PROTEINS: Structure, Function, and Bioinformatics 56:110–116 (2004)

# Folding Thermodynamics of Three β-Sheet Peptides: A Model Study

**Anders Irbäck*** **and Fredrik Sjunnesson**
*Complex Systems Division, Department of Theoretical Physics, Lund University, SE-223 62 Lund, Sweden*

***ABSTRACT*** **We study the folding thermodynamics of a β-hairpin and two three-stranded β-sheet peptides using a simplified sequence-based all-atom model, in which folding is driven mainly by backbone hydrogen bonding and effective hydrophobic attraction. The native populations obtained for these three sequences are in good agreement with experimental data. We also show that the apparent native population depends on which observable is studied; the hydrophobicity energy and the number of native hydrogen bonds give different results. The magnitude of this dependence matches well with the results obtained in two different experiments on the β-hairpin. Proteins 2004;56:110–116.**


**Key words: protein folding; β-hairpin; three-stranded β-sheet; all-atom model; Monte Carlo simulation**

## INTRODUCTION

Peptide folding is currently attracting considerable attention. Recent advances in this area include the de novo design of two monomeric three-stranded antiparallel β-sheet peptides, Betanova[1,2] and Beta3s.[3] Peptides that have the ability to fold on their own and are well characterized experimentally are valuable not least as a testbed for theoretical models for protein folding. β-sheet peptides are particularly interesting in this respect as β-sheet formation is more challenging to model than α-helix formation. Therefore, it is no surprise that both Betanova[4,5] and Beta3s[6] have become the subject of computational studies. Simulations of peptide sequences that are somewhat similar to these and occur in natural proteins, so-called WW domains, have been reported, too.[7] For a recent review of computational studies of peptide folding, see Granakaran et al.[8]

Here we present a study of the C-terminal β-hairpin from the protein G B1 domain and a triple mutant of Betanova called LLM.[2] The original Betanova, which is less stable than the peptide LLM,[2] is considered too. These different sequences are studied using an all-atom model with a simplified interaction potential. An earlier version of this model was tested[9] on the same β-hairpin and an α-helix, the designed so-called $F_S$.[10,11] The model was able to fold these two sequences and the folded population showed, in both cases, a temperature dependence comparable with experimental data. It should be pointed out that different sequences are studied using exactly the same parameters; the interaction potential is, like that of Kussell et al.[12] but unlike many other simplified potentials for protein folding, entirely sequence-based. This is of importance even if only one sequence is studied, because it ensures that the formation and breaking of nonnative bonds is not a neglected part of the dynamics.

## MATERIALS AND METHODS

The model we study is a revised version of an earlier model.[9] It contains all atoms of the polypeptide chain, including hydrogens, but no explicit water molecules. All bond lengths, bond angles, and peptide torsion angles (180°) are held fixed, so each amino acid has the Ramachandran torsion angles $\phi$, $\psi$, and a number of side-chain torsion angles as its degrees of freedom (for Pro, $\phi$ is held fixed at $-65°$). All bond lengths and bond angles are the same as in the original model.[9]

The potential function

$$E = E_{\mathrm{ev}} + E_{\mathrm{loc}} + E_{\mathrm{hp}} + E_{\mathrm{hb}} \tag{1}$$

is composed of four terms. The remaining part of this section describes these different terms, with emphasis on what is new compared with the earlier model. Energy parameters are quoted in dimensionless units. To set the energy scale of the model, we use the midpoint temperature for the β-hairpin as determined by Muñoz et al.,[13] $T_{\mathrm{m}} = 297$ K, which corresponds to $kT \approx 0.440$ in the model.

The first term in Eq. 1, $E_{\mathrm{ev}}$, represents excluded-volume effects and has the form

$$E_{\mathrm{ev}} = \kappa_{\mathrm{ev}} \sum_{i<j} \left[ \frac{\lambda_{ij}(\sigma_i + \sigma_j)}{r_{ij}} \right]^{12}, \tag{2}$$

where $\kappa_{\mathrm{ev}} = 0.10$ and $\sigma_i = 1.77$, 1.75, 1.55, 1.42, and 1.00 Å for S, C, N, O, and H atoms, respectively. The role of the parameter $\lambda_{ij}$ is to reduce the repulsion between nonlocal pairs; $\lambda_{ij} = 1$ for all pairs connected by three covalent bonds and $\lambda_{ij} = 0.75$ otherwise. The reason for using $\lambda_{ij} < 1$ for nonlocal pairs is partly computational efficiency and partly the restricted flexibility of chains with only tor-

Grant sponsor: Swedish Foundation for Strategic Research; Grant sponsor: Swedish Research Council.

*Correspondence to: A. Irbäck, Complex Systems Division, Department of Theoretical Physics, Lund University, Sölvegatan 14A, SE-223 62 Lund, Sweden. E-mail: anders@thep.lu.se

Received 29 September 2003; Accepted 13 November 2003

Published online 7 May 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20157

**TABLE I. The Interaction Matrix $M_{IJ}$ (see Eq. 4)**[†]

| | I | II | III |
|---|---|---|---|
| I Ala | 0.0 | 0.1 | 0.1 |
| II Ile, Leu, Met, Val | | 0.9 | 2.8 |
| III Phe, Trp, Tyr | | | 3.2 |

[†]All amino acid pairs not occurring in the table have $M_{IJ} = 0$.

sional degrees of freedom. To speed up the calculations, the sum in Eq. 2 is evaluated using a cutoff of $r_{ij}^c = 4.3\lambda_{ij}$ Å.

The second interaction term, $E_{\rm loc}$, is new compared with the earlier model. By introducing this term and modifying $\sigma_i$ for C and N, we slightly adjusted the shape of the Ramachandran $\phi$, $\psi$ distribution. $E_{\rm loc}$ is a local electrostatic energy given by

$$E_{\rm loc} = \kappa_{\rm loc} \sum_I \rho_I \left( \sum \frac{q_i q_j}{r_{ij}^{(I)}/\text{Å}} \right), \tag{3}$$

where the outer sum runs over all non-Pro amino acids along the chain and the inner sum represents the interaction between the partial charges of the backbone NH and C′O groups within one amino acid (the sum has four terms: NC′, NO, HC′, and HO). The partial charges are $q_i = \pm 0.20$ for H and N and $q_i = \pm 0.42$ for C′ and O.[14] We put $\kappa_{\rm loc} = 125$, which corresponds to a dielectric constant of $\epsilon_r \approx 2.0$ if $\rho_I = 1$. The factor $\rho_I$ reduces the interaction strength for the two end amino acids and Gly, which can be viewed as a simple form of context dependence; $\rho_I = 0.25$ for end amino acids, $\rho_I = 0.5$ for Gly, and $\rho_I = 1$ otherwise. A similar factor is used for $E_{\rm hb}$ (see below).

The third term in Eq. 1, $E_{\rm hp}$, is an effective attraction between hydrophobic side chains that are not nearest or next-nearest neighbors along the chain. It has the pairwise additive form

$$E_{\rm hp} = -\sum M_{IJ} C_{IJ}, \tag{4}$$

where $C_{IJ}$ is a measure of the degree of contact between side chains $I$ and $J$, and $M_{IJ}$ sets the energy that a pair in contact gets. The contact measure $C_{IJ}$ is a number between 0 and 1, defined as before.[9] The interaction matrix $M_{IJ}$ is given in Table I and differs from that used in our earlier study, which was based on the Miyazawa-Jernigan contact energies.[15] With an all-atom representation, this cannot be expected to be a good choice for more general sequences, since the Miyazawa-Jernigan contact energies were derived using a different, reduced chain representation.[15] The new matrix $M_{IJ}$ has a simplified structure in that the hydrophobic amino acids are grouped into three classes (see Table I). The $M_{IJ}$ values are taken to be large for the aromatic class (Phe, Trp, Tyr), which in part is an attempt to compensate for the fact that it is relatively difficult for these large side chains with few degrees of freedom to make proper contacts.

The last term of the potential, the hydrogen-bond energy $E_{\rm hb}$, is given by

$$E_{\rm hb} = \epsilon_{\rm hb}^{(1)} \sum_{\rm bb-bb} \rho_{ij} u(r_{ij}) v(\alpha_{ij}, \beta_{ij}) + \epsilon_{\rm hb}^{(2)} \sum_{\rm sc-bb} \rho_{ij} u(r_{ij}) v(\alpha_{ij}, \beta_{ij}), \tag{5}$$

where the two terms represent backbone–backbone interactions and interactions between the backbone and charged side chains, respectively. The second term in Eq. 5 does not include any side chain–side chain interactions, as it did in our earlier study. Apart from that, the only difference compared with the earlier model is the factor $\rho_{ij}$, which like $\rho_I$ in Eq. 3 can be seen as a simple form of context dependence. We put $\rho_{ij} = 0.25$ if any of the two amino acids involved is an end amino acid, $\rho_{ij} = 0.5$ if any of them is a Gly, and $\rho_{ij} = 1$ otherwise. The constants $\epsilon_{\rm hb}^{(1)} = 3.1$ and $\epsilon_{\rm hb}^{(2)} = 2.0$ as well as the functions $u$ and $v$ are exactly the same as before.[9]

To study the thermodynamic behavior of this model, we use simulated tempering,[16,17] in which the temperature is a dynamical variable. Details on our implementation of this method can be found elsewhere.[18] For a review of simulated tempering and other generalized-ensemble techniques for protein folding, see Hansmann and Okamoto.[19] Eight different temperatures are studied, ranging from 284 to 371 K. For the backbone degrees of freedom, we use three different elementary moves: first, the pivot move[20] in which a single torsion angle is turned; second, a semi-local method[21] that works with up to eight adjacent torsion angles, which are turned in a coordinated way; and third, a symmetry-based update of three randomly chosen backbone torsion angles.[9] For the side-chain degrees of freedom, we use simple Metropolis updates of individual angles.

For each peptide, eight independent Monte Carlo runs were performed, starting from random conformations. Each run required a few days on a standard desktop computer, and contained several folding/unfolding events. The similarity between the results from the different runs strongly suggest that the simulations did map out all relevant free-energy minima of the model. All statistical errors quoted are 1σ errors obtained from the variance between the runs. The fits of data discussed in the next section were carried out by using a Levenberg-Marquardt procedure.[22]

For a given protein structure, there generally exist alternative structures with similar secondary-structure content but different overall topologies. This holds true even for a small β-hairpin, for which a flip of the side chains gives rise to a topologically distinct structure. To make models discriminate between different topologies is a delicate task. To assess whether or not a model is able to do that, it is necessary to make a suitable choice of observables. In our calculations, we monitor two variables that can be used for this purpose: first, the root-mean-square deviation (rmsd) from the folded structure, $\Delta$, calculated over all non-H atoms (a backbone rmsd is much less informative); and second, the number of native backbone–backbone hydrogen bonds, $N_{\rm hb}^{\rm nat}$. Figure 1 illustrates which hydrogen bonds we take to be present in the native states of the peptides studied. In our calculations, a hydrogen bond is considered formed if the energy is less than $-\epsilon_{\rm hb}^{(1)}/3$ (see Eq. 5).

Using the original model, we studied the α-helical $F_S$ peptide and a β-hairpin.[9] Here, we study the same β-hair-

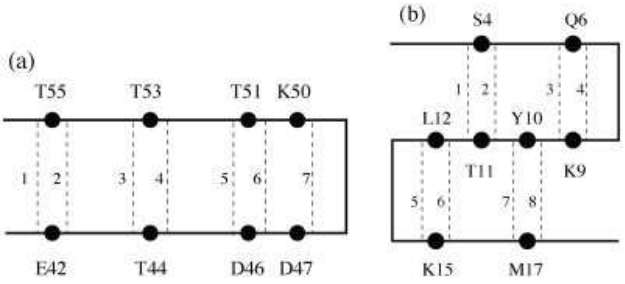


Fig. 1. Schematic illustration of the backbone–backbone hydrogen bonds taken as native for (**a**) the C-terminal β-hairpin from the protein G B1 domain,[13,23] and (**b**) the mutant LLM of Betanova.[2] (b) is used for the original Betanova, too (with L12 and M17 replaced by N12 and T17, respectively).

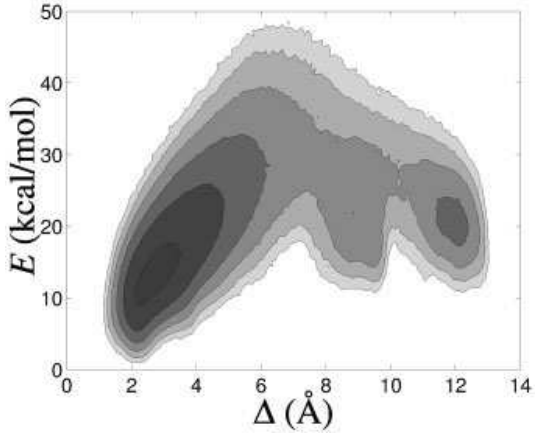


Fig. 2. Free energy $F(\Delta, E)$ for $F_S$ at $T = 284$ K, where $\Delta$ denotes heavy-atom rmsd from an ideal α-helix and $E$ is energy. The contours are spaced at intervals of 1 $kT$ and dark tone corresponds to low free energy. Contours more than 6 $kT$ above the minimum free energy are not shown.

pin and two three-sided β-sheet peptides, LLM and Betanova. Before turning to these results, it should be pointed out that the $F_S$ sequence still makes an α-helix in the revised model, as can be seen from the free energy $F(\Delta, E)$ in Figure 2. $F(\Delta, E)$ has a pronounced, dominating minimum at $\Delta \approx 2$–3.5 Å, which corresponds to α-helix. In addition, there are weakly populated minima corresponding to β-sheet structures at $\Delta \approx 9$ Å and $\Delta \approx 12$ Å.

## RESULTS AND DISCUSSION

### β-Hairpin

Using the model described in the previous section, we first study the 16–amino acid C-terminal β-hairpin from the protein G B1 domain. An important quantity monitored in our earlier study of this peptide[9] was the hydrophobicity energy $E_{hp}$. This variable should be strongly correlated with Trp fluorescence, which Muñoz et al.[13] used to characterize the melting behavior of this peptide. The temperature dependence of $E_{hp}$ was found to be in reasonable agreement with the data of Muñoz et al. Several other groups have performed atomic simulations of the same β-hairpin, with[24–28] or without[12,28–30] explicit water. In contrast to ours, most models seem to require further calibration in order not to show a temperature dependence much weaker than that of experimental data.

Figure 3(a) shows the temperature dependence of $E_{hp}$ in the revised model. The line is a fit of the data to a simple (first-order) two-state expression. The parameters of the fit are the midpoint temperature $T_m$, the energy difference $\Delta E$, and two baselines. We use the parameter $T_m$ to set the energy scale of the model; this parameter is taken as $T_m = 297$ K as determined by Muñoz et al.[13] For the energy difference, we then obtain $\Delta E = 13.1$ kcal/mol. These values of the two-state parameters $T_m$ and $\Delta E$ correspond to a native population of 74% at $T = 284$ K, which agrees well with the result of Muñoz et al., 72% at $T = 284$ K.[13] The NMR analysis of Blanco et al.[23] gave, by contrast, a lower native population, 42% at $T = 278$ K. A possible explanation of this discrepancy would be that this peptide does not show a clear two-state behavior; the apparent native population may then very well depend on which quantity is studied. At first glance, this explanation may seem unlikely, given that the temperature dependence of the Trp fluorescence data to a good approximation showed two-state character.[13] Let us, therefore, stress that, despite that the two-state fit in Figure 3(a) looks quite good, this sequence does not show ideal two-state behavior in our model. This can be seen, for example, from the energy distribution, which lacks a clear bimodal shape. This was shown in our earlier study,[9] and holds true in the revised model as well. Similar results have also been obtained in simulations of a designed, fast-folding three-helix-bundle protein.[31]

In Figure 3(b) we show the temperature dependence of the number of native hydrogen bonds, $N_{hb}^{nat}$, which we expect to be more strongly correlated than $E_{hp}$ with the NMR measurements of Blanco et al.[23] For $N_{hb}^{nat}$, a two-state fit is not meaningful; for that, further data at lower temperatures would be needed. On the other hand, the quantity $N_{hb}^{nat}$ can be used as a direct measure of nativeness. Based on inspection of many examples, we use as a criterion for nativeness that at most two of the native hydrogen bonds should be missing, which can be used for the two other peptides too (see below). For the β-hairpin with seven native hydrogen bonds [see Fig. 1(a)], this criterion ($N_{hb}^{nat} \geq 5$) gives a native population of 39% at $T = 284$ K. This value is close to the estimate of Blanco et al.,[23] 42% at $T = 278$ K. Due to uncertainties about the precise definitions of nativeness and of when a hydrogen bond is formed, this agreement could be somewhat accidental. There is no doubt, however, that the native population obtained using $N_{hb}^{nat}$ is significantly lower than that obtained above using $E_{hp}$. Figure 4 shows the probability distributions of $N_{hb}^{nat}$ at $T = 284$ K and $T = 306$ K. The number of native hydrogen bonds is seen to rapidly decrease with increasing $T$, as it should.

The two-state parameter $\Delta E$ extracted from the $E_{hp}$ data is somewhat smaller here, $\Delta E = 13.1$ kcal/mol, than it was in our earlier study, $\Delta E = 16.1$ kcal/mol.[9] The reason for this is not so much that the model has changed, but rather that the fits were done in different ways. In our previous study, $T_m$ was held fixed at the specific heat maximum.

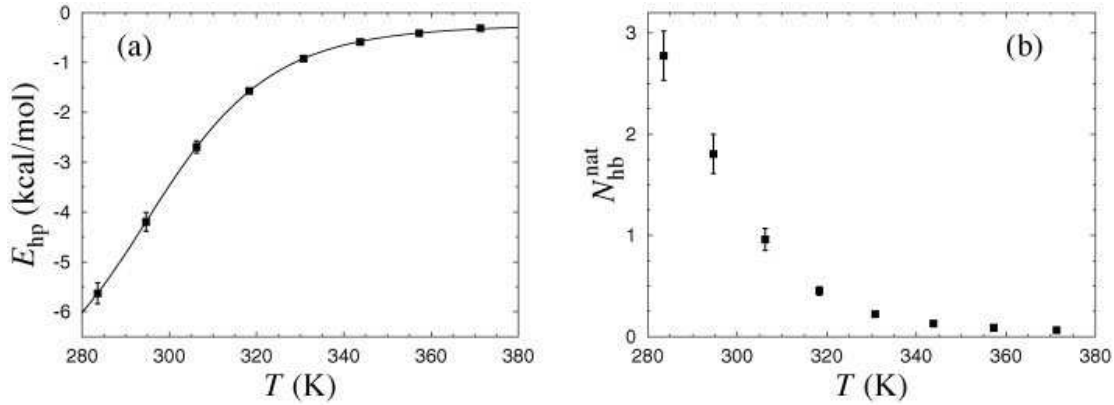


Fig. 3. The temperature dependence of (**a**) the hydrophobicity energy $E_{hp}$ and (**b**) the number of native hydrogen bonds, $N_{hb}^{nat}$, for the β-hairpin. The line in (a) is a fit to the two-state expression $E_{hp}(T) = (E_{hp}^{u} + E_{hp}^{n}K(T))/(1 + K(T))$, where $K(T) = P^{n}(T)/P^{u}(T)$, $P^{n}(T)$ and $P^{u}(T)$ are the populations of the native and unfolded states, respectively. The effective equilibrium constant $K(T)$ is assumed to have the first-order form $K(T) = \exp[(1/kT - 1/kT_m)\Delta E]$, where $T_m$ is the midpoint temperature and $\Delta E$ is the energy difference between the two states. The baselines $E_{hp}^{n}$ and $E_{hp}^{u}$ are taken as constants.

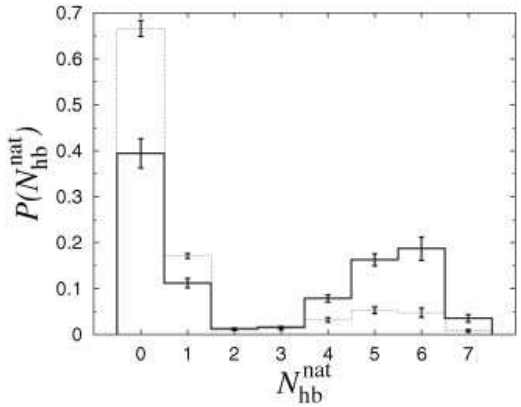


Fig. 4. Histogram of the number of native hydrogen bonds, $N_{hb}^{nat}$, at $T$ = 284 K (solid line) and $T$ = 306 K (dashed line) for the β-hairpin.

Here, following the analysis of Muñoz et al.[13] more closely, we take $T_m$ to be a parameter of the fit. The fitted value of $T_m$ turns out to lie slightly below (1–2%) the specific heat maximum. Our new analysis improves the agreement with the result of Muñoz et al.,[13] which was $\Delta E = 11.6$ kcal/mol.

Although the precise shape of the structures with lowest energy is sensitive to the details of the model, it is also interesting to make an rmsd-based comparison with experimental data. For this purpose, we use the NMR structure for the full protein G B1 domain (PDB code 1GB1, first model),[32] as the NMR restraints for the isolated β-hairpin were insufficient to determine a unique structure. Figure 5(a) shows the free energy $F(\Delta, E)$ calculated as a function of rmsd, Δ, and energy, $E$, at $T$ = 284 K. Three distinct, highly populated minima can be seen. The two minima with lowest $E$ are found at $\Delta \approx 2.0$ Å and $\Delta \approx 3.1$ Å, respectively. Both these correspond to β-hairpin structures with a high $N_{hb}^{nat}$. That $N_{hb}^{nat}$ is high implies, in particular, that the topology of the β-hairpin is the native one. The main difference between these two minima lies in the shape of the turn. The third minimum, at $\Delta \approx 4.0$ Å, is somewhat higher in $E$ than the first two. This minimum is also dominated by β-hairpin structures with the native topology and many hydrogen bonds, but the two strands tend to be out of register with each other, so $N_{hb}^{nat}$ is low. At least partly, it is the existence of this third minimum that makes the apparent native population depend on which of the observables $E_{hp}$ and $N_{hb}^{nat}$ we use. Finally, there are also two weakly populated free-energy minima corresponding to β-sheet structures with the nonnative topology ($\Delta \approx$ 5.3 Å) and α-helix ($\Delta \approx 8$–10 Å), respectively.

### Three-Stranded β-Sheets

The de novo design of the 20-amino acid three-stranded antiparallel β-sheet peptide Betanova was reported in 1998.[1] Recently, mutants of this peptide with higher stability were created by López de la Paz et al.[2] Among the most stable mutants found was the triple mutant LLM (Val5Leu, Asn12Leu, Thr17Met). The peptide LLM and the original Betanova were estimated[2] to have native populations of 36 and 9%, respectively, at $T$ = 283 K, based on NMR data. Melting curves have, as far as we know, not been reported for these peptides.

Our simulations of LLM show first of all that this sequence does make a three-stranded antiparallel β-sheet in this model. This can be seen from Figure 5(b), which shows the free energy $F(\Delta, E)$ at $T$ = 284 K. The free energy has a broad minimum at $\Delta \approx 3$–5 Å, corresponding to β-sheet structures with the native topology and a high $N_{hb}^{nat}$. The shape of the β-sheet varies within the minimum. At $\Delta \approx 3.4$ Å, where the free energy is lowest, the β-sheet has a bent shape, which enables the chain to make strong hydrophobic contacts. At $\Delta \approx 4.5$ Å, the β-sheet tends to be much flatter, which is hydrophobically disfavored but makes it possible for the chain to form more perfect hydrogen bonds. There is also a free-energy minimum at $\Delta \approx 6.5$ Å, which corresponds to three-stranded antiparallel β-sheet structures with the nonnative topology. However, the native topology is the thermodynamically favored one. Note that the native and nonnative topologies exhibit nonoverlapping sets of backbone–backbone hydrogen bonds, so $N_{hb}^{nat}$ is low at the $\Delta \approx 6.5$ Å minimum.

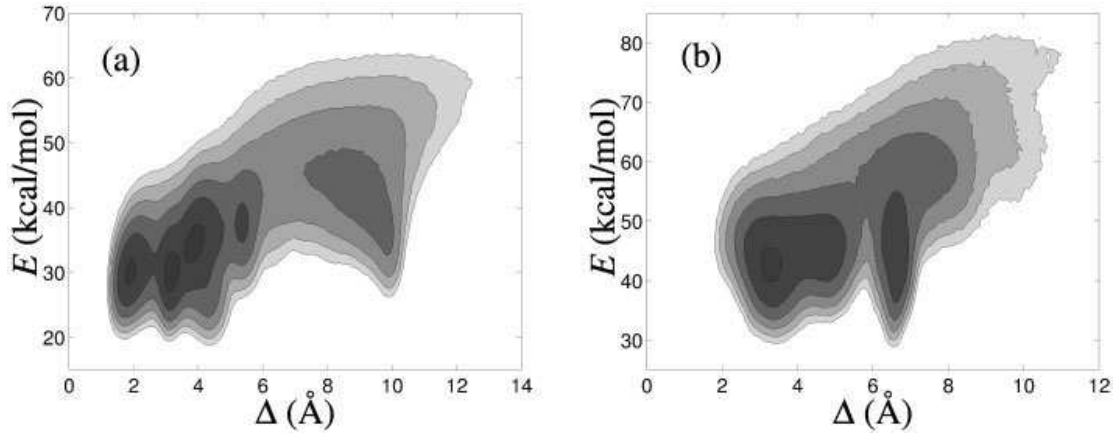


Fig. 5. Free energy $F(\Delta, E)$ at $T = 284$ K for (**a**) the β-hairpin and (**b**) the peptide LLM. $E$ is energy and $\Delta$ is a heavy-atom rmsd, calculated using all the 16 amino acids for the β-hairpin and amino acids 3–18 for LLM. The first two and last two amino acids of LLM do not take part in the β-sheet structure. The contour levels are as in Figure 2.

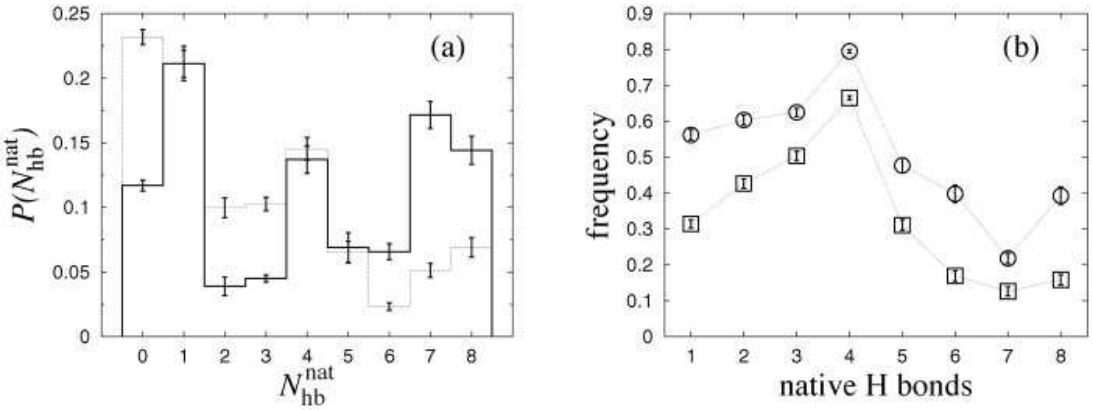


Fig. 6. (**a**) Histogram of the number of native hydrogen bonds, $N_{hb}^{nat}$, at $T = 284$ K for LLM (solid line) and Betanova (dahsed line). (**b**) The frequency of occurrence for the eight native hydrogen bonds [labeled according of Fig. 1(b)] for LLM (○) and Betanova (□) at $T = 284$ K.

The main reason why the model favors the native topology over the nonnative one lies in the side-chain orientations for the hydrophobic pairs Trp3–Leu12 and Leu5–Tyr10. The $C_\alpha$–$C_\beta$ vectors of these pairs point inwards in the nonnative topology, which makes it difficult to achieve proper contacts between the side chains. This is much easier to accomplish in the native topology, where the $C_\alpha$–$C_\beta$ vectors point outwards. Interestingly, the situation is similar for the β-hairpin above.[9] The β-hairpin also has two pairs of hydrophobic side chains that are "bow-legged" in the native topology and "knock-kneed" in the nonnative one.

Next we estimate the native population for LLM. As we want to compare with the NMR-based results of López de la Paz et al.,[2] we consider $N_{hb}^{nat}$ rather than $E_{hp}$. Figure 6(a) shows the $N_{hb}^{nat}$ distribution at $T = 284$ K. In addition to the native and nonnative peaks at high and low $N_{hb}^{nat}$, respectively, this distribution exhibits a third peak at $N_{hb}^{nat} = 4$. The typical conformation at this peak contains only the first of the two native β-turns [see Fig. 1(b)]. The second β-turn is less stable, as will be discussed below. Using the criterion that at most two native hydrogen bonds should be missing ($N_{hb}^{nat} \geq 6$), we obtain a native population of 38% at $T = 284$ K for LLM, which agrees well with the result of López de la Paz et al.,[2] 36% at $T = 283$ K. We also performed simulations of the original Betanova, and Figure 6(a) shows the result for this sequence too. From Figure 6 it is evident that Betanova is less stable than LLM. The probability that $N_{hb}^{nat} \geq 6$ is 14% for Betanova at $T = 284$ K, which means that this criterion gives a native population close to the NMR-based result of López de la Paz et al.[2] not only for LLM but also for Betanova.

That the model predicts LLM to be more stable than Betanova is not surprising because LLM has a more pronounced hydrophobic core. The agreement with experimental data is, nevertheless, remarkably good, especially since these calculations do not involve any adjustable parameter; the energy scale of the model is fixed using melting data for the β-hairpin and is then left unchanged.

Figure 6(b) shows the frequencies of occurrence for the different native hydrogen bonds [see Fig. 1(b)] for LLM and Betanova. For Betanova, there is a clear difference between the hydrogen bonds involved in the first β-turn (1–4) and those involved in the second β-turn (5–8). The

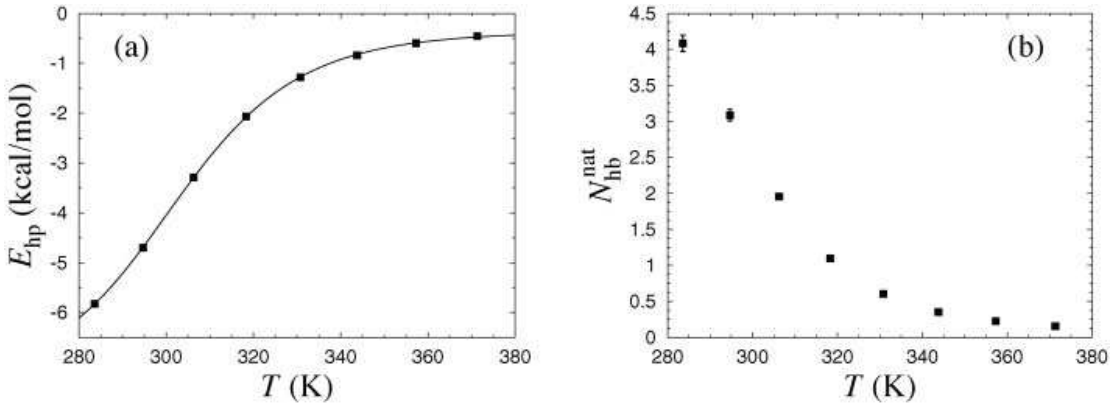


Fig. 7. The temperature dependence of (**a**) the hydrophobicity energy $E_{hp}$ and (**b**) the number of native hydrogen bonds, $N_{hb}^{nat}$, for the LLM peptide. The line in (a) is a first-order two-state fit (as in Fig. 3).

latter four occur infrequently, showing that the second β-turn is quite unstable, which is in line with the conclusions of López de la Paz et al.[2] For LLM, the difference in stability between the two β-turns is less pronounced. However, hydrogen bond 7, which connects Met17 to Tyr10 [see Fig. 1(b)], is quite unstable. The reason for this is that the side chain of Met17 can make better contacts with other hydrophobic side chains if the strand is slightly bent. This bend makes it difficult for hydrogen bond 7 to form.

Finally, in Figure 7 we show the temperature dependence of $E_{hp}$ and $N_{hb}^{nat}$ for LLM. As in the β-hairpin case, we find that a simple two-state fit provides a good description of the data for $E_{hp}$. The fitted values of the parameters $T_m$ and $\Delta E$ are $T_m$ = 303 K and $\Delta E$ = 13.0 kcal/mol, which means that the native population obtained from this fit is significantly higher than that obtained from the $N_{hb}^{nat}$ distribution [see Fig. 6(a)]. So, the model predicts that the apparent native population depends on which observable is used for this sequence, too. We are not aware of any existing experimental data that support, or refute, this conclusion for LLM.

## CONCLUSION

Using a novel all-atom model with a simplified sequence-based potential, we have investigated the equilibrium behaviors of three β-sheet peptides. We determined native populations for these peptides in two ways, from the distribution of the number of native hydrogen bonds ($N_{hb}^{nat}$) and from the temperature dependence of the hydrophobicity energy ($E_{hp}$). These two estimates were compared with experimental results based on NMR and Trp fluorescence, respectively. This comparison is summarized in Table II. The agreement with experimental data is good, which in particular means that the model to a good approximation is able to reproduce the relative stabilities of these three peptides, as obtained from the NMR measurements. In line with the experimental results on the β-hairpin, we find that the apparent native population depends on whether we use $N_{hb}^{nat}$ or $E_{hp}$. This reflects the fact that the melting transition is not a clear two-state transition in our model (for any of these three sequences). It is also worth noting that, despite that the two-state picture is an oversimplification, the temperature dependence of $E_{hp}$ is quite well described by a simple two-state expression [see Figs. 3(a) and 7(a)]. Computational studies of the β-hairpin have also been performed by many other groups, but the temperature dependence obtained was typically too weak, as has been pointed out by Zhou et al.[27] Our model shows a temperature dependence that is in good agreement with experimental data.

**TABLE II. Summary of Apparent Native Populations Obtained From Simulations and Experimental Data, Respectively**[†]

| | Model, 284 K (%) | | Experiment (%) | |
|---|---|---|---|---|
| | $N_{hb}^{nat}$ | $E_{hp}$ | NMR | Trp fluorescence |
| β-hairpin | 39 | 74 | 42 (278 K)[23] | 72 (284 K)[13] |
| LLM | 38 | | 36 (283 K)[2] | |
| Betanova | 14 | | 9 (283 K)[2] | |

[†]The model results have statistical errors of 1–4%.

Our study of these three different peptides was carried out using one and the same set of parameters. In addition, we showed that the $F_S$ peptide makes an α-helix for this choice of parameters. While these results are very encouraging, it is important to stress that we do not expect the model to be directly applicable to other sequences. However, by confronting the model with new sequences, we hope it will be possible to refine the potential, and thereby further extend its applicability. The present study was a first step in this direction, in which the model was improved by studying LLM and Betanova. To make the model able to fold these sequences, many changes were made, several of which were minor. The two most important changes were the replacement of the old hydrophobicity matrix ($M_{I,J}$) and the introduction of a simple form of context dependence for the hydrogen bonds. Whether it will be possible to carry on this process to a point where the model correctly reproduces the thermodynamics of small proteins remains to be seen. One thing that probably will

be necessary in order to achieve this goal is to include multibody effects in the hydrophobicity potential; the present pairwise additive potential is likely to become insufficient as the chains get larger. Computationally, there is room for extending the calculations to larger chains; the calculations presented here required about two weeks on a standard desktop computer for each peptide.

## ACKNOWLEDGMENTS

We thank Luis Serrano and Manuela López de la Paz for providing NMR data for LLM and Betanova. This work was in part supported by the Swedish Foundation for Strategic Research and the Swedish Research Council.

## REFERENCES

1. Kortemme T, Ramírez-Alvarado M, Serrano L. Design of a 20-amino acid, three-stranded β-sheet protein. Science 1998;281:253–256.
2. López de la Paz M, Lacroix E, Ramírez-Alvarado M, Serrano L. Computer-aided design of β-sheet peptides. J Mol Biol 2001;312:229–246.
3. de Alba E, Santaro J, Rico M, Jiménez MA. De novo design of a monomeric three-stranded antiparallel β-sheet. Protein Sci 1999;8:854–865.
4. Bursulaya BD, Brooks CL III. Folding free energy surface of a three-stranded β-sheet protein. J Am Chem Soc 1999;121:9947–9951.
5. Colombo C, Roccatano D, Mark AE. Folding and stability of the three-stranded β-sheet peptide betanova: insights from molecular dynamics simulations. Proteins 2002;46:380–392.
6. Cavalli A, Haberthür U, Paci E, Caflisch A. Fast protein folding on downhill energy landscape. Protein Sci 2003;12:1801–1803.
7. Karanicolas J, Brooks CL III. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design? Proc Natl Acad Sci USA 2003;100:3954–3959.
8. Granakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, García AE. Peptide folding simulations. Curr Opin Struct Biol 2003;13:168–174.
9. Irbäck A, Samuelsson B, Sjunnesson F, Wallin S. Thermodynamics of α- and β-structure formation in proteins. Biophys J 2003;85:1466–1473.
10. Lockhart DJ, Kim PS. Internal Stark effect measurement of the electric field at the amino terminus of an α helix. Science 1992;257:947–951.
11. Lockhart DJ, Kim PS. Electrostatic screening of charge and dipole interactions with the helix backbone. Science 1993;260:198–202.
12. Kussell E, Shimada J, Shakhnovich EI. A structure-based method for derivation of all-atom potentials for protein folding. Proc Natl Acad Sci USA 2002;99:5343–5348.
13. Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of β-hairpin formation. Nature 1997;390:196–199.
14. Branden C, Tooze J. Introduction to protein structure. New York: Garland Publishing; 1991.
15. Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density, for simulation and threading. J Mol Biol 1996;256:623–644.
16. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsov-Velyaminov PN. New approach to Monte Carlo calculation of the free energy: method of expanded ensembles. J Chem Phys 1992;96:1776–1783.
17. Marinari E, Parisi G. Simulated tempering: a new Monte Carlo scheme. Europhys Lett 1992;19:451–458.
18. Irbäck A, Potthast F. Studies of an off-lattice model for protein folding: sequence dependence and improved sampling at finite temperature. J Chem Phys 1995;103:10298–10305.
19. Hansmann UHE, Okamoto Y. New Monte Carlo algorithms for protein folding. Curr Opin Struct Biol 1999;9:177–183.
20. Lal M. Monte Carlo computer simulation of chain molecules. I. Mol Phys 1969;17:57–64.
21. Favrin G, Irbäck A, Sjunnesson F. Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. J Chem Phys 2001;114:8154–8158.
22. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes in C: the art of scientific computing. Cambridge: Cambridge University Press; 1992.
23. Blanco FJ, Rivas G, Serrano L. A short linear peptide that folds into a native stable β-hairpin in aqueous solution. Nat Struct Biol 1994;1:584–590.
24. Roccatano D, Amadei A, Di Nola A, Berendsen HJC. A molecular dynamics study of the 41–56 β-hairpin from B1 domain of protein G. Protein Sci 1999;8:2130–2143.
25. Pande VS, Rokhsar DS. Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment from protein G. Proc Natl Acad Sci USA 1999;96:9062–9067.
26. García AE, Sanbonmatsu KY. Exploring the energy landscape of a β hairpin in explicit solvent. Proteins 2001;42:345–354.
27. 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.
28. Zhou R. Free energy landscape of protein folding in water: explicit vs. implicit solvent. Proteins 2003;53:148–161.
29. Dinner AR, Lazaridis T, Karplus M. Understanding β-hairpin formation. Proc Natl Acad Sci USA 1999;96:9068–9073.
30. Zagrovic B, Sorin EJ, Pande V. β-hairpin folding simulations in atomistic detail using an implicit solvent model. J Mol Biol 2001;313:151–169.
31. Favrin G, Irbäck A, Samuelsson B, Wallin S. Two-state folding over a weak free-energy barrier. Biophys J 2003;85:1457–1465.
32. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM. A novel, highly stable fold of the immunoglobulin-binding domain of streptococcal protein G. Science 1991;253:657–661.