Theor Chem Acc (2006) 116: 262–273
DOI 10.1007/s00214-005-0041-9

REGULAR ARTICLE

Andrij Baumketner · Joan-Emma Shea

The thermodynamics of folding of a β hairpin peptide probed through replica exchange molecular dynamics simulations

Received: 7 April 2004 / Accepted: 30 August 2005 / Published online: 17 December 2005

Abstract Peptides that possess a well defined native state are ideal model systems to study the folding of proteins. They possess many of the complexities of larger proteins, yet their small size renders their study computationally tractable. Recent advances in sampling techniques, including replica exchange molecular dynamics, now permit a full characterization of the thermodynamics of folding of small peptides. These simulations not only yield insight into the folding of larger proteins, but equally importantly, they allow, through comparison with experiment, an objective test of the accuracy of force fields, water models and of different numerical schemes for dealing with electrostatic interactions. In this account, we present a molecular dynamics simulation of a small β-hairpin peptide using the replica exchange algorithm and illustrate how this enhanced sampling scheme enables a thorough characterization of the native and unfolded states, and sheds new light into its folding mechanism.

Keywords Protein folding · β-Hairpin · Replica exchange molecular dynamics · Thermodynamics of folding · Simulations

1 Introduction

Protein folding refers to the process by which a polypeptide chain reaches its biologically active three dimensional structure from a linear chain of amino acids. It is one of the most important processes in biology, yet despite nearly 40 years of experimental and theoretical research, the mechanisms and underlying principles behind folding remain incompletely understood. Simulations offer unique insight into folding by providing a microscopic description of the process and atomistic details unavailable from experiments. Simulations are challenging due to the large number of atoms present in biomolecular systems, coupled with the inherent complexity of the underlying multi-dimensional energy landscape [1–3]. The success of numerical simulations hinges on a number of elements, including (a) a sufficiently accurate representation of the protein and solvent and (b) efficient algorithms that allow proper exploration of the conformational space. Force fields used to describe proteins are continually being improved [4,5], and while still imperfect, they have been shown to adequately reproduce both structural features as well as the dynamics of polypeptide chains ranging from globular proteins [6–8] to di-, tri-peptide systems [9–11], and even including amino acid analogs [12]. In recent years, great strides have been made in the development of novel computer algorithms for biomolecular simulations [13,14]. Because of the high energy barriers and deep minima present in the energy landscape of most biomolecules, conventional molecular dynamics simulations are inadequate to yield a satisfactory sampling of conformational space. Insufficient sampling will lead to an incorrect statistical picture of conformational space and possibly to wrong conclusions being drawn about both folding mechanisms and conformational preferences. Hence much of the effort in algorithmic development has focused on the design of efficient sampling techniques that allow for accelerated equilibration of systems (such as biomolecules) with complex potential energy surfaces and slow relaxation times. While a complete characterization of the folding of large, solvated, atomically-detailed proteins remains a formidable challenge, considerable insight into the folding process can be gleaned from the study of peptides [15,16]. Peptides, in particular designed peptides with unique ground states, possess several features in common with proteins. While complex enough to provide a viable model for the statistical properties of proteins with longer sequences, they remain simple enough to be computationally tractable. Recent enhanced sampling techniques have allowed a thorough exploration of the thermodynamics of

A. Baumketner
Institute for Condensed Matter Physics,
1 Svientsitsky Str., Lviv 79011, Ukraine,
E-mail: andrij@icmp.lviv.ua

J.-E. Shea (✉) A. Baumketner
Department of Chemistry and Biochemistry,
University of California at Santa Barbara,
Santa Barbara, CA 93106, USA
E-mail: shea@chem.ucsb.edu

folding of peptides [17–19], and, through comparison with experiments, have not only shed light into aspects of folding of larger proteins, but also enabled the testing of the accuracy of force fields [20–22], water models [23], and more recently of different treatments of electrostatic interactions [24].

This account focuses on the folding of peptides, as model systems for protein folding. We begin by reviewing some of the common sampling techniques used for biomolecular systems and then introduce a recent advance in sampling known as replica exchange molecular dynamics [25]. We will apply this technique to study the folding of a small designed peptide, with a β-hairpin native state and illustrate how simulations are now capable of probing both the nature of the folded and unfolded states, as well a shed new insight into folding mechanism.

The earliest sampling techniques involved the use of umbrella potentials [26] combined with molecular dynamics simulations. This approach, pioneered by Brooks III and coworkers in the context of biomolecular simulations, provided the first insight into thermodynamics of folding of peptides and small proteins in explicit solvent [3,27,28]. This importance sampling approach involves running simulations with a quadratic biasing potential in a chosen reaction coordinate for a set of conformations spanning the folded to the unfolded state. Data collected from simulation is reweighted (using for example the weighted histogram analysis method (WHAM) [29,30]) to generate free energy surfaces for folding, as a function of temperature and given reaction coordinates. This method is efficient, as each biased simulation can be run independently. A more sophisticated sampling technique, known as multicanonical sampling [31], employs biasing potentials in the potential energy to "flatten" the energy landscape, leading to a random walk in energy space and allowing exploration of systems where the minima are separated by large energy barriers. Both Monte Carlo and molecular dynamics variants of this method have been used to explore the thermodynamics of folding of a number of biological systems [13,32–34]. The main drawback of this method, which has prevented its widespread use for biological systems, lies in the difficulty in determining the biasing potential. In the last few years, an improvement on the multicanonical method has been introduced which bypasses this problem. This method, known as replica exchange molecular dynamics (REMD) [25,35], was first developed in the context of spin glasses in an implementation using Monte Carlo simulations [36]. Replica exchange (REX) is also referred to as multiple markov chain [37] or parallel tempering [38]. REMD involves the simulation, in parallel, of a number of identical copies (or "replicas") of the original system at different temperatures. Two replicas i and j, adjacent in temperature T_i, T_j with energies E_i and E_j are periodically exchanged with a probability given by the Metropolis criterion:

$$p_{ij} = \begin{cases} 1, & \text{if } \Delta \leq 0 \\ \exp(-\Delta), & \text{if } \Delta > 0 \end{cases}$$

where $\Delta = [(\beta_i - \beta_j)(E_j - E_i)]$, $\beta = \frac{1}{k_B T}$ and k_B is Boltzmann's constant.

Since the escape time from local energy minima decreases significantly at elevated temperatures, the replica exchange method enables enhanced equilibration by treating temperature as a parameter that can change with time. In addition to leading to a more thorough exploration of conformational space, the algorithm also ensures that the conformations sampled at a given temperature belong to the canonical ensemble. This allows application of ensemble reweighting techniques to the simulation data to extract equilibrium thermodynamics quantities.

The replica exchange algorithm is a powerful tool to study the thermodynamics of folding, but care must be taken when interpreting the kinetic information obtained from REX simulations. Indeed, replica exchange simulations generate fictitious dynamics that cannot be directly related to the actual physical time. The interpretation of purely kinetic properties, such as temporal ordering of events, is hence difficult to extract from such simulations. One means of obtaining kinetic informations is by exploiting the fact that the REMD method produces an equilibrium canonical distribution and that free energy maps in appropriate coordinates can be generated and subsequently used to extract information relevant for folding kinetics. For instance, an L-shaped landscape along the radius of gyration and the fraction of native contacts would indicate a time ordering along these parameters in which collapse of the chain precedes ordering to the correct folded state [39]. Maps with two minima linked along the diagonal, on the other hand, suggest concurrent collapse and formation of native structure [3,28,40]. Limitations inherent to the free energy formalism must however be kept in mind when attempting to extract kinetic information from free energy surfaces. In particular, only the free energy minima have sound meaning. The location of the free energy barriers projected onto the chosen reaction coordinates may not correspond to actual kinetic bottlenecks (transition states). In addition, it is critical to ensure local connectivity in the map. In other words, a pathway leading from one point on the map to another must also visit all intermediate points between the two end points.

In what follows, we present a replica exchange investigation of the folding of a small peptide designed by Blanco et al. [41]. This peptide was shown through NMR studies by Constantine [42] and Friedrichs [43] to possess a native β-hairpin structure. Computer simulations on this peptide have been performed by Constantine et al. [44] and by Brooks et al. [45,46]. However, the lengths of the simulations performed were insufficient to yield equilibrium folding thermodynamics for this peptide. Here, we show that using REMD, we can fully characterize the equilibrium thermodynamics for a system in which the peptide is described in atomic detail using the OPLS/AA force field [47] with explicit TIP3P water molecules [48]. We generate free energy surfaces for folding as a function of a number of order parameters and proceed to analyze the nature of the folded and unfolded states. A detailed comparison with experiment is offered, as well as new insights into the mechanism of folding of this peptide.

2 Methods and models

Our simulations focus on the YQNPDGSQA peptide, a mutant of the 15-23 fragment of tendamistat (YQSWRYSQA) [41] which possesses enhanced turn-forming propensity. Experimentally, this peptide exhibits NOEs compatible with β-hairpin structure [42,43].

The peptide was modeled using the OPLS/AA all-atom empirical force-field [47], and solvated in a cubic box of 1149 TIP3P [48] water molecules. A single sodium ion was used to neutralize the -1 charge of the peptide. A total of 3,574 atoms were present in the simulation box. The size of the simulation box, 33.1 Å, was determined from short constant pressure simulations at T $=$ 280 K, equilibrated at a physiological external pressure of 1 atm. Simulations were carried out using GROMACS software [49–51]. Covalent bonds of the water molecules were held constant using the SETTLE algorithm [52] and the peptide bonds involving hydrogen atoms were constrained according to LINCS protocol [53]. This allowed use of a time step of $2\,fs$. Nonbonded Lennard–Jones interactions were tapered starting at 7 Å and extending to a 8 Å cut-off. Neighbor lists for the nonbonded interactions were updated every 10 simulation steps. Electrostatic interactions were included through the particle mesh Ewald (PME) [54] approach. A 8 Å cut-off for the real space force contributions and 1.2 Å grid spacing for the Fourier transform in the reciprocal space were used. Fourth order cubic interpolations were used for off-grid positions. The temperature was controlled by Nose–Hoover algorithm [55] with a 0.05 ps time constant.

The replica-exchange algorithm (REMD) [25,51] was used to enhance sampling and hence improve equilibration of the system. In total, 20 replicas of the original system were considered, at temperatures exponentially spaced between 280 and 540 K. Exchanges of replicas at adjacent replicas were attempted every 500 simulation steps. The same time interval was used to periodically save atomic coordinates. The simulations were started from a random extended conformation, the same one for all replicas, and were run for a total of $60\,ns$. The first $10\,ns$ were treated as an equilibration phase. All the statistical analyses reported here were performed on the last $50\,ns$ of the trajectory. Free energy maps were checked for statistical convergence by generating them over increasingly longer simulation times, 10, $20\,ns$, etc. No noticeable deviations were observed for maps obtained at 40 and $50\,ns$. In addition, maps generated using the first $20\,ns$ of the trajectory are similar to the maps generated from the statistically independent last $20\,ns$ portion of the trajectory. Maps from both parts of the trajectory are similar to the map averaged over the whole trajectory.

The principal component approach (PCA) [56,57] was used to analyze the conformational ensembles generated in the simulations. This method provides the best possible representation of the conformational space for grouping conformations into clusters. Conformations are projected onto a two-dimensional space spanned by the first two eigenvectors of a structural similarity matrix obtained by averaging over the atomic RMS deviations for all structures from a given reference conformation. In the present study, all the peptide atoms were used to construct the RMSD matrix, while the most frequently populated native β-hairpin was used as the reference state. Similar calculations on only the C_α atoms did not change the conclusions of this paper. In addition, a variant of PCA, the principal coordinate analysis [56], was carried out for 1,000 selected conformations recorded at evenly spaced time intervals along the entire trajectory. The matrix of C_α RMS deviations among all pairs of structures was used as the dissimilarity matrix. Results obtained by this method were consistent with those obtained from the straightforward principal component approach.

3 Results

Free energy surfaces obtained from the REMD simulations are plotted as a function of the first two principal components (PC1 and PC2) at T $=$ 280 K and T $=$ 370 K in Fig. 1 a and b. Both surfaces exhibit a global minimum at (PC1 $=$ -4.2, PC2 $=$ -1.0) as well as a number of shallow local minima in other regions of the map. Conformations belonging to the global minimum correspond to native state of the peptide, while local minima are associated with unfolded conformations. The native state is seen to be about 2–3 kT lower in free energy than typical denatured-state conformations irrespective of the temperature considered. The denatured state displays an additional conformational clustering at lower temperatures. The free-energy barriers separating these clusters are on the order of one kT and are thus unlikely to serve as significant kinetic barriers to folding. The native state in our simulations (corresponding to the most populated conformation) has a β hairpin topology (shown in Fig. 2) . It is characterized by specific inter-strand hydrogen bonds that are formed between backbone oxygens and nitrogens (three of which are shown in Fig. 2) and a turn at positions ASN3-GLY6.

We now turn to a more detailed analysis of the geometry of the native state, as well as a comparison with experiment.

3.1 The native state: Comparison with experiment

Nuclear Overhauser effect constraints

Nuclear Overhauser effect constraints (NOEs) are the primary source of experimental structural information about the native state of peptides. For our peptide, two NMR studies [41,43] yielded sets of NOEs corresponding to differing ranges of interatomic contacts along the peptide sequence. Short-range constraints are important for identifying local structure while long-range contacts determine the global structure of the system. Figure 3 shows a comparison of 23 long-range interstrand NOEs determined experimentally by Friedrichs et al. [43] with those obtained through simulation. NOEs from simulation were obtained by averaging over all the

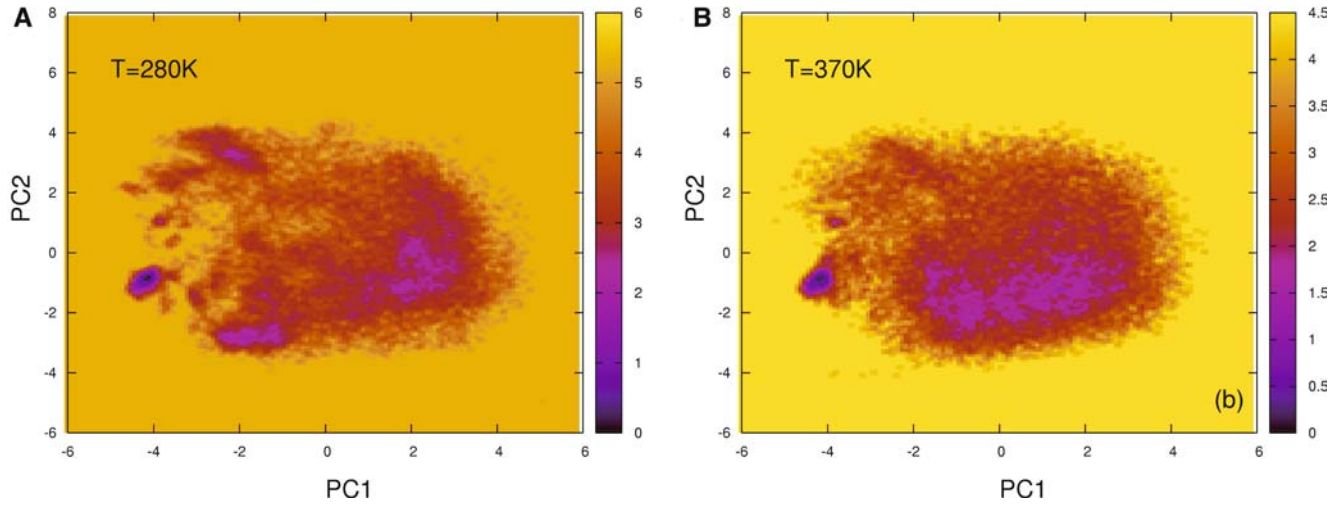

Fig. 1 Free energy map of the YQNPDGSQA peptide plotted as a function of the first two principal components at: (**a**) T = 280 K and (**b**) T = 370 K. Energies are shown in units of kT. The map exhibits a global minimum at (PC1 = −4.2, PC2 = −1.0) corresponding to β-hairpin conformations. The local minima present in the map are about 2–3 kT higher than the global minima and correspond to unfolded conformations

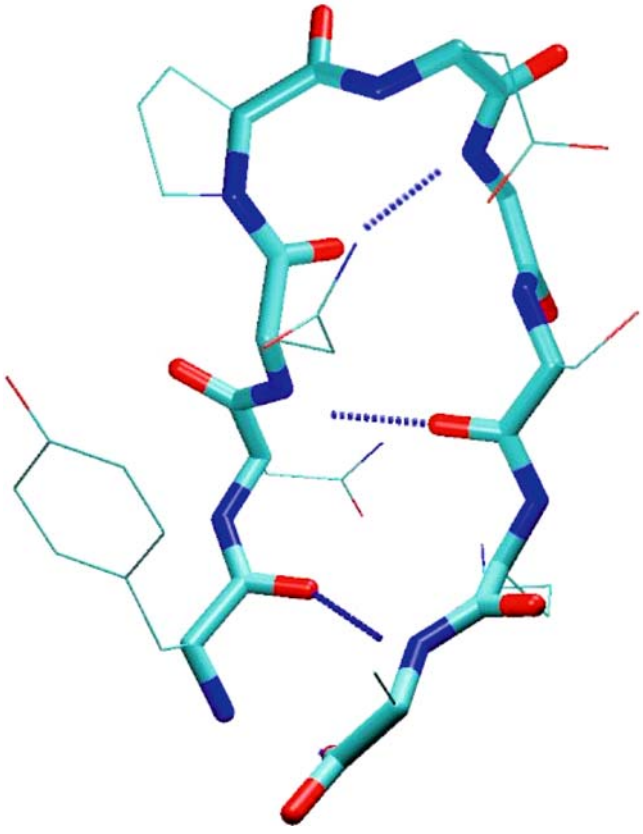

Fig. 2 A representative conformation of the most populated state observed in our simulations. Interstrand hydrogen bonds characteristic of the β-hairpin structure are depicted

conformations belonging to the native state minimum at T=280 K. The native state was delineated from the free energy surface in Figure 1 a as consisting of all conformations lying in a circle of radius 0.45 centered around the minimum point (PC1=−4.2,PC2=−1.0). The largest and smallest deviation recorded in the averaging are shown in Fig. 3 as a measure of the magnitude of the error present in the data. Aside from the NOEs involving hydrogens of the TYR1 side chain, the theoretical calculations are seen to be in good agreement with experiment. We note that the comparison of experimental and simulation NOEs is complicated by the fact that experiment and simulation deal with different statistical ensembles. For this peptide, Constantine et al. [42] showed that a single conformation could not satisfy all the NOE constraints. An assumption had to be introduced that an ensemble of equally probably conformations, rather than a single conformation, exists, and that the NOEs are satisfied as an average over this ensemble. This ensemble may not be representative of the one found in our simulations.

Hydrogen bonding in the native state

The NMR refinement simulations yielded a set of conformations satisfying the observed NOEs [42]. A common feature of these conformations is that they all possess a β-hairpin structure, characterized by a well defined network of hydrogen bonds.

Figure 4 shows a schematic of the native state of our peptide with a selected set of hydrogen bonds indicated by lines. In NMR refinement experiments, each hydrogen bond has a relative formation probability, or fractional occupation P_h, determined by the set of experimental conditions. For the purpose of comparison with our simulations, we included in Fig. 4 only those bonds that have the highest $P_{\mathrm{h}} > 0.3$ [42]. These are bonds formed between backbone atoms GLY6-H and ASN3-O (bond #1), ASN3-H and SER7-O (bond #2), ALA9-H and TYR1-O (bond #3) and ASP5-H and ASN3-O (bond #6). The other two bonds shown in Fig. 4 involve side-chain to backbone bonding between ASP5-H

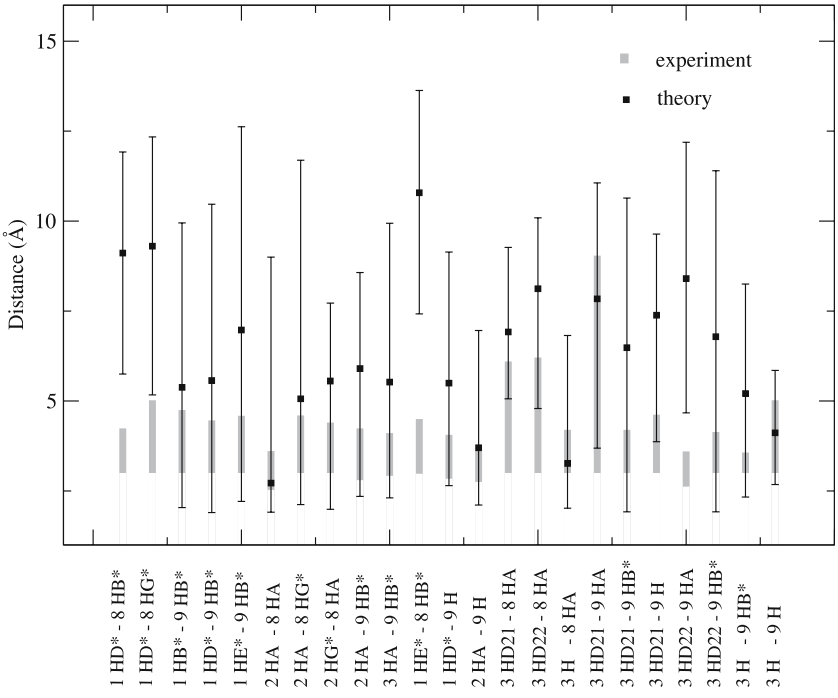

Fig. 3 Twenty-three long-range nuclear overhauser effect constraints (NOEs) spanning the two strands of the native peptide [42,43]. Theoretical values were obtained through $(\frac{1}{r^6})^{\frac{1}{6}}$ averaging over all native conformations. The error bars report the largest and smallest deviation from the mean values we observed

and ASN3-OD (bond #4) and GLN2-HE and GLY6-O (bond #5). Finally, hydrogen bond #7 is formed between side chain groups of ASN3-HD and ASP5-OD. Bonds 1, 4 and 6 are short-range in nature and characterize the local structure around the turn region, while bonds 2, 3 and 5 correspond to long-range bonds. The inter-strand bonds are best suited for describing how the native structure is acquired during folding. In a later section, we will use the hydrogen bonds involving backbone atoms (bonds 1, 2 and 3) to probe the mechanism of hairpin formation.

The experimentally and computationally determined fractional occupation probabilities for the hydrogen bonds displayed in Fig. 4 are listed in Table 1.

Following the definition of Friedrichs et al. [42], we consider that a hydrogen bond is formed between a pair of acceptors and donors if their mutual distance is less than 2.5 Å. Good agreement between simulation and experiment is observed for bonds 1, 2, 6 and 7, while bond 3 is overstabilized in our simulations and bonds 4 and 5 under stabilized. The failure of our simulations to describe properly bond 4 is surprising, since this bond is located in the turn region which is otherwise well represented by bonds 7, 1 and 6. We note that as was the case for the NOE constraints, the hydrogen bond probabilities in simulation and in experiment are computed over different ensembles. While we know that our simulations are averaged over a subset of the Boltzmann distribution of conformations at a given temperature, the ensemble of NMR simulations is generally poorly characterized [42]. It is hence unclear to what extent differences in experimental and computational hydrogen bond probabilities are due to inaccuracies in force field or solvent models alone.

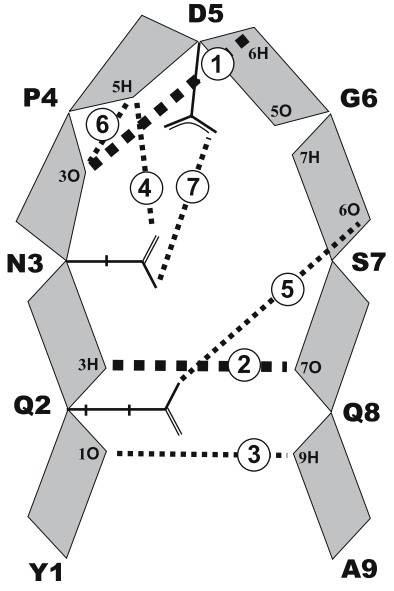

Fig. 4 A schematic representation of the native β-hairpin state of the peptide. A network of interstrand hydrogen bonds is indicated by dashed lines. The thickness of each line reflects the fractional population of the associated bond, or its "strength". Only the bonds most frequently populated in experiment [42] are shown here

Table 1 Fractional occupation probabilities for the hydrogen bonds that were suggested by the NMR experiments to be most stable [42]

Hydrogen bond #	1	2	3	4	5	6	7
Present simulation	0.97	0.99	0.97	0.08	0.01	0.26	0.35
NMR data	0.88	0.78	0.40	0.40	0.36	0.35	0.33

See Fig. 4 for the definition of the hydrogen bond numbers

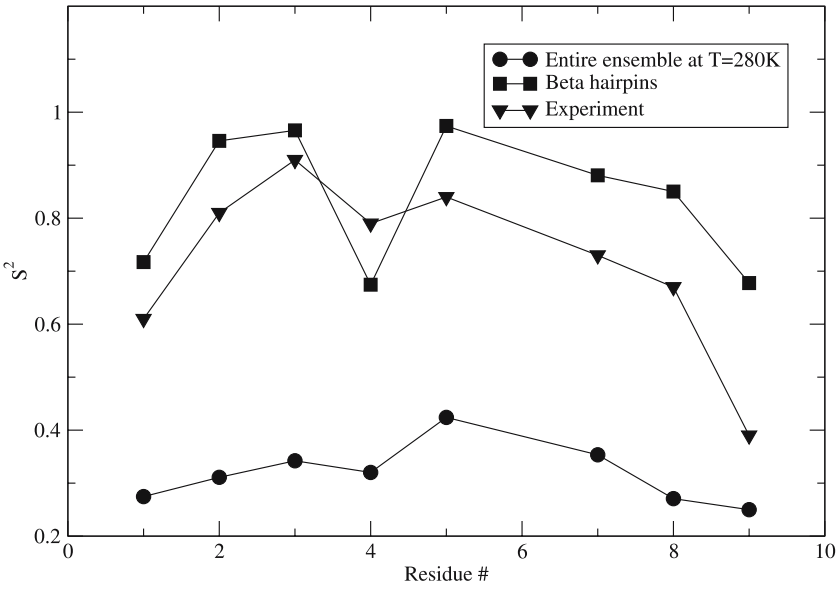

Fig. 5 Generalized order parameter S^2 of the Lipari-Szabo [58] model-free formalism obtained for $C_\alpha - H_\alpha$ vectors of the peptide. Theoretical data computed for the entire conformational ensemble at T=280 K and the native state conformations separately are compared to the NMR results [43]

Order parameters

The magnetic resonance relaxation constants offer another comparison with experiment. Following [42] we use the Lipari–Szabo model [58] to compute these constants. Of particular interest is the generalized order parameter, S^2 which quantifies the relaxation of the dipole–dipole interaction between two nuclei. In NMR studies of biological molecules, S^2 is used as a measure of the mobility of amino acid residues, in other words, of the degree of conformational order. Values of S^2 close to 0 indicate virtual absence of any stable structure while values close to 1 mean that the conformational ensemble is dominated by a single conformation. When applied to small peptides (as opposed to proteins), the generalized order parameter reflects only rapid motions of the nuclei and thus represents a lower limit of the actual degree of disorder [42, 43]. If slow motions with significant amplitudes occur, then a realistic model would produce lower S^2 values than experiment. And conversely, if the real system is subject to fast motions only, then theoretical and experimental S^2 values should closely match. Figure 5 displays S^2 values of $C_\alpha - H_\alpha$ vectors of all nine residues of the studied peptide. Higher theoretical S^2 values indicate that the simulated native state ensemble is slightly too ordered compared with experiment but, overall, in good agreement with it. In particular, the trends seen in the distribution of theoretical S^2 values across residues indicate that the terminal residues are less structured than all those in the center of the sequence, except for PRO4 which displays an enhanced mobility. Exactly the same trend is revealed in the experimental data. The entire conformational ensemble computed at T=280 K is seen to be considerably less structured than the native state alone. The lower S^2 values of this ensemble do not contradict experiment; rather, they indicate the presence of conformational transitions that occur on a longer than picosecond–nanosecond time scale.

3.2 Denatured state

We define the denatured state according to the free energy map shown in Fig. 1 as consisting of all conformation outside a circle of radius 0.45 centered at $(-4.2,-1)$. Secondary structural elements are assigned to these conformations following the PROSS classification scheme developed by Rose et al. [59]. This method assigns secondary structure based on backbone dihedral angles. The secondary structure computed here are summarized in Figs. 6 and 7. As shown in Fig. 6, the most common conformational state of the peptide in the unfolded state is a random coil, with residues GLN2, GLY6, SER7 and GLN8 in this state more than 50% of the time. A small amount (up to 20%) of β strand structure is observed for residues spanning GLN2 to ASP5. In addition,

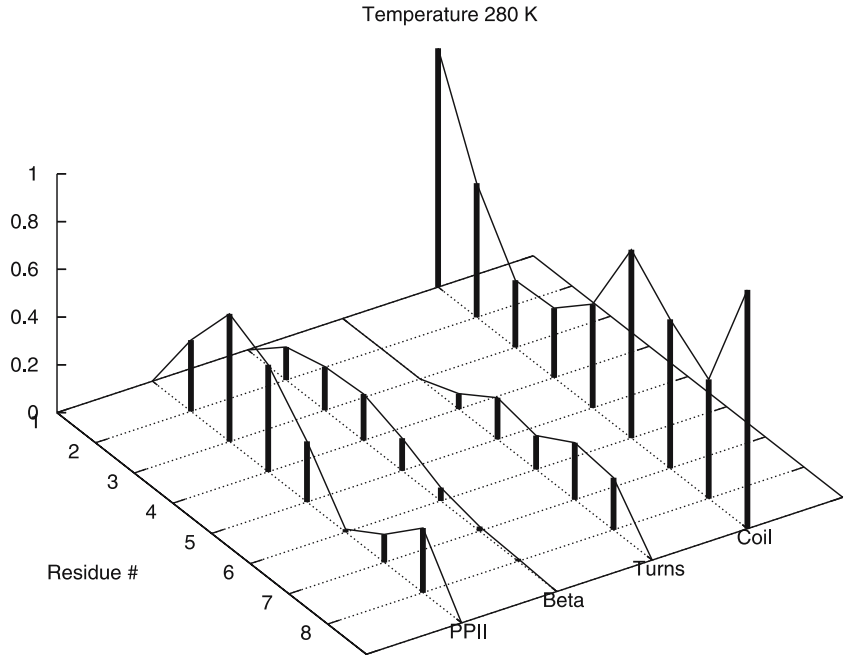

Fig. 6 Secondary structure across residues

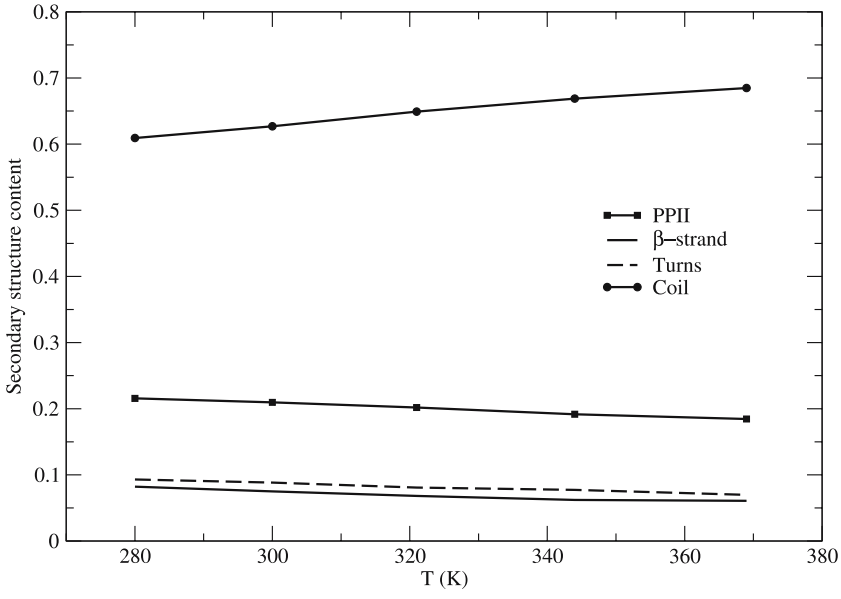

Fig. 7 Temperature dependence of the secondary structure content of the denatured state, averaged over all nine residues. A gradual vanishing of the ordered structure is observed as the temperature is raised

residues SER7 and GLN8 show an enhanced propensity to form turns in the unfolded state, a propensity that disappears completely in the folded state. Perhaps surprisingly, residues PRO4 and ASP5 which form a turn in the native hairpin, do not exhibit any preference for this secondary structure in the unfolded state. The most structured residue found in the unfolded state is ASN3, which displays 54% PPII in the unfolded state. The second most structured residue is PRO4 with 44% of PPII conformations. The most structurally stable cluster in the denatured ensemble appears to be ASN3-GLY6 segment.

Figure 7 shows the temperature dependence of the secondary structure content. The ordered secondary structure content is seen to gradually diminish as the temperature is elevated. The population of random coils increases from an average of 60% at T=280 K to 70% at T =370 K, with a resulting drop in population of PPII, β-strand and turns structures.

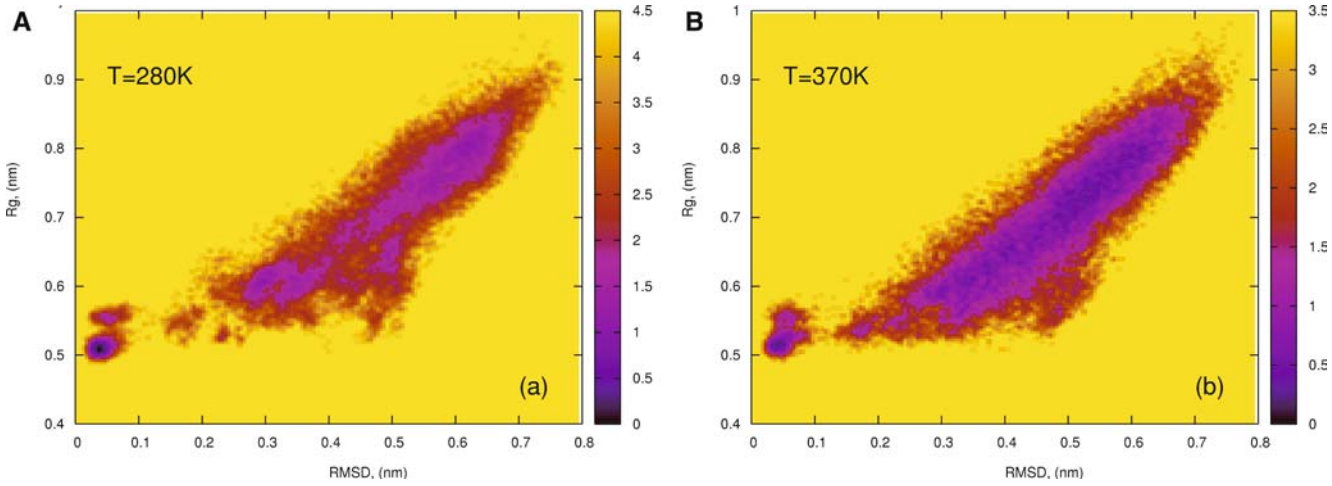

Fig. 8 Free energy map as a function of the RMSD deviation from native conformation (shown in Fig. 2) and the radius of gyration Rg at temperatures (a) T=280 K and (b) T=370 K. The energies are displayed in units of kT

3.3 Folding mechanism

Insights into the folding mechanism can be obtained from an analysis of the free energy surfaces projected onto appropriately chosen order parameters. Figure 8 shows the free energy maps at T=280 K (a) and at T=370 K (b), plotted as a function of the gyration radius R_g and C_α RMS deviation from the native β-hairpin. These two order parameters provide complementary information about conformational changes that the peptide undergoes as it folds. RMSD describes the structural similarity of a given conformation to the native state, while Rg is a measure of the peptide's overall size. The free energy maps at both temperatures show two major minima corresponding to the folded and unfolded states. The peptide is seen to undergo in a first step a substantial reduction in size, from $Rg \sim 8$ Å to about 6 Å, and starts to adopt some resemblance to the native state, with the RMSD dropping from 6 to about 2 Å. A second step involves further compaction of the protein (from $Rg \sim 6-5$ Å) with additional formation of native contacts. This mechanism does not follow the classical two-step "collapse followed by rearrangement to the folded state" often observed for proteins with β-sheet content [39].

In order to determine the order in which the backbone hydrogens bonds are formed, we examine the free energy surfaces as a function of the distances $d1$, $d2$ and $d3$ between the atoms forming the hydrogen bonds #1, #2 and #3 defined in Figure 4. To further probe side-chain formation, we plot the free energy surface as a function of the distance $d4$ between the C_δ atoms of GLN8 and GLN2. These two residues form the only long-range contact in the native state that is not involved in the formation of the hairpin turn. This contact serves as a measure of the time ordering of tertiary and secondary structure formation. The free energy maps at T = 280 K for four distance pairs are plotted in Fig. 9: (a) for $d1 - d2$, (b) for $d1 - d3$, (c) for $d1 - d4$ and (d) for $d2 - d3$.

Figure 9a indicates that for the majority of folding pathways, hydrogen bond #1, located in the turn region, forms before hydrogen bond #2. Figure 9d shows that the terminal hydrogen bond #3 is formed after the middle bond #2. Taken together, these two maps suggest the following ordering of hydrogen bonds: bond #1 is formed first, followed by #2 and finally by #3. This ordering is confirmed by Figure 9b. The overall folding mechanism emerging from the analysis of Figure 9 a, b and d is that of a classical hairpin zipper. Folding is initiated in the turn region and propagated onwards along the sequence until the native hairpin is formed. The interpretation of the time ordering between hydrogen bond formation and the formation of side-chain contacts is less clear-cut. Indeed, Fig. 9c suggests that the following two scenarios can occur: (a) side chains contacts are formed first, followed by backbone hydrogen bonds, (b) hydrogen bonds form first, then side chain contacts. This dual mechanism is supported by the data shown in Fig. 10, in which two folding trajectories from our simulations are projected onto the two-dimensional $d1 - d4$ free energy surface In the first REX trajectory, (shown in black), the side chains come into contact only after the turn hydrogen bond #1 is formed. Zipping of the peptide backbone precedes the formation of long-range contacts in this case. The second REX trajectory (shown in green), follows a different folding route. The side-chain contact for this trajectory is formed before formation of the hydrogen bond #1. In this case, collapse followed by folding/zipping is a more apt description of the folding process. An examination of the temperature dependence of the free energy maps of Fig. 9 (data not shown) reveals that the relative population of the available folding pathways strongly depends on the denaturing conditions. In particular at T=300 K, a new channel opens up that allows for the middle hydrogen bond #2 to be formed after the terminal bond #3. At a more elevated temperature, T = 344 K, the terminal bond #3 can exist even if the turn bond #1 is

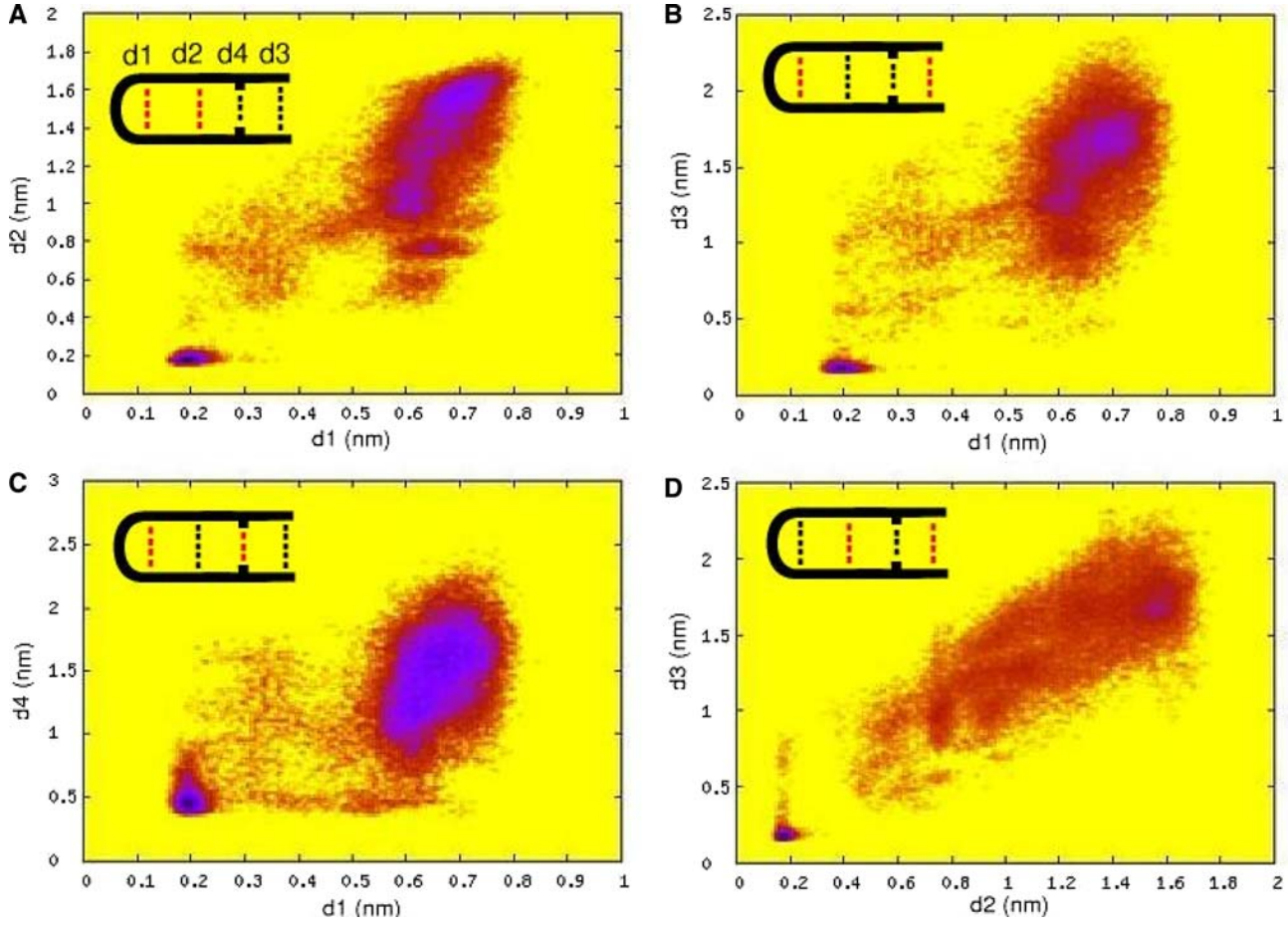

Fig. 9 Free energy maps at T=280 K for four pairs of order parameters ($d1$, $d2$, $d3$ and $d4$ defined in the main body of the text)

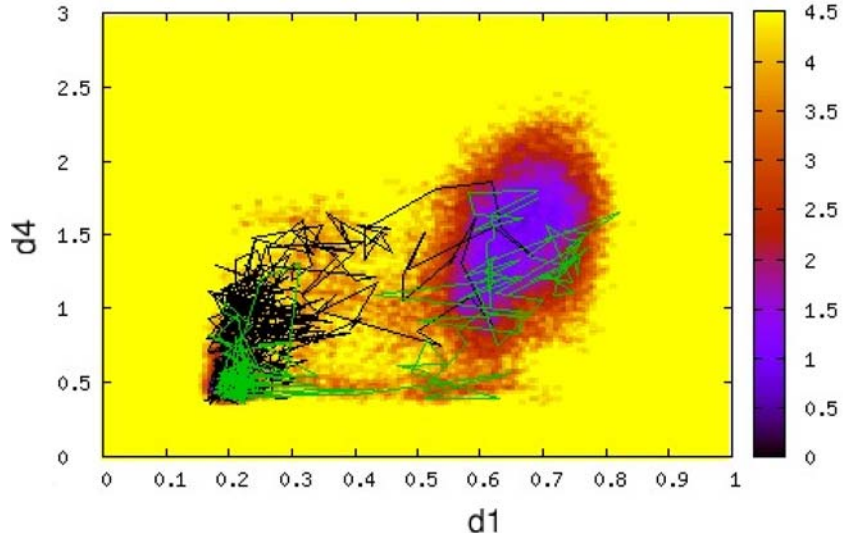

Fig. 10 An illustration of the collapse and zipping initiated folding mechanisms observed for the β-hairpin peptide in our simulations. Two folding trajectories obtained from the replica exchange (REX) simulations are projected onto two-dimensional subspace spanned by the distances $d1$ and $d4$. The *black curve* shows a trajectory in which the backbone hydrogen bonds that define the β-hairpin topology are formed before the side-chain atoms make contact. The *green line* represents a trajectory in which the side chain contact is made first and only then do the hydrogen bonds form

not formed. In sum, our data indicate a greater variability of kinetic pathways for folding at high temperatures.

4 Discussion

In this account, we have presented a comprehensive computational study of the folding thermodynamics of a small β-hairpin peptide. The peptide is a mutant of the 15–23 fragment of tendamistat, in which the S17-Y20 residues were replaced by the turn-inducing NPDG segment [41]. Separate NMR studies by Constantine et al. [42] and Friedrichs et al. [43] demonstrated that this peptide exhibits NOE signals compatible with stable β-hairpin conformations [41].

Molecular dynamics simulations of this peptide were reported by Constantine et al. [43] and later by Brooks et al. [45, 46]. Using short simulations on the order of a few nanoseconds, Constantine probed the local conformational dynamics of the peptide around one of the native conformations derived from NMR experiments. Brooks was able to observe a number of folding events through self-guided molecular dynamics simulations [45,46], with the peptide described by the all-atom AMBER force field and using an explicit solvent model. Although β-hairpin conformations were observed, the simulations of Brooks did not provide conclusive evidence that they constituted the ground state structure. The present work is the first to unambiguously demonstrate that the β-hairpin conformation is the native state of this peptide. This was possible because we used the enhanced sampling REMD technique [25]. We were able to run our simulations over a sufficiently long period of 50 *ns* (significantly longer than recent REMD studies reported for other, larger systems [19,60,61]) to generate *equilibrium* free energy maps of the studied peptide. The quality of equilibration was checked through thorough statistical analysis. The free energy surfaces in Fig. 1 clearly show that β-hairpins have lower free energy than any other sampled conformation.

The native state obtained in the present study agrees well with the one determined from experiment. We were able to satisfactorily reproduce a number of experimental parameters including NOE constraints and generalized order parameters of nuclear magnetic relaxation. In addition, our simulations correctly predict the network of hydrogen bonds that defines the topology of the native state. The native conformations generated in this work have low (<1 Å) C_α RMS deviations from the β-hairpin conformation designated by Brooks et al. [45] as the native state. Our simulations suggest that the unfolded state of the peptide is about 2 kCal/mol higher in free energy than the native state. This disagrees with the value of 12.7 kCal/mol reported in the simulations by Brooks [45]. Differences in the force-fields employed as well as insufficient equilibration in the work by Brooks [45] likely account for this discrepancy.

In addition to identifying the native state, our simulations allow us to characterize the nature of the unfolded state ensemble [62]. Whether or not the unfolded state can be described as a random coil has recently been the subject of much controversy. An argument in favor of a random character of the unfolded state is that the size distribution of denatured conformations obeys random-coil statistics [63]. On the other hand, the observation of considerable amounts of secondary structure [63,64] as well as the ability of some proteins to retain native-like topology in the unfolded state [65, 66], support non-randomness of the denatured state. Further evidence of structuring in unfolded conformations comes from the finding that local steric effects considerably restrict the size of the accessible conformational space [67], leading to a breakdown of the Flory isolated-pair hypothesis for polypeptide chains. Polyproline II (PPII) conformations have been proposed as the dominant motif in the unfolded states of proteins and peptides to explain the observed persistent structure in the denature state [68]. We probed the structure of the unfolded conformations of our peptide using the PROSS algorithm of Rose [59]. Our results indicate that the unfolded peptide is only about 20% PPII at $T = 280$ K and more than 60% random coil. This ratio further drops to 15%/70% at higher temperatures. Hence, our data do not support the idea that PPII conformations represent the "universal" structure of the denatured state. We also find that secondary structure is not evenly distributed along the sequence. The maximum of non-coil structure is observed around PRO4. Surprisingly, the pre-proline residue ASN3 contains more PPII structure, 54%, than proline residue PRO4, 45%. ASN3 and PRO4 residues were found to have the lowest content of random coil($<30\,\%$). In addition, the pre-proline residue ASN3 exhibited restricted conformational flexibility, in line with the recent reassessment of allowed regions in the ϕ, ψ dihedrals space [69]. Furthermore, we find that the amount of secondary structure in the unfolded state depends on both residue type and position along the sequence. For instance, glutamine residues 2 and 8 display different amounts of secondary structure despite being of the same type. This observation is at odds with the recent work of Eker et al. [70] which suggests that conformational preferences of amino acids in the unfolded state are specific to their type. Finally, our data also indicates a slight decrease in the amounts of extended β conformations as the temperature is raised.

Our simulations further allowed us to generate new insights into the folding mechanism of this peptide. Two main models have been proposed to explain the folding of β-hairpin forming peptides and proteins [71]. The first one, known as the hairpin zipper model, predicts that folding is initiated through formation of the hydrogen bond located nearest to the β-turn. Once this turn hydrogen bond is in place, the native structure then further propagates along both strands, forming each consecutive inter-strand hydrogen bond in sequence until the native hairpin is complete. The second model stipulates that the protein first undergoes a collapse transition (either hydrophobic or through side-chain contacts) and then proceeds to the native state from these collapsed conformations. There is experimental and theoretical support for both models (see [71,72] for recent reviews). More recently, a third mechanism of β-hairpin formation has been described [73] that involves a lateral motion of strands with respect to each other.

For our peptide, an analysis of the free energy surfaces for folding projected onto different coordinates suggests that the peptide can fold either by a zipper or by a collapse-initiated mechanism. The existence of multiple pathways for folding has also been noted in simulations by Higo et al. [74] on a 10-aa long β-hairpin peptide and by Wei et al. [73] on the C-terminal β-hairpin of protein G. Simulations by Brooks et al. [46] on the peptide studied in our work, only identified pathways corresponding to the collapse-initiated model. While this difference can be attributed to the use of different force fields (implying a dependence of folding mechanism on force fields), it is likely that insufficient statistics in the work of Brooks [46] are to blame. This highlights the importance of using enhanced sampling techniques that allow all possible pathways for folding to be visited. Finally, we note that recent work by Swope et al. [75] suggests that the distinction between the two folding mechanisms may be a matter of pure semantics. Indeed, using different definitions of a hydrogen bond lead to the favoring of one folding model over the other. Our simulations indicate that the relative population of folding routes is temperature dependent, with more pathways linking the folded and unfolded states at higher temperatures and the distinction between the two folding models becoming increasingly blurred.

5 Conclusion

The advent of enhanced sampling protocols such as replica exchange molecular dynamics, coupled with the availability of reasonably priced, fast processors, is opening the way for a full characterization of the thermodynamics of folding of proteins. As shown in this account, we are already able to efficiently generate equilibrium free energy maps for small peptides and characterize their folding mechanisms. We anticipate that future research in the field of folding will shift from a characterization of the native state to the investigation of the nature of the "other half" of the folding problem, the unfolded state. Because of its vastness, the unfolded state is difficult to probe using experimental and theoretical techniques. In addition, unfolding studies are complicated by the fact that different denaturing conditions yield different denatured ensembles. While simulating the high temperature unfolding of peptides and proteins is reasonably facile, the simulation of other means of denaturation (chemical, pressure, surface-induced, and so forth) poses a significant challenge. Pioneering work in the field of unfolding simulations has recently been undertaken by Paschek et al. [76], who used replica exchange molecular dynamics to probe the mechanisms of both temperature and pressure denaturation of the β-hairpin C-terminal domain of protein G. The mechanisms involved in chemical (using for instance urea or guanidium chloride) or surface induced unfolding, and how they differ from temperature and pressure denaturation, remain open questions. Equally intriguing is the process of cold denaturation, a challenging matter to tackle via simulation, as trapping in energy minima becomes a particularly serious issue at low temperatures. Only through the use of enhanced sampling protocols will it be possible to characterize, at a microscopic level, the mechanism by which proteins unfold at low temperatures [77].

Extending the work done on peptides to proteins (both with respect of folding and unfolding simulations) remains a challenge: as the size of the system grows, so do the number of replicas and the length of time that each replica needs to run. Recently, Garcia and Onuchic [61] have successfully used REMD to study the thermodynamics of folding of fragment B of protein A, a small 46-residue three helix bundle protein. This work currently constitutes the state of the art in protein folding simulations for systems described in fully atomic detail, with explicit solvent molecules. We envision that in the near future, similar studies will be possible on other, larger proteins and that REMD will emerge as an important technique for the study of the peptide and protein aggregation processes implicated in amyloid diseases.

Still out of our grasp is a full kinetic analysis of folding. In order to ensure reliable statistics, hundreds of folding events would need to be recorded. Bearing in mind that a small protein folds on the order of milliseconds to seconds, a straightforward molecular dynamics simulation from an unfolded state to the folded state would be prohibitive. In addition, such an exercise would only be useful if the force fields are sufficiently accurate to find the native state from a random starting configuration. Here again is where enhanced sampling methodologies can play a role. With reliable sampling of conformational space, one is able to assess and refine, through comparison with experiment, the quality of force fields. With improved force fields, it is likely that we will be able in coming years to extract meaningful kinetic information from simulation, either from free energy surfaces generated by enhanced sampling methods, or using distributed computing schemes such as the one implemented in "Folding@Home" [16,78].

Acknowledgements Support from the National Science Foundation (CAREER MCB #0133504), the David and Lucile Packard Foundation and the A. P. Sloan Foundation are gratefully acknowledged. Simulations were performed using the computational resources of the California NanoSystems Institute (NSF grant CHE-0321368).

References

1. Shea J.E, Friedel M, Baumketner A (2006) Reviews in Computational Chemistry, vol 22, Chapter 3
2. Onuchic JN, Nymeyer H, García AE, Chahine J, Socci ND (2000) Adv Prot Chem 53:87
3. Shea J-E, Brooks CL III (2001) Annu Rev Phys Chem 52:499
4. Ponder JW, Case DA (2003) Advances in protein chemistry. vol 66, Force fields for protein simulations, pp 27–85, Academic Press, New York
5. Feig M, Brooks CL III (2004) Curr Opin Str Biol 14:217
6. Price DJ, Brooks CL III (2002) J Comp Chem 23:1045
7. Karplus M, McCammon JA (2002) Nat Struc Biol 9:646
8. Hansson T, Oostenbrink C, van Gunsteren WF (2002) Curr Opin Str Biol 12:190
9. Hu H, Elstner M, Hermans J (2003) Proteins: Str Fun Gen 50:451
10. Mu Y, Kosov DS, Stock G (2003) J Phys Chem B 107:5064

11. Zaman MH, Shen M-Y, Berry RS, Freed KF, Sosnick TR (2003) J Mol Biol 331:693
12. Shirts MR, Pitera JW, Swope WC, Pande VS (2003) J Chem Phys 119:5740
13. Hansmann UHE, Okamoto Y (1999) Curr Opin Str Biol 9:177
14. Okamoto Y (2004) J Mol Graph Mod 22:425
15. Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, Garcia AE (2003) Curr Opin Struct Biol 13:168
16. Snow CD, Nguyen N, Pande VS, Gruebele M (2002) Nature 42:102
17. Yang WY, Pitera JW, Swope WC, Gruebele M (2004) J Mol Biol 336:241
18. Pitera JW, Swope W (2003) Proc Natl Acad Sci USA 100:7587
19. Zhou R (2003) PNAS 100:13280
20. Garcia AE, Sanbonmatsu KY (2002) Proc Natl Acad Sci 99:2783
21. Yoda T, Sugita Y, Okamoto Y (2004) Chem Phys Let 386:460
22. Gnanakaran S, Garcia AE (2003) Phys Chem B 107:12555
23. Zhou RH, Berne BJ (2002) Proc Natl Acad Sci 99:12777
24. Baumketner A, Shea J-E (2005) J Phys Chem B 109:21322
25. Sugita Y, Okamoto Y (1999) Chem Phys Lett 314:141
26. Torrie GM, Valleau JP (1977) J Comput Phys 23:187
27. Bursulaya B, Brooks CL III (2000) J Phys Chem B 104:12378
28. Boczko E, Brooks CL III (1995) Science 269:393
29. Ferrenberg AM, Swendsen RH (1989) Phys Rev Lett 63:1195
30. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JH (1992) J Comp Chem 13:1011
31. Berg BA, Neuhaus T (1992) Phys Rev Lett 68:9
32. Hansmann UHE, Okamoto Y (1999) J Chem Phys 110:1267
33. Hansmann UHE, Okamoto Y (1998) J Phys Chem B 102:653
34. Nakajima N, Nakamura H, Kidera A (1997) J Phys Chem B 101:817
35. Nymeyer H, Gnanakaran S, Garcia AE (2004) Methods in Enzymology 383:119
36. Hukushima K, Nemoto K (1996) J Phys Soc Jpn 65:1604
37. Tesi MS, van Rensburg EJJ, Orlandini E, Whittington SG (1996) J Stat Phys 82:155
38. Marinari E, Parisi G (1992) Europhys. Let 19:451
39. Guo Z, Brooks CL III (1997) Biopolymers 42:745
40. Shea J-E, Nochomovitz YD, Guo Z, Brooks CL III (1998) J Chem Phys 109:2895
41. Blanco FJ et al. (1993) J Am Chem Soc 115:5887
42. Constantine KL et al. (1995) J Am Chem Soc 117:10841
43. Friedrichs MS, Stouch TR, Bruccoleri RE, Mueller L, Constantine KL (1995) J Am Chem Soc 117:10855
44. Constantine KL, Friedrichs MS, Stouch TR (1996) Biopolymers 39:591
45. Wu X, Wang S, Brooks BR (2002) J Am Chem Soc 124:5282
46. Wu X, Brooks BR (2004) Biophys J 86:1946
47. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) J Phys Chem B 105:6474
48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926
49. Lindahl E, Hess B, van der Spoel D (2001) J Mol Mod 7:306
50. Berendsen HJC, van der Spoel D, van Drunen R (1995) Comp Phys Comm 91:43
51. Replica exchange facility for GROMACS was written by Luca Monticelli and Walter Ash, University of Calgary
52. Miyamoto S, Kollman PA (1992) J Comp Chem 13:952
53. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) J Comp Chem 18:1463
54. Essmann U et al. (1995) J Chem Phys 103:8577
55. Nosé S (1991) Prog Theor Phys 103:1
56. Jolliffe IT (1986) Principal component analysis. Springer, New York Berlin Heidelberg Tokyo
57. van der Spoel D et al. (2004) Gromacs User Manual version 3.2, www.gromacs.org
58. Lipari G, Szabo A (1982) J Am Chem Soc 104:4546
59. Srinivasan R, Rose GD (1999) PNAS 96:14258
60. Zhou R (2003) Proteins: Str Fun Gen 53:148
61. Garcia AE, Onuchic JN (2003) PNAS 100:13898
62. Shortle D (1996) FASEB J 10:27
63. Kohn JE et al. (2004) Proc Nat Acad Sci USA 101:12491
64. Sosnick TR, Trewhella J (1992) Biochemistry 31:8329
65. Shortle D, Ackerman MS (2001) Science 293:487
66. Snow CD, Zagrovic B, Pande VS (2002) J Am Chem Soc 124:14548
67. Pappu RV, Srinivasan R, Rose GD (2000) PNAS 97:12565
68. Shi Z, Woody RW, Kallenbach NR (2002) Adv Prot Chem 62:163
69. Lovell SC et al. (2003) Proteins: Str Fun Gen 50:437
70. Eker F, Griebenow K, Cao X, Nafie LA, Schweitzer-Stenner R (2004) PNAS 101:10054
71. Galzitskaya OV, Higo J, Finkelstein AV (2002) Curr Prot Peptide Sci 3:191
72. Snow CD, Sorin EJ, Rhee YM, Pande VS (2005) Annu Rev Biophys Biomol Str 34:43
73. Wei G, Derreumaux P, Mousseau N (2003) J Chem Phys 119: 6403
74. Kamiya N, Higo J, Nakamura H (2002) Protein Sci 11:2297
75. Swope WC et al. (2004) J Phys Chem B 108:6582
76. Paschek D, Garcia AE (2004) Phys Rev Let 93:238105
77. Baumketner A et al. (in preparation)
78. Snow CD et al. (2004) Proc Natl Acad Sci 101:4077