bar{3}]$		0.158	0.151	0.228	0.191	0.185	0.099	0.151
	$[4-\bar{4}]$		0.101	0.087	0.093	0.123	0.126	0.153	0.104
	$[5-\bar{5}]$		0.053	0.051	0.064	0.062	0.062	0.079	0.188
$\tau_{\text{lifetime}}^{\text{Eq}}$	all	0.12	0.036	0.038	0.035	0.037	0.033	0.041	0.093
	$[2-\bar{2}]$		0.366	0.858	0.431	0.325	0.278	0.324	0.339
	$[3-\bar{3}]$		0.245	0.257	0.265	0.255	0.295	0.305	0.450
	$[4-\bar{4}]$		0.121	0.121	0.110	0.130	0.128	0.218	0.253
	$[5-\bar{5}]$		0.046	0.041	0.040	0.042	0.046	0.060	0.152
τ_{unfold}		0.12	1.54	1.95	1.38	1.44	1.68	8.35	9.19

Table 4.1: Characteristic times for β-hairpin folding in MCS measured in the computer simulation of our model. τ_{fold} is the first-passage time between the starting point of folding from a denatured state to the first approach of the native state. τ_{zip} is the time duration between a native BHB nucleation and the first approach of the native state. To analyze the system further, we have also measured $\tau_{\text{first}}^{\text{HP}}$, he time duration for the system to collapse from a denatured state to reach a compact state where the first contact of the hydrophobic pair occurs and τ_{first}, where the first contact of a native BHB occurs. To characterize the molten globular stage of the folding, we measured the average lifetime of a particular native BHB during the folding, τ_{lifetime}. which can be compared with the average lifetime of the same native BHB in an equilibrium state $\tau_{\text{lifetime}}^{\text{Eq}}$. τ_{unfold} is the average transition time the system takes to reach a denatured state starting from a native state.

Probability	Bonding site	$\tilde{T}$	Sequence S0	S1	S2	S3	S4	S5	S6
$P_{\text{nucleation}}$	$[2-\bar{2}]$	0.12	63.5	83.6	64.0	64.2	45.3	42.5	39.8
[%]	$[3-\bar{3}]$		26.5	10.9	32.0	27.8	38.7	32.3	30.3
	$[4-\bar{4}]$		7.8	3.3	3.8	7.0	13.0	20.2	22.4
	$[5-\bar{5}]$		2.2	2.2	0.3	1.0	3.0	5.0	7.5
P_{last}	$[2-\bar{2}]$	0.12	4.0	0.5	4.8	4.0	5.8	8.0	20.2
	$[3-\bar{3}]$		1.5	1.8	1.0	0.5	2.2	4.3	5.5
	$[4-\bar{4}]$		0.7	1.0	0.2	1.5	1.7	2.7	3.0
	$[5-\bar{5}]$		93.8	96.7	94.0	94.0	90.3	85.0	71.3
P_{first}	$[2-\bar{2}]$	0.12	30.5	51.7	46.0	25.8	11.3	10.5	10.8
	$[3-\bar{3}]$		11.2	5.5	20.5	17.7	5.7	3.5	5.2
	$[4-\bar{4}]$		13.3	9.3	9.7	26.7	17.5	16.0	6.2
	$[5-\bar{5}]$		45.0	33.5	23.8	29.8	65.5	70.0	77.8

Table 4.2: Characteristic probability: $P_{\text{nucleation}}$ is the probability that a certain native BHB would form as a nucleation site in the final approaching to the native state. P_{last} is the probability that a particular native BHB would form as the final step in the folding. To analyze the initial stage of folding, we have also collected P_{first}, the probability a certain native BHB forms first when the chain collapse from a denatured state.

4.3.3 Nucleation sites and the zipping time

We now examine the nucleation process that leads to the final formation of the native structure. The nucleation sites that initiate this final stage of folding are dependent on the sequence.

For the case of no hydrophobic interaction (S0), nucleation most frequently starts near the β-turn; the probability of forming the first native BHB ($[2-\bar{2}]$) dominates the rest, with a significant probability of forming the second BHB pair ($[3-\bar{3}]$) as well — see Table 4.2 for details. Consequently the bonding that first takes place near the β-turn is the most favorable pathway in S0 (the "zipping-out" scenario in Fig. 4.1).

From the same table, we see that the additional hydrophobic interaction in the hydrophobic chains, S1 to S6, is particularly effective in promoting and stabilizing the formation of a native BHB near the hydrophobic site, and consequently the nucleation probability increased near the hydrophobic pair. In S2, S3, and S4, the "middle-out" folding pathway is approximately 30%, which is significant in comparison with the relative occurrence of the "zipping-out" folding pathway — still one of the major folding pathways. In the case of S5 and S6, there is a significant increase of nucleation events near the ends of the chain, with comparable probabilities of native BHB nucleation at sites $[2-\bar{2}]$, $[3-\bar{3}]$, and $[4-\bar{4}]$. This is very different from the scenario in S0 — the "zipping-in" events are almost equally important as the "zipping-out" and "middle-out" events.

We have also observed the zipping time, starting from the formation of a nucleation site, to the formation of a native state, for different sequences. For a given sequence, the overall zipping time for nucleation starting at any sites, and the zipping time for nucleation starting at a particular site, are listed in Table 4.1 for comparison. Generally, the time scale is less than 5 percent of the folding time listed in Table 4.1, which indicates that the chain spends most of its time searching in the conformational space for the right nucleation event to happen; once nucleates are formed, the chain folds into the native state relatively quickly. This rapid growth of the interstrand native contacts in the final stage of the β-hairpin formation was also observed in the model without the Gō-approximation.[26] In the

Gō-type models, on the other hand, the final growth stage appears much earlier and the number of native contacts grows more smoothly [33].

In Table 4.1, we see that the folding time for the non-hydrophobic sequence S0 is much longer than that of a hydrophobic chain, at $\tilde{T} = 0.12$. The zipping time is also relatively longer than those with hydrophobic interactions. This suggests that the existence of a hydrophobic pair, placed at the correct position, can enhance the earlier formation of a nucleation site by bringing the corresponding monomers together and in the final zipping stage, it also helps the formation of a correct hydrogen bond.

Although the zipping times in Table 4.1 are similar, there is a subtle difference in magnitude as well. Two features can be deduced from this table. First, Consider the occurrence of a nucleation site $[2, \bar{2}]$ for all sequences, which would correspond to a zipping-out scenario. For sequences S0 and S1, the chains spend more time to zip out than, say, sequences S5 and S6. The latter sequence contains a hydrophobic pair near the ends of the chain, which encourages the zipping-out process, hence shortens the zipping time. Second, consider the zipping-in scenario that starts with a nucleation site at $[4, \bar{4}]$ near the end. Sequence S6 spends more time in the final stage in comparison with others. The earlier bonding of monomers 4 and $\bar{4}$ might have introduced improper orientations of the bonds that actually prevent the sequence from easily relaxing into a native structure. Sequences S1, S2, S3, S4, and S5 contain additional hydrophobic interactions in the middle of the chain, which enhance the interaction of those monomers. As the result, the zipping time is reduced.

To further examine the physical picture of the final folding process, we identified the last native BHB that forms during the entire folding process. The percentage of occurrence of a particular native BHB as the last site is summarized in Table 4.2 as P_{last}.

Taking sequence S0 as an example, we see that the majority of folding events are "zipping-out", that they start with the correct native BHB near the β-turn (see $P_{\text{nucleation}}$), and they end with the last native BHB at $[5 - \bar{5}]$. Now taking sequence S6 as an example, we see that there is a significant probability that the chain starts with a nucleation site near the ends (see $P_{\text{nucleation}}$ for $[4 - \bar{4}]$ and S6 a "zipping-in"

scenario). The percentage that native BHB $[2-\bar{2}]$ is the last pair to form is almost the same as the percentage that $[4-\bar{4}]$ is the nucleation site.

In all sequences, $[5-\bar{5}]$ is always more likely to be the final native BHB to form. This is largely associated with the fact that the terminal ends are more mobile (high local entropy) in the system, therefore the $[5-\bar{5}]$ native BHB can be easily detached in the middle of the folding process. Theoretical studies of thermodynamics of the β-hairpin of GB1, whose hydrophobic pair is located at the middle section of the strands, have shown that the hydrogen bonding probability of the terminal ends is less than 10%, even at low temperatures [23,26,28]. The instability of the hydrogen bonds at the terminal end of a β-hairpin was also suggested based on experimental observations [11].

4.3.4 The collapsing stage of folding

In a typical folding process, the chain starts with a denatured state and then folds into a more compact conformation; the initial collapsing is often driven by the formation of a hydrophobic core in a hydrophobic sequence and by the formation of a non-native hydrogen bond in the non-hydrophobic sequence. The chain, then, spends most of its time re-arranging the positions of the monomers, until the final, relatively short folding (zipping) stage begins. During this rearrangement period, the chain has an adequate probability of forming, detaching and reforming individual BHB without initiating the final folding process. For a given sequence, the nucleation site discussed in the previous section is not necessarily the first BHB to form because the first BHB can be detached in the molten globular stage. In this section, we examine the initial collapsing stage.

In the initial collapsing stage of hydrophobic sequences, typically a hydrophobic core forms and then the first native BHB properly forms. The time duration for the chain to make the first hydrophobic contact, $\tau_{\text{first}}^{\text{HP}}$, and the time duration for the chain to make the first native BHB contact, τ_{first}, are characteristic of the beginning stage of the process. As can be seen in Table 4.1, $\tau_{\text{first}}^{\text{HP}}$ is less than several percent and τ_{first} is approximately 10% of the entire process. This is consistent with the collapsing scenario that has been discussed previously [23–29].

The time scale of $\tau_{\text{first}}^{\text{HP}}$ depends on the proximity and the mobility of the hydrophobic residues; $\tau_{\text{first}}^{\text{HP}}$ of S1 is the shortest because the hydrophobic residues are located close to each other; the hydrophobic residues in S6 are not located close to each other, but these residues are more mobile hence have more chance to find each other. $\tau_{\text{first}}^{\text{HP}}$ of the middle section is longer because these hydrophobic residues are less mobile and other monomers prevent them from moving closer during the collapsing. In this initial collapsing stage of the folding, hydrophobic interactions introduce additional attractions between monomers; hence, τ_{first} for hydrophobic sequences (S1-S6) is generally shorter than that of S0, which is consistent with expected polymer behavior. Note that the overall τ_{first} (the first row of τ_{first} in Table 4.1) has been calculated from the average time of the shortest τ_{first} of all conditional first contact times (the other rows of τ_{first} where a certain pairing [...] was considered); hence the overall τ_{first} can be shorter than τ_{first} of a typical conditioned first contact time.

In Table 4.2 we have summarized the probability that a certain native BHB forms first during folding, P_{first}. For the non-hydrophobic sequence S0, we see that the terminal $[5-\bar{5}]$ has a relatively large probability to form; this can again be explained by the fact that the terminal ends are more mobile, so that they have a greater probability to sample a larger conformational space in a short time, in order to make the contact. The larger P_{first} for $[5-\bar{5}]$ can be contrasted with the smaller $P_{\text{nucleation}}$ for the same site; the $[5-\bar{5}]$ bond is short lived, again, because of the very reason that makes the first contact possible: larger mobility.

Hydrophobic interactions at a particular site generally enhance the probability that the neighboring monomers form the first native BHB. This can be viewed from Table 4.2 for S1-S6 where we can see a shifting pattern of enhanced first-contact probability to the terminal ends, as the location of the hydrophobic interaction moves the same way.

4.3.5 The molten globular stage of folding

The formation of a first native BHB symbolizes the finishing of the initial collapsing of the chain. In the molten globular stage, monomers move around, in a compact

conformation, to adjust the positions, by possibly dissociating a certain BHB. At this stage, the average lifetime of a native BHB, τ_{lifetime}, is characteristic of the system. As can be seen from Table 4.1, τ_{lifetime} is approximately 2 to 3 orders of magnitude smaller than the entire folding time. Though this time scale is heavily influenced by the temperature used in the simulation, it is a good indication of the number of times that a particular native BHB would form, detach and re-reform, before the final folding stage.

For a given sequence, the relative lifetime of a particular native BHB is an indication of the stability of native BHB. As an example, for sequence S0, the $[5-\bar{5}]$ bond is short lived in comparison with the $[2-\bar{2}]$ bond, consistent with the general picture that the end terminals are more mobile. As another example, for sequence S6, the $[5-\bar{5}]$ bond is highly stabilized by the hydrophobic interaction, hence has a longer lifetime. In general, there is an one-to-one correspondence between the location of the hydrophobic monomers and the stability of the native BHB near those sites. The only exception to this is the $[2-\bar{2}]$ bond; the fluctuations are reduced by the properly formed β-turn and hence τ_{lifetime} of $[2-\bar{2}]$ is significantly longer.

The statistics collected in Table 4.1 are based on folding simulations, a dynamic process. The molten globular stage, however, can be related to a thermodynamically equilibrium system. After the beginning stage, the system is no longer in an extreme non-equilibrium state. We have performed separate computer simulations for these sequences at thermal equilibrium as well, in order to measure the lifetime of a particular native BHB in an equilibrium state, $\tau^{\text{Eq}}_{\text{lifetime}}$, summarized in Table 4.1. These lifetimes can be compared with τ_{lifetime} in the molten globular stage of the folding — they have similar order of magnitude.

4.3.6 Free energy

To further explore the folding pathway of the 7 sequences studied above, we characterize the free energy landscape associated with the systems. For this purpose, we collected the histogram of visited states, specified by the structural overlap parameter x and root mean square radius of gyration R, from the computer simulations.

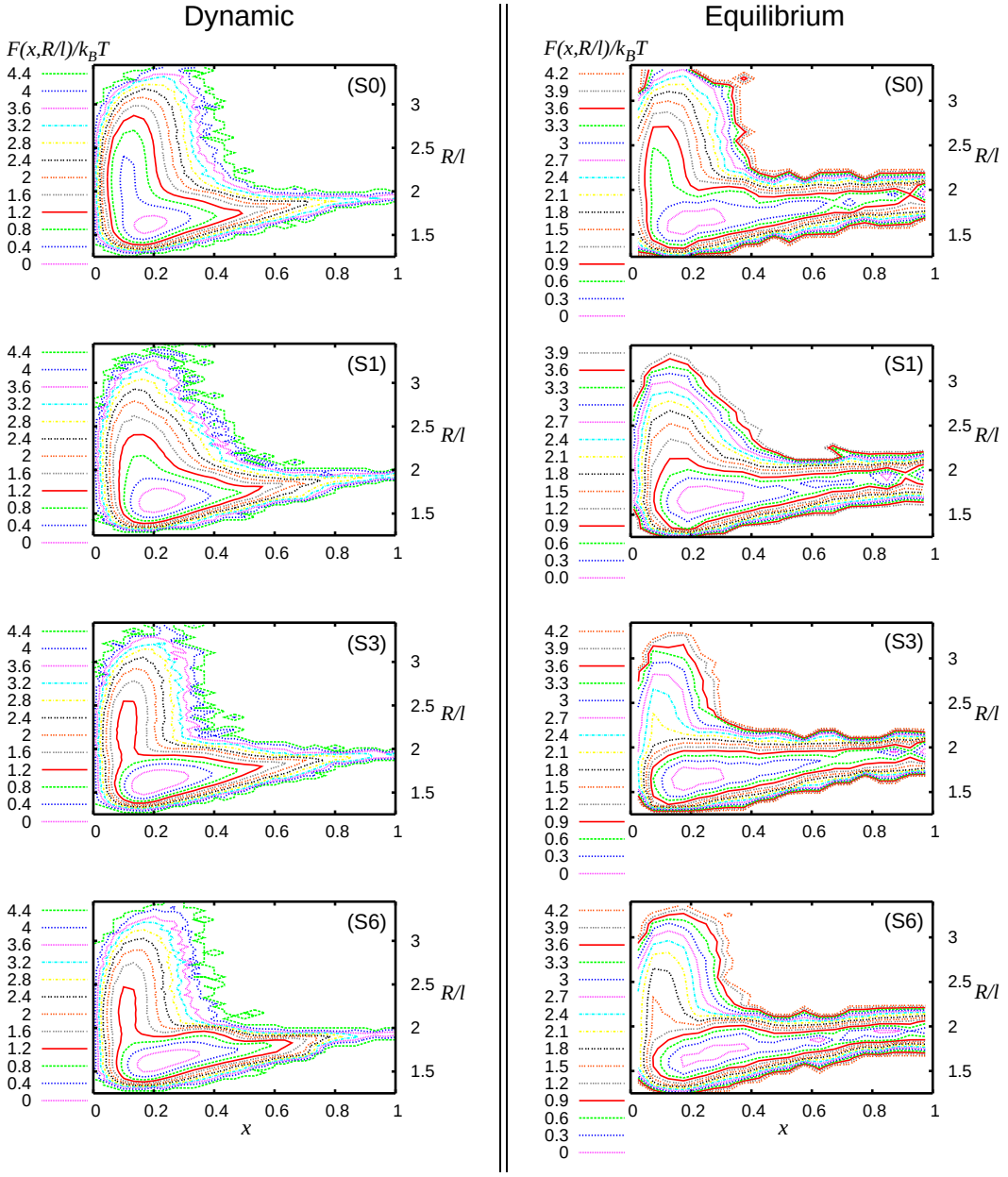

Figure 4.4: The "Dynamic" free energy as a function of the structural overlap parameter x and root mean square radius of gyration R observed in the folding simulations, for sequence S0, S1, S3, and S6 (left panel). For comparison, the free energy is also computed by a thermodynamic Monte Carlo algorithm at $\tilde{T}$, 0.108, 0.114, 0.114 and 0.115 in S0, S1, S3, and S6, respectively. The free energy is shown in contour plots where the value is specified by the side bars.

The free energy landscape can then be illustrated according to Eq. 4.6. Strictly speaking, the free energy is a concept of a thermodynamically equilibrium system and therefore the one calculated from dynamic folding differs from the free energy obtained in a fully equilibrated system. As we argued in the previous section, however, the molten globular stage of the folding, which dominates the folding process, can be regarded as a quasi-equilibrated state where the concept of the free energy still provides the insight into the properties of the system. For example, the dynamic free energy has been applied to describe the dynamic folding of a β-hairpin previously [25, 27]. Conversely, in related systems, the free energy landscape, computed thermodynamically, has been used to characterize the folding dynamics [23, 26, 28], which is a non-equilibrium phenomenon. Here, we have examined both free energy defined thermodynamically, and a dynamic free energy, defined through Eq. 4.6. The results are shown in Fig. 4.4 by using contour plots.

For sequence S0, the dynamic free energy displays a low value in areas corresponding to denatured states ($x < 1/3$), which is an indication that the system needs to explore a wide range of conformation space before it reaches the native state. The contour curves of the low free energy area do not extend to the native conformation region near $x = 1$ and $R = 1.8\ell$ and the pathways to the native state become narrower as x increases (Fig. 4.4(S0)). For comparison, we have also shown the free energy computed in a separate simulation after the system reached thermodynamic equilibrium (Fig. 4.4(S0)), which displays a similar structure. This confirms the previous analysis that the chain spends most of its time in the molten globular stage, and this stage is close to a thermodynamically equilibrium state. To study the final stage of folding, we have also plotted the dynamic free energies, Fig. 4.5(S0-A), computed from the MC data covering steps 3.0×10^3 MC to 1.5×10^3 MC steps before the completion of the folding, and Fig. 4.5(S0-B) for the last 1.5×10^3 MC steps before the completion of the folding. The low free energy region now shifts gradually towards the native state. To reach there, the chain must elongate itself, reflected in these plots by the motion of the low free energy region to a higher R/ℓ.

For sequences with hydrophobic interactions, the most prominent feature of the free energies in Fig. 4.4, is that the R has smaller fluctuations about a value

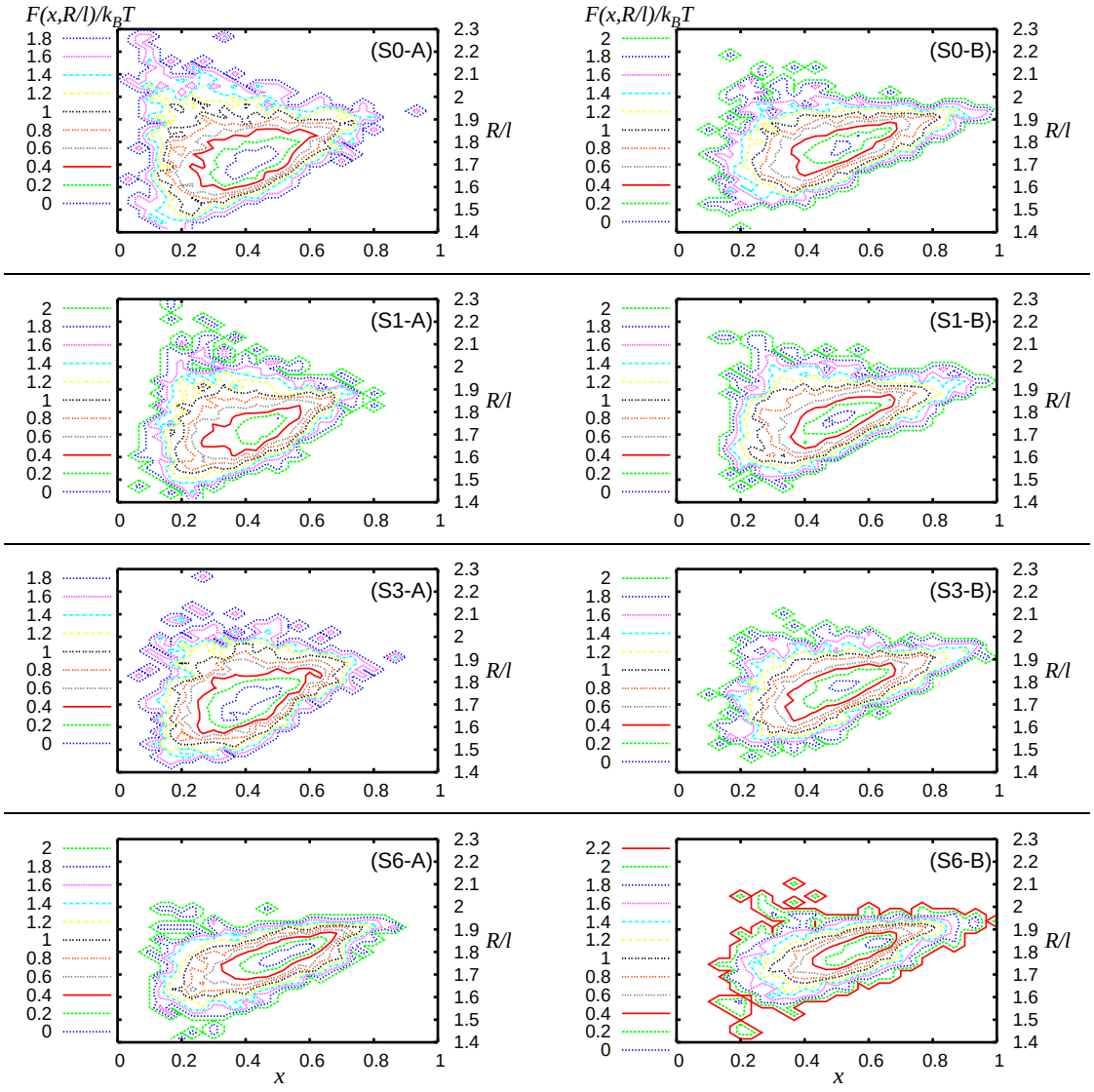

Figure 4.5: The "Dynamic" free energy as a function of the structural overlap parameter x and root mean square radius of gyration R observed near the final stage of the folding simulations, for sequence S0, S1, S3, and S6. The left plots labelled (S0-A), (S1-A), (S3-A), and (S6-A), represent Eq. 4.6 generated from the histogram covering the MC steps 3.0×10^3 to 1.5×10^3 before the completion of the folding. The right plots labelled (S0-B), (S1-B), (S3-B), and (S6-B), represent Eq. 4.6 generated from the histogram covering the last 1.5×10^3 MC steps before the completion of the folding.

corresponding to the collapsed state. Additional hydrophobic interactions drive the system into a more compact form. Within this compact form, x fluctuates in a wider region, leading to states that are closer to the native state. The chains spend most of the time visiting these states, which is a process that can be compared with that in a thermodynamically equilibrated state whose free energies are also shown on the right-hand side of Fig. 4.4. The widened fluctuation regions in x enable the systems to find the pathway to the final folding stage faster. Because this molten globular stage dominantly contributes to the folding time, $\tau_{\rm fold}$ is reduced in a chain containing a hydrophobic pair(also see Table 4.1).

In sequence S6 where a hydrophobic pair is located near the chain ends, the hydrophobic interaction encourages configurations which contain bounded terminal ends and restricts the conformational freedom of the system. This is related to the bottle neck around $x = 0.44$ and $R = 2.0\ell$ which can be seen in Fig. 4.4(S6). This bottle neck causes slowing down of the molten globular folding process.

The theory of protein folding based on statistical mechanics suggests that protein folding involves a reduction in the configurational entropy of a chain, as stabilizing interactions within the developing fold are formed. Therefore the folding should be more efficient if the conformational space available to the system during the folding is more limited [74]. The results obtained above support this general idea. The conformational space of the non-hydrophobic β-hairpin S0 is broader in R than that of a sequence containing a hydrophobic pair. On the other hand, the formation of excessively strong attraction within a polypeptide chain may produce local minima on the energy surface; these can act as kinetic traps that prevent a protein from proceeding efficiently to the native state [68, 74]. Sequence S6, for example, shows this behavior to a certain degree. The small region in R (Fig. 4.5) indicates tight binding and the loss of conformational freedom at the ends too early, before the proper formation of native BHB near the turn; as a result, it becomes a time consuming process to rearrange native BHBs in order to reach the native state. The high stability at the ends of the strands, however, prolongs the stability and lifetime of the β-hairpin once the correct conformation is achieved, as seen in an experiment [13].

L	$\bar{1}$	m_1	m_2	m_3	1	2	3	4	5
$\bar{6}$	0	2	5	9	14	20	27	35	44
$\bar{5}$	1	3	6	10	15	21	28	*36*	**45**
$\bar{4}$		4	7	11	16	22	*29*	**37**	*46*
$\bar{3}$			8	12	17	*23*	**30**	*38*	47
$\bar{2}$				13	18	**24**	*31*	39	48
$\bar{1}$					19	25	32	40	49
m_1						26	33	41	50
m_2							34	42	51
m_3								43	52
1									53

Table 4.3: The label L for possible binding pairs, native and non-native, in our model, used in Fig. 4.6. Native BHB pairs are given in bold font, and the pairings that gives rise to displaced β-hairpins are given in italic font.

4.3.7 Formation of non-native hydrogen bonds

The model used in this chapter allows for interactions between non-native residues, which is characteristically different from a Gō-type model, where non-native hydrogen bond interaction are often entirely or partially suppressed [4, 32, 33]; ignoring the non-native hydrogen bonding artificially reduces the energy frustration of the system and increases cooperativity of the folding process. The significance of non-native hydrogen bonds is explored in this section. In order to consider all interacting pairs, we have used a labelling parameter L defined in Table 4.3.

The histograms of occurrence of all possible BHBs, native and non-native, are displayed in Fig. 4.6 for all 7 sequences, observed in our folding simulations. Not surprisingly, native BHBs, shown in black, indeed occur most often. Next to the native BHBs, binding between a monomer and the nearest neighbor of the native counterpart is also very significant. A partially folded chain would bring monomers close enough to form possible native pairs; however, this also gives rise to a possibility that a hydrogen bond might form between wrong neighboring pairs. For S0,

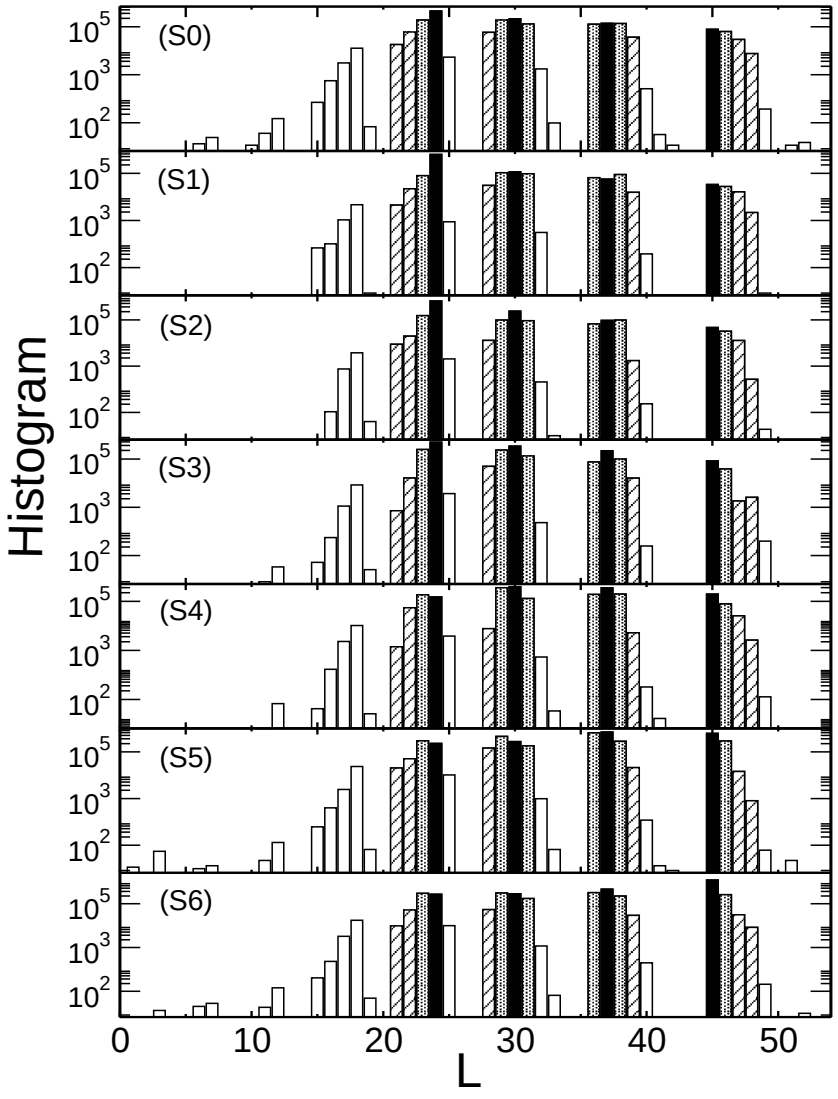

Figure 4.6: The unnormalized histogram of occurrence of all possible β-like hydrogen bonds, both native and non-native, collected from 400 folding simulations. The horizontal axis is the pairing label L defined in Table 4.3. The histogram columns in black indicate native β-like hydrogen bonds. Grey columns correspond to displaced β-hairpins whose hydrogen bonds are between a residue and the nearest neighbor of a native counterpart. The columns with tilted shades are for displaced β-hairpins whose hydrogen bonds are between a residue and the next nearest neighbor of a native counterpart.

Transition time	Sequence						
τ_{rep} $(10^4 MCS)$	S0	S1	S2	S3	S4	S5	S6
$1.5 \leq \tau_{\text{rep}} \leq 3.0$	3.8	2.7	5.0	4.7	4.2	7.7	3.5
$0 \leq \tau_{\text{rep}} \leq 1.5$	3.5	6.3	7.0	4.8	6.3	7.3	5.8

Table 4.4: Reptational transition: The percentage of reptationnal-folding events (according to a less restrictive definition — see text) observed in 400 folding simulations. For comparison, see the last raw of Table 4.1 for the unfolding time.

the $[2 - \bar{2}]$ bond, the native BHB next to the turn, is the most stable, which can be seen from the highest column in the figure. Adding hydrophobic interactions at various monomer pairs enhances the stability of native BHB between these pairs and this can also be seen from the figure; as a hydrophobic pair is placed further away from the β-turn, the location of maxima of the histograms moves to the right — towards the terminal ends.

The white histogram columns in Fig. 4.6 are indication of hydrogen bonds occurring in the neighborhood of the native monomers. These bonds indicate improperly formed structures that are fairly different from the native structure. Any improper bonding needs to be overcome in folding and can be related to the folding time itself. For a given sequence, the summed area of the white columns is representative of the time spent in resolving these wrong configurations, hence is related to the folding time of a given sequence. Sequence S0, for example, has the largest total white area in Fig. 4.6, therefore has the longest folding time; sequence S1 and S2 have the smallest total white areas, therefore have the shortest folding times. Going from S1 to S6 sequentially, we see an increase in the total white area, which can be related to the increase of the folding time displayed in Table 4.1.

4.3.8 The reptational folding pathway

The statistics in Fig. 4.6 show significant non-native neighboring bonds. There is a possibility that these non-native hydrogen bonds reflect a process in a specific folding pathway — the reptational folding pathway (Fig. 4.1D), as suggested by

Wei and coworkers [29–31]. However, the significance of the statistics of these bonds is necessary but not sufficient for the existence of the reptational folding pathway. In such a pathway, a chain would form displaced BHBs where every monomer on one strand binds the neighbor of the native counterpart as the first stage. Following that, the binding moves to a closer neighbor of the native counterpart, and so on, until finally reaches the native state.

To further clarify the possibility of this folding pathway, we concentrate on quantifying the transition from a displaced β-hairpin, whose strands are shifted by one residue in comparison with the native β-hairpin, to the native structure (the final step of a possible reptation move). A restrictive definition of reptation necessitates the disassociation of three BHBs, which are constantly formed before the transition, followed by the *immediate* formation of 4 native BHBs in the next step; such a transition has not been observed in our folding simulations.

Taking a less restrictive (and more realistic) definition, we monitor the simulated trajectories that contain a three-BHBs to four-BHBs (native) transition, where the transition period may contain other types of bindings. During the observation, if the displaced β-hairpin unfolds and refolds back to a conformation with three BHBs, we take the newly formed structure as the starting point of the observation of a reptation event.

The time taken to make this transition, however, has to be relatively short, shorter than a typical unfolding time of a near native structure — otherwise such a three-BHBs to native transition would simply correspond to a misfold-unfold-refold process, which is not characteristic of reptation. For this purpose, we have independently computed the average unfolding time, τ_{unfold}, based on 400 unfolding simulations for each sequence. The last raw of Table 4.1 displays the average unfolding time, for the 7 sequences considered in this chapter.

We have selected to observe two transition times, $1.5 \times 10^4 MCS \leq \tau_{\text{rep}} \leq 3.0 \times 10^4 MCS$, and $0 \leq \tau_{\text{rep}} \leq 1.5 \times 10^4 MCS$, and collected the statistics of folding events that contain a reptation-type transition. The percentages of folding events that correspond to this less restrictive definition of reptational folding are listed in Table 4.4. From the table, we can conclude that the reptation represents a significant possibility for β-hairpin. We also see that the reptational pathway

is more frequent in hydrophobic sequences (S1-S6) than in the non-hydrophobic sequence S0. In hydrophobic sequences, the hydrophobic attraction provides additional force along the β-hairpin strands that promotes the next stage of a reptational, anti-parallel crawl.

4.3.9 Limitation of the model

In this model, the hydrogen bonding in the middle section of the chain, where the β-hairpin makes an anti-parallel turn, is erased. In real peptide systems, these monomers (such as Pro and Aib) may have weak hydrogen-bonding propensities, which are not identically zero. In our model, because the middle monomers do not actively seek to form the β-turn by themselves, the formation of the β-hairpin depends on the interactions between long-range monomers. This type of approximations can also be found in other models for the β-sheet [33, 34, 47]. The incorporation of some, probably weaker, middle-section hydrogen-bonding propensities in the model may have a few impacts in the dynamics and even the native structure observed here. First, some unknown molten globular states could be stabilized because of the participation of hydrogen bonding associated with these monomers in these states. Second, the reptational folding pathway could be encouraged because the favorable hydrogen bonding in a shifted β-hairpin. Third, with a sizable hydrogen-bonding propensity in the middle section, the native structure can be converted to an α-helix [39, 54]. The further inclusion of an effective, middle-section hydrogen-bonding parameter in our model is beyond the scope of the current chapter but discussed in the references [39, 54].

4.4 Conclusion

The dynamic properties of simple β-hairpins, where a hydrophobic pair has been placed at various positions, have been thoroughly investigated by using a simple protein model. Extensive Monte Carlo simulations have enabled us to observe the characteristic transition times, binding probabilities and the free energy landscapes.

We have found that the folding time is shorter in the sequences where the hydrophobic pair is placed closer to the β-turn. This result agrees well with the theoretical prediction made by Muñoz et al. [4, 22] and the corresponding experimental observation reported by Dyer et al. [18]. Four types of folding pathways, zipping-out, middle-out, zipping-in, and reptation, dominate the folding scenario. For a specific sequence, the former three may present with comparable probabilities. The reptational folding pathway is less frequently observed in our model, but significant in all sequences (Fig. 4.1). We also found that non-native hydrogen bondings can not be completely ignored and contribute significantly to the molten globular stage of folding, and are the physical mechanism that yields the reptational folding pathway. The "dynamic" free energy landscape of the entire folding process is similar to that of the equilibrium state. This justifies the use of the free energy for the study of β-hairpin.

Bibliography

[1] Creighton TE. Proteins. New York: W H Freeman; 1993. p507.

[2] Munoz V, Serrano L. Helix design, prediction and stability. Curr Opin Biotechnol. 1995;6:382-386. Chakrabartty A, Baldwin RL. Stability of alpha-helices. Adv Protein Chem. 1995;46:141-176.

[3] Poland D, Scheraga HA. Theory of Helix-Coil Transitions in Biopolymers. New York: Academic Press; 1970. p797. Qian H and Schellman JA. Helix/coil theories: a comparative study for finite length polypeptides. J Phys Chem 1992;96:3987-3994. Doig AJ. Recent advances in helix-coil theory. Biophys Chem 2002;101-102:281-93.

[4] Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of β-hairpin formation. Nature 1997;390:196-199.

[5] Aurora R, Creamer TP, Srinivasan R, Rose GD. Local interactions in protein folding: Lessons from the α-helix. J Biol Chem 1997;272:1413-1416.

[6] Searle MS. Peptide models of protein β-sheets: design, folding and insights into stabilising weak interactions. J Chem Soc Perkin Trans 2001;2:1011-1020. Searle MS. Insights into stabilizing weak interactions in designed peptide β-hairpins. Biopolymer 2004;76:185-195. Searle MS, Ciani B. Design of β-sheet systems for understanding the thermodynamics and kinetics of protein folding. Curr Opin Struc Biol 2004;14:458-464.

[7] Gellman SH. Minimal model systems for β-sheet secondary structure in proteins. Curr Opin Chem Biol 1998;2:717-725.

[8] Blanco F, Ramírez-Alvarado M, Serrano L. Formation and stability of β-hairpin structures in polypeptides. Curr Opin Struc Biol 1998;8:107-111.

[9] Lacroix E, Kortemme T, de la Paz ML, Serrano L. The design of linear peptides that fold as monomeric β-sheet structures. Curr Opin Struc Biol 1999;9:487-493.

[10] Blanco FJ, Rivas G, Serrano L. A short linear peptide that folds into a native stable bold β-hairpin in aqueous solution. Nat Struct Biol 1994;1:584-590.

[11] Espinosa JF, Muñoz V, Gellman SH. Interplay between hydrophobic cluster and loop propensity in β-hairpin formation. J Mol Biol 2001;306:397-402.

[12] Santiveri CM, Rico M, Jiménez MA. Position effect of cross-strand side-chain interactions on β-hairpin formation. Protein Sci 2000;9:2151-2160.

[13] Kiehna SE, Waters ML. Sequence dependence of β-hairpin structure: Comparison of a salt bridge and an aromatic interaction. Protein Sci 2003;12:2657-2667.

[14] Griffiths-Jones SR, Maynard AJ, Searle MS. Dissecting the stability of a β-hairpin peptide that folds in water: NMR and molecular dynamics analysis of the β-turn and β-strand contributions to folding. J Mol Biol 1999;292:1051-1069.

[15] Ciani B, Jourdan M, Searle MS. Stabilization of β-hairpin peptides by salt bridges: Role of preorganization in the energetic contribution of weak interactions. J Am Chem Soc 2003;125:9038-9047.

[16] Fesinmeyer RM, Hudson FM, Andersen NH. Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin. J Am Chem Soc. 2004 Jun 16;126(23):7238-7243.

[17] Xu Y, Oyola R, Gai F. Infrared study of the stability and folding kinetics of a 15-residue β-hairpin. J Am Chem Soc 2003;125:15388-15394.

[18] Dyer RB, Maness SJ, Peterson ES, Franzen S, Fesinmeyer RM, Andersen NH. The mechanism of β-hairpin formation. Biochemstry 2004;43:11560-11566.

[19] Du DG, Zhu YJ, Huang CY, Gai F. Understanding the key factors that control the rate of beta-hairpin folding. Proc Natl Acad Sci USA 2004;1:15915-15920.

[20] Santiveri CM, Santoro J, Rico M, Jimenez MA. Thermodynamic analysis of beta-hairpin-forming peptides from the thermal dependence of (1)H NMR chemical shifts. J Am Chem Soc. 2002 Dec 18;124:14903-14909

[21] Hilario J, Kubelka J, Syud FA, Gellman SH, Keiderling TA. Spectroscopic characterization of selected β-sheet hairpin models. Biopolymers. 2002;67:233-236.

[22] Muñoz V, Henry ER, Hofrichter J, Eaton WA. A statistical mechanical model for β-hairpin kinetics. Proc Natl Acad Sci USA 1998;95:5872-5879.

[23] Dinner AR, Lazaridis T, Karplus M. Understanding β-hairpin formation. Proc Natl Acad Sci USA 1999;96:9068-9073.

[24] Pande VS, Rokhsar DS. Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. Proc Natl Acad Sci USA 1999;96:9062-9067.

[25] Zagrobic B, Sorin EJ, Pande V. β-hairpin folding simulations in atomistic detail using an implicit solvent mode. J Mol Biol 2001;313:151-169.

[26] García AE, Sanbonmatsu KY. Exploring the energy landscape of a β hairpin in explicit solvent. Proteins 2001;42:345-354.

[27] Zhou Y, Linhananta A. Role of hydrophilic and hydrophobic contacts in folding of the second β-hairpin fragment of protein G: molecular dynamics simulation studies of an all-atom model. Proteins 2002;47:154-162.

[28] Zhou R, Berne BJ, Germain R. The free energy landscape for β-hairpin folding in explicit water. Proc Natl Acad Sci USA 2001;98:14931-14936. Zhou R, Berne BJ. Can a continuum solvent model reproduce the free energy landscape of a beta -hairpin folding in water? Proc Natl Acad Sci USA 2002;99:12777-12782.

[29] Wei GH, Derreumaux P, Mousseau N. Exploring the energy landscape of proteins: A characterization of the activation-relaxation technique. J Chem Phys 2002;117:11379-11387.

[30] Wei GH, Derreumaux P, Mousseau N. Sampling the complex energy landscape of a simple β-hairpin. J Chem Phys 2003;119:6403-6406.

[31] Wei GH, Mousseau N, Derreumaux P. Complex folding pathways in a simple β-hairpin. Proteins 2004;56:464-474.

[32] Kolinski A, Ilkowski B, Skolnick J. Dynamics and thermodynamics of β-hairpin assembly: Insights from various simulation techniques. Biophys J 1999;77:2942-2952.

[33] Klimov DK, Thirumalai D. Mechanisms and kinetics of β-hairpin formation. Proc Natl Acad Sci USA. 2000;97:2544-2549.

[34] Klimov DK, Thirumalai D. Stiffness of the distal loop restricts the structural heterogeneity of the transition state ensemble in SH3 domains. J Mol Biol 2002;317:721-737.

[35] Irbäck A, Samuelsson B, Sjunnesson F, Wallin S. Thermodynamics of α- and β-structure formation in proteins. Biophys J 2003;85:1466-1473.

[36] Guo C, Levine H, Kessler DA. How does a β-hairpin fold/unfold? Competition between topology and heterogeneity in a solvable model. Proc Natl Acad Sci USA 2000;97:10775-10779.

[37] Chen JZY, Imamura H. Universal model for alpha-helix and beta-sheet structures in protein. Physica A 2003;321:181-188.

[38] Chen JZY, Lemak AS, Lepock JR, Kemp JP. Minimal model for studying prion-like folding pathways. Proteins 2003;51:283-288.

[39] Imamura H, Chen JZY. Conformational conversion of proteins due to mutation. Europhys Lett 2004;67:491-497.

[40] Head-Gordon T, Brown S. Minimalist models for protein folding and design. Curr Opin Str Bio 2003;13:160-167.

[41] Kolinski A, Skolnick J. Reduced models of proteins and their applications. Polymer 2004;45:511-524.

[42] Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and uctuations by computer simulations. Int J Peptide Protein Res 1975;7:445-459. Gō N. Theoretical studies of protein folding. Ann Rev Biophy BioEng 1983;12:183-210. Takada S. Gō-ing for the prediction of protein folding mechanisms. Proc Natl Acad Sci USA 1999;96:11698-11700.

[43] Dill KA. Folding proteins: Finding a needle in a haystack. Curr Opin Stru Bio 1993;3:99-103.

[44] Kelly JW. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Curr Opin Struct Biol 1998;8:101-106. Harrison PM, Chan HS, Prusiner SB, Cohen FE. Conformational propagation with prion-like characteristics in a simple model of protein folding Protein Sci. 2001;10:819-835. Carrell RW, Gooptu B. Conformational changes and disease - serpins, prions and Alzheimer's. Curr Opin Struct Biol 1998;8:799-809.

[45] Barkema GT, Mousseau N. Identification of relaxation and diffusion mechanisms in amorphous silicon. Phys Rev Lett 1998;81:1865-1868. Malek R, Mousseau N. Dynamics of Lennard-Jones clusters: A characterization of the activation-relaxation technique. Phys Rev E 2000;62:7723-7728.

[46] Eaton WA, Thompson PA, Chan CK, Hage SJ, Hofrichter J. Fast events in protein folding. Structure 1996;4:1133-1139.

[47] Honeycutt JD, Thirumalai D. The nature of folded states of globular proteins. Biopolymers 1992;32:695-709.

[48] Jang H, Hall CK, Zhou Y. Folding thermodynamics of model four-strand antiparallel β-sheet proteins. Biophys J 2002;82:646-659.

[49] Kemp JP, Chen ZY. Formation of helical states in wormlike polymer chains. Phys Rev Lett 1998;81:3880-3883.

[50] Favrin G, Irbäck A, Wallin S. Folding of a small helical protein using hydrogen bonds and hydrophobicity forces. Proteins 2002;47:99-105.

[51] Dill KA. Dominant forces in protein folding. Biochemistry 1990;29:7133-7155.

[52] Wood SJ, Wetzel R, Martin JD, Hurle MR. Prolines and amyloidogenicity in fragments of the Alzheimer's peptide β/A4. Biochemistry 1995;34:724-730.

[53] Awasthi SK, Shankaramma SC, Raghothama S, Balaram P. Solvent-induced beta-hairpin to helix conformational transition in a designed peptide. Biopolymers 2001;58:465-476.

[54] Imamura H, Chen JZY. Conformational conversion of proteins due to mutation and insertion of a hydrophobic pair. (To be published).

[55] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087-1092.

[56] Hansmann UHE, Okamoto Y. Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. Physica A 1994;212:415-437.

[57] Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg J. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comp Chem 1992;13:1011-1021. Lemak AS, Gunn JR. Rotamer-specific potentials of mean force for residue pair interactions. J Phys Chem B 2000;104:1097-1107.

[58] Binder K. Monte Carlo simulation in statistical physics. Berlin, New York: Springer-Verlag; 1988. p127. Newman MEJ, Barkema GT. Monte Carlo methods in statistical physics. New York: Oxford University Press; 1999. p475. Binder K. Applications of Monte Carlo methods to statistical physics Rep Prog Phys 1997;60:487-559.

[59] Kremer K and Binder K. Monte Carlo simulation of lattice models for macromolecules. Comp Phys Rep 1988;7:259-310.

[60] Kikuchi K, Yoshida M, Maekawa T, Watanabe H. Metropolis Monte Carlo method as a numerical technique to solve the Fokker-Planck equation. Chem Phys Lett;185:335-338. Chubykalo O, Nowak U, Smirnov-Rueda R, Wongsam MA, Chantrell RW, Gonzalez JM. Monte Carlo technique with a quantified time step: Application to the motion of magnetic moments. Phys Rev B 2003;67:No.64422.

[61] Rey A, Skolnick J. Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins. Chem Phys. 158:199-219, 1991.

[62] Cieplak M, Henkel M, Karbowski J, Banavar JR. Master equation approach to protein folding and kinetic traps. Phys Rev Lett 1998;80:3654-3657.

[63] Šali A, Shakhnovich E, Karplus M. Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. J Mol Biol 1994;235:1614-1636.

[64] Chan HS, Dill KA. Transition states and folding dynamics of proteins and heteropolymers J Chem Phys 1994;100:9238-9257.

[65] Chan HS, Dill KA Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics Proteins 1998;30:2-33.

[66] Karplus M, Shakhnovich E. Protein folding: Theoretical studies of thermodynamics and dynamics. In: Protein Folding. Creighton TE. (ed.). New York: W. H. Freeman 1992:127-196.

[67] Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding - A perspective from simple exact models. Protein Sci 1995;4:561-602.

[68] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167-195.

[69] Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. San Diego: Academic Press; 1996. p443. Mountain R, Thirumalai D. Quantitative measure of efficiency of Monte Carlo simulations. Physica A 1994;210:453-460.

[70] Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a de Novo designed four-helix bundle protein. J Mol Bio 1996;263:323-343.

[71] Viguera AR, Serrano L. Bergerac-SH3: "frustation" induced by stabilizing the folding nucleus. J Mol Biol 2001;311:357-371.

[72] Hoang TX, Cieplak M. Molecular dynamics of folding of secondary structures in Go-type models of proteins. J Chem Phys 2000;112:6851-6862.

[73] Honda S, Kobayashi N, Munekata E. Thermodynamics of a β-hairpin structure: Evidence for cooperative formation of folding nucleus. J Mol Biol 2000;295:269-278.

[74] Dobson CM, Săli A, Karplus M. Protein folding: A perspective from theory and experiment. Angew Chem Int Ed 1998;37:868-893.

Chapter 5

Conformational conversion of proteins due to mutation: hydrogen bonding effect

We investigate the mechanism of conformational conversion, specifically between an α-helix and a β-hairpin, due to sequence perturbations mainly in non-hydrophobic sequence. The dynamics and thermodynamics of the folding behavior were analyzed by the canonical and multi-canonical type Monte Carlo simulation techniques, respectively.

5.1 Introduction

Perturbations such as mutation or environmental change can be tolerated by some proteins and this is the key to the survival of biological stability from the evolutionary perspective [1]. In some cases, however, such perturbations to protein sequences can destabilize the protein three-dimensional conformation and this mystifies the researchers in the field. Cordes et al. [2] found that a two-residue mutation can drastically transform the local secondary structures in peptides. Other examples of conformational change include chaperone proteins that adopt different conformations related to interaction with their environment [3] and muscle acylphosphatase

that converts to a different conformation in amyloid aggregation [4]. Grishin and Kinch have systematically described various recurrently observed mechanisms of conformational conversion in proteins [5]. Conformational misfolding caused by minor or unimportant alteration in proteins or the environment is suspected to lead to the formation of beta amyloids associated with neurodegenerative diseases, including fatal prion diseases and Alzheimer disease [6, 7].

The physical mechanism of conformational conversion in proteins has not been well established. All-atom computer simulations, for example, are currently too computationally expensive to address this issue as a whole even though these models have become increasingly powerful in addressing the detail description of the equilibrium and short time scale properties of protein folding [8]. Most minimal models have been designed to accommodate a pre-assumed target structure, hence cannot be used to simultaneously model both α-helix and β-hairpin structures (see structural sketch in Fig. 5.1), the commonly seen secondary structures affected in conformational conversion. However, some simple polymer models have recently been developed to handle the α-helix and β-hairpin structural duality [9,10]. For a concise informative review of the recent development of simple models, see Kolinski and Skolnick [11].

In this chapter, we present a full characterization of the conformational stability of a simple protein model, based on our earlier unified potential model [10], to study the α-helix and β-hairpin conformational dependence on protein sequences. We show that the competition between the local hydrogen bonding (that prefers α-helix) and non-local hydrogen bonding (that prefers β-hairpin) gives rise to a conformational transition when the residues in the middle are affected by hypothetical mutations. We also show that the efficiency of folding critically relies on key residues along the sequence.

5.2 The Model and method

The polymer chain model consists of 13 monomers connected linearly by bonds of fixed length l. The bond angle is fixed at 105° by utilizing a stiff harmonic

potential [12]. No further torsional constraint involving three adjacent bonds is imposed in this model which differs from others [9]. Non-neighboring monomers along the backbone interact with each other in the form of an excluded volume of diameter $1.2l$ and an effective potential representing hydrogen bonding [13]. The effective diameter of the residue is slightly larger than the bond length and the backbone residues lightly overlap each other. This excluded volume incorporates the existence of a virtual oxygen O and a virtual hydrogen H which by themselves do not have any excluded volume. The hydrogen bonding is defined between these virtual residues, in particular for the ith monomer whose position is described by $\vec{r}_i$, an O_i and H_i were placed at the tips of the vectors $\vec{r}_i^{(O)}$ and $\vec{r}_i^{(H)}$, directed away from $\vec{r}_i$:

$$\vec{r}_i^{(O)} = \frac{1}{3}(\vec{r}_{i+1} - \vec{r}_i) + 0.6l\vec{n}_i \tag{5.1}$$

$$\vec{r}_i^{(H)} = \frac{2}{3}(\vec{r}_{i+1} - \vec{r}_i) - 0.6l\vec{n}_i \tag{5.2}$$

where $\vec{n}_i$ is the normal direction of the plane defined by the $(i-1)$th, ith, and $(i+1)$th atoms. Except for the first and second nearest neighbors, every pair of O and H interact with each other through a shifted Lennard-Jones potential,

$$V_{OH} = 4\epsilon(l^{12}/(r+r_0)^{12} - l^6/(r+r_0)^6) \tag{5.3}$$

where $r_0 = 2^{1/6} \times 1.2l$ is the optimal bonding distance, and r is the OH distance. The effective hydrogen bonding reaches its optimal energy $-\epsilon$ as $r = 0$. This hydrogen bonding depends significantly on the orientation of adjacent residues and is the determinant factor for the stabilization of the secondary structures; other minimal models have also demonstrated the importance of hydrogen bonding in a similar fashion [14]. Some minimal models showed that secondary structures can be attained without the explicit description of hydrogen bonding; rather, a torsional angle potential was required for each type of secondary structure [15].

We apply perturbations on the strength of the hydrogen bonding interaction of the 7th and 8th monomers. The strength of the hydrogen bonding interaction of virtual atoms in the two nearest neighboring bonds of these monomers, with any other virtual atoms along the chain, is set to ηV_{OH}. That is Equation 5.3 now takes the form

$$V_{OH}^{\eta} = 4\eta\epsilon(l^{12}/(r+r_0)^{12} - l^6/(r+r_0)^6). \tag{5.4}$$

The key parameter η determines the effective strength of hydrogen bonding of these particularly specified monomers. By varying the parameter η, we can study the stability of the three-dimensional native structures of the modelled chain. The upper right corner of Fig. 5.1 shows the native α-helix structure, obtained by an energy minimization procedure for $\eta = 1$; the lower left corner shows the native β-hairpin structure, for $\eta = 0$. The native state changes from a β-hairpin to an α-helix at the conversion point $\eta_c = 0.627$ as η increases from 0. The native state energy of a β-hairpin depends on η weakly as shown by the solid line in Fig. 5.1, because the 7th and 8th monomers are not extensively involved in forming hydrogen bonds in a β-hairpin conformation. On the other hand, the native energy of the α-helix depends on the magnitude of η more significantly, as shown by the long dashed line.

Physically, the optimal strength of a backbone hydrogen bond in different amino acids does not vary widely [16]. Some amino acids, however, has significantly different conformation than others and therefore they can promote or disrupt the formation of hydrogen bonding in the secondary structures; for example, glycine is very flexible while prolin is twisted and these residues often act as a breaker of hydrogen bond network of an α-helix and β-strands. These residues often make up a reverse turn of β-sheets where backbone hydrogen bonds are missing [17]. To incorporate these characteristics of amino acids into our model, we introduced η to control the effective strength of the hydrogen bonds of key mutational residues. The variation of η from 0 to a higher value can be regarded as coarse grained mutation to the middle residues in a β-hairpin by different amino acids that has higher tendency to form a hydrogen bond. As a result of the mutation, a native α-helix conformation is obtained in our model. Therefore, this model provides an opportunity for investigation of conformational instability due to a hypothetical idealized mutation in a simple framework.

Next we analyze the foldability and its relationship with the mean folding time for various η. Klimov and Thirumalai argued that folding rates are related to the underlying thermodynamic properties and they described several methods that can be used to relate these kinetic and thermodynamic characteristics of a system [18].

The foldability parameter measures the closeness of the collapsing and folding

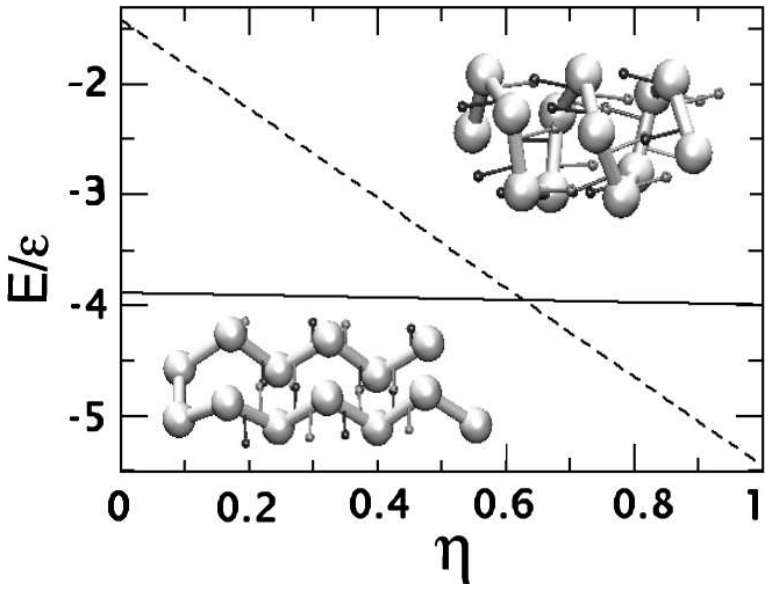

Figure 5.1: The energies of the native β-hairpin at lower left corner and α-helix at upper right corner conformations as a function of η. Large and small spheres represent the backbone monomers and virtual atoms, respectively. The size of backbone monomers is reduced for visual aid. Dotted line and solid line represent the energies of the α-helix and β-hairpin conformations. Below $\eta_c = 0.627$, the β-hairpin is stable and above η_c, the α-helix is stable.

temperatures and is defined by

$$\sigma = |T_\theta - T_f|/T_\theta \tag{5.5}$$

where T_θ is the collapsing temperature, above which coil configurations dominate the polymer conformation, and T_f is the folding transition temperature, below which the polymer reaches the native conformation. In the following we use a reduced temperature $\tilde{T} = k_B T/\epsilon$. Computationally, $\tilde{T}_\theta$ can be determined from the maximal peak of the reduced heat capacity $C_V/k_B = (\langle (E/\epsilon)^2 \rangle - \langle E/\epsilon \rangle^2)/\tilde{T}^2$; $\tilde{T}_f$ can be estimated from the temperature at which the variance of the structural overlap function $\delta\chi = \langle \chi^2(\tilde{T}) \rangle - \langle \chi(\tilde{T}) \rangle^2$ is maximum, where χ measures the fraction of the

number of the residue pairs that are closer than a conditional distance. In general, a small σ implies that a sequence folds into a native state as it collapses from a denatured state, without stable intermediate states that prevent the formation of the native bonds. A large σ indicates the presence of stable intermediate states or another competing native state.

This measurement is useful for comparing the systems whose native conformations are nearly identical but some of their residue types are different. These systems have nearly the same contact order [19], which measures the locality and non-locality of the native bonding and is consider to be a reasonable characteristic measurement of folding properties. When the systems are perturbed by the mutation and σ in these systems changes, the change is closely related to the alteration of the folding mechanism and σ is considered to be a good measurement connecting kinetic and thermodynamics of the system examined here. One must be cautious to apply this measure on more complicated systems. For example, a protein of completely different conformation composed of different secondary structures has a different folding mechanism that can not be sufficiently characterized by a simple foldability index. If the folding mechanism is very different, σ can not reflect the complexities of the folding mechanisms of proteins, whose conformations and chemical compositions are vastly different. For example Millet et al. [20] argued that a variant of the foldability index that is defined as a function of the concentrations of denaturant instead of a function of temperatures does not correlate with the folding timescale of five small proteins whose native conformations and size are significantly different.

Computational results for the thermodynamics were obtained from an improved multi-canonical algorithm, in which the weights determined in the weight-determination runs [21] were directly used for thermodynamic-property computation according to the weighted histogram analysis [22].

5.3 Results and discussion

As shown in Fig. 5.2a, the β-hairpin conformation is relatively robust with respect to sequence variations within a certain small range of σ. σ increases as η is increased

from 0.0 to η_c, indicating less stability as η approaches the conversion value η_c. Local interactions that prefer to settle into structural domains typical to α-helix start to cause frustrations for the non-local bonding in β hairpin. Near $\eta_c = 0.627$ where a β-hairpin and α-helix yield approximately the same energy -3.95ϵ, two competing iso-energy native states complicate the folding pathways. (See Fig. 5.2A). Going beyond η_c, σ decreases rapidly as the tendency to form local hydrogen bonds increases.

In order to confirm this physical picture, we have directly conducted dynamic folding MC simulations, performed at a relatively low $\tilde{T}$ where the configurations are stable in comparison with the native structures. We measured the first-passage time $t_{\rm first}$ at a temperature slightly above $\tilde{T}_f$, below which the simulations tend to be trapped in local minima. A structure was considered to reach the native state when all native OH pairs have formed. The mean folding time weakly depends on the selected temperature used in the folding simulations [23] and a different choice of $\tilde{T}$ in the range $[\tilde{T}_f, \tilde{T}_\theta]$ does not significantly affect the order of magnitude estimate of $\langle t_{\rm first} \rangle$ presented in Fig. 5.2B. For each data point shown in Fig. 5.2B, 80 folding simulations were performed separately from different denatured initial conditions.

The overall qualitative shapes of the two plots in Fig. 5.2 are similar and σ indeed is a useful indicator for the folding timescale; a smaller σ implies a faster folding and vice versa. A notable feature of Fig. 5.2B is that at low η, a β-hairpin folds approximately 15 times more slowly than an α-helix of the same polymer length. This finding agrees with previous experimental and theoretical conclusions that the folding of a β-hairpin is considerably slower than that of an α-helix of similar peptide length [24]. It is known that the helical formation is driven mainly by local contacts and opposed by conformational entropy that arises from the formation of correct consecutive backbone structure. On the contrary the formation of a β-hairpin requires multiple local and nonlocal hydrogen bonding that includes the formation of the turn and the bonding across the two β stands. As will be demonstrated elsewhere, we found that the folding exponent λ in $\langle t_{\rm first} \rangle \propto N^\lambda$, where N is the number of residues, is much greater in the β-hairpin formation in comparison with that in the α-helix formation [26]. This finding qualitatively

agrees with the result obtained by Cieplak and coworkers who have been extensively investigating scaling in realistic conformations in simple models [27].

The folding properties and landscape can be further examined by thermodynamic quantities such as the heat capacity, radius of gyration, bond formation rate as well as the order-parameter-dependent free energy F. The heat capacity C_v/k_B as a function of $\tilde{T}$ for three η values shows that the characteristic transitions arise differently in Fig. 5.3. At $\eta = 0$, C_v/k_B has a large sharp peak around $\tilde{T} = 0.1$. The shape of the C_v/k_B line indicates that the collapsing and folding arise almost simultaneously. For $\eta = \eta_c$, C_v/k_B has two moderately elevated peaks with a broad skirt spanning wider $\tilde{T}$ due to a complex folding mechanism arising at different $\tilde{T}$s. For $\eta = 1$, the overall shape of the C_v/k_B is similar to that of $\eta = 0$ but is shifted to higher $\tilde{T}$ and its skirt becomes wider. The plot of the radius of gyration of the systems (Fig. 5.4) shows the characteristic volume change in the chain is also different in these systems. For the system with $\eta = 0$, the chain collapses into a compact conformation with some β-hairpin contents around $\tilde{T} = 0.1$ and then expanded its strands to reach the native β-hairpin. For the system with $\eta = \eta_c$, the radius of gyration settle around $1.45l$, but the radius gyration, in this case fails to reveal accurate conformational compositions at lower temperature as we will see more clear pictures in the free energy plots of the same system. For the system with $\eta = 1.0$, the chain gradually reaches to α-helix conformation as its volume continuously decreases. In Fig. 5.5 the bond formation rates of right handed helical hydrogen bonds and β-hairpin hydrogen bonding describe similar physical pictures described above for each $\eta = 0$ and 1. For $\eta = \eta_c$ the bond formation rates imply there exists a complicated mix of β-hairpin and α-helical conformations at lower $\tilde{T}$.

For the free energy, we define an order parameter x that characterizes the conformational similarity to two reference states, and it is defined as

$$x = (D^\beta - D^\alpha)/D_0$$

where D_0 is a normalization factor and D^β and D^α are the root-mean-square-deviation distances between a conformation of interest and the ideal β-hairpin and α-helix, respectively. The root-mean-square-deviation distance is the measure of the conformational deviation between two conformations that are optimally aligned

with each other. This order parameter x is normalized and yields approximately 1 and -1 for an ideal α-helix and β-hairpin respectively, and it is close to 0 when a conformation substantially differs from both an α-helix or a β-hairpin. Using this order parameter we can calculate the reduced free energy

$$F(x, \tilde{T})/k_b T = A - k_{\mathrm{B}} T \ln \Omega(x, \tilde{T})$$

where $\Omega(x, \tilde{T})$ is the the histogram distribution at $x, \tilde{T}$. A is a constant for the base line of the reduced free energy to become zero. Figure 5.6 shows the reduce free energy $F(x, \tilde{T})/k_B T$ as a function of x for various $\tilde{T}$. For $\eta = 0$ (Fig. 5.6A) and $\eta = 1$ (Fig. 5.6C), as $\tilde{T}$ is lowered, $F(x, \tilde{T})$ shows the development of a minimum around $x = -1$ and $x = 1$, respectively. The descend to the minimum experiences small rugged areas, but on average the slope is directing the system to the native states.

The short folding time and low σ in the low η (0.0 to 0.3) and high η (0.8-1.0) regimes reflect the fact that both a β-hairpin and an α-helix are frequently appearing, relatively stable conformations of protein with respect to sequence variations. These conformations are highly designable by nature, and survived the evolution process [28]. The initial rapid raising of σ in η (form 0 to 0.2) also indicates that a β-hairpin turn is somewhat sensitive to sequence variations; only limited numbers of types of amino acids (for example, glycins and prolines) can form a β-turn easily. In contrast, the minuscule change of σ and its persistently small value in the regime from 0.8 to 1 suggest that an α-helix is generally more stable and less sensitive to sequence variations, which is consistent with experimental observations by Balaram's group [29].

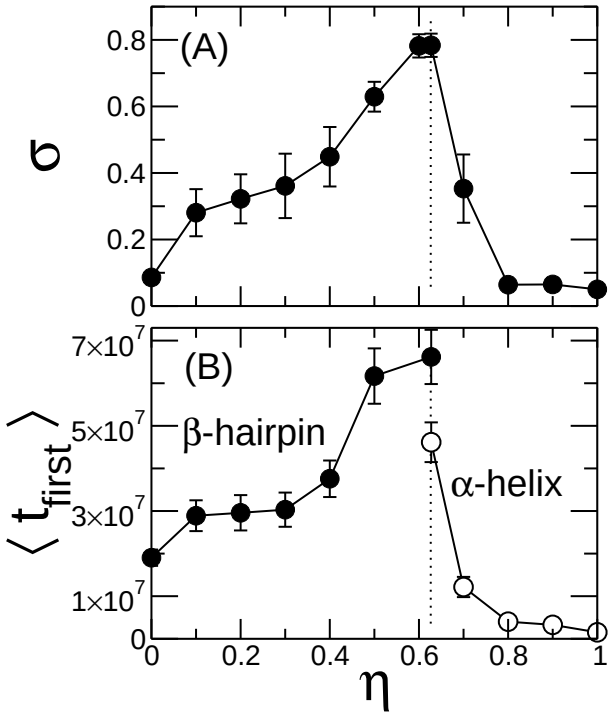

Figure 5.2: (A) The foldability parameter σ and (B) the mean first passage folding time $\langle t_{\rm first} \rangle$ versus η. σ was calculated by using an analysis of the thermodynamic properties and $\langle t_{\rm first} \rangle$ was calculated by direct folding simulations. Eighty folding simulations were performed at a selected temperature for each data point. These temperatures were: $\tilde{T} = 0.095$ for $\eta = 0, 0.1, 0.2, 0.4$ and 0.5, $\tilde{T} = 0.105$ for $\eta = 0.627$, $\tilde{T} = 0.115$ for $\eta = 0.7, 0.8$ and 0.9, and $\tilde{T} = 0.135$ for $\eta = 1.0$.

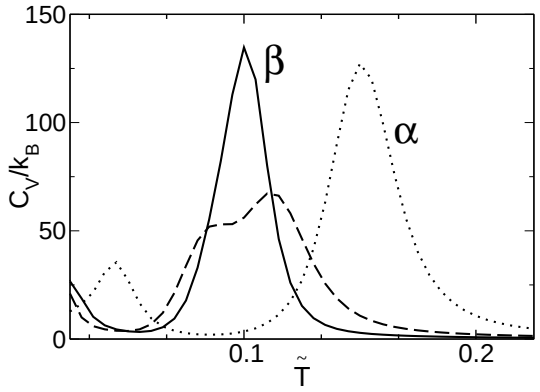

Figure 5.3: The heat capacity. Solid, dash, and dotted lines correspond to the system with $\eta = 0$, $\eta = \eta_c$, and $\eta = 1$. Sequences with η close to $\eta = 1$ or $\eta = 0$ have relatively sharp peak heat capacity peak, and sequences with intermediate value of η have more complex folding pathways that leads to multiple peaks in heat capacity.

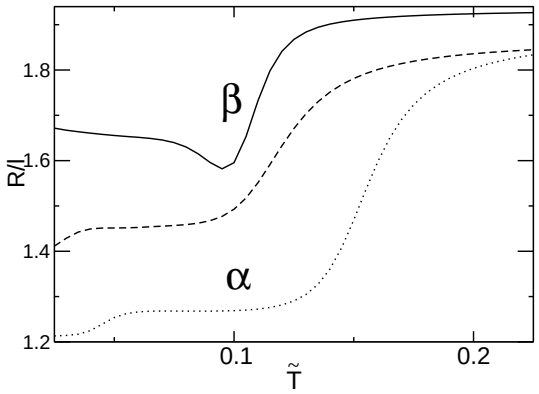

Figure 5.4: The radius of gyration: Solid, dash and dot lines denote the systems at $\eta = 0$, $\eta = \eta_c$ and $\eta = 1$.

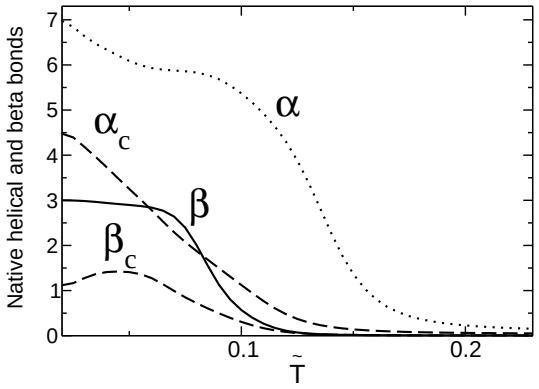

Figure 5.5: The bond formation rates: Solid and dot lines denote the formation rate of the native β-hairpin hydrogen bonds at $\eta = 0$ and the formation rate of α-helical hydrogen bonds at $\eta = 1$. For the system with $\eta = \eta_c$, the formation rate of both α-helical and native β-hairpin hydrogen bonds are shown as marked with the letter α_c and β_c, respectively.

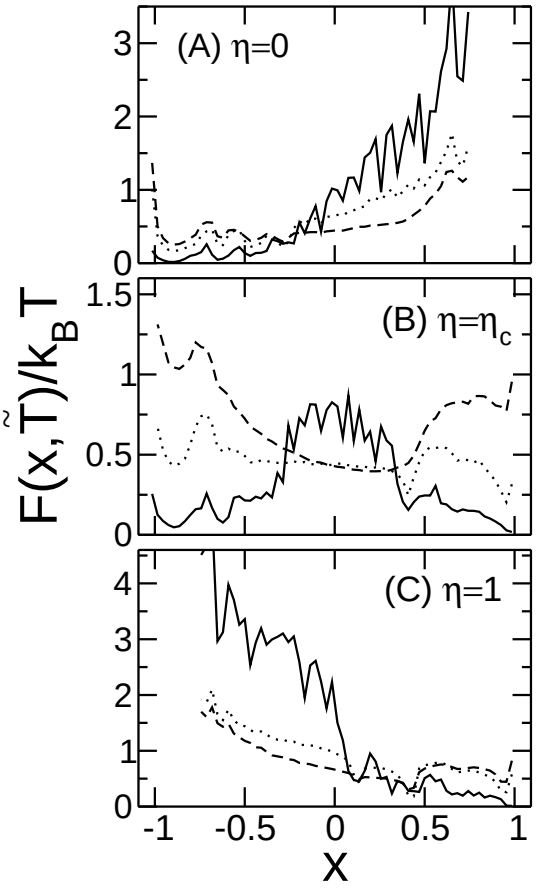

Figure 5.6: Reduced free energy $F(x, \tilde{T})$ as a function of the order parameter x and the reduced temperature $\tilde{T}$ for (A) $\eta = 0$, (B) $\eta = \eta_c = 0.627$ and (C) $\eta = 1$. In plot (A), dashed, dotted and solid lines correspond to $\tilde{T} = 0.10$, 0.08, and 0.015. In plot (B) dashed, dotted, and solid lines correspond to $\tilde{T} = 0.13$, 0.10 and 0.02. In plot (C) dashed, dotted, and solid lines correspond to $\tilde{T} = 0.13$, 0.10 and 0.02.

The folding of the β-hairpin is critically dependent on the formation of the hairpin turn near the middle of the polymer sequence, which can be identified as a hot nucleation site in folding. For low η, such a hot site exits, but for $\eta < \eta_c$ near η_c this hot site is no longer dominating, thus prompting the long folding time in this regime. At $\eta = \eta_c$, both structures are relatively stable. $F(x, \tilde{T})$ develops two minima at $x = \pm 1$ as the temperature is lowered. These two states are separated by a partition barrier, which clearly identifies the two separate folding funnels. In the vicinity of $\eta = \eta_c$, the free energy landscape is similar to that in Fig. 5.6b; the only difference is that one of the minima is lower. As the protein folds, it may take the funnel to the metastable state. The proximity of η_c, thus, corresponds to much longer folding times. The sequences are highly frustrated due to the existence of the uncertainty to fold into an α-helix or a β-hairpin. The low foldability of these sequences implies that the sequences are more likely to be eliminated by evolution — we observe proteins with the ability to make α-helix and β-sheet transition only in rare systems.

So far our model presented in this chapter has been controlled by one parameter only: η. By introducing the hydrophobic attraction into the system and shifting the energy balance in a conformation, we also observe similar conformational conversion discussed above. In a sequence with $\eta = 0.7$, for instance, the native conformation is transformed into an α-helix from a β-hairpin which has a hydrophobic pair near the turn when the magnitude of hydrophobic force is increased and the energy balance is shifted to favor the α-helix. It is also possible to transform the ground state into a β-hairpin from an α-helix, which has a hydrophobic pair near the ends by increasing the magnitude of the hydrophobic attraction. The β-hairpin in this case becomes a native conformation. It is also observed that the native conformation shifts from a β-hairpin to an α-helix by increasing the strength of the hydrophobic attraction near the turn.

Similar conversions have been seen in experimental observation by Takahashi, Mihara and coworkers [30]. They have found that hydrophobic defects would cause the formation of helical aggregates and subsequently led to the formation of β-sheet structures, and they have suggested that the key mechanism of the α-β transition is the conversion from local-monomer attractions that stabilize an α-helix to the

non-local interactions that stabilize the β-sheet conformation.

5.4 Conclusion

In summary, we have demonstrated a coarse-grained minimal model that is suitable for studying the structural stability of the secondary structures in proteins, α-helices and β-hairpins, and structural transformation between these two as the effective strength of mutation in the sequence varies. The perturbative change onto the hot site (the middle monomers) in the sequence drastically converts the entire structure of the modelled protein. Our model offers a first step towards the understanding of structural transformation due to mutation recently observed experimentally [2, 30] within the framework of the simple model. In the next chapter we will illustrate conformational conversions in the presence of the hydrophobic interaction.

Bibliography

[1] Nelson ED, Onuchic JN. Proposed mechanism for stability of proteins to evolutionary mutations. Proc Natl Acad Sci USA 1998;95:10682-10686.

[2] Cordes MHJ, Burton RE, Walsh NP, McKnight CJ, Sauer RT. An evolutionary bridge to a new protein fold Nat Struct Bio 2000;7:1129-1132. Cordes MHJ, Walsh NP, McKnight CJ, Sauer RT. Evolution of a protein fold in vitro. Science 1999;284:325-327.

[3] Minor DL, Kim PS. Context-dependent secondary structure formation of a designed protein sequence. Nature 1994;367:660-663.

[4] Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, Dobson CM. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc Natl Acad Sci USA 1999;96:3590-3594.

[5] Grishin NV. Fold change in evolution of protein structures. J Struct Biol 2001;134:167-185. Kinch LN, Grishin NV. Evolution of protein structures and functions. Curr Opin Struct Biol 2002;12:400-408.

[6] Vendruscolo M, Zurdo J, MacPhee C, Dobson CM. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. Philos Trans R Soc Lond Ser A 2003;365:1205-1222.

[7] Kelly JW. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Curr Opin Struct Biol 1998;8:101-106. Harrison PM, Chan HS, Prusiner SB, Cohen FE. Conformational propagation with prion-like characteristics in a simple model of protein folding. Protein Sci

2001;10:819-835. Carrell RW, Gooptu B. Conformational changes and disease - serpins. prions and Alzheimer's. Curr Opin Struct Biol 1998;8:799-809.

[8] Day R, Daggett V. All-atom simulations of protein folding and unfolding. Adv Prot Chem 2003;66:373-403. Duan Y, Kollman PA. Computational protein folding: From lattice to all-atom. IBM Syst J 2001;40:297-309.

[9] Irback A. A minimalistic all-atom approach to protein folding. J Phys Cond Matt 2003;15:S1797-S1807. Banavar JR, Maritan A. Geometrical approach to protein folding: a tube picture. Rev Mod Phys 2003;75:23-34.

[10] Chen JZY, Imamura H. Universal model for α-helix and β-sheet structures in protein. Physica A 2003:321:181-188.

[11] Kolinski A, Skolnick J. Reduced models of proteins and their applications. Polymer 2004;45:511-524.

[12] Honeycutt JD, Thirumalai D. The nature of folded states of globular proteins. Biopolymer 1992;32:695-709.

[13] The importance of a directionally biased potential for helix formation and the main effect of hydrogen bonding was stressed in Kemp JP, Chen ZY. Phys Rev Lett 1998;81:3880.

[14] Klimov DK, Betancourt MR, Thirumalai D. Virtual atom representation of hydrogen bonds in minimal off-lattice models of α helices: effect on stability, cooperativity and kinetics. Fold Des 1998;3:481-496. Veitshans T, Klimov D, Thirumalai D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. Fold Des 1997;2:1-22. Borg J, Jensen MH, Sneppen K, Tiana G. Hydrogen bonds in polymer folding Phys. Rev. Lett. 2001;86:1031-1033. Irback A, Sjunnesson F, Wallin S. Hydrogen bonds, hydrophobicity forces and the character of the collapse transition. 2001;27:169-179.

[15] Guo Z, Thirumalai D. kinetics and thermodynamics of folding of a de novo designed four-helix bundle protein. J Mol Biol 1996;263:323-343.

[16] Dill KA. Dominant forces in protein folding. Biochemistry 1990;29:7133-7155.

[17] Creighton TE. Proteins: Structures and Molecular Properties. New York : WH Freeman, 1993.

[18] Klimov DK, Thirumalai D. Criterion that determines the foldability of proteins. Phys Rev Lett 1996;76;4070-4073. Klimov DK, Thirumalai D. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. J Chem Phys 1998;109;4119-4125.

[19] Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985-994.

[20] Millet IS, Townsley LE, Chiti F, Doniach S, Plaxco KW, Equilibrium collapse and the kinetic 'foldability' of proteins. Biochemistry 2002;41:321-325.

[21] Hansmann UHE, Okamoto Y. Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. Physica A 1994;212:415-437.

[22] Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg J.The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comp Chem 1992;13:1011-1021.

[23] Socci ND, Onuchic JN. Folding kinetics of proteinlike heteropolymers. J Chem Phys 1994;101:1519-1528.

[24] Bieri O, Kiefhaber T. Elementary steps in protein folding. Biol Chem 1999;380:923-929.

[25] Chen JZY, Lemak AS, Lepock JR, Kemp JP. Minimal model for studying prion-like folding pathways. Proteins 2003;51:283-288.

[26] Kemp JP, Chen JZY. Folding dynamics of the helical structure observed in a minimal model. Europhys Lett 2002;59:721-727.

[27] Cieplak M, Hoang TX. Universality classes in folding times of proteins. Biophys J 2003;84:475-488, and references therein.

[28] Li H, Helling R, Tang C, Wingreen N. Emergence of preferred structures in a simple model of protein folding. Science 1996;273:666-669.

[29] Awasthi SK, Shankaramma SC, Raghothama S, Balaram P. Solvent-induced beta-hairpin to helix conformational transition in a designed peptide. Biopolymers 2001;58:465-476.

[30] Takahashi Y, Yamashita T, Ueno A, Mihara H. Construction of peptides that undergo structural transition from α-helix to β-sheet and amyloid fibril formation by the introduction of N-terminal hydrophobic amino acids. Tetrahedron 2000;56:7011-7018. Mihara H, Takahashi Y. Engineering peptides and proteins that undergo α-to-β transitions. Curr Opin Struct Biol 1997;7:501-508.

Chapter 6

Conformational conversion between the α-helix and β-hairpin in proteins: hydrophobic effect

We investigate the conformational conversion between a β hairpin and an α-helix conformation induced by the perturbation on hydrogen bonds at the middle of the chains and by the insertion of hydrophobic pair at various position using a reduced model. The conformational stability is examined by the phase diagram maps and the transition mechanisms are also studied by the heat capacity. These thermodynamic properties were computed by Multicanonical Monte Carlo and wighted histogram analysis method algorithms. The conformational phases are found to depend on the strength of perturbation on the hydrogen bonds in the middle and on the position of inserted hydrophobic pair. Kinetic accessibility of a β hairpin and an α-helix conformation are also examined in the systems where these two conformations have same the energies. The β-hairpin is found to be more accessible in most of the system. The detail conformational thermodynamic and kinetic analyses obtained by the reduced model provide insights to the mechanism of the conformational interconversion between a β-hairpin and an α-helix.

6.1 Introduction

Structural transitions in proteins that involves major conformational change can be induced by varying the chemical or physical conditions in the system. As have been observed experimentally, sometimes the transition can be caused by perturbative variations in the chemical or physical conditions: the mutational replacement of a few monomers or small changes in the environment surrounding the molecules [1–4]. Such conformational transitions in proteins are relatively rare but can be crucial to their normal functions if they arise, and can be contrasted with the fact that most proteins have relatively stable conformations. In general, the three dimensional structural information of the protein is encoded in the primary sequence [5, 6], and the common belief is that various perturbations in proteins can be tolerated by proteins and this tolerance is considered to be a key to the survival for the biological stability from the evolutionary perspective; the native conformation must have evolved and adjusted to overcome or accommodate various perturbations in chemical or physical conditions with only minor local conformational changes if any [7, 8]. In other cases, however, such perturbations can reshape a global three-dimensional conformation. In experiments, various types of mechanisms that cause the conformational conversion in proteins [9, 10] and peptides [11] have been observed.

One of the most drastic conformational conversion involves the transition between an α-helix domain and a β-sheet domain found in natural and synthesized sequences. In peptides, the conversion can be initiated by adjusting solvent conditions such as pH [12, 13], temperature [14], salt or organic concentration [13, 15], and oxidation-reduction reaction [16]. Such conformational conversion in other systems can be induced also by other factors directly associated with chain systems; a change in the length of the polypeptide [22], a mutation replacement of residues in the sequence [23–25], interchange of residues within a polypeptide [26]. It can be also induced by the interaction of a chain with a solid substrate [27], and it can arise in the normal folding process of a mainly β-content dominant β-lactoglobulin [28].

The effort to advance the understanding of conformational diseases [29, 30] has been a major driving force of recent studies in the conformational conversions between α-helix and β-sheet structures. A minor alteration in protein sequences or

a change in the environment is suspected to be a mechanism that causes protein misfolding leading to a conformational conversion resulting in irreversible amyloid protofilaments [29, 30]. The formation of amyloid protofilaments is linked to neurodegenerative diseases, including fatal prion diseases such as bovine spongiform encephalopathy, Creutzfeldt-Jakob disease, Kuru and Alzheimer's disease. The formation of ordered amyloids is observed in a wide range of systems; it can be fatal while harmless in other cases [31]. The diversity in sequences of known amyloid fibrils suggests that the amyloid fibril formation is not driven by the sequence specific side chain interactions, rather it should be driven by the sequence non-specific backbone hydrogen bonding and hydrophobic interactions [31]. For exploring the interconversion between an α-helix and β-sheet and the formation of amyloid fibrils from normal polypeptides, polyalanine chains provide unique simple test systems because alanine residues have the highest propensity for the helix conformation [45–47] and yet polyalanine chains can form β-sheet conformations. This observation suggests that the environment such as a solvent significantly influences on the conformation of the system. In experiments, for example, the interconversion of polyalanine chain induced by the implicit solvent effect was observed by using circular dichroism, Fourier-transform infrared spectroscopy and reversed phase liquid chromatography [17, 18].

In this work, we present an off-lattice protein model that can portrait the dependence of the structures on the effective hydrogen-bonding formation propensity or the hydrophobicity on a few key residues; depending on the values of these parameters, the overall native structure of the modelled peptide can form one of the two ground states: an α-helix (Figs. 6.1-A1 and 6.1-A2) or a β-hairpin (Figs. 6.1-B1 and 6.1-B2). We consider the impacts of the hydrogen bonding propensity of the middle section of the chain systems (at the locations labelled by m_1, m_2 and m_3 in Fig. 6.1) and the hydrophobicity of a residue pair (at the locations labelled by n and $\bar{n}$ where $n = 1, 2, 3, 5$ in Fig. 6.1), on stabilizing the native structure, α-helix or β-hairpin, of the model. Three themes can be illustrated on the thermodynamics of the systems. The first theme is modifying the effective strength of middle-section hydrogen bonding, which can be accomplished experimentally by replacing the corresponding residues with de novo amino acid residues, tailored to

control the ability for a residue to form the backbone hydrogen bonding or/and varying the solvent quality toward the peptide [32]. The second is modifying the hydrophobic interaction strength of a special pair of residues, which can be accomplished experimentally by controlling the solvent condition directly [33, 34] or by replacing them with another natural or synthesized residue pair of different hydrophobicity, properly avoiding or minimizing any undesirable change in physical properties other than the hydrophobicity [35, 36]. The third is on possible conformational conversions initiated by temperature variation; we show that in a few special regions of the interaction parameter space, i.e., for specially designed sequences, our model portraits a series of structural transitions: from coil (stable at a denaturing temperature) to β-hairpin (stable at an intermediate temperature) and then from β-hairpin to α-helix (stable at a lower temperature). The preliminary results of the first theme above were reported in [37]. In addition to these thermodynamic pictures, we consider the kinetic accessibility of an α-helical to a β-hairpin conformation when the two conformations have same energies in the systems.

To investigate the conformational conversion between a β-hairpin and an α-helix which have totally different arrangement of the backbone hydrogen bonding, a suitable model must accommodate the α- and β-conformations solely based on a given sequence. Many conventional simple models use conformational specific dihedral angle potentials that impose steric bias toward an α- or a β-conformation [52, 55] and/or use Gō type potentials that are based on predesigned native contact potentials for a single reference conformation [38] and these models are not suitable to explore conformational interconversions in proteins. All-atom models can in principle define a native conformation based on its sequence and accommodate multiple native conformations. Hence they are suitable to investigate conformational conversions [19, 39–41] and can provide microscopic folding trajectories of a small protein providing a unique perspective on detailed physical properties of protein folding under various idealized environments. However, to thoroughly characterize the thermodynamics and kinetics and explore the various of perturbative effects such as mutation and change in a solvent conditions, the complexity of all-atom models is still computationally too demanding, particularly in the time scale involved in the conformational conversion or aggregation, which can be sub-

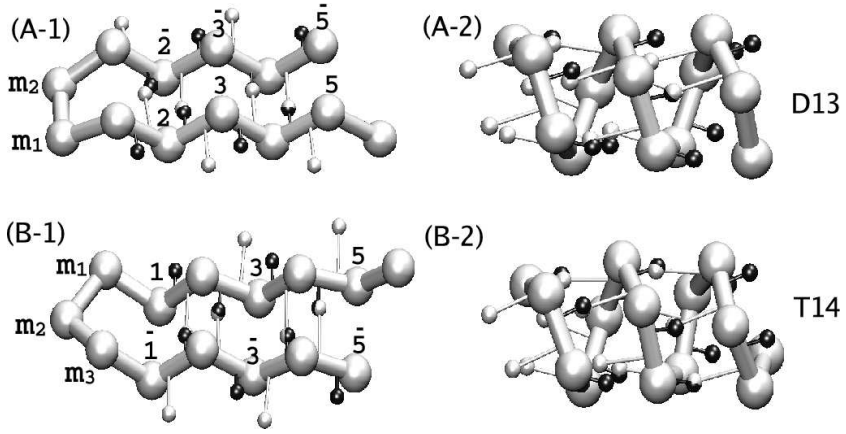

Figure 6.1: Two typical native conformations for the 13mer and 14mer models considered in this work. For this figure, we have considered the results from a global energy minimization of the model for $\eta = 0$ (left two plots) and $\eta = 1$ for the sequences D13S0 and T14S0 (see Table 6.1 for definition). Large gray monomers represent α-carbons, small white spheres virtual oxygen atoms, and small black spheres virtual hydrogen atoms. In the 13mer model, the α-helix is a stable structure when all monomers interact with each other through a hydrogen-bonding interaction; after the two middle monomers (labelled m_1 and m_2) being replaced by β-turn formers, the displayed β-hairpin is stabilized. Such a mutational replacement of monomers is realized in our model by changing the mutation parameter η from 1 to 0.

stantially larger than the time scale commonly observed in normal protein folding process. To overcome the difficulty of the long time scale, several intermediate resolution models are developed along with efficient computational techniques and are applied to describe the interconversion phenomena. All of these models notably incorporate the hydrogen bonding potential explicitly, which is the essential physical parameter for the conformational interconversion of between helical and β-sheet conformations [20, 21, 37, 42]. A reduced model, such as our current model, constructed from a judicious parameterization of essential physical features, can be used to characterize the generic properties the conformational conversion in proteins. The advantages of reduced models lie in their ability to rapidly accumulate meaningful statistics on folding pathways, thermodynamics and dynamics, and in ease of controlling the most physically relevant parameters of the systems [43, 44].

On the conformational interconversion, a number of computer simulations have been performed on polyalanine chain systems [19–21] because polyalanine provides a simple test system that display the conformational conversion between α- and *beta*-conformations [17, 18]. Levy et al. [19] calculated the free energy landscapes of polyalanine in vacuum, representing a hydrophobic medium, and in an aqueous hydrophilic solution in an all-atom model with molecular dynamics — large conformational spaces were surveyed at higher temperature and then the conformational areas of interest were sampled at physiologically feasible temperatures. They found that an α-helical and β-hairpin conformations with some coils are more stable in non-polar and polar solutions, respectively. Ding et al. [20] examined the conformational interconversion between an α-helix and a β-hairpin for a 16mer polyalanine chain and for a sequence having hydrophobic and polar interactions using an intermediate resolution model with an efficient discrete molecular dynamics method. Performing ten simulations at various temperatures, they observed that intermediate β-hairpin conformations arise in a system where the α-helix is the true native state and a stable conformation in the system depends on the simulating temperature. They also argued that weakening the relative sidechain interactions, simulating environment changes induced by the solvent quality, may lead to a β-hairpin intermediate and the stronger hydrophobic interaction tends to drive the chain into a compact globular state. Nguyen et al. [21] have investigated the thermodynamic

properties of a polyalanine chain for various temperatures and for various effective strengths of hydrophobic interactions using an intermediate resolution model with a discrete molecular dynamics algorithm. They found that the environmental factor has a significant impact on the stability of various structures such as α and β-conformations at various temperatures.

Other peptide chain systems such as prion proteins that undergo the conformational interconversion are also investigated in computational studies. Daidone et al. [40] investigated the interconversion in a prion system — the spontaneous transition to the β-hairpin conformations of the Syrian hamster PrP peptide H1 and of the A (12-28) fragment without change in environments or any mutational effect by using long timescale MD simulations in an all-atom model with explicit water and they suggested that the system undergoes the spontaneous interconversion because the system is highly frustrated and contains the hydrophobic residues that are susceptible to further intermolecular aggregation. Ikeda et al. [41] performed multicanonical molecular dynamics on an all-atom model of a sequence in the yeast MAT 2/MCM1/DNA complex, known as the chameleon sequence, that folds into an α-helix or a β-hairpin depending on the surrounding and they observed such conformational conversions in their simulations. Malolepsza et al. [48] have simulated possible conformational autocatalysis propagation of β-sheet structures as a protocol for amyloid formation using a reduced, high resolution continuous model through the replica exchange Monte Carlo sampling method. Hoang et al. [49] investigated a simple tube model emphasizing geometry and symmetry imposed by the hydrogen bond steric effect and hydrophobic attraction, which explored the phase diagram of the system in terms of geometric parameters and solvent mediated energies. They found that the compactness and directionality induced by the hydrogen bonding are found to be the key determinant factors for the conformations. Our group [42] has demonstrated helix/β-hairpin interconversions in the presence of the hydrophobic attractive wall by using canonical and multicanonical Monte Carlo algorithms using the model similar to the current model Peng and Hansmann also investigated similar conformational interconversion in the presence of another β-sheet structure by using an all-atom model and multicanonical Monte Carlo method [39].

In this chapter we investigate the conformational conversions of the simple systems that can form β-hairpin and α-helical conformations depending on the effective strength of hydrogen bonding η and the position of hydrophobic pair and reduced temperature. We examine the phase compositions of the systems and explore implications to possible conformational conversions in proteins. To consider the underlining physical mechanism behind the phase diagram, we examine the heat capacity and, further, the free energy maps are used to unearth the more detail phase compositions that cannot be captured by the phase diagram maps, and they show complex phase conformations, mostly either β-hairpin or α-helix and few others at a conversion point η_c where a β-hairpin and an α-helix have same energies. In addition, the folding rate to a β- or α -conformation from a random coil is examined by using canonical Monte Carlo algorithm for the systems with η_c where β- and α-conformations have same energies. The transition rate from the β-hairpin to an α-helix and from the α-helix to a β-hairpin is also examined. We found that theses dynamic results agree with the physical pictures obtained by thermodynamic analysis.

6.2 The model

The reduced model used in this chapter is an extension of a previous off-lattice model, capable of describing both α-helix and β-sheet structures, proposed by our group [42, 50, 51]. The advantage of this built-in duality is particularly important, which allowed us to study the physical properties of structural conversion between these two native structures [37, 42]. The preliminary report in Chapter 5 [37] on the study of structural conversion have shown that the conformational propensity to α-helix and β-hairpin conformations depends on the effective strength of the hydrogen bonding in the middle of the chain, which is controlled by the parameter η. The dynamic and thermodynamic properties are also shown to be modified by η. We extend the previous chapter to include the hydrophobic interaction and also examine the impact of the size of mutations by studying the triple mutant chains in addition to double mutant chains in this work. The sequence we examine here is listed in Table 6.1 where residue types P, H and B represent for a polar,

hydrophobic and β-turn former (perturbed residue) — these residue type specific interactions are explained below. In this chapter, we mainly focus on two model sequences, one consisting of 13 monomers, and the other of 14 monomers and the sequences we examine is listed in Table 6.1. the D13 sequence consists of two middle residues, whose adjacent hydrogen bonding strength are controlled by η, and similarly the T14 sequence consists of the three middle residues. The sequence without a hydrophobic pair is named as D13S0 and T14S0. The hydrophobic sequence, whose hydrophobic pair is located at $(2, \bar{2})$, $(3, \bar{3})$, $(5, \bar{5})$ for 13mer is named as D13S2, D13S3, and D13S5 respectively, and at $(1, \bar{1})$, $(3, \bar{3})$, $(5, \bar{5})$ for the 14mer named as T14S1, T14S3, and T14S5 respectively (Table 6.1).

A reduced peptide chain system is governed by bonded and non-bonded potential interactions. The monomers along the chain are linearly connected by bonds of fixed length l, and non-terminal monomers are allowed to rotate about the axis connecting the two adjacent monomers. Each bond angle is controlled by a stiff harmonic potential energy

$$V_\theta = k_\theta/2(\theta - \theta_0)^2$$

where $k_\theta = 20\epsilon/(rad)^2$ where ϵ is the effective strength of a hydrogen bond in units of energy in this model and θ_0 is $1.833\, rad$ (105^o). Many other reduced models, based on residue-level approximations, contain conformational specific potentials and/or an additional potential energy that deal with a preferred torsional angle formed by the three consecutive bonds, relating four consecutive residues; such an energy is usually proposed to assist the model in driving the local bond arrangement to form one of the two typical secondary structures, either α-helix or β-strand conformation [52–58]. The model does not adopt this type of torsional energy in order to efficiently encompass the conformational conversion between an α-helix and a β-hairpin. In our model the final torsional angles formed in a native structure follow directly from a combination of directionally biased hydrogen-bonding type potential energies between monomers [59] as formulated below. The formation of an α-helix, β hairpin, or combined configurations, for example, is the result of the hydrogen bonding between local and non-local monomers [37, 51].

In addition to the bond-angle potential energy, three types of non-bounded interactions are invoked between any pair of monomers along the chain. First, any

two monomers interact with each other through an excluded volume, characterized by a monomer diameter of $6l/5$. This diameter is somewhat greater than the bond length l and reflects the geometry of the resulting α-helix, to match that of real systems [51]. As exceptions, a hydrophobic residue interact with another hydrophobic one through a hydrophobic interaction defined below and two nearest-neighbor residues are not subject to this excluded-volume interaction since these residues are already constrained by the bond-angle potentials. Second, any two monomers can interact through an effective hydrogen bonding, generated by the interaction between a virtual oxygen (O) and hydrogen (H), separately associated with each of the two monomers. In particular, for the ith monomer that is described by the position vector $\vec{r}_i$, an O_i and a H_i were placed at the tips of the vectors $\vec{r}_i^{\,(\mathrm{O})}$ and $\vec{r}_i^{\,(\mathrm{H})}$ respectively, directed away from $\vec{r}_i$:

$$\vec{r}_i^{\,(\mathrm{O})} = \frac{1}{3}(\vec{r}_{i+1} - \vec{r}_i) + 3\zeta_O l/5\vec{n}_i \tag{6.1}$$

$$\vec{r}_i^{\,(\mathrm{H})} = \frac{2}{3}(\vec{r}_{i+1} - \vec{r}_i) - 3\zeta_H l/5\vec{n}_i \tag{6.2}$$

where $\vec{n}_i$ is the normal direction of the plane defined by the $(i-1)$th, ith, and $(i+1)$th monomers and expressed as $\vec{n}_i = (\vec{r}_i - \vec{r}_{i-1}) \times (\vec{r}_{i+1} - \vec{r}_i)/|(\vec{r}_i - \vec{r}_{i-1}) \times (\vec{r}_{i+1} - \vec{r}_i)|$. The coefficients, $\zeta_O = 1.3$ and $\zeta_H = 0.7$, are chirality factors, introduced to mimic the specific chirality of a right handed α-helix in real systems. These chirality biases were explicitly used in a similar intermediate resolution model [60]. Except for the first and second nearest-neighbors, a pair of a virtual oxygen and hydrogen associated with unperturbed residue types, hydrophobic H and polar P, interact with each other through a shifted Lennard-Jones potential,

$$V_{OH} = 4\epsilon \left[(l/(r + r_0))^{12} - (l/(r + r_0))^6 \right] \tag{6.3}$$

and where $r_0 = 2^{1/6} \times 5l/6$ is the optimal bonding distance, and r is the OH distance. Any pair of a virtual oxygen and hydrogen associated with the perturbed residue type, β-turn former B, interacts with any type of a residue through a potential,

$$V_{OH}^{\eta} = 4\eta\epsilon \left[(l/(r + r_0))^{12} - (l/(r + r_0))^6 \right]. \tag{6.4}$$

An effective hydrogen bonding is reached as $r = 0$, that allows the two involved monomers to acquire a bonding energy $V_{OH} = -\epsilon$ or $V_{OH}^{\eta} = -\eta\epsilon$ (for $\eta \neq 0$)

depending on types of residues involved. Previous models based on a similar parameterization have been used to study the dependence of structural stability and folding dynamics of α-helices and β-hairpins, on chain length, temperature (k_BT/ϵ), and the effective hydrogen bond strength of the middle residues [37, 42, 51]. One of the two strands is longer than the other by one residue because the directionality of the hydrogen bonding for the ith residue is defined by the $(i-1)$th residue therefore an additional residue is attached to define the directionality for the first active residue.

Third, in our model, a pair of hydrophobic monomers can interact through an effective hydrophobic interaction. We use a simple Lennard-Jones parameterization,

$$V_H = 4\epsilon_H[(6l/5r)^{12} - (6l/5r)^6] \tag{6.5}$$

where $\epsilon_H = 0.7\epsilon$ represents the effective strength of a hydrophobic potential, and r is the distance between the two interacting monomers. Irbäck and coworkers [61] have investigated the effect of relative strength of hydrophobic and hydrogen attractions, ϵ_H/ϵ in a similar approach; for comparison, they have used a similar directionally biased potential designed for α-helical bundles and gave example results for the case of $\epsilon_H = 0.78\epsilon$. By using a similar model to that described in this section, we have recently examined the physical properties of β hairpins that contain a pair of hydrophobic monomers placed on selective monomers, by varying ϵ_H from 0.3ϵ to 1.2ϵ (unpublished results). In the current work, the value $\epsilon_H = 0.7\epsilon$ is used to investigate the impact of a hydrophobic pair on the α-β conversion.

The middle segment of the peptide chain significantly influences its folding and stability. Two factors can affect on the middle section of systems discussed: one is the mutational effect on the strength of the hydrogen bonding and another is a hydrophobic pair. The strength of a backbone hydrogen bonding is difficult to control without altering other physical properties that should be preserved. It can be, however, altered by residue replacement using a natural and synthesized amino acid or/and by modification on solvent condition [32] . Recently site-specific amide-to-ester mutations were applied to perturb the backbone hydrogen bonds at a specific site of various small proteins to control the strength of the backbone hydrogen bonding, whose strength in natural amino acids does not significantly

Sequence	Residue type	η_c
	$5\ 4\ 3\ 2\ 1\ m_1\ m_2\ \bar{1}\ \bar{2}\ \bar{3}\ \bar{4}\ \bar{5}$	
D13S0	$P\ P\ P\ P\ P\ P\ B\ \ B\ \ P\ P\ P\ P\ P$	0.543
D13S2	$P\ P\ P\ P\ \mathsf{H}\ P\ B\ \ B\ \ P\ \mathsf{H}\ P\ P\ P$	0.482
D13S3	$P\ P\ P\ \mathsf{H}\ P\ P\ B\ \ B\ \ P\ P\ \mathsf{H}\ P\ P$	0.669
D13S5	$P\ \mathsf{H}\ P\ P\ P\ P\ B\ \ B\ \ P\ P\ P\ P\ \mathsf{H}$	0.670
	$5\ 4\ 3\ 2\ 1\ m_1\ m_2\ m_3\ \bar{1}\ \bar{2}\ \bar{3}\ \bar{4}\ \bar{5}$	
T14S0	$P\ P\ P\ P\ P\ P\ P\ B\ \ B\ \ B\ P\ P\ P\ P\ P$	0.576
T14S1	$P\ P\ P\ P\ P\ P\ \mathsf{H}\ B\ \ B\ \ B\ \mathsf{H}\ P\ P\ P\ P$	0.563
T14S3	$P\ P\ P\ \mathsf{H}\ P\ P\ P\ B\ \ B\ \ B\ P\ P\ \mathsf{H}\ P\ P$	0.644
T14S5	$P\ \mathsf{H}\ P\ P\ P\ P\ P\ B\ \ B\ \ B\ P\ P\ P\ P\ \mathsf{H}$	0.666

Table 6.1: The sequences: P, H and B stand for a polar, hydrophobic and β-turn former (perturbed residue). η_c is a η value at which α-helix and β-hairpin become the same energy.

vary [62]. For example amide-to-ester mutations were used to examine the site dependency of the backbone hydrogen bonding in a small three stranded β-sheet protein of the PIN WW domain [63]. Some amino acids, also, have significantly different side-chain conformation than others and they can promote or disrupt the formation of backbone hydrogen bonding in the secondary structures. For example, glycine, whose side chain is just a proton, can take a wide range of steric conformations, and proline is sterically constrained because of its unique ring structure and its backbone hydrogen bonds that can not contribute to the β-sheet hydrogen bond network [64]; these residues often disturb and break the hydrogen-bond network of an α-helix and a β-sheet, and they frequently comprise a reverse turn of β-sheets where no intra-strand backbone hydrogen bonding is formed [34]. These direct and indirect perturbative effects on the backbone hydrogen bonding can be reflected in our model by effectively perturbing V_{OH} for these residues. Schematically, the residues at the β-turn, labelled m_1, m_2 and m_3, are shown in Fig. 6.1 without the associated virtual O and H, and these perturbed residues are also present in the helices of Fig. 6.1, but are not shown for clarity. The strength of hydrogen bonding

of all O's and H's associated with the perturbed middle residues is controlled by the parameter η — see Fig. reffig:model2, where small, dark (light) spheres representing virtual O's (H's) are shown for other monomers. For $\eta = 0$, a β-hairpin is the dominant conformation, and as η increases, then the helical conformation, consisting of consecutive local hydrogen bonding, becomes a more preferred conformation.

Hydrophobic interactions play significant role in proteins [62] and the position of the hydrophobic pair crucially influences the stability of a system. We investigate how the hydrophobic interaction modify the conformations in a simple system by examining the change in phase diagrams. The hydrophobic pair placed on various positions can stabilize the β-hairpin conformation differently as we will demonstrate and also it can stabilize the α-helix content but it can also destabilize α-helix content when the hydrophobic pair is placed at an unfavorable location for the α-helical conformation. The middle residues can become the effective β-turn formers or a part of a regular α-helix, depending on the value of η and the position of the hydrophobic pair if it is present. For example for $\eta = 0$ they become most effective β-turn formers (helix breakers) and for $\eta = 1$ a regular helical segment. For intermediate values they may participate in both conformations. As the value of η changes from zero to one, the native conformation of the system switches from a β-hairpin to an α-helix, and somewhere in the middle each system will go through a conversion point for some η value where both β-hairpin and α-helix conformations have the same energies. That value of η is called the conversion value denoted by η_c and its value depends on the position of the hydrophobic pair because the position of the hydrophobic pair shifts the energy balance of the β-hairpin and the α-helix. The value of η_c is listed next to its corresponding sequence in Table 6.1.

6.3 Simulation method

The above model contains coarse-graining approximations to reduce atomic complexities in all-atom models. Focusing on the essential features of proteins, we are able to examine the most important structural properties and to investigate the entire folding process with sufficient statistics generated by a computer simulation.

To obtain the thermodynamic properties with sufficient conformational samples, especially at low temperatures where the Metropolis Monte Carlo method [65]

becomes increasingly inefficient, we computed trajectories according to a biased weighting factor (in the so-called multicanonical scheme), which enables a system to sample the energy space uniformly without being trapped in local minima [66]. The bias in the weighting scheme is then removed by the weighted histogram analysis method (WHAM) [67], which translates the physical observables measured with the biased non-Boltzmann weight into those normalized by the Boltzmann weight [67]. For the folding simulations, we have used a Boltzmann weight at a fixed temperature and used a local rotational move on a randomly selected monomer. The Metropolis Monte Carlo method was originally developed for calculating equilibrium properties of physical systems; [65, 68] the generalization to describing a stochastic processes of nonequilibrate system is also possible once a physically feasible move is adopted [68–72]. For example, Kikuchi et al. [70] have directly shown that MMC can be used to solve the Fokker-Planck equation, an approximation to the the master equation for Markov processes and Rey et al. [71] have shown that MMC can reproduce the folding mechanism obtained by Brownian dynamics. Chubykalo et al. [72] have shown that MC time step can be converted to real time steps in a magnetic system. As for local moves in the model, if the selected monomer is terminal, which has only one neighboring monomer, the monomer was rotated around the axis defined by the nearest and second nearest monomers. If the selected monomer is not terminal, the monomer was rotated around the axis connecting the two adjacent monomers. The rotational angle is a random number selected from the range between $-\pi/2$ and $\pi/2$, which leads to an acceptance rate of approximately $20-30\%$ [73]. A Monte Carlo step (MCS) in the simulation amounts to 13 or 14 moves for a system containing 13 or 14 monomers. Whenever the concept of time is discussed below, MCS is used as the basic units. The relationship between the time measured here and the real physical time can be estimated up to an undetermined constant factor. The dynamic folding rates are determined by more than 23000 full folding events through canonical Monte Carlo simulations in total.

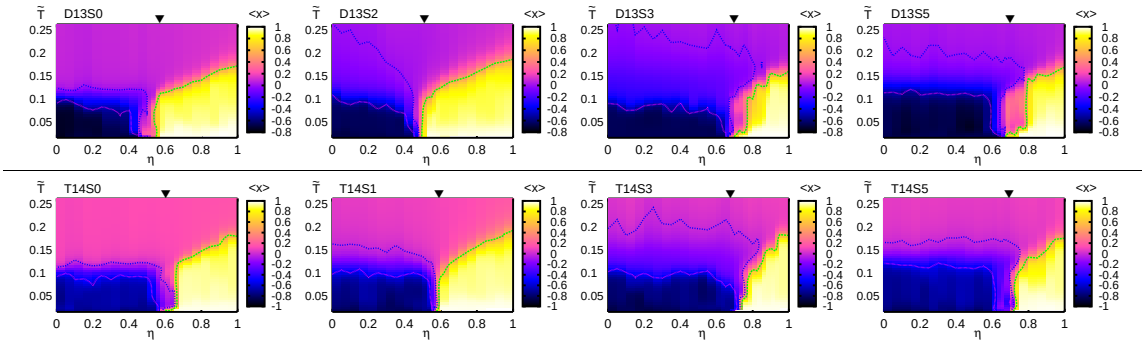

Figure 6.2: State diagrams for the eight sequences considered in this work. The parameter space filled with a dark blue (or light yellow) color represents the regime where a β hairpin (or α-helix) is stable, as characterized by the structural order parameter $\langle x \rangle$. The vertical side bar indicates the value of $\langle x \rangle$ in the state diagram, varying from light yellow ($\langle x \rangle \approx 1$, α-helix), through purple ($\langle x \rangle \approx 0$, random coil), to dark blue ($\langle x \rangle \approx -1$, β-hairpin). The down triangles on top of each diagram indicate the structural conversion point, whose values are listed in Table 6.1. The contour lines in pink, dark blue, and green, outline $\langle x \rangle = -0.5, 0$, and 0.5, respectively.

6.4 Results and discussions

6.4.1 State diagrams

The physical properties, in particular the structural conformation of the current model, are completely determined by parameters such as the number of monomers $N(= 13 \text{ or } 14)$, reduced temperature $\tilde{T} = k_B T/\epsilon$ where k_B is the Boltzmann constant, and the mutation parameter η. For fixed N and $\tilde{T} = 0$, depending on the magnitude of the parameter η, there could be two clearly identifiable native conformations, α-helix and β-hairpin, in our model; to quantitatively describe the coordinates of the monomers, two characteristic native structures have been separately obtained, for $\eta = 0$ and $\eta = 1$ respectively, through an energy minimization scheme on the systems and these two reference conformations are displayed in Fig. 6.1 for each chain length, $N = 13$ and 14.

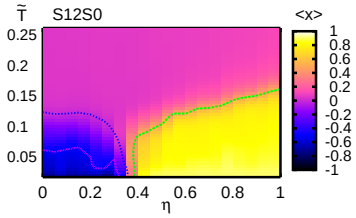

Figure 6.3: The state diagram for a non-hydrophobic sequence of 12 residues with a single mutation.

To describe the "distance" or similarity between a given conformation and that of a specific reference native conformation, one usually uses the root mean square deviation between all involved coordinates of the two as a measure. The problem at hand, however, requires to compare a given conformation with not one, but two reference native conformations. For this purpose, we introduce an order parameter x,

$$x = (D^{\beta} - D^{\alpha})/D_0 \tag{6.6}$$

where D^{β} (or D^{α}) is the root mean square deviation of monomer distances between a certain conformation and the reference coordinates in a β-hairpin (or α-helix). Note that D^{β} (or D^{α}) is computed after minimizing itself for all possible relative rotations and translations of the two conformations. D_0 above is a normalization factor. The order parameter x can properly handle the distinction between the two target conformations; for a structure similar to a β-hairpin, $x \to -1$ and for a structure similar to an α-helix, $x \to 1$. Otherwise, if the considered conformation is very different from both α-helix and β-hairpin, x approaches 0. For any point in the parameter space, we consider $\langle x \rangle$ where angular brackets $\langle ... \rangle$ represent the thermodynamic average with respect to all conformations.

Using the averaged order parameter $\langle x \rangle$ as a measure, we can plot a state diagram, displayed in Fig. 6.2, as a function of $\tilde{T}$ and η for each sequence defined in Table 6.1. In such a state diagram, the parameter regime of a particularly stable conformation is measured by the color code, corresponding to the magnitude of

$\langle x \rangle$ displayed on the side bar. Taking sequence D13S0 for example, at $\eta = 0$ a β-hairpin is stable below $\tilde{T} \approx 0.12$, reflected by the deep blue color, which, by the side bar, indicates that $\langle x \rangle$ is close to -1; for fixed $\tilde{T} = 0$, the stability of phases depends on the magnitude of η — a β-hairpin is stable below $\eta_c \approx 0.4$ and an α-helix (represented by light yellow color) is stable beyond this η_c. The stable β-turn regime is confined approximately in the regime $0 \leq \eta \leq \eta_c$ and $0 \leq \tilde{T} \leq 0.12$. Between $\eta = 0.4$ and $\eta = 0.57$, $\langle x \rangle$ is close to zero. The free energy maps discussed in the coming sections indicate that β-hairpin, α-helical and also some small portion of non-specific conformations coexist in this regime.

Applying the backbone hydrogen bond and steric perturbation in the middle of a designed helix, and changing the condition of the solvent, Awasthi et al. have investigated the conformational interconversion between a well defined $\alpha/3_{10}$ helical and β-hairpin conformation. This conformational interconversion is possible because the conformational propensity of the peptide chain is not so strong to either a helix nor a β-hairpin and can adapt one of two conformations depending on the solvent [32]. They showed that backbone hydrogen bonding can be perturbed by changing the properties of the solution as well as introducing an ester-to-amide replacement.

When a hydrophobic pair is introduced in the sequence (see sequences D13S3, D13S5, and D13S7), the overall physical pictures of the existence of two separated β-hairpin and α helix regimes are not significantly changed, but we do observe some changes. One of the major changes is that the stable regimes of β-hairpin and α-helix shift in the space of temperature and η. Since the anisotropic hydrophobic effect starts interacting at higher temperature than the directionally specific hydrogen bonding interaction, the β-hairpins become stable at higher $\tilde{T}$ for all hydrophobic systems D13S2, D13S3 and D13S5. The position of the hydrophobic pair in the sequence also affects the stability of the β-hairpin and also the α-helix. The hydrophobic pair at any location enhances the overall β-hairpin contents and this leads to the expanded β-hairpin regions shown in blue and dark blue in the phase diagram of D13S2, D13S3 and D13S5. The hydrophobic pair, however, enhances α-hairpin contents only in D13S2 where the hydrophobic pair can form the folding core of an α-helix conformation at the middle — this can be observed in

the enlarged yellow region in D13S2. In D13S3 and D13S5, the hydrophobic pair stabilizes β-hairpin contents and consequently hinders the helical formation. This leads to the dwindled domains of the α-helical conformation in the temperature space seen for D13S3 and D13S5. In other words, the α-helix becomes stable only at lower $\tilde{T}$ in these systems.

The state diagrams for the 14mers (the second row of Fig. 6.2) share many qualitatively common features of the 13mers; in particular there is close similarity between the sequences whose hydrophobic pair is placed at a similar location. A major difference from the 13mersis the expansion of the stable β-turn regime to a higher value of η in the 14mers. This is because the number of perturbed hydrogen bonds those with a hydrogen bonding strength of $\eta\epsilon$ in the middle section is increased to six from four, which leads to a major interruption of the helix in the core. In the meantime, the energy of the native β-hairpin is kept almost the same. In addition, the overall β-hairpin propensity in the 14mer model is slightly increased because a proper β-turn can now be formed more easily with three residues in the middle section, in comparison with two residues in the 13mers. Consequently, the β-hairpin conformation becomes more stable in the 14mers, in comparison with their counterpart 13mers. We have further tested the influence of the number of the perturbed residues in the middle using a sequence of 12mers that contains only a single perturbed residue in the middle (Fig. 6.3). Its phase diagram indicates a more unstable β-hairpin conformation, that is, a smaller dark blue region, because the the β-turn of one residue is unstable and a helix is not so seriously disturbed as in the D13 and D14 systems. Note that the 12mer differs from the D13 and T14 systems because the 12mer system has a ground state that does not belong to either β- or α-conformation even at the lowest temperature for $0.25 < \eta < 0.38$ (Fig. 6.3).

6.4.2 Biological interpretation of the state diagrams

What structural implications can we draw based on these state diagrams? Consider a sequence whose native conformation is discordant to the conformation predicted from the residue secondary structure propensity. This sequence, under some

perturbation, might seek to switch the conformation to one that is close to the conformational propensity of the residues [1]. Given another sequence that has no overwhelming preference for specific local backbone conformation at individual residues, then an overall resulting conformation can be determined by balancing between intrapeptide and solvent-peptide hydrogen bonds [32]. In the current systems, the stability of various conformations is determined by the effective strength of the backbone hydrogen bond η, the position of the hydrophobic pair and the number of perturbed residues in the middle, and temperature. It is possible to envision that the current model can be used to explore at least three types of structural-conversion experiments.

First suppose that we sample a real peptide sequence that contains two middle monomers which do not carry strong backbone conformational propensity to a specific secondary structure. An example of such a peptide is an octapeptide containing -Aib-Gly- residues in the middle investigated by Awasthi et al. [32]. The overall native state of the peptide falls into the light yellow regime of a state diagram and the stable structure is an α-helix. Now, if the two middle residues have been replaced by residues that contain a weaker hydrogen-bonding propensity (i.e., a sufficiently small η), the native state of the peptide would fall in the deep blue regime. Proline and α-amino isobutyric acid (Aib), for example, are known to have less hydrogen-bonding propensity and to be β-turn formers as well. Hence, the mutational replacement of the middle residues with different hydrogen-bonding propensity can yield a structural conversion between an α-helix and a β-hairpin. Indeed, in a recent experiment, Balaram's group has demonstrated such a conformational conversion [32]. The current model offers a simple modelling tool for the theoretical study of these mutational experiments.

Second, the current model also predicts another possible structural conversion scenario based on mutations. Take sequence D13S0 and select two middle monomers that have an η value of 0.55. The native structure of the peptide would fall into the light yellow regime of the state diagram in Fig. 6.2 at low $\tilde{T}$ (a point approximately at $\eta = 0.55$ and $\tilde{T} = 0$). The native structure is hence the α-helix. Instead of mutational replacement of residues at the middle (m_1 and m_2) to enforce a β-turn, we can consider the replacement of one pair of residues, at positions that

correspond to the native contact across the two proposed β-strands, by residues with a stronger hydrophobicity [for instance, residues 3 and $\bar{3}$ (see Fig. 6.1)]. Such a replacement would take a point at $\eta = 0.55$ and $\tilde{T} \approx 0$ in the state diagram Fig. 6.2 (D13S0) directly to the same point in the state diagram Fig. 6.2 (D13S5). As can be observed in the figure, the latter point corresponds to a deep blue color, so that the stable state is a β-hairpin. Hence, the enhancement of attraction on residues native to a proposed β-hairpin by either mutating the residues with more hydrophobic residues or amplifying the hydrophobicity of the residues with solvent modification can cause a structural conversion between an α-helix and a β-hairpin as well. The current model can be used as a tool to understand this provisional theoretical interconversion. Such a structural conversion has not been observed so far in experiments.

Finally, for a fixed sequence that contains a certain range of the hydrogen strength parameter, the current model also predicts possible structural transition induced by temperature change. An example of this can be seen for sequences D13S3, D13S5, T14S3, and T14S5. Assume that one can design a sequence with an interaction parameter $\eta \approx 0.75$. As the reduced temperature $\tilde{T}$ is increased from a low value to approximately $\tilde{T} = 0.15$, the current model predicts that the structural state of this peptide undergoes a structural conversion from an α-helix to a β-hairpin, through an intermediate regime where no definitive structure is stable. Similar structural conversion between α- and β-conformations induced by temperature change is observed in various peptide systems [14, 74–78] though the interconversion between a helix and β-hairpin has yet to be observed because of the difficulties in the experiments associated with a β-hairpin system.

In experiments, designing a specific peptide system that undergoes interconversion under controlled environmental parameters has paramount importance to characterize the physical mechanism of conformational interconversion. It is desirable to have a residue that specifically affects the only physical property that one likes to examine without disturbing the other physical quantities of the system. For this goal, the twenty natural amino acids provide only limited physical variation and non-standard amino acids provide far more wider tools to manipulate the properties of targeted physical quantities without undesirable side effects. For example,

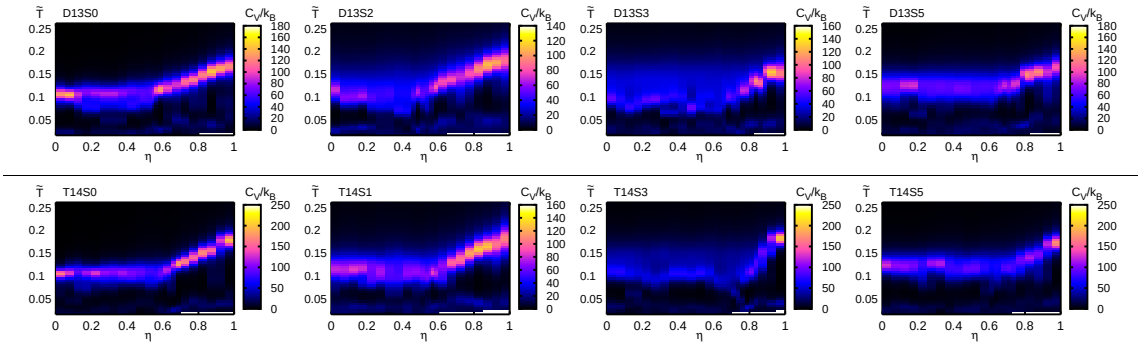

Figure 6.4: Heat capacity maps for the 8 protein sequences considered in this work. The vertical side bar on the right side indicates the value of the reduced heat capacity $\tilde{C}_v$ for every given η and $\tilde{T}$.

non-standard amino acids can endow a wide range of specific streochemical constrains and also flexible sidechain functionality [79–81]. In addition, the backbone hydrogen bond strength can directly be controlled by ester-to-amid replacement without affecting the steric and sidechain properties, which is not possible by using natural amino acids. This technique was applied to investigate the role of the backbone hydrogen bonding in β-sheet of a segment of WW at various sites and they found the degree of destabilization due to elimination of the backbone hydrogen bond depends on the site [63], and in helices composed of depsipeptides, peptides originated from natural antibacterials, it is found that a small backbone hydrogen perturbation, caused by the replacement of an alanine by a lactic acid containing an ester, can be tolerated with minor conformational adjustment [82, 83]. With a set of flexible non-standard amino acid, it is interesting to see a designed de novo sequence that exhibits similar conformational phase diagram shown in this chapter by controlling the solvent condition.

6.4.3 Heat capacity map

The state diagrams presented in Sect. 6.4.1, however, are not proper indicators of the intensity of the state transition across the state boundaries. The location of the

maximum in a heat-capacity curve, C_v, as a function of $\tilde{T}$ for a fixed sequence, is normally used to characterize the location of a major structural transition. Because we are dealing with finite systems of size N here the variation of C_v is smoother and cannot be compared with that of a system undergoing a real phase transition. Nevertheless, the intensity of the heat capacity peak indicates the nature of the structural transition between two well-defined state regimes. In this work, the reduced heat capacity, C_v/k_B, was computed based on energy fluctuations [84]

$$C_v/k_B = \tilde{C}_v = (\langle (E/\epsilon)^2 \rangle - \langle (E/\epsilon) \rangle^2)/\tilde{T}^2. \tag{6.7}$$

The heat capacity maps in Fig. 6.4 display a few characteristic features of the structural transition in the system. First, the magnitude of the $\tilde{C}_v$ peaks tends to be larger around $\eta = 0$ (or $\eta = 1$) where a β-hairpin (or α-helix) is much more stable in comparison with the occurrence of the same structure in other sequences that have $\eta \neq 0$ (or $\eta \neq 1$). This is reflected in the heat capacity maps by the highlighted narrow string of regimes. The transition between the coil state, above the highlighted curve, and the β-hairpin (or α-helix), below the highlighted curve, as $\tilde{T}$ is lowered for $\eta = 0$ (or $\eta = 1$), displays a more definitive structural transition signature across the corresponding transition temperature. We have previously focused on the coil - β-hairpin transition for a system with $\eta = 0$ [50] and the coil - α-helix transition for an model that is equivalent to $\eta = 1$ [85, 86].

The coil - β-hairpin transition temperature remains nearly the same as that of the $\eta = 0$ sequence. This is because the magnitude of the hydrogen bonding energy of the β-turn formers marked with B in Table 6.1 does not influence the total potential energy of the native β-hairpin. In contrast, the coil - α-helix transition, where, in the current model, the total potential energy of the native α-helix depends on all those participating in hydrogen-bond formation — hence the hydrogen-bonding propensity η of the middle monomers. This dependence yields the variation of the coil - α-helix transition temperatures in η (Fig. 6.4).

Finally, for sequences D13S3, D13S5, T14S3 and T14S5 at $\eta \approx 0.75$, two perceivable stages of transitions could be observed in Fig. 6.2, as $\tilde{T}$ is lowered. The transition from the β-content dominated regime in blue to the intermediate stage, which occupies between the blue and green contour lines in the figure, is marked by

a weak, broad yet perceivable peak in heat capacity in Fig. 6.4. On the other hand there is no clear heat capacity peak in Fig. 6.4 for the conformational change from this intermediate stage to the regime of complete α-helix conformation depicted in the light yellow, which arises only at very low $\tilde{T}$. Hence this second conformational change is smooth and cannot be clearly identified as a structural transition.

6.4.4 Free energy

To further analyze the conformation properties of a particular sequence for a fixed η, we consider the free energy landscape $F(\tilde{T}, x)$, a function of both $\tilde{T}$ and order parameter x. Computationally, the reduced free energy F/k_BT is obtained by

$$\tilde{F}(\tilde{T}, x) \equiv F(\tilde{T}, x)/k_BT = A - \ln \Omega(\tilde{T}, x) \tag{6.8}$$

where $\Omega(\tilde{T}, x)$ is the histogram for occurrence of a given $\tilde{T}$ with a $0.002\tilde{T}$ increment and x with a $0.03x$ increment, found in the computer simulation of a sequence with a specified η. A constant A has been introduced to set the base line of the free energy zero, so that the reduced free energy is always positive.

Three characteristic values of η in the systems, $\eta = 0$, $\eta = \eta_c$ and $\eta = 1$, are considered for all 13mers and also for all 14mers for comparison. As can be seen in Fig. 6.2, at sufficiently low temperature, β-hairpin and α-helix conformations are stable at $\eta = 0$ and 1, respectively, reflected by $\langle x \rangle = -1$ (β-hairpin) and $\langle x \rangle = 1$ (α-helix). Also at sufficiently low temperature, the average order parameter $\langle x \rangle$ vanishes at $\eta = \eta_c$. However, $\langle x \rangle = 0$ is no indication of definitive structural information and two scenarios may exist; this may imply that all the conformations at this parameter point in the state diagram is very different from both α-helix and β-sheet (hence $x = 0$) and this may also imply that there is a coexistence of conformations corresponding to $x \approx 1$ and $x \approx -1$. By examining the free energy as a function of x itself, we are able to characterize the system more unambiguously.

The reduced free energies are shown in Fig. 6.5 in a map form again, with color contours imposed on top. A few interesting features can be seen in these plots. First, for any η at a high temperature, say $\tilde{T} = 0.25$, the random coil conformation, characterized by $x \approx 0$, occurs most frequently; both α-helix with

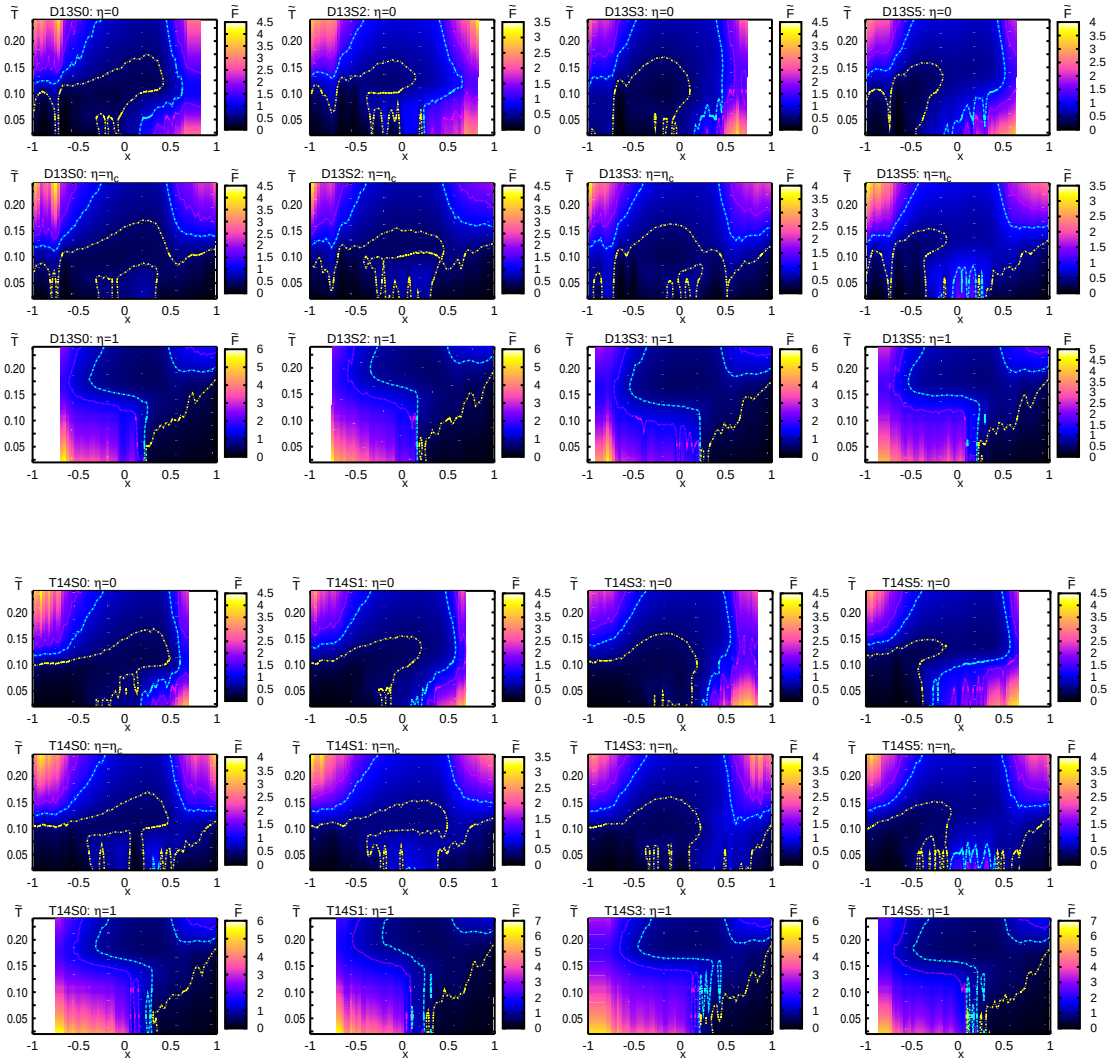

Figure 6.5: Maps of free energy as a function of the reduced temperature $\tilde{T}$ and order parameter x for a fixed sequence. The first (fourth), second (fifth) and third (sixth) rows correspond to the 13mers (14mers) considered in the current chapter for $\eta = 0$, η_c and 1, respectively. See Table 6.1 for the value of η_c used in these plots. According to the side bars, dark areas correspond to low free energies and light areas high free energies.

$x = 1$ and β-hairpin with $x = -1$ are structures rarely present, as indicated by the relatively high (light yellow) free energies.

Second, as the temperature is lowered, approaching the reduced temperature approximately corresponding to the coil - α-helix or coil - β-hairpin transition, characterized by the heat capacity peaks in Fig. 6.4, a wide range of conformations become accessible to the system, indicated by the narrow range of color across x at those transition temperatures. The system undergoes rapid fluctuations between all types of conformations at these temperatures, hence creates large energy fluctuations or large C_v.

Third, for systems corresponding to $\eta = 0$ (or $\eta = 1$), below $\tilde{T} \approx 0.10$, a deep free energy minimum develops, located near $x = -1$ (or $x = 1$ for $\eta = 1$). The system is dominated by conformations with structures close to the native β-hairpin (or α-helix for $\eta = 1$). This observation based on the reduced free energy agrees well with that based on the state diagrams specified by $\langle x \rangle$, in Fig. 6.2.

Fourth, at low temperatures for systems corresponding to η_c (second and fifth rows in Fig. 6.5), the reduced free energies show double minima at $x \approx 1$ and $x \approx -1$, separated by an energy barrier around $x \approx 0$. This is an indication that at a low temperature such as $\tilde{T} = 0.02$ the intermediate regimes having $\langle x \rangle \approx 0$, shown in Fig. 6.2 between the α-helix and β-hairpin regimes, contain a mixing of α-helix and β-sheet conformations. The comparable probabilities of the occurrence of both α-helix and β-hairpin represent a frustration in the energy landscape — the system cannot clearly choose which conformation should be the native structure. The window for such a frustrated system is restricted to the vicinity of η_c and is small. Interactions in real proteins, between residues that produce an equivalent effective η_c in our model, would rarely fall into this window in their normal condition. However, there are cases where two energy minima, one corresponding to a normal α-content dominated $\mathrm{PrP^C}$ and the other to a β-content dominated $\mathrm{PrP^{Sc}}$, are believed to be the cause of conformational conversion in prion proteins [3,4,87]. There is also some evidence that some proteins are intrinsically unstructured by design and hence a system of η_c may carry some advantage on its own [88].

The free energy maps for the 14mers are qualitatively similar to those of the 13 mers. There are, however, a few differences in the maps. First, at $\eta = 0$, the

lower free energy area expands toward the native β-conformation regime $x = -1$. This physical pictures are also observed in the state diagrams Fig. 6.2. Second, at $\eta = \eta_c$, the β-conformation becomes more dominant than the α-conformation, and consequently there is no free energy bridge connecting β- and α-conformations that existed for in some 13mer systems (Fig. 6.2 $\eta = \eta_c$ of 13mers and 14mers).

Frustration in a protein arises when the local propensity and the environmental condition does not match. If the local propensity overwhelmingly surpasses or matches the environmental condition then the sequence folds into the structure governed by their propensity without frustration and it is called a well-behaved system. If the environmental condition overwhelmingly surpasses the local propensity then the local residues should settle to the conformation imposed by the environment condition. If, however, the local propensity does not match the environment condition, neither parts can overwhelm the other, then the system remains frustrated and this systems is called an equivocal sequence. It is known that the secondary structure of the well-behaved sequence is easily predicted in knowledge based prediction algorithms [45,91–93], but often these algorithms fail on equivocal sequences because they only explicitly consider the interaction between local amino acids close in sequence but not global interactions. For a primary sequence to be a fast folder and to have stable native conformation, it is advocated that the sequence must satisfy the "principle of minimal frustration" [94–96]. A system which undergoes the conformational interconversion is far from the minimally frustrated state. The frustration in our systems can be increased or reduced by perturbation on the hydrogen bond strength and the hydrophobic interaction at various positions and the present model is suitable to examine this type of a frustrated system.

6.4.5 Folding rate

When two conformations have same energies in systems, which correspond the systems at η_c, it is not clear which conformation is dominant. It became clear only after the phase maps of these systems were drawn. The stability of the β-hairpin and α-helix are not same in the systems at η_c. The stability depends on the presence of the hydrophobic pair and its position, and also the number of mutated residues.

Probability (%)	$\tilde{T}$	Sequence D13 S0	S2	S3	S5
P_β	0.09	58.7	72.0	88.0	86.7
	0.10	48.0	68.3	87.0	80.8
	0.11	44.2	63.3	87.3	73.0
	0.12	**39.7**	**59.8**	**83.5**	**68.7**
	0.13	42.8	64.3	88.7	61.3
	0.14	44.0	64.8	82.3	59.5
	0.15	47.3	78.0	73.7	53.0
P_α	0.09	41.3	28.0	12.0	13.3
	0.10	52.0	31.8	13.0	19.3
	0.11	55.8	36.8	12.8	27.0
	0.12	**60.3**	**40.2**	**16.5**	**31.3**
	0.13	57.2	35.6	11.3	38.7
	0.14	56.0	35.3	17.8	40.5
	0.15	52.7	22.0	26.3	47.0

Table 6.2: The probability to reach the α or β conformation in D13: P_α and P_β denote the probability to reach the α or β conformation starting from a random coil for the first time, respectively.

Probability (%)	$\tilde{T}$	Sequence T14 S0	S1	S3	S5
P_β	0.09	69.7	63.5	81.3	82.7
	0.10	63.7	70.7	87.3	84.4
	0.11	65.7	63.3	85.7	82.3
	0.12	**65.3**	**65.3**	**89.7**	**76.3**
	0.13	63.7	68.0	95.0	74.3
	0.14	70.3	73.3	90.0	73.7
	0.15	82.3	84.3	90.3	77.0
P_α	0.09	30.3	36.5	18.7	17.3
	0.10	36.3	29.3	12.7	15.6
	0.11	34.3	36.7	14.3	17.7
	0.12	**34.7**	**34.7**	**10.3**	**23.7**
	0.13	36.3	32.0	5.0	25.7
	0.14	29.7	26.7	10.0	26.3
	0.15	17.7	15.7	9.7	23.0

Table 6.3: The probability to reach the α or β conformation in T14.

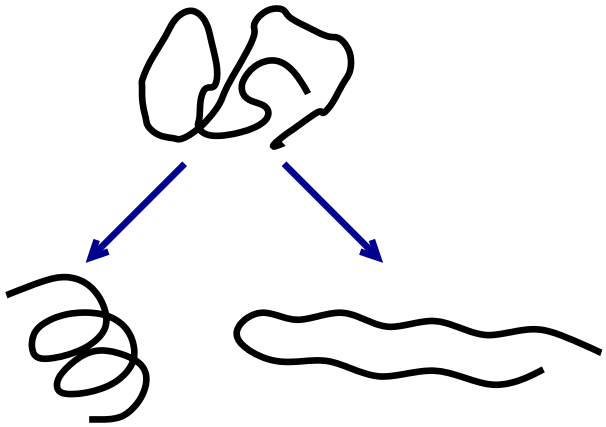

Figure 6.6: The schematic diagram of conformational transition from a random-coil state to the α- or β-conformation.

Now we like to determine which conformation is more kinetically accessible. If we consider their energies only, either one of two ground state conformations can be equally accessible. Considering these results, we expect that kinetic accessibility of these two conformations depends on the position of the hydrophobic pair.

To study the kinetic accessibility of these two conformations, two types of folding rates are examined One is the average folding probability and time from a random conformation to the β-hairpin and the α-helix denoted by P_β and τ_β and P_α and τ_α, respectively. Another is the average interconversion time from β-hairpin to α-helix and from α-helix and β-hairpin denoted by $\tau_{\alpha\to\beta}$ and $\tau_{\beta\to\alpha}$, respectively. The latter is to calculate the transition time between two conformations explicitly since the folding time τ_β and τ_α do not explicitly quantify the interconversion transition time between the two conformations while the probability can quantify the time scale of the relative interconversion time.

Folding rate from a random coil to a β-hairpin or an α-helix

Overall, in a non-hydrophobic chain D13S0, the folding rate to an α-helix is higher than that of a β-hairpin and in other hydrophobic systems, the folding probability to a β-hairpin is higher than in the non-hydrophobic system. This indicates a significant shift in the kinetic folding rate in the presence of a hydrophobic pair. Temperature dependency of the folding probability is complex and reflects the entangled phase composition of the system at η_c shown in the state diagrams in Fig. 6.2 and also in the free energy maps in the free energy maps in Fig. 6.5.

We now consider the folding rate in more details. For D13S0, the α-helix and β-hairpin is first reached in 60% and 40% of all cases, respectively, and this is similar to the results which we have observed in the previous analysis on a similar non-hydrophobic chain [37]. The α-helix is always reached first in this system except at lower temperatures (Table 6.2). In D13S2, the β-hairpin is first reached more frequently while P_α is comparably large compared to other hydrophobic systems. Its hydrophobic pair contributes to stabilize both α-helix and β-hairpin and also it reduces the populations of intermediate conformations since few mixed conformational states are seen in the phase map of D13S2. This results in a shorter folding

time τ_α and τ_β because the systems are not trapped in mixed non-specific compact conformations. In D13S3 and D13S5, P_β increased due to the hydrophobic pair that contributes to lower the energy of the β-hairpin but not the α-helix. In more than 80% of the cases, the β-hairpin is first reached in D13S3 and 68.7% in D13S5. The domination of the β-hairpin is apparent from the phase map of D13S3 and D13S5 (Fig. 6.2). The folding time for both conformations in D13S3 and D13S5 increase because there are more intermediate conformations that prevent the chain to reach either the α-helix or the β-hairpin.

In T14 systems, β-hairpin are more dominant in the systems for η_c. P_β is larger in all the T14 systems because of the increased number of the perturbed residues in the middle that severally disrupt the helical bonds. In the T14 systems, the β-hairpin is reached first in more than 65% of the cases and further the β-hairpin is reached first in all temperatures.

In the current systems, the β-hairpin is the more kinetically accessible conformation, whereas the previous study [42] showed that a β-hairpin conformation is kinetically difficult to reach and conformations with helix contents are more accessible. These differences arise because the perturbative mechanisms for the conformational conversion in the two systems are different. In the current systems, α-helical bonding was disrupted in the middle, which makes the propagation of the helical conformation difficult. The hydrophobic interaction stabilizes the β-hairpin and make it more accessible, but it does not stabilize the α-helix except for D13S2 and T14S1, in which the hydrophobic interaction contributes stabilizing both of the conformations. On the other hand, in the previous study [42], a β-hairpin was induced by the interaction between the attractive wall and the chain system. Helical hydrogen bonds at the middle of the chain were not disturbed and consequently helical conformations was more accessible, even when the true ground state was the β-hairpin attached to the wall.

Interconversion time between a β-hairpin and α-helix

In the previous section, we saw the probabilities that the system reaches either the β- or α-conformation, and from this probabilities we expect that if P_β is higher

Transition Time (10^4 MCS)		$\tilde{T}$	Sequence S0	S2	S3	S5
D13	$\tau_{\alpha\to\beta}$	0.12	58.0	34.7	81.4	105.7
	$\tau_{\beta\to\alpha}$	0.12	41.2	58.4	765.9	147.0
			S0	S1	S3	S5
T14	$\tau_{\alpha\to\beta}$	0.12	53.0	34.6	55.7	93.9
	$\tau_{\beta\to\alpha}$	0.12	83.3	83.1	1263.9	359.2

Table 6.4: The average transition time between two conformations: $\tau_{\alpha\to\beta}$ and $\tau_{\beta\to\alpha}$ denote the average transition time from β-hairpin to α-helix and from α-helix to β-hairpin, respectively.

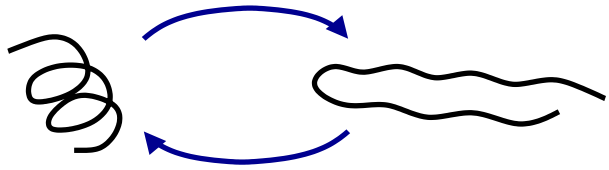

Figure 6.7: The transition between an α-helix and a β-hairpin.

than P_α then $\tau_{\alpha\to\beta}$ should be shorter than $\tau_{\beta\to\alpha}$ because the higher P_β indicates that the β-hairpin is reached faster than the α-helix. The transition rates in Table 6.4 qualitatively agree with the relative transition time scale that can be deduced from the corresponding probabilities in Table 6.2 and 6.3.

In D13S0, $\tau_{\beta\to\alpha}$ is shorter than $\tau_{\alpha\to\beta}$ and this reflects that P_α is larger than P_β — an α-helix is a more easily accessible conformation than a β-hairpin at $\tilde{T} = 0.12$. In D13S2, $\tau_{\beta\to\alpha}$ is slightly longer than $\tau_{\alpha\to\beta}$. In D13S3 and D13S5, however, $\tau_{\alpha\to\beta}$ is much shorter than $\tau_{\beta\to\alpha}$ which agrees with the expected interconversion rate implied by the fact that P_β is substantially larger than P_α. In T14 systems, a β-hairpin becomes more stable due to the ease of formation of the turn because of less angle potential constrain in the turn section, and an α-helix is more perturbed in the helical hydrogen bonding. $\tau_{\alpha\to\beta}$ is shorter in most systems and $\tau_{\beta\to\alpha}$ becomes longer than that of the D13 systems. The dependence in the transition time on the

position of a hydrophobic pair is similar to the D13 systems. The results for the interconversion are in good agreement with the physical picture we obtained from the phase diagram maps, free energy maps, folding time and folding probability.

Inspecting these results more closely, we see that the interconversion rate between a β-hairpin and an α-helix is influenced by mainly three factors. First, the stability of the starting native conformation — the chain can be trapped around the starting conformation. Second, intermediate conformations are more stable than the two native conformations so that the chain is trapped in the intermediate conformational basin. Third, the accessibility and stability of the target native conformation. To see which factors are influencing the interconversion time, we have also examined the dynamic free energy maps of the D13 systems (data not shown) that are similar to the ones extensively used in chapter 4 on the β-hairpin folding. The transition time $\tau_{\beta\to\alpha}$ is large in D13S3 because the compact globular conformation is highly stable and spends most of its time being trapped in these conformations and fluctuating around without reaching the α-helix. On the other hand in D13S5, $\tau_{\beta\to\alpha}$ is relatively large for different reasons: the chain fluctuates around the conformations close to the β-hairpin rather than being trapped in the random conformations. As for the transition time $\tau_{\alpha\to\beta}$, $\tau_{\alpha\to\beta}$ of the D13S3 and D13S5 are larger than the D13S0 and D13S2 systems. This is partly because the α-conformation is more stable because the values of η is larger in these systems which results in stabilizing the helical conformation.

In Chapter 4, we found that in the hydrophobic system, a system whose hydrophobic pair is placed closer to the turn folds to a β-hairpin from a random coil faster rate, and we observed similar trends in the folding rate $\tau_{\alpha\to\beta}$ of the D13 and T14 systems. The folding rate was measured differently in Chapter 4 and this chapter so the direct comparison of these folding rates is not simple, but the results here suggest that the main folding mechanisms would not be completely altered in the presence of perturbation in the middle of the chain, and the folding rate is more influenced by the position of the hydrophobic pair. These folding rates obtained in this section and the results obtained in Chapter 4 [50] show that conformational preference significantly change one over another owing to the position of the hydrophobic pair and the number of the perturbed hydrogen bondings in the

middle.

6.5 Conclusion

We have demonstrated that simple peptide systems can be designed to form a β-hairpin and an α-helix depending on the strength of perturbation of the hydrogen bonds in the middle of the chains, depending on the position of a hydrophobic pair and depending on reduced temperature. We illustrated by using the thermodynamic phase maps that conformational conversions between a β-hairpin and an α-helix can be induced by various changes in a system itself and/or by in its environmental conditions. These modelled systems provide unique physical insights to the conformational conversion observed in experiments [32] and also in similar conversions in various contexts, including protein folding, misfolding and conformational switches. In addition we have investigated the kinetic folding rates from a random coil to a β-hairpin or an α-helix in the systems with η_c where the β-hairpin and α-helix have same energies. We found that the folding probability to a β-hairpin is higher in the most systems except in the 13mer non-hydrophobic sequence. In addition the transition time from an α-helix to a β-hairpin is faster than the transition time from a β-hairpin to an α-helix.

Bibliography

[1] Waterhous DV, Johnson WC Jr. Importance of environment in determining secondary structure in proteins. Biochemistry. 1994;33:2121-2128

[2] Dobson CM. The structural basis of protein folding and its links with human disease. Philos Trans R Soc London Ser B 2001;356:133-145.

[3] Prusiner SB. Molecular biology and genetics of prion disease. Philos Trans R Soc London, Ser. B 1994;343:447-463.

[4] Jackson GS, Clarke AR. Mammalian prion proteins. Curr Opin Struct Biol. 2000;10:69-74.

[5] Haber E, Anfinsen CB. Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. J Biol Chem 1961;236:422-424.

[6] Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223-230.

[7] Bornberg-Bauer E, Chan HS. Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space. Proc Natl Acad Sci USA 1999;96:10689-10694.

[8] Nelson ED, Onuchic JN. Proposed mechanism for stability of proteins to evolutionary mutations. Proc Natl Acad Sci USA 1998;95:10682-10686.

[9] Damaschun G, Damaschun H, Gast K, Zirwer D. Proteins can adopt totally different folded conformations. J Mol Biol 1999;291(3):715-725.

[10] Grishin NV. Fold change in evolution of protein structures. J Struct Biol 2001;134:167-185. Kinch LN, Grishin NV. Evolution of protein structures and functions. Curr Opin Struct Biol. 2002;12:400-408.

[11] GD Fasman, Ed., Poly-a-amino acids. Marcel Dekker, New York, 1967.

[12] Mutter M, Hersperger R. Peptides as conformational switch - medium-induced conformational transitions of designed peptides. Angew Chem-Int Ed Engl 1990;29:185-187.

[13] Fukushima Y. Sequence effects on helix-sheet conformational transitions of designed amphiphilic peptides. Bull Chem Soc Japan 1996;69:701-708.

[14] Zhang SG, Rich A. Direct conversion of an oligopeptide from a β-sheet to an α-helix: A model for amyloid formation. Proc Natl Acad Sci USA 1997;94:23-28.

[15] Reed J, Kinzel V. A conformational switch is associated with receptor affinity in peptides derived from the CD4-binding domain of gp120 from HIV I. Biochemistry. 1991;30:4521-4528.

[16] Schenck HL, Dado GP, Gellman SH. Redox-triggered secondary structure changes in the aggregated states of a designed methionine-rich peptide. J Am Chem Soc 1996;118:12487-12494.

[17] Forood B, Perez-Paya E, Houghten RA, Blondelle SE. Formation of an extremely stable polyalanine β-sheet macromolecule. Biochem Biophys Res Commun 1995;211:7-13.

[18] Blondelle SE, Forood B, Houghten RA, Perez-Paya E. Poly-alanine-based peptides as models for self-associated β-sheet complexes. Biochemistry 1997;36:8393-8400.

[19] Levy Y, Jortner J, Becker OM. Solvent effects on the energy landscapes and folding kinetics of polyalanine. Proc Natl Acad Sci USA 2001;98:2188-2193.

[20] Ding F, Borreguero JM, Buldyrey SV, Stanley HE, Dokholyan NV. Mechanism for the α-helix to β-hairpin transition. Proteins 2003;53:220-228.

[21] Nguyen HD, Marchut AJ, Hall CK. Solvent effects on the conformational transition of a model polyalanine peptide. Protein Sci 2004;13:2909-2924.

[22] Castano S, Desbat B, Cornut I, Meleard P, Dufourcq J α-helix to β-sheet transition within the Leu(i)Lys(j) (i=2j) series of lytic amphipathic peptides by decreasing their size. Lett Pept Sci 1997;4:195-200.

[23] Cordes MHJ, Burton RE, Walsh NP, McKnight CJ, Sauer RT. An evolutionary bridge to a new protein fold Nat Struct Bio 2000 7 1129-1132. Cordes MHJ, Walsh NP, McKnight CJ, Sauer RT. Evolution of a protein fold in vitro. Science 199;284:325-327.

[24] Takahashi Y, Yamashita T, Ueno A, Mihara H. Construction of peptides that undergo structural transition from α-helix to β-sheet and amyloid fibril formation by the introduction of N-terminal hydrophobic amino acids. Tetrahedron 2000;56:7011-7018.

[25] Mihara H, Takahashi Y, Ueno A. Design of peptides undergoing self-catalytic α-to-beta transition and amyloidogenesis. Biopolymers 1998;47:83-92.

[26] Minor DL, Kim PS. Context-dependent secondary structure formation of a designed protein sequence. Nature 1994;367:660-663.

[27] Sethuraman A, Vedantham G, Imoto T, Przybycien T, Belfort G. Protein unfolding at interfaces: Slow dynamics of α-helix to β-sheet transition. Proteins 2004;56:669-678.

[28] Hamada D, Segawa S, Goto Y. Non-native α-helical intermediate in the refolding of β-lactoglobulin, a predominantly β-sheet protein. Nat Struc Bio 1996;3:868-873.

[29] Vendruscolo M, Zurdo J, MacPhee C, Dobson CM. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. Philos. Trans. R. Soc. Lond. Ser. A 2003;365:1205-1222.

[30] Dobson CM. Protein folding and misfolding. Nature 2003;426:884-890.

[31] Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, Dobson CM. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc Natl Acad Sci USA 1999;96:3590-3594.

[32] Awasthi SK, Shankaramma SC, Raghothama S, Balaram P. Solvent-induced β-hairpin to helix conformational transition in a designed peptide. Biopolymers 2001;58:465-476.

[33] Fersht A. Structure and mechanism in protein science. New York: WH Freeman, 1999.

[34] Creighton TE. Proteins. New York: W H Freeman; 1993. p507.

[35] Fersht AR, Serrano L. Principles of protein stability derived from protein engineering experiments. Curr Opin Struct Biol 1993;3:73-83.

[36] Cornish VW, Schultz PG. A new tool for studying protein structure and function. Curr Opin Struct Biol 1994;4:601-607.

[37] Imamura H, Chen JZY. Conformational conversion of proteins due to mutation. Europhys Lett 2004;67:491-497.

[38] Gō N. Theoretical studies of protein folding. Ann Rev Biophy BioEng 1983;12:183-210. Miyazawa S, Jernigan RL. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. Macromolecules 1985;18:534-552.

[39] Peng Y, Hansmann UHE. Helix versus sheet formation in a small peptide. Phys Rev E 2003;68:No.041911

[40] Daidone I, Simona F, Roccatano D, Broglia RA, Tiana G, Colombo G, Di Nola A. β-hairpin conformation of fibrillogenic peptides: Structure and $\alpha - \beta$ transition mechanism revealed by molecular dynamics simulations. Proteins 2004;57:198-204.

[41] Ikeda K, Higo J. Free-energy landscape of a chameleon sequence in explicit water and its inherent $\alpha\beta$ bifacial property. Protein Sci 2003;12:2542-2548.

[42] Chen JZY, Lemak AS, Lepock JR, Kemp JP. Minimal model for studying prion-like folding pathways. Proteins 2003;51:283-288.

[43] Head-Gordon T, Brown S. Minimalist models for protein folding and design. Curr Opin Str Bio 2003;13:160-167.

[44] Kolinski A, Skolnick J. Reduced models of proteins and their applications. Polymer 2004;45:511-524.

[45] Chou PY, Fasman GD. Empirical predictions of protein conformation. Annu Rev Biochem. 1978;47:251-276.

[46] O'Neil KT, DeGrado WF. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science. 1990;250:646-651.

[47] Chakrabartty A, Kortemme T, Baldwin RL. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci 1994;3:843-852.

[48] Malolepsza E, Boniecki M, Kolinski A, Piela L. Theoretical model of prion propagation: A misfolded protein induces misfolding. Proc Natl Acad Sci USA 2005;102:7835-7840.

[49] Hoang TX, Trovato A, Seno F, Banavar JR, Maritan A. Geometry and symmetry presculpt the free-energy landscape of proteins. Proc Natl Acad Sci USA 2004;101:7960-7964. Geometrical model for the native-state folds of proteins. Hoang TX, Trovato A, Seno F, Banavar JR, Maritan A. Biophys Chem 2005;115:289-294.

[50] Imamura H, Chen JZY. Dependence of folding dynamics and structural stability on the location of a hydrophobic pair in β-hairpins. (Submitted.)

[51] Chen JZY, Imamura H. Universal model for α-helix and β-sheet structures in protein. Physica A 2003;321:181-188.

[52] Honeycutt JD, Thirumalai D. The nature of folded states of globular proteins. Biopolymer 1992;32:695-709.

[53] Klimov DK, Thirumalai D. Mechanisms and kinetics of β-hairpin formation. Proc Natl Acad Sci USA. 2000;97:2544-2549.

[54] Klimov DK, Thirumalai D. Stiffness of the distal loop restricts the structural heterogeneity of the transition state ensemble in SH3 domains. J Mol Biol 2002;317:721-737.

[55] Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a de novo designed four helix bundle protein. J Mol Biol 1996;263:323-343.

[56] Klimov DK, Betancourt MR, Thirumalai D. Virtual atom representation of hydrogen bonds in minimal off-lattice models of α helices: effect on stability, cooperativity and kinetics. Fold Des 1998;3:481-496.

[57] Shea JE, Nochomovitz YD, Guo ZY, Brooks CL. Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist β-barrel model. J Chem Phys 1998;109:2895-2903.

[58] Sorenson JM, Head-Gordon T. Matching simulation and experiment: A new simplified model for simulating protein. folding. J Comp Biol 2000;7:469-481.

[59] Kemp JP, Chen ZY. Formation of helical states in wormlike polymer chains. Phys Rev Lett 1998;81:3880-3883.

[60] Knott M, Chan HS. Exploring the Effects of Hydrogen Bonding and Hydrophobic Interactions on the Foldability and Cooperativity of Helical Proteins Using a Simplifed Atomic Model. Chem Phys 2004;307:187-199.

[61] Favrin G, Irback A, Wallin S. Folding of a small helical protein using hydrogen bonds and hydrophobicity forces. Proteins 2002;47:99-105.

[62] Dill KA. Dominant Forces in Protein Folding. Biochemistry 1990;29:7133-7155.

[63] Deechongkit S, Nguyen H, Powers ET, Dawson PE, Gruebele M, Kelly JW. Context-dependent contributions of backbone hydrogen bonding to β-sheet folding energetics. Nature 2004;430:101-105.

[64] Wood SJ, Wetzel R, Martin JD, Hurle MR. Prolines and Aamyloidogenicity in Fragments of the Alzheimer's Peptide β/A Biochemistry 1995;34:724-730.

[65] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087-1092.

[66] Hansmann UHE, Okamoto Y. Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. Physica A 1994;212:415-437.

[67] Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg J. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comp Chem 1992;13:1011-1021. Lemak AS, Gunn JR. Rotamer-specific potentials of mean force for residue pair interactions. J Phys Chem B 2000;104:1097-1107.

[68] Binder K. Monte Carlo simulation in statistical physics. New York: Springer-Verlag, 1988, p127. Newman MEJ, Barkema GT. Monte Carlo methods in statistical physics. New York: Oxford University Press, 1999,

[69] Chan HS, Dill KA. Transition states and folding dynamics of proteins and heteropolymers. J Chem Phys 1994;100:9238-9257.

[70] Kikuchi K, Yoshida M, Maekawa T, Watanabe H. Metropolis Monte Carlo method as a numerical technique to solve the Fokker-Planck equation. Chem Phys Lett 1991;185:335-338.

[71] Rey A, Skolnick J. Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins. Chem Phys 158:199-219, 1991.

[72] Chubykalo O, Nowak U, Smirnov-Rueda R, Wongsam MA, Chantrell RW, Gonzalez JM. Monte Carlo

[73] Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. San Diego: Academic Press, 1996. p443. Mountain R, Thirumalai D. Quantitative measure ofefficiency of Monte Carlo simulations. Physica A 1994;210:453-460.

[74] Cerpa R, Cohen FE, Kuntz ID. Conformational switching in designed peptides: the helix/sheet transition. Fold Des 1996;1:91-101.

[75] Dzwolak W, Smirnovas V. A conformational α-helix to β-sheet transition accompanies racemic self-assembly of polylysine: an FT-IR spectroscopic study. Biophys Chem 2005;115:49-54.

[76] Spach G, Brack A. Multiconformational synthetic polypeptides. J Am Chem Soc 1981;103:6319-6323.

[77] Welch WH Jr, Fasman GD. Hydrogen-tritium exchange in polypeptides. Models of α-helical and β conformations. Biochemistry. 1974;13:2455-2466.

[78] Davidson B, Fasman GD. The conformational transitions of uncharged poly-L-lysine. α helix-random coil-β structure. Biochemistry 1967;6:1616-1629.

[79] Balaram P. Non-standard amino acids in peptide design and protein engineering. Curr Opin Struct Biol 1992;2:845-850.

[80] Mendel D, Ellman JA, Chang Z, Veenstra DL, Kollman PA, Schultz PG. Probing protein stability with unnatural amino acids. Science. 1992;256:1798-802.

[81] Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. A general method for site-specific incorporation of unnatural amino acids into proteins. Science. 1989;244:182-188.

[82] Aravinda S, Shamala N, Das C, Balaram P. Structural analysis of peptide helices containing centrally positioned lactic acid residues. Biopolymers 2002;64:255-267.

[83] Karle IL, Das C, Balaram P. Effects of hydrogen-bond deletion on peptide helices: Structural characterization of depsipeptides containing lactic acid. Biopolymers 2001;59:276-289.

[84] Dill KA, Sarina B. Molecular driving forces. New York:Garland Science, 2003, p666.

[85] Kemp JP, Chen JZY. Formation of helical states in wormlike polymer chains. Phys Rev Lett 1998;81:3880-3883.

[86] Kemp JP, Chen JZY. Helical structures in proteins. Biomacromolecules 2001;2:389-401.

[87] Kelly JW. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Curr Opin Struct Biol 1998;8:101-106. Harrison PM, Chan HS, Prusiner SB, Cohen FE. Conformational propagation with prion-like characteristics in a simple model of protein folding. Protein Sci. 2001;10:819-835. Carrell RW, Gooptu B. Conformational changes and disease - serpins, prions and Alzheimer's. Curr Opin Struct Biol 1998;8:799-809.

[88] Tompa P. Intrinsically unstructured proteins evolve by repeat expansion. Bioessays 2003;25:847-55.

[89] Pathria RK. Statistical mechanics. Boston: Butterworth-Heinemann, 1996, p529.

[90] Finkelstein AV, Ptitsyn OB. Protein physics. San Diego: Academic Press, 2002, p354.

[91] Burgess AW, Ponnuswamy PK. and Scheraga HA. Analysis of conformations of amino acid residues and prediction of backbone topography in proteins. Isr J Chem 1974;12:239-286.

[92] Lim VI. Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure. J Mol Biol 1974;88:857-872.

[93] Garnier J, Osguthorpe DJ, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 1978;120:97-120.

[94] Socci ND, Onuchic JN, Folding kinetics of proteinlike heteropolymersJ. Chem Phys 1994;101:1519-1528.

[95] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167-195.

[96] Bryngelson JD, Wolynes PG. Spin Glasses and the Statistical Mechanics of Protein Folding. Proc Natl Acad Sci USA 1987;84:7524-7528.

Chapter 7

Conclusion

7.1 Summary

We have developed a simple protein model by improving a simple polymer model that was designed for a helix conformation [1]. The current model can accommodate the α- and β-conformations, and therefore the model can illustrate a wide range of conformations in proteins such as helices, helix bundles, β-hairpins, β-sheets of various sizes and also a combination of these conformations as a ground state of a given sequence. These conformations were described in Chapter 3.

In Chapter 4, the folding mechanism of a β-hairpin was discussed. β-sheet conformations in proteins are almost as abundant as α-helix conformations but the experimental and theoretical studies on β-sheet conformations had not been performed as intensely as these studies on α-helices until Muñoz et al. [2] provided the experimental folding kinetics for a β-hairpin system. Experiments cannot yet characterize the microscopic folding mechanisms of β-hairpins, hence currently only computer simulations can provide physical insights of the microscopic folding mechanisms. No simulation has exhaustively simulated the impact of the hydrophobic pair in complete folding processes so far. We have investigated the folding mechanism of β-hairpins by the MC dynamics and found that the hydrophobic pair significantly affect the folding mechanism, and also found that the folding time dependence on the hydrophobic pair position agrees with experiments and theoretical

prediction, but further experimental investigations are required to determine their dependence conclusively.

In Chapter 5 and 6, we discussed the conformational conversion between a β-hairpin and an α-helix conformation induced by the perturbation on the hydrogen bonds at the middle of the chains and by the insertion of hydrophobic pair at a various position. The phase diagrams showed that the conformational conversion is possible by various perturbative effects such as the change in the effective hydrogen bonding strength at the middle part of the chain, the change in the hydrophobic pair location, the change in the temperature, and change in the number of the perturbed residues in the middle. The closely related conformational change was observed by Baralam's group in their experiments [3]. Conformational conversions in proteins discussed in thesis are considered to be closely related to the conformational diseases such as Alzheimer's and prion diseases for which currently no adequate treatment has been developed [4].

The current model developed in the thesis can now explore biologically relevant phenomena. It should be possible to improve the model by implementing some additional factors into the model. First, weak, general dihedral potentials may reduce the conformational space, and overall folding behavior of a system in the model may improve. Reference independent dihedral potentials are often used in medium and high resolution models. Second, introducing the variation in residue diameters, reflecting the geometric steric constrain, may provide better packing and hence stabilize native state while destabilize other nonnative conformations. Third, a side chain can be also incorporated a monomer. This can be achieved either by a virtual type (non-excluded volume by itself) or a normal residue type approach. Fourth, Miyazawa-Jarnigan type potentials — knowledge based residue specific contact interactions — make it possible to investigate residue specific mutational effect. It is worth while to try to incorporate these factors into the model for improvement. The design of a model is depends on physical and time scale of a system that one is interested in therefore it is possible to adjust many parameters to a desirable model for a given problem. Unfortunately, there is no reference that covers the details of various peptide models and one often has to check its original paper for a specific model.

7.2 Future work

There are many phenomena that remain to be answered in protein folding and biological processes. The current model, with some additions and modifications, should be able to investigate the following problems.

1. Proteins interacting with a hydrophobic wall: An attractive wall was demonstrated to induce the conversion from an α- to a β-conformation [5]. To explore this process further we can simulate the conformational conversion on the system discussed in Chapter 5 and 6 in the presence of an attractive wall. This simulation should yield different phase diagram obtained in the previous paper [5] and provide a further insight to the conversion mechanism.

2. Chaperon proteins: Some large chaperon molecular such as a GroEL-GroES complex perform an annealing process that provides the environment for the nascent protein chain to reconfigure misfolded configuration. How does the chaperon identify the misfolded conformation? A misfolded conformation or unfolded are recognize by their exposed hydrophobic areas [8] and the GroEL cavity can change the properties of the inner wall between hydrophobic and hydrophilic. When the inner wall becomes hydrophobic, a conformation exposing the hydrophobic residue is captured by the wall of the chaperon, and therefore misfolded/unfolded chains will stay in the chaperon as long as their hydrophobic core is exposed to the surrounding. Once the hydrophobic core is properly concealed and the chains have properly reached the native conformation, they are released from the chaperon. To simulate this protein folding process in a chaperon-like cage, we can turn on and off the hydrophobic attraction on the wall of a spherical cavity and see if this process helps misfolded chain to reach the native conformation without being trapped in misfolded states.

3. Chain aggregation and prion propagation: There are competing interactions when more than one chain is present in a system; inter and intra chain interactions. Aggregation in proteins is considered to be a key factor in the amyloid formation [6]. The simple model that includes the chain stiffness

and electrostatic interaction was used to attempt to explain images obtained by the atomic force microscopy technique along with a polymer kinetic theory [7]. Similar interaction between peptide chains is observed in prion proteins. Prion proteins undergo the conformational conversion by autocatalytic process but the conformational conversion can be induced by other infectious form of prion proteins. To study this prion propagation [9], we can insert a rigid seed like β-hairpin structure among α-helices and test if we can see the propagation of the β-structure either intra or/and inter chain interaction.

Significant progress has been achieved in experimental and theoretical works of protein folding in recent years. We hope simple models will continue to provide valuable insights in many aspects of protein folding and also biological systems. We believe that the models discussed here will find wider applications in proteins and other peptide systems.

Bibliography

[1] Kemp JP, Chen JZY. Formation of helical states in wormlike polymer chains. Phys Rev Lett 1998;81:3880-3883.

[2] Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of β-hairpin formation. Nature 1997;390:196-199.

[3] Awasthi SK, Shankaramma SC, Raghothama S, Balaram P. Solvent-induced beta-hairpin to helix conformational transition in a designed peptide. Biopolymers 2001;58:465-476.

[4] Dobson CM, The structural basis of protein folding and its links with human disease, Philos Trans R Soc London Ser B 2001;356:133-145.

[5] Chen JZY, Lemak AS, Lepock JR, Kemp JP. Minimal model for studying prion-like folding pathways. Proteins 2003;51:283-288.

[6] Kelly JW.] The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Curr Opin Struct Biol 1998;8:101-106.

[7] Jun S, Hong Y, Imamura H, Ha BY, Bechhoefer J, Chen P. Self-assembly of the ionic peptide EAK16: The effect of charge distributions on self-assembly. Biophys J 2004;87:1249-1259.

[8] Mayhew M, da Silva ACR, Martin J, Erdjument-Bromage H, Tempst P, Hartl FU. Protein folding in the central cavity of the GroEL-GroES chaperonin complex. Nature 1996;379:420-426.

[9] Prusiner SB. Molecular biology and genetics of prion disease. Philos Trans R Soc London Ser B 1994;343:447-463.

Publications

"Universal model for α-helix and β-sheet structures in protein." Chen JZY, Imamura H. Physica A 2003;321:181-188.

"Conformational conversion of proteins due to mutation." Imamura H, Chen JZY. Europhys Lett 2004;67:491-497.

"Dependence of folding dynamics and structural stability on the location of a hydrophobic pair in β-hairpins." Imamura H, Chen JZY. Accepted to Proteins.

"Conformational conversion due to mutation and a hydrophobic pair." Imamura H, Chen JZY. (To be submitted)

"Self-assembly of the ionic peptide EAK16: The effect of charge distributions on self-assembly." Jun S, Hong Y, Imamura H, Ha BY, Bechhoefer J, Chen P. Biophys J 2004;87:1249-1259.

"On the excitation of large aspect angle Farley-Buneman echoes via three-wave coupling: A dynamical system model." Hamza AM, Imamura H. J Geophy Res Space Phys 2001;106:24745-24754.