

Folding of Proteins with Diverse Folds

Sandipan Mohanty* and Ulrich H. E. Hansmann*[†]

*John von Neumann Institut für Computing, Forschungszentrum Jülich, Jülich, Germany; and [†]Department of Physics, Michigan Technological University, Houghton, Michigan

ABSTRACT Using parallel tempering simulations with high statistics, we investigate the folding and thermodynamic properties of three small proteins with distinct native folds: the all-helical 1RIJ, the all-sheet beta3s, and BBA5, which has a mixed helix-sheet fold. In all three cases, simulations with our energy function find the native structures as global minima in free energy at experimentally relevant temperatures. However, the folding process strongly differs for the three molecules, indicating that the folding mechanism is correlated with the form of the native structure.

INTRODUCTION

Despite considerable efforts, numerical studies of proteins have remained a formidable challenge in computational science. This is in part because it is often difficult to judge whether the failure of a simulation to fold a protein is because of inaccuracies of the potential (1) or because the simulation did not yet converge and better algorithms are needed (2,3). Only for small molecules is it computationally feasible to exhaustively explore the conformational space. However, not only are small proteins harder to study experimentally, but there are only a few small peptides that spontaneously fold into well-defined native structures. Still, a number of such polypeptides do exist, and offer an opportunity to study the mechanism of folding in computer simulations. The three proteins in this article (see Fig. 1) were chosen because, despite their small size (≈ 23 residues) and without any disulphide bridges to stabilize them, they are known to form monomers. They span three diverse classes: one all-helical, one three-stranded sheet, and the last one mixed. This is because it is still a challenge to correctly fold both proteins with helical structures, and those with β -sheet structures, as most force fields show a bias toward one particular class of structures. The first molecule is the α -helical 1RIJ (4), which is similar to the often-studied tryptophan cage peptide (PDB id: 1L2Y) (see, for instance, (5) and literature quoted there), with one extra turn of the α -helix. The designed three-stranded β -sheet beta3s (6) is also interesting from the point of force-field development, as our model in the form presented in Irbäck and Mohanty (7) is unable to fold it. Such peptides give useful clues for refinements of the force fields, and with minimal improvements of the force field (S. Mohanty, unpublished), we can indeed also fold this peptide. The third peptide here is a mixed $\beta\beta\alpha$ peptide, called BBA5 (PDB id: 1T8J) (8). Peptides of this size, i.e., 23 residues, with both helical and β -sheet secondary structure elements, are very rare

and have not before been successfully simulated with our force field.

With the investigation of these three small proteins, our article has two goals. First, we want to ensure that the feasibility of our energy function for simulations is not restricted to certain folds. This is made possible by comparing our simulation results with experimentally obtained data. Secondly, we go beyond the experiments and explore the mechanism of folding for these three proteins. Consistent with previous work (see for instance, (9,10)) that relied on minimal protein models, we find in our all-atom simulations that the form of the folding process is related to the final fold.

MODEL AND METHODS

The model used for this study belongs to the class of all-atom protein models with fixed bond-lengths and bond-angles, and implicit water. It was developed by Irbäck et al. (7,11,12), and was modified by us to treat D-amino acids, since BBA5 contains a D-proline. The force field consists of four simple terms. These represent excluded volume effects, a local electrostatic term, a hydrogen-bond term, and an effective hydrophobic attraction:

$$E = E_{\text{ev}} + E_{\text{loc}} + E_{\text{hb}} + E_{\text{hp}}. \quad (1)$$

The excluded volume effects are represented by the E_{ev} term as a strong r^{12} repulsion between the atoms. The second term in Eq. 1 represents a limited electrostatic term. Only the partial charges on the backbone N (and attached H) and C (and attached O) are considered. Atoms in one amino acid only interact electrostatically with the other atoms of the same amino acid. For all other cases, the term E_{loc} ignores partial charges on the atoms. The hydrogen bonds are represented by the E_{hb} term. The hydrophobicity term, E_{hp} , is a pairwise attraction term between hydrophobic amino-acid side chains. Each hydrophobic amino acid has a designated set of hydrophobic atoms. For each pair of such amino acids, the fraction of hydrophobic atoms in contact with a hydrophobic atom of the other amino acid is computed. This is then multiplied by a pairwise strength depending on the type of the two amino acids. Details on the form of the force field can be found in the literature (7,11,12). The simulations were carried out using the protein folding and aggregation program, PROFASI (13), which implements the above-mentioned model.

We use parallel tempering simulations (14,15), a technique first introduced to protein science in the literature (16), to sample the energy landscape of these molecules. In this article, we have used eight temperatures in the range 274–369 K distributed in a simple geometric series. The energy distributions obtained at the different temperatures used indicate a substantial

Submitted April 21, 2006, and accepted for publication July 28, 2006.

Address reprint requests to U. H. E. Hansmann, Tel.: 906-487-4552; E-mail: hansmann@mtu.edu.

© 2006 by the Biophysical Society

0006-3495/06/11/3573/06 \$2.00

doi: 10.1529/biophysj.106.087668

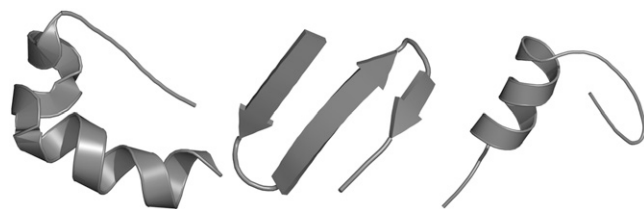


FIGURE 1 The three proteins studied in this article have very distinct folds. 1RIJ (*left*) is mostly α -helical, with a tail consisting of three prolines at one end which folds back to make hydrophobic contacts with the helix. The beta3s (*center*) form a three-stranded β -sheet. BBA5 (*right*) has both an α -helix as well as a small β -hairpin at the N-terminal. Drawn with PYMOL (18).

overlap for each of the peptides considered—which facilitates a smooth flow of each replica through all the different temperatures. We observed from trial runs that a typical folding event takes on the order of 10,000 Monte Carlo sweeps (10^6 elementary Monte Carlo updates) from a state with high energy and few native contacts, to one which is nearly completely folded and back. In case of our production runs, each of the 64 runs was long enough to contain several folding events.

To explore the space of conformations of a peptide we use two categories of conformational updates. Single-angle updates are used for both backbone and side-chain degrees of freedom, where a degree of freedom is chosen randomly and set to any value between 0 and 2π . A single backbone move of this kind can cause a very large change in the structure of the molecule, but has also a high probability for rejection. The second kind of conformational update we use is a semilocal move (17) of the backbone degrees of freedom. This so-called biased-Gaussian step is a concerted rotation of a set of up to eight consecutive backbone degrees of freedom with only small rigid body changes outside the area of update, which can be implemented as a fast algorithm without any equation solving during the update.

Each of the simulations, consisting of $\approx 8 \times 10^9$ elementary Monte Carlo updates on total, takes ≈ 3.5 h of computation time on 64 processors of the supercomputer JUMP (IBM Regatta, IBM, Armonk, NY) at the Forschungszentrum Jülich, Jülich, Germany.

RESULTS

We now discuss the observed folding behavior of the three proteins in our simulations. Of the three proteins studied, the α -helical E6-binding-Trp-cage or 1RIJ peptide (sequence: ALQEL LGQWL KDGGP SSGRP PPS) has the simplest behavior.

In Fig. 2 *a* we show the fraction of residues with their backbone angles in the helical and β -strand regions, as functions of temperature. Two regions can be clearly distinguished in this plot. At high temperatures both the average helicity and strand content are small, and the protein is in a random coil state. As the temperature decreases, the helix content increases while the β -strand content goes well below the random value. Note that the midpoint of the increase of helicity corresponds to the peak in the specific heat at $T = 306$ K shown in the inset. At the same temperature we also observe the midpoint of the increase in the probability of obtaining a nativelike state, with at least 80% of its native contacts formed, which is displayed in Fig. 2 *b*. Hence, the folding of the protein is mainly driven by helix formation. This can be also seen in the plot of the free energy landscape (at $T = 274$ K)

displayed in Fig. 3 as a function of helicity and radius of gyration r_g . Like the similar tryptophan cage peptide 1L2Y, the free energy landscape of 1RIJ is characterized by a funnel-like topology around a dominating minimum, indicating a broad native attractor-basin. This funnel topology indicates that collapse of the protein chain and helix formation is synchronous. The corresponding lowest energy conformer has a backbone root mean-square deviation (RMSD) of $r_{\text{rmsd}} \approx 2.7$ Å and is also shown in the figure, overlaid on the native structure. This conformer appears at $T = 274$ K with a frequency of $\sim 50\%$ (see Fig. 2 *b*), which is smaller than the experimentally observed probability of 90%.

Inspection of the histogram of the number of native contacts at $T = 274$ K reveals there is only a very small fraction ($< 3\%$) of structures that do not share 40% or more of contacts with the native state. This means that at the lowest temperature, there are no fundamentally different competing states, but rather, the nativelike ground state is in equilibrium

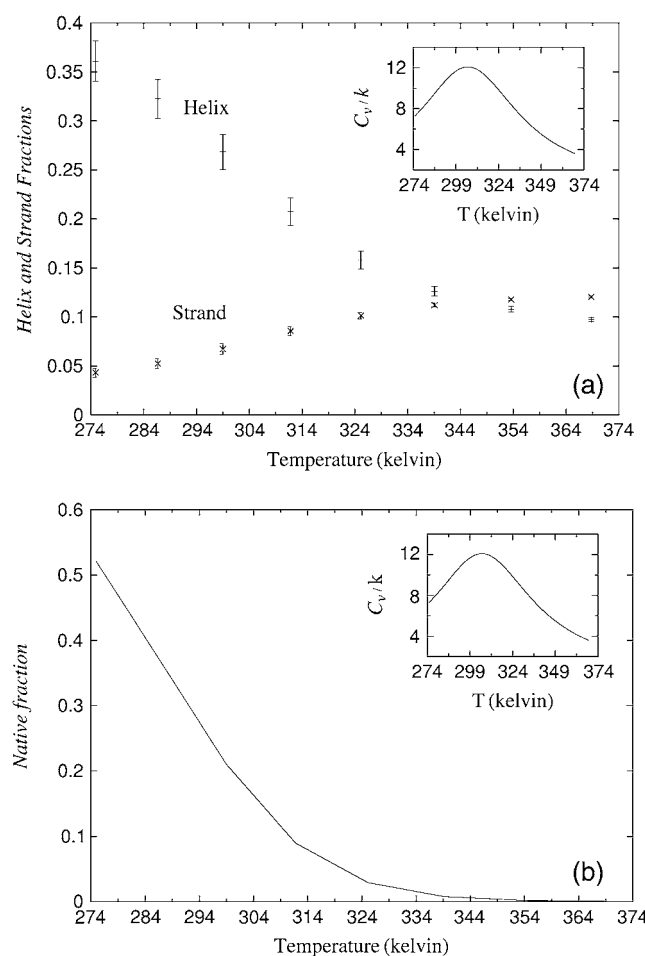


FIGURE 2 (a) Average helix and β -strand fractions for 1RIJ as functions of temperature. The specific heat curve is shown in the inset. (b) Estimate of the probability of obtaining a nativelike state as a function of T , where “nativelike” is defined as states with at least 80% of native contacts formed. The inset once again shows the specific heat curve, calculated from energy histograms using histogram-reweighting techniques (19).

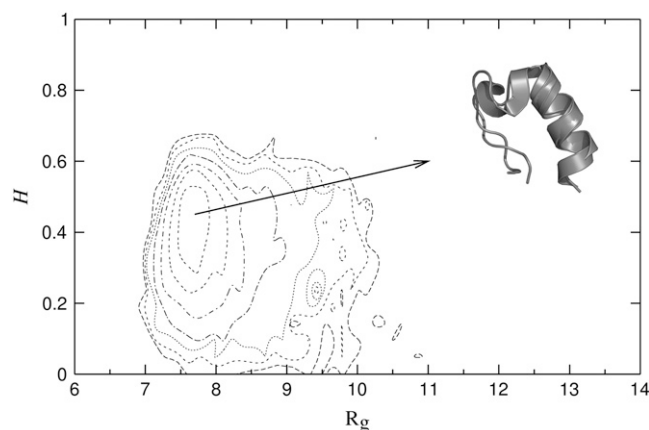


FIGURE 3 Free-energy F as a function of radius of gyration R_g and helix content H . The structure obtained as the free energy minimum in the simulations is shown superimposed on the PDB structure (1RIJ) in the inset. The contour lines in this figure are separated by $0.25 k_B T$.

with a multitude of excited states, in which the protein is partially unfolded from the ground state. The energy of these partially folded states are within $\sim 2 K_B T$ of the ground state. This is consistent with the simple funnel-like landscape, somewhat flat at the bottom, that one observes in Fig. 3. The different native contacts and native hydrogen bonds do not show any particular order of formation as a function of temperature. However, the contacts and the hydrogen bonds corresponding to the ends of the helix are more likely to be broken than those at the center of the helix.

Compared to the α -helical peptides, β -sheet peptides are known to show more complex free energy landscapes. Fig. 4 shows the free energy as a function of energy and backbone RMSD for the artificial molecule beta3s (sequence: TWIQN GSTKW YQNGS TKIYT). As there is no PDB entry for this molecule, the reference structure was chosen to be one of the 20 structures obtained from the NMR study of the molecule by de Alba et al. (6).

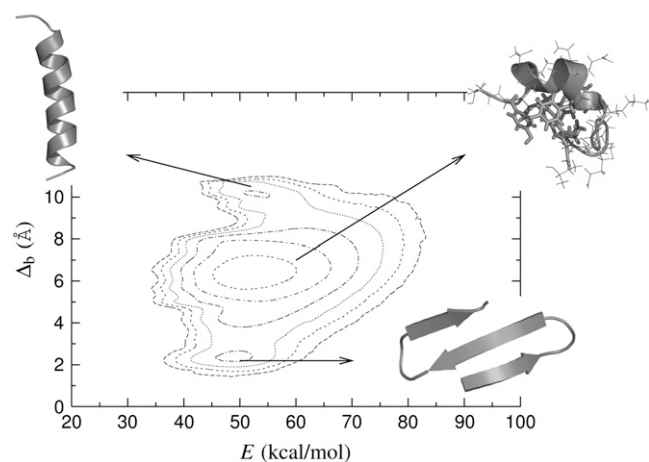


FIGURE 4 The free-energy landscape for beta3s as a function of total energy and backbone RMSD.

A nativelike state appears as a global minimum $\sim 2 \text{ \AA}$ backbone RMSD in both plots. However, this structure is energetically slightly disfavored over the competing ones seen in Fig. 4, while also being entropically disfavored compared to the population of competing structures.

The free energy minimum, at RMSD of $\sim 6 \text{ \AA}$ seen in Fig. 4, is not a very well-defined state. Several conformational characteristics of the molecule, such as the radius of gyration, vary quite a lot among structures contributing to that minimum. One common characteristic is, however, that they all form a tight hydrophobic core with all the tryptophan, isoleucine, and tyrosine groups packed close together. This is a known side effect of the pairwise additive form of the hydrophobicity term in the potential, which fails to take into account the multibody effects that become important for such conformations. Since this molecule has a large number of highly hydrophobic residues, the potential has a built in weakness for the study of this molecule. Despite this, the simulations succeed in escaping that minimum and finding the native state as the second-most significant minimum even in this case.

As a consequence, beta3s has only a small probability in the model to be in the experimentally reported native state at $T = 274 \text{ K}$. We show in Fig. 5 the frequency of configurations as a function of RMSD to the native structure. Clearly, most configurations found have RMSDs $> 6 \text{ \AA}$. Only 10% of the configurations have RMSDs $< 3 \text{ \AA}$, i.e., can be considered to be similar to the native state. This compares with an experimentally observed propensity of 13–31% at 284 K for this structure (6).

Unlike 1RIJ, the peak in specific heat, shown in Fig. 6, does not mark a folding transition but only the collapse into dense structures as can be seen from the plot of the radius of gyration as a function of temperature, shown in the same figure. The peak in specific heat marks the decrease of the

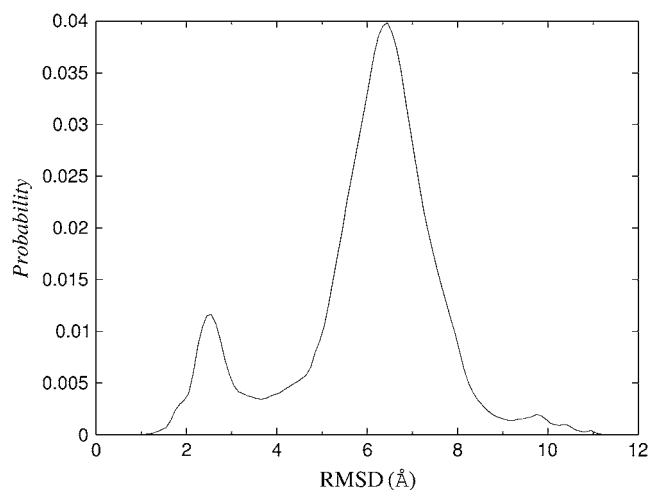


FIGURE 5 For beta3s, the histogram of backbone RMSD shows a two-peak character, and an estimate of the native population was obtained from the population of the nativelike peak.

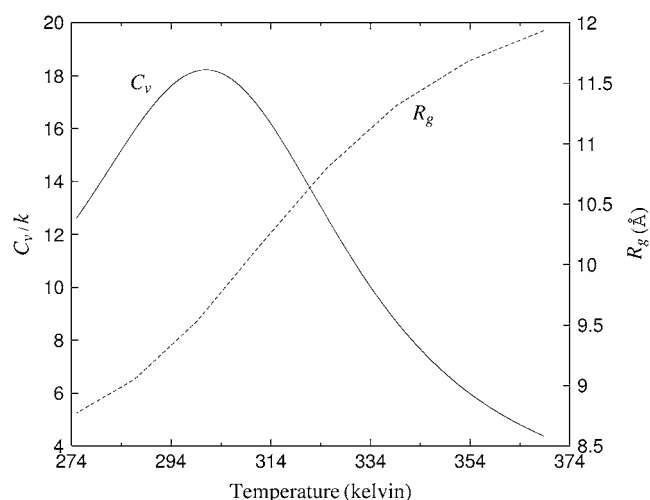


FIGURE 6 The specific heat (C_v) and mean radius of gyration (R_g) for beta3s as functions of temperature.

radius of gyration, but the population of native states only reaches a value of $\sim 10\%$ at the lowest temperature considered in this study.

Several other features of the folding mechanism can be observed from our simulations. For instance, the probability of finding individual native hydrogen bonds as a function of temperature shown in Fig. 7, has a clear pattern. These bonds are mostly absent for the hydrophobic-pit structures discussed above, and also for the helical structure shown in Fig. 4, and therefore represent only those events in which the molecule folds into its native structure. There is a well-defined order for the formation of native hydrogen bonds as a function of

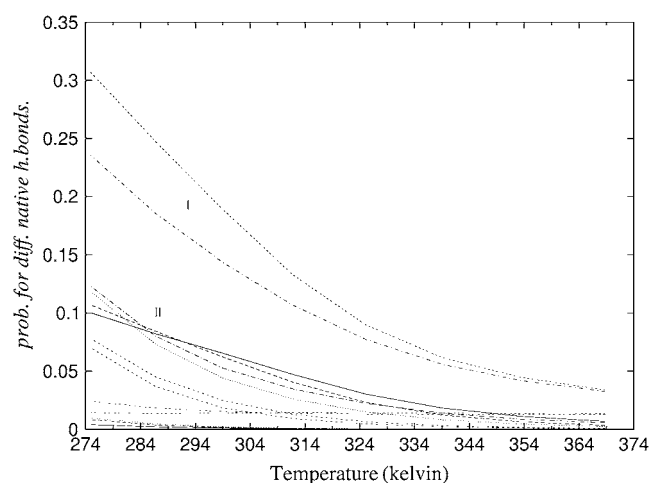


FIGURE 7 Probability of occurrence of different native hydrogen bonds of beta3s as functions of temperature indicates a certain order for their formation. The two curves labeled *I* represent the hydrogen bonds closest to the two turns of the three-stranded β -sheet. The curves labeled *II* represent the next four hydrogen bonds away from the turn along the two β -hairpins. We interpret that folding for beta3s starts with the formation of the turns and proceeds along the hairpins in a zipper mechanism.

decreasing temperature. The hydrogen bonds (NH–CO) THR¹⁶–ASN¹³ and THR⁸–ASN⁵ are the first to form, and are also the ones located at the turns of the three-stranded β -sheet. The next four bonds along the β -hairpins form a cluster with similar temperature behavior. They are the next-most-probable hydrogen bonds. The two bonds after that along the hairpins form a third cluster with somewhat lower probability. Motivated by this pattern, we conjecture a zipperlike mechanism for the folding of beta3s.

We also find that it takes very few Monte Carlo cycles to fold the molecule from an apparently disordered state with a large energy to the native state, even though the Monte Carlo search might spend a very long time in states with high RMSD. This suggests a narrow native attractor basin, unlike that of 1RIJ. We interpret it as the formation of a transition state (perhaps consisting of the turn-hydrogen bonds) which leads to quick folding, even though the transition states themselves are not easily reached in the model. We also observe that the formation of one hairpin provides a template for the other strand to quickly attach with the well-formed hairpin to make the three-stranded β -sheet. But there is no particular order as to whether one or the other hairpin is formed first.

Our third protein is the 23-residue BBA5 (sequence: YRVPS YDFSR SDELA KLLRQ HAG, where \bar{P} means D-Proline). Unlike 1RIJ and beta3s this polypeptide has a mixed structure with both α -helix and β -sheet secondary structure elements. Residues Glu¹³ through Gln²⁰ form a well-defined α -helix, while residues Tyr¹ through Phe⁸ form a hairpin structure. A small hydrophobic core is formed between these two structures with several strongly hydrophobic groups. This mini-protein is one of the several related small $\beta\beta\alpha$ motifs designed and experimentally studied by Imperiali et al. (8), containing only one nonstandard amino acid, the D-proline⁴.

While our force field has been developed for proteins with only either α -helices or β -sheets as secondary structure elements, it identifies the native state of this protein of mixed secondary structure as the lowest-energy minimum, with an RMSD of 2.2 Å to the PDB structure. However, there exist several competing configurations with RMSD ≈ 6 Å, which are energetically within $\sim 1\text{--}3 k_B T$ of the ground state, according to our force field. The energy landscape as a function of energy and RMSD is plotted in Fig. 8, with the corresponding configurations (and their overlap with the PDB structure, when relevant) also displayed.

The free energy landscape of BBA5 is characterized by a broad valley connecting the nativelike minimum to another minimum at $\sim \text{RMSD } 6$ Å. Although the native minimum consists of rather similar conformations, the minimum at 6 Å RMSD consists of several distinct structures with roughly the same energy, so that the projection of the free energy landscape onto the energy-RMSD plane does not distinguish them as separate minima. In Fig. 8, we have shown the conformations corresponding to the nonnative minima. One

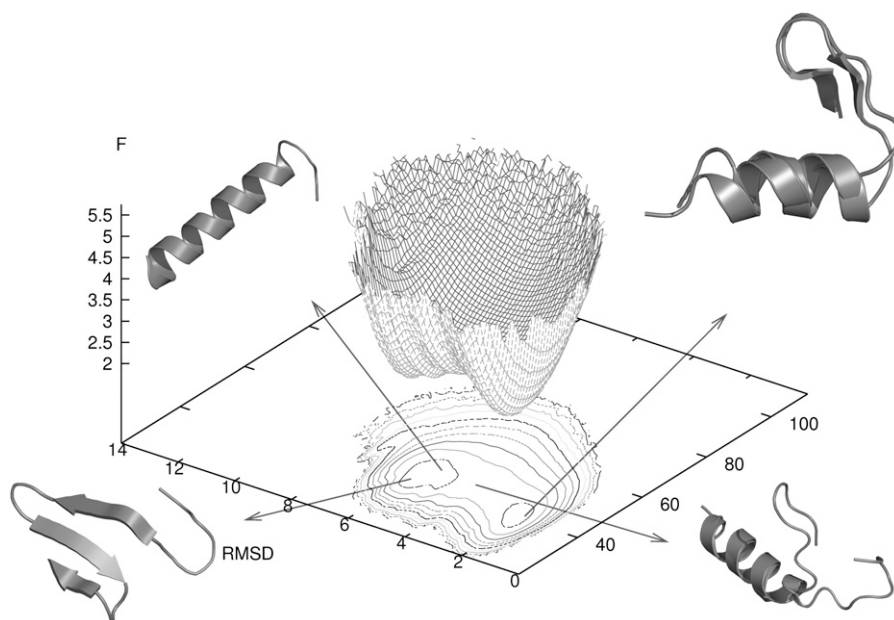


FIGURE 8 Free energy as a function of RMSD and energy for BBA5. The structures corresponding to two local minima are shown, with the nativelike conformation also superimposed on the experimental structure.

of these structures is very similar to the native one, with both the C-terminal helix and the N-terminal hairpin formed, though without the hydrophobic contacts between the two structures. Another such structure contributing to the same minimum is a helical structure, where the C-terminal helix is broken only at the D-proline. A third, rather different looking local minimum is a four-stranded β -sheet, where the C-terminal helix folds into a hairpin instead, and joins with the N-terminal hairpin with hydrogen bonds. The native state is energetically favorable compared to each of these competitors, but the combined population of the three competing minima is greater. The population of the native minimum is $\sim 30\%$ and appears somehow smaller than the experimentally observed NOE spectrum would suggest.

The specific heat peak for this protein, shown in Fig. 9 *c*, corresponds to the collapse of the structures, but is not a signal for folding.

Studying the probability of finding each individual native contact and hydrogen bond at different temperatures, we find that the native contacts fall into two categories. Those contacts corresponding to the helix and those to the hairpin show similar behavior to their counterparts in native hydrogen bonds. We find no particular temperature order for the formation of these contacts. Although some bonds are more likely to form than others, there is no clear temperature dependence indicated by this difference. But the native contacts corresponding to the hydrophobic contacts between the two secondary structure elements show a slightly different temperature behavior. For all temperatures they have a lower probability of being formed compared to the secondary structure elements, and this difference grows with decreasing temperature. This is interpreted by us as a greater propensity for the formation of the secondary structure elements compared to the tertiary structure of this molecule.

DISCUSSION

Comparing our results for the three proteins, we find that the observed folding mechanism is related to their final fold. For 1RIJ, collapse of the protein and the formation of native hydrogen bonds are synchronous. Helical structures are natural compact configurations of chain molecules, and if a protein segment is stable as a helix, compaction and formation of stabilizing hydrogen bonds can take place simultaneously. For beta3s, we observed that collapse precedes the formation of the native hydrogen bonds. Formation of one of the two hairpins in a zipperlike fashion catalyzes the formation of the second, by acting as a template for the remaining part. We observed no local minimum of energy with only one of the

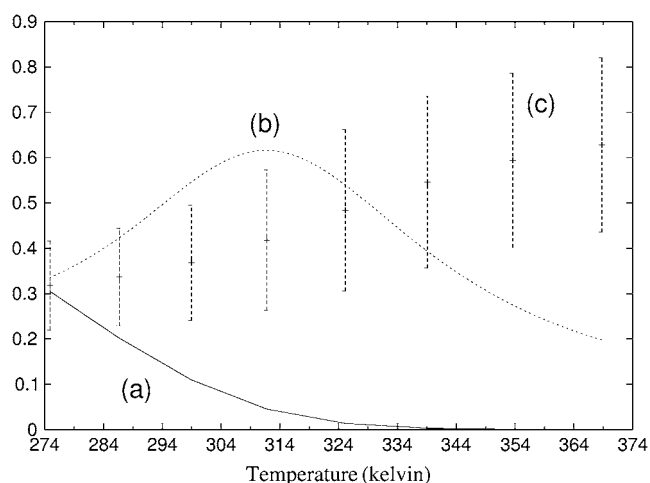


FIGURE 9 We show here, as functions of temperature, (*a*) the native fraction, based on formation of 80% of native contacts; (*b*) the specific heat curve, scaled to the same y-range; and (*c*) radius of gyration, scaled and shifted to fit in the range 0–1.

hairpins folded. This is in contrast with the folding of the two other three-stranded β -sheets studied earlier with this model (7), β -nova, and LLM. For those peptides, the histogram of native hydrogen bonds showed a three peaked structure, with a clearly identified central peak corresponding to the formation of only one of the two hairpins. For BBA5, we find that the two secondary structure elements form independently, while their relative arrangement in the specific manner of the experimentally reported PDB structure is less probable. The N-terminal hairpin has a D-proline which facilitates the formation of the type II' turn. But yet the hydrogen bonds corresponding to a good β -hairpin structure have a slightly lower probability to form than the C-terminal helix. This probability also varies less with temperature compared to the probability for the helix hydrogen bonds. Contacts between the helix and the hairpin seem to be less stable than the secondary structures themselves, which indicates a preferred folding mechanism in which the secondary structures form first and then assemble into their nativelike arrangement.

CONCLUSION

We have performed simulations of three small proteins using a recently developed force field. In all three cases we find the native structures as the dominant configurations at biologically relevant temperatures but the observed frequencies of nativelike states appear to be smaller than the ones observed in experiment. Interestingly, we observe a specific folding mechanism that differs for the three proteins and are strongly correlated with their specific fold. The all-helical 1RIJ collapse and helix-formation happen in parallel whereas, for beta3s (built out of only β -sheets), the collapse of the polypeptide chain precedes the secondary structure formation. BBA5 has both an α -helix and a β -sheet as secondary structure elements, both of which have to be formed before the peptide can assume its final shape. Finally, the observed independence of correlated folding success and thermodynamics from the final fold and folding mechanism suggest a certain universality of the force field. Future studies will have to test whether this observation persists for larger proteins.

All simulations were done on the IBM Regatta supercomputer JUMP at the John v. Neumann Institute for Computing in Jülich, Germany.

This work was supported in part by the National Institutes of Health grant No. GM62838.

REFERENCES

1. Lin, C. Y., C. K. Hu, and U. Hansmann. 2003. Parallel tempering simulations of HP-36. *Proteins*. 52:436–445.
2. Hansmann, U. H. E., and Y. Okamoto. 1999. New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* 9:177–183.
3. Kwak, W., and U. Hansmann. 2005. Efficient sampling of protein structures by model hopping. *Phys. Rev. Lett.* 95:138102.
4. Liu, Y., Z. Liu, E. Androphy, J. Chen, and J. D. Baleja. 2004. Design and characterization of helical peptides that inhibit the E6 protein of papillomavirus. *Biochemistry*. 43:7421–7431.
5. Schug, A., W. Wenzel, and U. Hansmann. 2005. Energy landscape paving simulations of the Trp-cage protein. *J. Chem. Phys.* 122:194711.
6. de Alba, E., J. Santoro, M. Rico, and M. A. Jiménez. 1999. De novo design of a monomeric three-stranded antiparallel β -sheet. *Protein Sci.* 8:854–865.
7. Irbäck, A., and S. Mohanty. 2005. Folding thermodynamics of peptides. *Biophys. J.* 88:1560–1569.
8. Struthers, M., J. Ottesen, and B. Imperiali. 1998. Design and NMR analyses of compact, independently folded BBA motifs. *Fold. Des.* 3:95–103.
9. Brooks III, C. L., M. Gruebele, J. N. Onuchic, and P. G. Wolynes. 1998. Chemical physics of protein folding. *Proc. Natl. Acad. Sci. USA*. 95:11037–11038.
10. Head-Gordon, T., and S. Brown. 2003. Minimalist models for protein folding and design. *Curr. Opin. Struct. Biol.* 13:160–167.
11. Irbäck, A., B. Samuelsson, F. Sjunnesson, and S. Wallin. 2003. Thermodynamics of α - and β -structure formation in proteins. *Biophys. J.* 85:1466–1473.
12. Irbäck, A., and F. Sjunnesson. 2004. Folding thermodynamics of three β -sheet peptides: a model study. *Proteins*. 56:110–116.
13. Irbäck, A., and S. Mohanty. 2006. PROFASI: a Monte Carlo simulation package for protein folding and aggregation. *J. Comput. Chem.* 27:1548–1555.
14. Hukushima, K., and K. Nemoto. 1996. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. (Jap)*. 65:1604–1608.
15. Geyer, C. J., and E. A. Thompson. 1995. Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Am. Stat. Assn.* 90:909–920.
16. Hansmann, U. 1997. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* 281:140.
17. Favrin, G., A. Irbäck, and F. Sjunnesson. 2001. Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. *J. Chem. Phys.* 114:8154–8158.
18. DeLano, W. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.
19. Ferrenberg, A. M., and R. H. Swendsen. 1988. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61:2635–2638.