Folding Thermodynamics of Peptides

Anders Irbäck and Sandipan Mohanty[*]

Complex Systems Division, Department of Theoretical Physics
Lund University, Sölvegatan 14A, SE-223 62 Lund, Sweden
`http://www.thep.lu.se/complex/`

Abstract:
A simplified interaction potential for protein folding studies at the atomic level is discussed and tested on a set of peptides with about 20 residues each. The test set contains both α-helical (Trp cage, F_s) and β-sheet (GB1p, GB1m2, GB1m3, Betanova, LLM) peptides. The model, which is entirely sequence-based, is able to fold these different peptides for one and the same choice of model parameters. Furthermore, the melting behavior of the peptides is in good quantitative agreement with experimental data. Apparent folded populations obtained using different observables are compared, and are found to be very different for some of the peptides (e.g., Betanova). In other cases (in particular, GB1m2 and GB1m3), the different estimates agree reasonably well, indicating a more two-state-like melting behavior.

[*]E-mail: anders, sandipan@thep.lu.se

1 Introduction

The function of peptides and proteins is inextricably connected to their folding behavior, as is underlined by the facts that many neuro-degenerative disorders are being linked to misfolding and aggregation [1], and that coupled folding and binding seems to be a more common phenomenon than previously thought [2]. It is therefore an important development that folding simulations at the atomic level are now becoming feasible for short polypeptide chains [3], thanks to faster computers, more efficient algorithms and improved force fields.

There are, however, questions about the interaction potentials used in the simulations that need further investigation. One difficulty is that different potentials give very different relative weights to the α-helix and β-strand regions of the Ramachandran space [4]. A potential that successfully folds an α-helical peptide might therefore have problems with β-sheet peptides, and *vice versa*. Another difficulty is with the temperature dependence of observable quantities. As pointed out by Zhou et al. [5], it seems that most current models need further calibration in order to give a temperature dependence that is not too weak; as a result, calculated melting temperatures are often unrealistically high. A systematic study of these thermodynamic questions requires extensive conformational sampling and is a challenge, especially in models with explicit water.

Here we study a model that contains all atoms of the polypeptide chains but no explicit solvent molecules. Formally, such a model is obtained by integrating out the solvent degrees of freedom. Finding an accurate and computationally tractable approximation of the resulting effective potential is, however, a highly non-trivial problem. Examples of implicit solvent models that have been used in folding studies with some success, include the generalized Born approach [6], the method based on screened Coulomb potentials by Hassan et al. [7], and the method based on solvent accessible surface areas by Ferrara et al. [8]. In this paper, we study a minimalistic model in which the effects of the solvent are represented by an effective attraction between nonpolar side chains. Our study focuses on the thermodynamic behavior of this model, which we investigate using efficient Monte Carlo methods rather than molecular dynamics. This choice is made for computational convenience; with some minor modifications, it would be possible to study the same model using molecular dynamics. Promising computational techniques have recently been proposed by Hansmann and Wille [9] and Schug et al. [10], but these methods are for energy minimization, which is insufficient for our purposes.

In addition to effective hydrophobic attraction, the interaction potential of our model

contains two major terms, representing excluded-volume effects and hydrogen bonding. The potential is deliberately kept simple, partly for the sake of clarity but also for practical reasons; any potential requires careful calibration, and this task is easier with a simple potential like ours with fewer parameters to tune. In the future, the potential may be further developed with the inclusion of new terms such as Coulomb interactions between side-chain charges, but not before it becomes clear that they are needed. The different terms of the potential represent either the interaction between two individual atoms (excluded volume), or two pairs of atoms (e.g., hydrogen bonds), or an effective interaction between a pair of side chains (hydrophobicity). The largest units playing a role in the potential are the amino acids, and no information about the sequence as a whole or its native structure is used in the potential.

Our approach towards the problem of determining the interaction potential is phenomenological. The shape of individual terms is inspired by intuitive notions rather than being rigorously derived from a microscopic picture. Their exact functional forms and relative sizes are constrained by the effectiveness of the model in describing the folding behavior of more and more sequences. When such a potential evolves to a point where it can successfully fold a significant number of peptides of different native geometries, and capture the thermodynamic behavior of all those peptides, it would be useful on its own as a working potential for thermodynamic studies of new sequences, and also provide hints about the relative importance of different physical effects in protein folding.

We have previously shown that earlier versions of this model are able to fold both α-helix and β-sheet peptides [11, 12]. In this paper we present a further development of this model. We test the new model on the following set of peptides (see Fig. 1): the α-helical Trp cage [14] and F_s [15, 16], and the β-sheet peptides GB1p [17, 18], GB1m2 and GB1m3 [19], Betanova [20] and LLM [21]. Here GB1p denotes the C-terminal β-hairpin from the protein G B1 domain, while Betanova is a designed three-stranded β-sheet peptide. GB1m2 and GB1m3 are mutants of GB1p, while LLM is a mutant of Betanova, with enhanced stabilities. We find that our model provides a good description of the thermodynamic behavior of all these peptides. The same model was furthermore used in a recent study of the oligomerization properties of the amyloid $A\beta_{16-22}$ peptide [22], with very promising results.
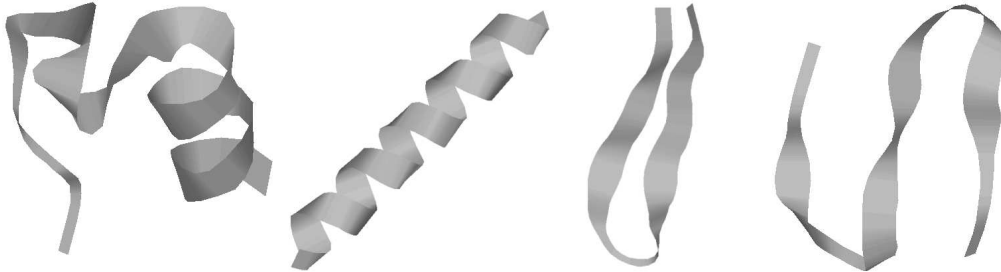
Figure 1: Schematic illustration of the different geometries of the peptides studied. Shown from left to right are the reference structures (see below) used for the Trp cage, F_s, GB1m3 and Betanova. Drawn with RasMol [13].

2 Model and Methods

2.1 Model

Our model contains all atoms of the polypeptide chains, including hydrogen atoms. The model assumes fixed bond lengths, bond angles and peptide torsion angles (180°), so that each amino acid only has the Ramachandran torsion angles ϕ, ψ and a number of side-chain torsion angles as its degrees of freedom. Numerical values of the geometrical parameters held constant can be found elsewhere [11].

In the simulations we internally use a dimensionless energy scale. The correspondence (constant factor) of this scale to the physical energy scale is determined by using the model prediction of the dimensionless energy value for an observable and the experimental value for the same. We use the melting temperature $T_m = 315\,\mathrm{K}$ of the Trp cage [14] for this purpose (see below), which is found to correspond to a dimensionless energy kT_m of 0.470 in the model (k is Boltzmann's constant). Energy parameters of the model (such as the κ_{ev}, κ_{loc}, $\epsilon_{hb}^{(1)}$, etc. below) are given in our internal energy scale. It must be emphasized that this energy scale is left unchanged when analyzing the other peptides.

The interaction potential

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

is composed of four terms. The first term in Eq. 1, E_{ev}, represents excluded-volume

effects and has the form

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

where the summation is over pairs of atoms (i, j), $\kappa_{\mathrm{ev}} = 0.10$, 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 values of the radii σ_i agree reasonably well with the statistical analysis of Tsai et al. [23]. The σ_i values for C, N and O strongly influence the shape of the Ramachandran ϕ, ψ distribution, and must therefore be carefully chosen. The parameter λ_{ij} in Eq. 2 has the value 0.75 for all pairs except those connected by three covalent bonds, for which $\lambda_{ij} = 1$. The reason why we use a reduction factor $\lambda_{ij} < 1$ for all non-local pairs is both computational efficiency and the restricted flexibility of a chain with only torsional degrees of freedom, which could create artificial traps. To speed up the calculations, Eq. 2 is evaluated using a cutoff of $r_{ij}^{\mathrm{c}} = 4.3\lambda_{ij}$ Å, and pairs with fixed separation are omitted.

The second energy term, E_{loc}, has the form

$$E_{\mathrm{loc}} = \kappa_{\mathrm{loc}} \sum_{I} \left(\sum \frac{q_i q_j}{r_{ij}^{(I)}/\mathrm{\AA}} \right),$$ (3)

where the inner sum represents the interactions between the partial charges of the backbone NH and C'O groups in one amino acid, I. This potential is not used for Pro which lacks the NH group, or Gly which tends to be more exposed to water than other amino acids, due to the missing side chain. Neither is it used for the two end amino acids, unless these are protected by capping groups. The inner sum in Eq. 3 has four terms (NO, NC', HC' and HO) which depend only on the ϕ and ψ angles for amino acid I. The partial charges are taken as $q_i = \pm 0.20$ for H and N and $q_i = \pm 0.42$ for C' and O [24], and we put $\kappa_{\mathrm{loc}} = 100$, corresponding to a dielectric constant of $\epsilon_r \approx 2.5$.

The third term of the energy function is the hydrogen-bond energy E_{hb}, which has the form

$$E_{\mathrm{hb}} = \epsilon_{\mathrm{hb}}^{(1)} \sum_{\mathrm{bb-bb}} u(r_{ij})v(\alpha_{ij}, \beta_{ij}) + \epsilon_{\mathrm{hb}}^{(2)} \sum_{\mathrm{sc-bb}} u(r_{ij})v(\alpha_{ij}, \beta_{ij}),$$ (4)

where the two functions $u(r)$ and $v(\alpha, \beta)$ are given by

$$u(r) = 5\left(\frac{\sigma_{\mathrm{hb}}}{r}\right)^{12} - 6\left(\frac{\sigma_{\mathrm{hb}}}{r}\right)^{10}$$ (5)

$$v(\alpha, \beta) = \begin{cases} (\cos\alpha \cos\beta)^{1/2} & \text{if } \alpha, \beta > 90° \\ 0 & \text{otherwise} \end{cases}$$ (6)

We consider only hydrogen bonds between NH and CO groups, and r_{ij} denotes the HO distance, α_{ij} the NHO angle and β_{ij} the HOC angle. The parameters $\epsilon_{\text{hb}}^{(1)}$, $\epsilon_{\text{hb}}^{(2)}$ and σ_{hb} are taken as 3.1, 2.0 and 2.0 Å, respectively. The function $u(r)$ is calculated using a cutoff of $r^{\text{c}} = 4.5$ Å. The first sum in Eq. 4 contains backbone-backbone interactions, while the second sum contains interactions between charged side chains (Asp, Glu, Lys and Arg) and the backbone. The latter type of interaction is taken to be effectively weak ($\epsilon_{\text{hb}}^{(2)} < \epsilon_{\text{hb}}^{(1)}$), because there are competing interactions between the side-chain charges and the surrounding water that are omitted in the model. For the same reason, we do not include any term in E_{hb} corresponding to side chain-side chain interactions. It is possible that the effective strength $\epsilon_{\text{hb}}^{(2)}$ should be made stronger in case the side-chain charge gets shielded from the water. This context dependence is ignored in the model, which should be a reasonable approximation for small peptides. Hydrogen bonds between parts that are very close in sequence are rare in protein structures and therefore disregarded in the model; specifically, we disallow backbone NH (C′O) groups to make hydrogen bonds with the two nearest backbone C′O (NH) groups on each side of them, and we also forbid hydrogen bonds between the side chain of one amino acid with the nearest donor or acceptor on either side of its C_α. As a simple form of context dependence, we assign a reduced strength to hydrogen bonds involving chain ends, which tend to be exposed to water. A hydrogen bond involving one or two end groups is reduced in strength by factors of 2 and 4, respectively. If there are capping groups, these groups are taken to be the end groups; otherwise, the two end amino acids take this role.

The fourth energy term, E_{hp}, represents an effective hydrophobic attraction between nonpolar side chains. It has the pair-wise additive form

$$E_{\text{hp}} = -\sum_{I<J} M_{IJ} C_{IJ} \,, \tag{7}$$

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 full contact gets. The matrix M_{IJ} is defined in Table 1. To calculate C_{IJ} we use a predetermined set of atoms, A_I, for each side chain I. We define C_{IJ} as

$$C_{IJ} = \frac{1}{N_I + N_J} \Big[\sum_{i \in A_I} f(\min_{j \in A_J} r_{ij}^2) + \sum_{j \in A_J} f(\min_{i \in A_I} r_{ij}^2) \Big] \,, \tag{8}$$

where the function $f(x)$ is given by $f(x) = 1$ if $x < A$, $f(x) = 0$ if $x > B$, and $f(x) = (B - x)/(B - A)$ if $A < x < B$ [$A = (3.5$ Å$)^2$ and $B = (4.5$ Å$)^2$]. Roughly speaking, C_{IJ} is the fraction of atoms in A_I or A_J that are in contact with some atom from the other side chain. For Pro, the set A_I consists of the C_β, C_γ and C_δ atoms. The definition of A_I for the other hydrophobic side chains has been given

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

Table 1: The hydrophobicity matrix M_{IJ}. Hydrophobic amino acids are divided into three categories. The matrix M_{IJ} represents the size of hydrophobicity interaction when an amino acid of type I is in contact with an amino acid of type J.

elsewhere [11]. We expect the gain in forming a hydrophobic contact to be smaller if the two side chains are close in sequence, because such a pair is partly protected by the backbone. Therefore, we reduce the strength of the hydrophobic attraction for pairs that are nearest or next-nearest neighbors along the sequence; M_{IJ} is reduced by a factor of 2 for next-nearest neighbors, and taken to be 0 for nearest neighbors.

The parameters of this potential were essentially determined by a somewhat tedious trial and error procedure, involving parallel simulations of the different peptides. The target was to have native-like free-energy minima for all the peptides at low temperature, whereas the temperature dependence was not considered at all. It is interesting to note that this criterion alone was sufficiently discriminating to yield parameter values that appear physically reasonable, as well as a realistic temperature dependence (see below). Some parameters, such as $\epsilon_{\mathrm{hb}}^{(1)}$, strongly influence the folding properties of the model, and are therefore well determined. Others, such as $\epsilon_{\mathrm{hb}}^{(2)}$, are less important and, as a result of this, quite poorly determined.

The new version of the model differs from earlier versions in the precise form of the simple context dependence of E_{loc} and E_{hb}. Also, the reduction factor for the hydrophobic attraction between next-nearest neighbors along the chain has been changed. Furthermore, we have added Pro, which does not occur in any of our previously studied sequences, to the list of hydrophobic amino acids. All other parameters of the potential are the same as in the last version of the model, except for a slight reduction in strength of the local potential (κ_{loc}).

It should be stressed that this potential is not expected to provide a good description of general amino acid sequences. For example, it is likely that the pair-wise additive hydrophobicity potential is inadequate for long chains, due to double-counting effects. For long chains, anti-cooperative multibody effects might play a significant role [25]. By extending the present calculations in the future to new and longer sequences, we hope that it will be possible to refine the potential and thereby make it more general.

2.2   Computational methods

To study the thermodynamic behavior of this model, we use simulated tempering [26–28]. in which the temperature is a dynamical variable. For a review of simulated tempering and other generalized-ensemble techniques for protein folding, see Hansmann and Okamoto [29]. We study eight different temperatures T_k, which range from $T_{\min} = 275\,\mathrm{K}$ to $T_{\max} = 369\,\mathrm{K}$ and are given by $T_k = T_{\min}(T_{\max}/T_{\min})^{(k-1)/7}$ ($k = 1,\ldots,8$). The average acceptance rate for the temperature jumps is about 70 %.

Our simulations are carried out using two different elementary moves for the backbone degrees of freedom: first, the highly non-local pivot move in which a single backbone torsion angle is turned; and second, a semi-local method [30] that works with up to eight adjacent backbone degrees of freedom, which are turned in a coordinated manner. Side-chain angles are updated one by one. Every update involves a Metropolis accept/reject step, thus ensuring detailed balance. All our simulations are started from random configurations. All statistical errors quoted are 1σ errors obtained from the variation between independent runs. For each peptide, we performed about 10 independent runs. Each run contained 10^9 elementary Monte Carlo steps ($1.5 \cdot 10^9$ steps for GB1p) and required 1–2 CPU days on a 1.6 GHz computer.

To characterize the folding behavior of the different peptides, we monitor several quantities. For a peptide with N amino acids, we define the α-helix content H as the fraction of the $N - 2$ inner amino acids with their Ramachandran (ϕ, ψ) pair in the region $-90° < \phi < -30°$, $-77° < \psi < -17°$. We calculate the radius of gyration, R_{g}, over the backbone atoms, with unit mass for all atoms. We also study root-mean-square deviations (RMSD) from folded reference structures, calculated over either the backbone atoms or all heavy atoms. A backbone RMSD is denoted by Δ_{b} and a heavy-atom RMSD by Δ. For the β-sheet peptides, there exist topologically distinct states that the backbone RMSD cannot discriminate between, which makes it necessary to use the heavy-atom RMSD.

In our analysis of the results from the simulations, it turns out that the temperature dependence of a quantity X in many cases can be well described by the simple two-state expression

$$X(T) = \frac{X_{\mathrm{u}} + X_{\mathrm{n}}K(T)}{1 + K(T)} \, . \tag{9}$$

Our fits to this equation are carried out by using a Levenberg-Marquardt procedure [31]. Throughout the paper, the baselines X_{u} and X_{n} are taken to be temperature independent, whereas the effective equilibrium constant $K(T)$ is assumed to

have the first-order form $K(T) = \exp[(1/kT - 1/kT_\mathrm{m})\Delta E]$, where T_m is the midpoint temperature and $\Delta E = E_\mathrm{u} - E_\mathrm{n}$ is the energy difference between the unfolded and native states. With these assumptions, a fit to Eq. 9 has four parameters: ΔE, T_m, X_u and X_n.

3 Results and Discussion

Using the model and methods described in the previous section, we performed high-statistics thermodynamic simulations of the peptides mentioned in the introduction, namely the Trp cage, F_s, GB1p, GB1m2, GB1m3, Betanova and LLM. In this section we present the results of these calculations.

3.1 Trp cage

The optimized 20-residue Trp cage (NLYIQWLKDGGPSSGRPPPS) is a "miniprotein" with a compact folded state and a melting temperature of $315\,\mathrm{K}$, as determined by circular dichroism (CD) and NMR measurements [14]. The NMR-derived native structure [14] contains a short α-helix (residues 2–8), a single turn of 3_{10}-helix (residues 11-14), and a hydrophobic core consisting of three proline residues (Pro12, Pro18, Pro19) and two aromatic residues (Tyr3, Trp6). The folding time is a few μs at room temperature [32]. Its small size, fast folding and relative stability makes the Trp cage an ideal testbed for computational methods, and folding simulations of this peptide were reported by several groups [10, 33–36]. Two of these groups performed thermodynamic studies [35, 36]. Both groups made detailed comparisons with raw NMR data with very good results, but the calculated melting temperatures were too high ($\gtrsim 400\,\mathrm{K}$).

In our model the melting temperature of the Trp cage is, by definition, equal to its experimental value, since we use this quantity to set the energy scale of the model. For this purpose, we consider the helix content H, as defined in the previous section, which should be strongly correlated with the CD signal studied experimentally. Fig. 2a shows our results for H against temperature. A fit to the data with the two-state expression in Eq. 9 is also shown. As can be seen in the figure, the two-state fit provides an excellent description of the data. The midpoint temperature from this fit, T_m, is set to $315\,\mathrm{K}$, the experimental melting temperature. Having done that, there is no free parameter left in the model. The fitted value of the parameter
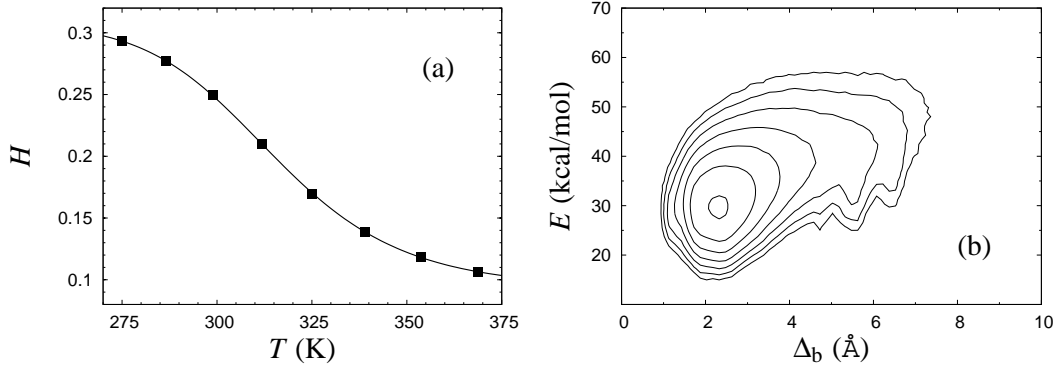
Figure 2: The Trp cage. (a) Helix content against temperature. The line is a fit to Eq. 9 ($T_\mathrm{m} = 315\,\mathrm{K}$, $\Delta E = 11.5 \pm 0.2\,\mathrm{kcal/mol}$). Statistical errors are smaller than the plot symbols. (b) Contour plot of the free energy $F(\Delta_\mathrm{b}, E)$ at $275\,\mathrm{K}$. The contours are spaced at intervals of $1\,kT$. Contours more than $6\,kT$ above the minimum free energy are not shown. The free energy $F(\Delta_\mathrm{b}, E)$ is defined by $\exp[-F(\Delta_\mathrm{b}, E)/kT] \propto P(\Delta_\mathrm{b}, E)$, where $P(\Delta_\mathrm{b}, E)$ denotes the joint probability distribution of Δ_b and E at temperature T.
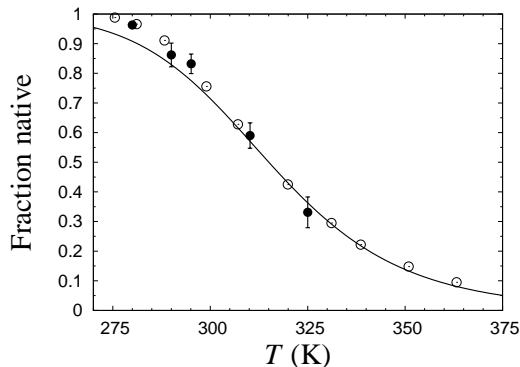

Figure 3: Native population against temperature for the Trp cage. The line is the result obtained from the model, through the fit shown in Fig. 2a. Plot symbols show experimental results [14] based on CD ($\circ$) and NMR ($\bullet$), respectively.

$\Delta E = 11.5 \pm 0.2\,\mathrm{kcal/mol}$ is, in contrast to that of T_m, not used for calibration, but is rather a prediction of the model.

In the two-state picture (Eq. 9), the native population at temperature T is given by $1/\{1+\exp[-(1/kT - 1/kT_\mathrm{m})\Delta E]\}$. Fig. 3 shows the native population obtained using the above mentioned ΔE and T_m, against temperature, along with experimental values based on CD and NMR [14]. We see that the results obtained from the model

are in good agreement with the experimental data over the entire temperature range, with a maximum deviation of $\sim 5\%$ at the lowest temperatures. With the overall energy scale properly determined, we thus find that the melting behavior of this peptide is well described by the model.

At low temperature, we find a helix content similar to that of the NMR structure, $\sim 30\%$ (see Fig. 2a). An RMSD analysis confirms that the typical low-temperature structure is similar to the NMR structure (PDB code 1L2Y, first model), as illustrated in Fig. 2b. This figure shows the free energy $F(\Delta_b, E)$ calculated as a function of the backbone RMSD Δ_b (residues 2–19) and the energy E, at 275 K. We see that $F(\Delta_b, E)$ has a simple shape with one dominating minimum, which is located at $\Delta_b \approx 2.3\,\text{Å}$.

3.2 F_s

The designed 21-residue F_s peptide is given by Suc-$A_5(AAARA)_3$A-NH_2, (where Suc is succinylic acid) and makes an α-helix [15, 16]. Other N-capping groups than Suc have also been used in the experiments on this peptide. The melting behavior of F_s was studied using CD as well as infrared (IR) spectroscopy. The melting temperature measured by IR was 334 K [37], whereas the CD-based studies obtained $T_m = 308\,\text{K}$ [16] and $T_m = 303\,\text{K}$ [38]. Computational studies of F_s have also been reported [39–41]. By explicit water simulations, García and Sanbonmatsu [40] obtained a T_m of 345 K, which is in reasonable agreement with the IR-based value. Using an earlier version of our model and ignoring the capping groups, a T_m of 310 K was obtained [11]. In the present calculations, we include the Suc and NH_2 groups.

Fig. 4a shows the helix content versus temperature as obtained from our F_s calculations. A two-state fit of the data gives $T_m = 304 \pm 1\,\text{K}$, which is significantly lower than the IR-based result mentioned above but in perfect agreement with the CD studies, especially that of Thompson et al. [38]. For the energy difference, we obtain $\Delta E = 11.9 \pm 0.3\,\text{kcal/mol}$, which also agrees with what Thompson et al. found, namely $\Delta E = 12 \pm 2\,\text{kcal/mol}$. It may be worth noting that the experimental data that we compared with in the Trp cage case were based on CD rather than IR.

In Fig. 4b we show the free energy $F(\Delta_b, E)$ at 275 K. In the absence of a precise experimental structure for F_s, we define Δ_b as the (backbone) RMSD from an ideal α-helix (all residues). From the figure we see that the free energy has its global minimum at $\Delta_b \approx 0.5\,\text{Å}$, which indeed corresponds to the α-helix. There are also
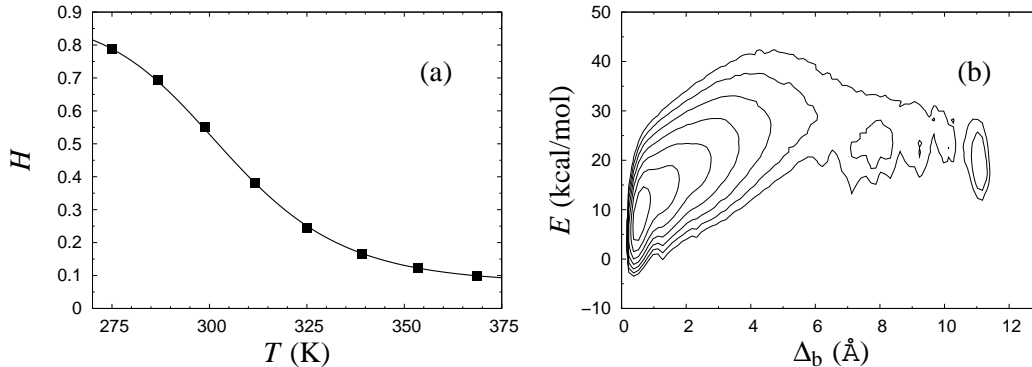
Figure 4: Same as Fig. 2 for the F_s peptide ($T_m = 304 \pm 1\,$K, $\Delta E = 11.9 \pm 0.3\,$kcal/mol).

two local minima at $\Delta_b \approx 7\,$Å and $\Delta_b \approx 11\,$Å, both of which correspond to β-sheet structures. These two minima are very weakly populated compared to the α-helix minimum.

3.3   GB1p and GB1m2/GB1m3

Using exactly the same model, we now turn to β-sheet peptides. That GB1p (GEW-TYDDATKTFTVTE), the 41–56-residue fragment from the protein G B1 domain, makes a β-hairpin on its own was a breakthrough discovery [18] that has been followed by numerous atomic simulations of this particular sequence [5, 42–50]. Recently, two mutants of GB1p with enhanced stability were designed [19], GB1m2 and GB1m3, by replacing the turn segment DDATKT by NPATGK. The mutant GB1m2 (GEW-TYNPATGKFTVTE) is identical to GB1p except for this change, while GB1m3 (KKWTYNPATGKFTVQE) differs from GB1p at the chain ends as well. By CD and NMR, GB1m3 was estimated to be $86 \pm 3\%$ folded at $298\,$K and to have a T_m of $333 \pm 2\,$K, whereas GB1m2 was found to have a slightly lower folded population, $74 \pm 5\,\%$ at $298\,$K, and a T_m of $320 \pm 2\,$K [19]. In the same study, GB1p was estimated to be $\sim 30\,\%$ folded at $298\,$K. An earlier NMR study found GB1p to be $42\,\%$ folded at $278\,$K [18]. Both these estimates of native population for GB1p are low compared to the result of a Trp fluorescence study [51]; a two-state analysis of these data gave $T_m = 297\,$K and $\Delta E = 11.6\,$kcal/mol [51].

It turns out that our model fails to reproduce the experimental difference in stability between GB1m2 and GB1m3. In fact, GB1m2 and GB1m3 show nearly identical
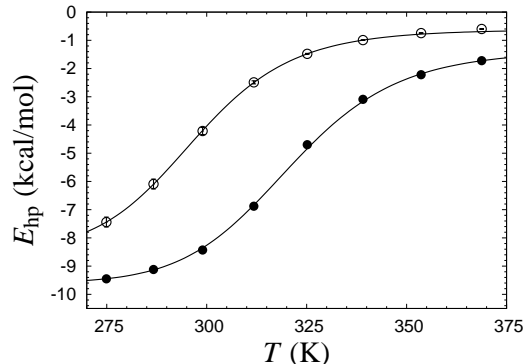
Figure 5: The hydrophobicity energy E_{hp} against temperature for GB1p ($\circ$) and GB1m3 ($\bullet$). The lines are fits to Eq. 9 ($T_m = 297 \pm 1\,\mathrm{K}$, $\Delta E = 14.2 \pm 0.2\,\mathrm{kcal/mol}$ for GB1p; $T_m = 321 \pm 1\,\mathrm{K}$, $\Delta E = 15.0 \pm 0.4\,\mathrm{kcal/mol}$ for GB1m3). The points corresponding to the two highest temperatures were omitted for GB1p, as removing them resulted in a significantly better fit in terms of χ^2 per degree of freedom.

behavior in our model. For clarity, we therefore show results only for one of these peptides, GB1m3, in the figures below.

Fig. 5 shows the hydrophobicity energy E_{hp} against temperature for GB1p and GB1m3 in the model. We expect E_{hp} to be strongly correlated with Trp fluorescence for these peptides, as Trp43 forms a hydrophobic cluster together with Tyr45, Phe52 and Val54. A two-state fit to our data for GB1p gives $T_m = 297 \pm 1\,\mathrm{K}$ and $\Delta E = 14.2 \pm 0.2\,\mathrm{kcal/mol}$, which indeed is in good agreement with the Trp fluorescence results for this peptide ($T_m = 297\,\mathrm{K}$, $\Delta E = 11.6\,\mathrm{kcal/mol}$). The same type of fit gives $T_m = 321 \pm 1\,\mathrm{K}$ and $\Delta E = 15.0 \pm 0.4\,\mathrm{kcal/mol}$ for GB1m3, and $T_m = 322 \pm 2\,\mathrm{K}$ and $\Delta E = 15.1 \pm 0.4\,\mathrm{kcal/mol}$ for GB1m2. These two very similar T_m estimates lie close to the experimental result for GB1m2 ($320 \pm 2\,\mathrm{K}$) and somewhat below that for GB1m3 ($333 \pm 2\,\mathrm{K}$). Our E_{hp} data indicate that GB1m2 and GB1m3 indeed are markedly more stable than GB1p in the model, which is confirmed by the results discussed next.

Fig. 6a shows our data for the free energy $F(\Delta, E)$ for GB1p, at $275\,\mathrm{K}$. On its own the GB1p fragment is believed to adopt a folded structure similar to that it has as part of the native protein G B1 domain, although the NMR restraints were insufficient to determine a unique structure for the excised fragment. As reference structure in the calculation of Δ, we therefore use the corresponding fragment of the NMR structure for the full protein G B1 domain (PDB code 1GB1, residues 41–56, first model) [52]. The heavy-atom RMSD Δ is used instead of the backbone RMSD Δ_b,

13
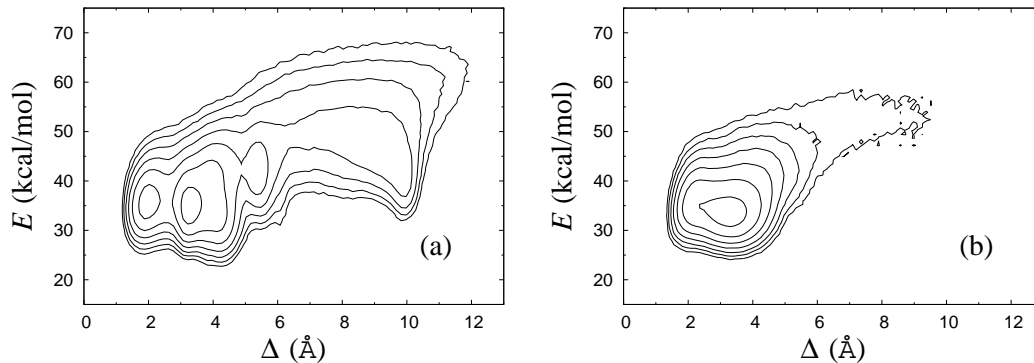
Figure 6: Contour plot of the free energy $F(\Delta, E)$ for (a) GB1p and (b) GB1m3, at 275 K. Contour levels are as in Fig. 2b.

because Δ_b cannot distinguish between the two possible β-hairpin topologies (with similar backbone folds but oppositely oriented side chains). We find that the two lowest minima of $F(\Delta, E)$, at $\Delta \approx 2.0$ Å and $\Delta \approx 3.2$ Å, both correspond to a β-hairpin with the same topology and the same set of backbone hydrogen bonds as the reference structure. The main difference between these two minima lies in the shape of the turn region. In addition to these minima, there are two weakly populated local minima at $\Delta \approx 5.3$ Å and $\Delta \approx 8$–10 Å, which correspond to a β-hairpin with the opposite topology and α-helix, respectively. The shape of $F(\Delta, E)$ for GB1p was also studied using earlier versions of our model [11, 12]. The present model yields very similar results, with a minor enhancement of the two native-like minima at the expense of the two other local minima mentioned above.

Fig. 6b shows the corresponding free-energy plot for GB1m3. As reference structure for GB1m3, we use a mutated and relaxed version of the GB1p reference structure. We see that $F(\Delta, E)$ has a simpler shape for GB1m3 than for GB1p. There is only one detectable free-energy minimum for GB1m3, and this minimum corresponds to a structure similar to the favored one for GB1p.

Different experiments on GB1p have, as mentioned above, obtained different β-hairpin populations. One way of estimating folded populations in the model is by two-state fits like those in Fig. 5. An independent and more direct estimate can be obtained by counting native backbone hydrogen bonds. To this end, we consider a hydrogen bond formed if its energy is less than $-\epsilon_\mathrm{hb}^{(1)}/3$. The number of native backbone hydrogen bonds in a given conformation is denoted by $N_\mathrm{hb}^\mathrm{nat}$. Fig. 7 shows the probability distribution of $N_\mathrm{hb}^\mathrm{nat}$ for GB1p and GB1m3 at 299 K, which is very close to the temperature (298 K) at which the folded populations of these two peptides were
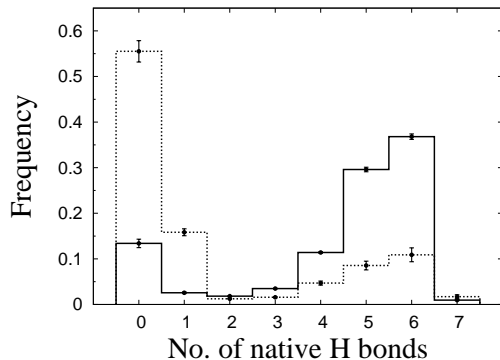
14

Figure 7: Probability distribution of the number of native hydrogen bonds, N_{hb}^{nat}, for GB1m3 (full line) and GB1p (dotted line) at 299 K. The hydrogen bonds taken as native are the same for both peptides. In GB1p notation, the native hydrogen bonds are Glu42(N)-Thr55(O), Glu42(O)-Thr55(N), Thr44(N)-Thr53(O), Thr44(O)-Thr53(N), Asp46(N)-Thr51(O), Asp46(O)-Thr51(N) and Asp47(O)-Lys50(N).

compared by CD and NMR [19]. We find that the probability distribution $P(N_{hb}^{nat})$ has a clear bimodal shape for both peptides, with one native and one unfolded peak. The native peak is, as expected from the results above, significantly larger for the mutant GB1m3 than for GB1p. Taking conformations with $N_{hb}^{nat} \geq 3$ as native and those with $N_{hb}^{nat} \leq 2$ as unfolded, we obtain native populations of $82 \pm 1\%$ for GB1m3, $84 \pm 1\%$ for GB1m2, and $27 \pm 2\%$ for GB1p. The overall agreement between these results and the experimental data ($86 \pm 3\%$ for GB1m3, $74 \pm 5\%$ for GB1m2, $\sim 30\%$ for GB1p) is very good, although the model slightly overestimates the folded fraction for GB1m2. Note that the native populations estimated from $P(N_{hb}^{nat})$, thanks to the bimodality, are quite well determined, despite that the precise definition of native in terms of N_{hb}^{nat} is somewhat arbitrary.

For GB1m3, we find that one of the hydrogen bonds taken as native is very unlikely to form in our model, namely Pro47(O)-Gly50(N). As a result, conformations with $N_{hb}^{nat} = 7$ are very rare (see Fig. 7).

Our E_{hp}- and N_{hb}^{nat}-based native populations for GB1p are different; from the E_{hp} data we obtain a native population of 46% at 299 K, where the N_{hb}^{nat} analysis gives 27%. The magnitude of this difference is similar to that between different experiments. The N_{hb}^{nat}-based result is is good agreement with CD and NMR data, whereas the E_{hp}-based result agrees with Trp fluorescence data. For GB1m3 (and GB1m2), we do not know of any Trp fluorescence study. Our model suggests that the difference between different methods would be smaller in this case. Our E_{hp}-based folded population at
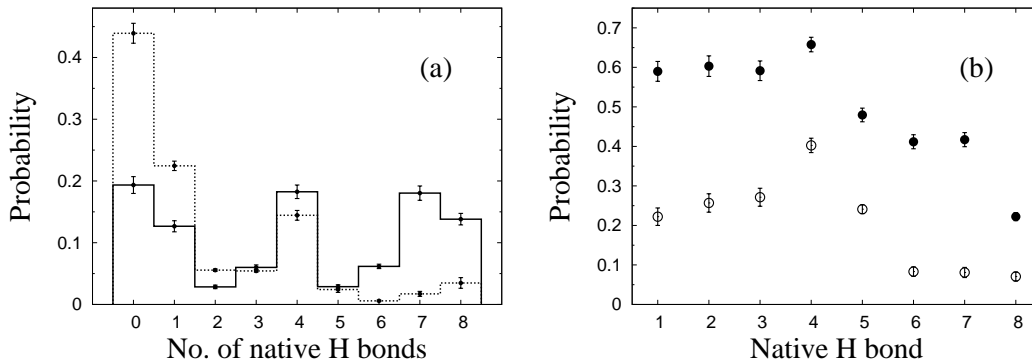
Figure 8: (a) Probability distribution of the number of native backbone hydrogen bonds, N_{hb}^{nat}, for LLM (full line) and Betanova (dotted line) at 287 K. (b) Frequencies of occurrence for the different native hydrogen bonds for Betanova (○) and LLM (●) at 287 K. In Betanova notation, the native hydrogen bonds are 1: Ser4(N)-Thr11(O), 2: Ser4(O)-Thr11(N), 3: Gln6(N)-Lys9(O), 4: Gln6(O)-Lys9(N), 5: Tyr10(N)-Thr17(O), 6: Tyr10(O)-Thr17(N), 7: Asn12(N)-Lys15(O) and 8: Asn12(O)-Lys15(N).

299 K is 85 % for GB1m3, which is close to our N_{hb}^{nat}-based result of 82 %.

3.4   Betanova and LLM

Betanova is a designed antiparallel three-stranded β-sheet peptide with 20 residues (RGWSVQNGKYTNNGKTTEGR) [20], which is only marginally stable [21]. Recently, Betanova mutants with higher stability were developed [21], such as the triple mutant LLM (Val5Leu, Asn12Leu, Thr17Met). The NMR-based native populations of LLM and Betanova are 36 % and 9 %, respectively, at 283 K [21]. Results in good agreement with these estimates were obtained when testing an earlier version of our model on these two peptides [12]. Folding simulations of Betanova have also been performed by other groups, using coarse-grained [53] and atomic [54, 55] models.

The folded structure of Betanova and LLM contains eight backbone hydrogen bonds, four in each of the two β-hairpins. Fig. 8a shows the probability distribution of the number of native backbone hydrogen bonds, N_{hb}^{nat}, in our model for LLM and Betanova, at 287 K. The distributions have three peaks. In addition to the folded and unfolded peaks at high and low N_{hb}^{nat}, there is also a peak at $N_{hb}^{nat} = 4$. Visual inspection of snapshots from the simulations reveals that conformations at this peak

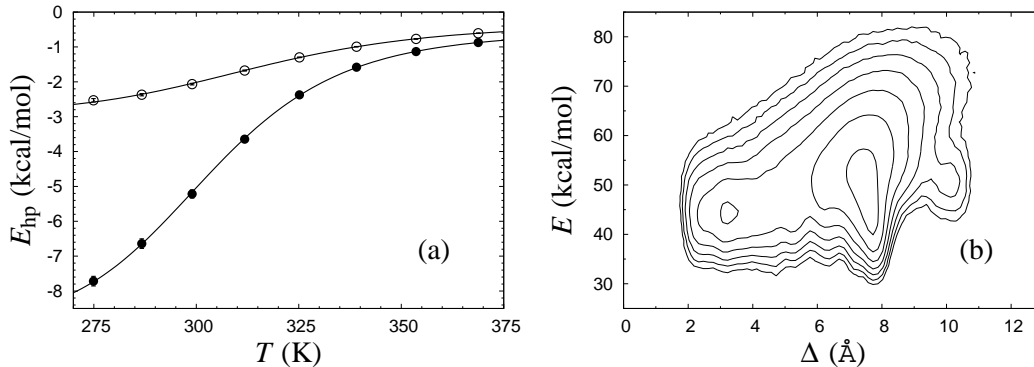Figure 9: (a) The hydrophobicity energy E_{hp} against temperature for Betanova ($\circ$) and LLM ($\bullet$). The lines are fits to Eq. 9 ($T_{\mathrm{m}} = 314 \pm 1$, $\Delta E = 8.9 \pm 0.1\,\mathrm{kcal/mol}$ for Betanova; $T_{\mathrm{m}} = 302 \pm 1\,\mathrm{K}$, $\Delta E = 10.9 \pm 0.2\,\mathrm{kcal/mol}$ for LLM). (b) Free energy $F(\Delta, E)$ for Betanova at 275 K. Contour levels are as in Fig. 2b.

tend to contain the first (N-terminal) β-hairpin but not the second (C-terminal) one. This conclusion, which is in agreement with experimental data [21], is confirmed by the frequencies of occurrence of the individual hydrogen bonds, shown in Fig. 8b. We see that the hydrogen bonds of the first β-hairpin (1–4) occur more frequently than those of the second β-hairpin (5–8), especially for Betanova. For a conformation to be counted as folded, we require that $N_{\mathrm{hb}}^{\mathrm{nat}} \geq 6$. With this definition, we find that Betanova and LLM are $6 \pm 1\,\%$ and $38 \pm 2\,\%$ folded, respectively, at 287 K, which is in good agreement with the experimental results ($9\,\%$ and 36% at 283 K).

The melting behavior has, as far as we know, not been studied experimentally for Betanova or LLM. In Fig. 9a we show melting curves for these peptides in our model. As in the β-hairpin case, we consider the hydrophobicity energy E_{hp}. Betanova has fewer hydrophobic residues than LLM, and we see that E_{hp} is much lower in absolute value for Betanova than for LLM. In our model, the difference in hydrophobicity is the main reason why LLM is more stable than Betanova. A two-state analysis of our E_{hp} data gives $T_{\mathrm{m}} = 314\pm1$ and $\Delta E = 8.9\pm0.1\,\mathrm{kcal/mol}$ for Betanova, and $T_{\mathrm{m}} = 302\pm1\,\mathrm{K}$ and $\Delta E = 10.9 \pm 0.2\,\mathrm{kcal/mol}$ for LLM. These fitted two-state parameters contrast sharply with the results of the $N_{\mathrm{hb}}^{\mathrm{nat}}$ analysis above, especially for Betanova. In fact, for Betanova, the fitted two-state parameters correspond to a native population of $80\,\%$ at the temperature 287 K, at which Betanova was estimated above to be only $6\,\%$ folded. This discrepancy between the native populations obtained using E_{hp} and $N_{\mathrm{hb}}^{\mathrm{nat}}$ data clearly show that, in our model, these two peptides do not behave as ideal two-state systems. It is worth noting that the quality of the two-state fits in Fig. 9a, nevertheless, is very good, which illustrates that deviations from the simple

two-state picture can be very hard to detect from the temperature dependence of a single quantity [56].

Fig. 9b shows the free energy $F(\Delta, E)$ for Betanova at 275 K. Like for the β-hairpins, we use all the heavy atoms in the RMSD, but limit the comparison to the residues 3-18. The residues 1, 2, 19 and 20 do not participate in the β-sheet structure. There is a local minimum at $\Delta \approx 3.2\,\text{Å}$ representing the state obtained in our model that most resembles the NMR structure. That this state is not the most probable state in the model is consistent with the low native population found experimentally for this peptide. The corresponding graph for LLM shows a much more prominent minimum representing the native conformation.

3.5   The character of the melting transition

For GB1p, Betanova and LLM, we saw above that the apparent native population depends on which quantity we study. This dependence reflects the fact that these peptides do not show ideal two-state behavior in the model. A quantity for which we obtain a relatively high apparent melting temperature not only for these three peptides but for all the peptides studied, is the radius of gyration, R_g. The T_m values obtained from our R_g data for F_s and the Trp cage are 29 K and 9 K higher, respectively, than what we found above using the helix content. For GB1m3, our R_g data gives a T_m that is 6 K higher than that obtained above using the hydrophobicity energy. These comparisons show that none of the peptides studied behaves as a perfect two-state system in our model, although the deviations from this behavior might be relatively small for some of them, such as GB1m3.

One measure of the sharpness of the melting transition is the height of the peak in the specific heat, C_v. In Fig. 10, we show specific heat curves for the different peptides studied. The results for GB1m2 are again very similar to those for GB1m3 and therefore omitted. The specific heat exhibits a clear peak for all the peptides studied, but the height of the peak varies. The peak is highest for GB1m3, indicating that the melting transition is most two-state-like for this peptide. A comparison of the energy distributions of the different peptides (not shown) supports this conclusion. For GB1m3, we find that the energy distribution has a bimodal shape, although not very pronounced. The other peptides all have wide but single-peaked distributions. The distribution is particularly wide, virtually flat, for GB1p, which has the next highest peak in C_v.
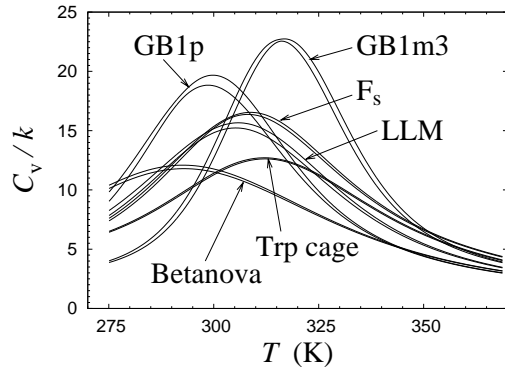
Figure 10: The specific heat C_v against temperature for the different peptides, as obtained using histogram reweighting techniques [57]. For each peptide, a band is shown. The band is centered around the expected value and shows statistical 1σ errors. C_v is defined as $C_\mathrm{v} = N^{-1}d\langle E\rangle/dT = (NkT^2)^{-1}(\langle E^2\rangle - \langle E\rangle^2)$, where N is the number of amino acids and $\langle O\rangle$ denotes a Boltzmann average of variable O.

For the peptide with the sharpest transition, GB1m3, we find that the specific heat maximum, $316\,\mathrm{K}$, is located near the temperature at which its folded population is $50\,\%$. The other peptides are less than $50\,\%$ folded at their specific heat maxima, especially Betanova. Betanova was estimated above to be 6% folded at $287\,\mathrm{K}$ in the model, and has its specific heat maximum at a temperature higher than that, $293\,\mathrm{K}$.

4 Conclusion

We have developed an atomic model with a simplified phenomenological potential for folding studies of polypeptide chains, which was tested on a set of peptides with about 20 amino acids each, namely the Trp cage, F_s, GB1p, GB1m2, GB1m3, Betanova and LLM. First of all, our study shows that the model folds these different sequences to structures similar to their experimental structures, for one and the same choice of model parameters. In addition, we investigated the stability and melting behavior of the peptides. The following list is a brief summary of these calculations, focusing on the observables expected to be correlated with the corresponding experimental probes.

- The helix content of the Trp cage shows a temperature dependence that is in good agreement with experimental data based on CD and NMR (see Fig. 3).

	Exp.	Model
GB1p	$\sim 30\,\%$ (298 K)	$27 \pm 2\,\%$ (299 K)
GB1m2	$74 \pm 5\,\%$ (298 K)	$84 \pm 1\,\%$ (299 K)
GB1m3	$86 \pm 3\,\%$ (298 K)	$82 \pm 1\,\%$ (299 K)
Betanova	$9\,\%$ (283 K)	$6 \pm 1\,\%$ (287 K)
LLM	36% (283 K)	$38 \pm 2\,\%$ (287 K)

Table 2: Folded populations of the different β-sheet peptides in the model, along with experimental results. The experimental data on GB1p, GB1m2 and GB1m3 are from Fesinmeyer et al. [19], whereas those on Betanova and LLM are from López de la Paz et al. [21].

- A two-state analysis of the helix content of F_s gives T_m and ΔE values that are in good agreement with CD data, while the T_m value is somewhat low compared to its IR-based value.

- Estimates of folded populations based on native hydrogen bond data for the β-sheet peptides GB1p, GB1m2, GB1m3, Betanova and LLM are in good agreement with CD- and NMR-based experimental results, as is summarized in Table 2. Recall that the energy scale was set using the α-helical Trp cage.

- Experimentally, GB1p has been studied using Trp fluorescence as well, which gave a folded population higher than that in Table 2. Our results based on hydrophobicity energy data are in good agreement with those from the Trp fluorescence study.

The model fails to reproduce the difference in folded population between the two stable mutants of GB1p (see Table 2), which in part may be due to the fact that Coulomb interactions between side-chain charges are ignored; GB1m3 contains some charged residues that are missing in GB1m2. The overall quantitative agreement with experimental data is, nevertheless, excellent. This agreement indicates that factors such as Coulomb interactions between charged residues play a quite limited role in the folding thermodynamics of these peptides, compared to hydrogen bonding and hydrophobic attraction, which are the main driving forces of the model.

The temperature dependence of the model is, to us, surprisingly good, for two reasons. First, the temperature dependence was not considered at all when calibrating the model, except in the determination of the energy scale. A considerable amount of fine-tuning was required in order to obtain proper folded structures, but no further fine-tuning was performed once that goal had been achieved. Second, our calculations

do not involve any reparametrization of the energy function. In other words, the parameters of the energy function are temperature independent, which is a simplifying assumption rather than a controlled approximation. On the other hand, it should be noted that the melting transition is not triggered by a sudden change in, for example, the strength of the hydrophobic attraction.

In the development of this model, we have taken a purely phenomenological approach. The model will be further developed by studying new amino acid sequences, which will impose new conditions on the interaction potential. As before, the challenge will be to do this in a backwards compatible manner; the model must not lose its ability to fold previously studied sequences. As to limitations of the current version of the model, we know that it is unable to properly fold the so-called trpzip β-hairpins [58], which make β-hairpins in the model but with the wrong topology. We also expect that refinement of the model will be needed as the chains get larger. For example, as mentioned earlier, it is likely that our pair-wise additive hydrophobicity potential will have to be supplemented with multibody terms for large chains. Finding out how to change the model in order to make it more general without losing computational efficiency will not be an easy task, but the results obtained so far makes it tempting to try.

Acknowledgments

References

[1] Dobson, C.M. 2003. Protein folding and misfolding. *Nature* 426: 884–890.

[2] Dyson, H.J., and P.E. Wright. 2002. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12: 54–60.

[3] Gnanakaran, S., H. Nymeyer, J. Portman, K.Y. Sanbonmatsu, and A.E. García. 2003. Peptide folding simulations. *Curr. Opin. Struct. Biol.* 13: 168–174.

[4] Zaman, M.H., M.Y. Shen, R.S. Berry, K.F. Freed and T.R. Sosnick. 2003. Investigations into sequence and conformational dependence of backbone entropy, inter-basin dynamics and the Flory isolated-pair hypothesis for peptides. *J. Mol. Biol.* 331: 693–711.

[5] Zhou, R., B.J. Berne, and R. Germain. 2001. The free energy landscape for β hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA* 98: 14931–14936.

[6] Still, W.C., A. Tempczyk, R.C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112: 6127–6129.

[7] Hassan, S.A., E.L. Mehler, D. Zhang and H. Weinstein. 2003. Molecular dynamics simulations of peptides and proteins with a continuum electrostatic model based on screened Coulomb potentials. *Proteins* 51: 109–125.

[8] Ferrara, P., J. Apostolakis, and A. Caflisch. 2002. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins* 46: 24–33.

[9] Hansmann, U.H.E., and L.T. Wille. 2002. Global optimization by energy landscape paving. *Phys. Rev. Lett.* 88: 068105.

[10] Schug, A., T. Herges, and W. Wenzel. 2003. Reproducible protein folding with the stochastic tunneling method. *Phys. Rev. Lett.* 91: 158102.

[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] Sayle, R., and E.J. Milner-White. 1995. RasMol: biomolecular graphics for all. *Trends Biochem. Sci.* 20: 374–376.

[14] Neidigh, J.W., R.M. Fesinmeyer, and N.H. Andersen. 2002. Desiging a 20-residue protein. *Nat. Struct. Biol.* 9: 425–430.

[15] Lockhart, D.J., and P.S. Kim. 1992. Internal Stark effect measurement of the electric field at the amino acid terminus of an α helix. *Science* 257: 947–951.

[16] Lockhart, D.J., and P.S. Kim. 1993. Electrostatic screening of charge and dipole interactions with the helix backbone. *Science* 260: 198–202.

[17] Kobayashi, N., S. Endo, and E. Munekata. 1993. Conformational study on the IgG binding domain of protein G. In Peptide Chemistry 1992. N. Yanaihara, editor. ESCOM, Leiden. 278–281.

[18] Blanco, F.J., G. Rivas, L. Serrano. 1994. A short linear peptide that folds into a native stable β-hairpin in aqueous solution. *Nat. Struct. Biol.* 1: 584–590.

[19] Fesinmeyer, R.M., F.M. Hudson, and N.H. Andersen. 2004. Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin. *J. Am. Chem. Soc.* 126: 7238–7243.

[20] Kortemme, T., M. Ramírez-Alvarado, and L. Serrano. 1998. Design of a 20-amino acid, three-stranded β-sheet protein. *Science* 281: 253–256.

[21] López de la Paz, M., E. Lacroix, M. Ramírez-Alvarado, and L. Serrano. 2001. Computer-aided design of β-sheet peptides. *J. Mol. Biol.* 312: 229–246.

[22] Favrin, G., A. Irbäck, and S. Mohanty. 2004. Oligomerization of amyloid Aβ_{16-22} peptides using hydrogen bonds and hydrophobicity forces. *Biophys. J.* 87: 3657–3664.

[23] Tsai, J., R. Taylor, C. Chothia, and M. Gerstein. 1999. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* 290: 253–266.

[24] Brändén, C., and J. Tooze. 1991. *Introduction to Protein Structure.* Garland Publishing, New York.

[25] Shimizu, S., and H.S. Chan. 2001. Anti-cooperativity in hydrophobic interactions: a simulation study of spatial dependence of three-body effects and beyond. *J. Chem. Phys.* 115: 1414–1421.

[26] Lyubartsev, A.P., A.A. Martsinovski, S.V. Shevkunov, and P.N. Vorontsov-Velyaminov. 1992. New approach to Monte Carlo calculation of the free energy: method of expanded ensembles. *J. Chem. Phys.* 96: 1776–1783.

[27] Marinari, E., and G. Parisi. 1992. Simulated tempering: A new Monte Carlo scheme. *Europhys. Lett.* 19: 451–458.

[28] Irbäck, A., and F. Potthast. 1995. Studies of an off-lattice model for protein folding: sequence dependence and improved sampling at finite temperature. *J. Chem. Phys.* 103: 10298–10305.

[29] Hansmann, U.H.E., and Y. Okamoto. 1999. New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* 9: 177–183.

[30] 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.

[31] Press W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge.

[32] Qiu, L., S.A. Pabit, A.E. Roitberg, and S.J. Hagen. 2002. Smaller and faster: The 20-residue Trp-cage protein folds in $4\,\mu$s. *J. Am. Chem. Soc.* 124: 12952–12953.

[33] Snow, C.D., B. Zagrovic, and V.S. Pande. 2002. The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. *J. Am. Chem. Soc.* 124: 14548–14549.

[34] Simmerling, C., B. Strockbine, and A.E. Roitberg. 2002. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 124: 11258–11259.

[35] Pitera, J.W., and W. Swope. 2003. Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. *Proc. Natl. Acad. Sci. USA* 100: 7587–7592.

[36] Zhou, R. 2003. Trp-cage: folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. USA* 100: 13280–13285.

[37] Williams, S., T.P. Causgrove, R. Gilmanshin, K.S. Fang, R.H. Callender, W.H. Woodruff, and R.B. Dyer. 1996. Fast events in protein folding: Helix melting and formation in a small peptide. *Biochemistry* 35: 691–697.

[38] Thompson, P.A., W.A. Eaton, and J. Hofrichter. 1997. Laser temperature jump study of the helix⇌coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* 36: 9200–9210.

[39] Vila, J.A., D.R. Ripoll, and H.A. Scheraga. 2000. Physical reasons for the unusual α-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides. *Proc. Natl. Acad. Sci. USA* 97: 13075–13079.

[40] García, A.E., and K.Y. Sanbonmatsu. 2002. α-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. USA* 99: 2782–2787.

[41] Nymeyer, H., and A.E. García. 2003. Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. USA* 100: 13934–13939.

[42] Roccatano, D., A. Amadei, A. Di Nola, and H.J.C. Berendsen. 1999. A molecular dynamics study of the 41–56 β-hairpin from B1 domain of protein G. *Protein Sci.* 8: 2130–2143.

[43] Pande, V.S., and D.S. Rokhsar. 1999. Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. *Proc. Natl. Acad. Sci. USA* 96: 9062–9067.

[44] Dinner, A.R., T. Lazaridis, and M. Karplus. 1999. Understanding β-hairpin formation. *Proc. Natl. Acad. Sci. USA* 96: 9068–9073.

[45] García, A.E., K.Y. Sanbonmatsu. 2001. Exploring the energy landscape of a β hairpin in explicit solvent. *Proteins* 42: 345–354.

[46] Zagrovic, B., E.J. Sorin, and V. Pande. 2001. β-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* 313: 151–169.

[47] Kussell, E., J. Shimada, and E.I. Shakhnovich. 2002. A structure-based method for derivation of all-atom potentials for protein folding. *Proc. Natl. Acad. Sci. USA* 99: 5343–5348.

[48] Zhou, R. 2003. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins* 53: 148–161.

[49] Bolhuis, P.G. 2003. Transition-path sampling of β-hairpin folding. *Proc. Natl. Acad. Sci. USA* 14: 12129–12134.

[50] Wei, G., N. Mousseau, and P. Derreumaux. 2004. Complex folding pathways in a simple β-hairpin. *Proteins* 56: 464–474.

[51] Muñoz, V., P.A. Thompson, J. Hofrichter, and W.A. Eaton. 1997. Folding dynamics and mechanism of β-hairpin formation. *Nature* 390: 196–199.

[52] Gronenborn, A.M., D.R. Filpula, N.Z. Essig, A. Achari, M. Whitlow, P.T. Wing-field, and G.M. Clore. 1991. A novel, highly stable fold of the immunoglobulin-binding domain of streptococcal protein G. *Science* 253: 657–661.

[53] Kim, S.Y., J. Lee, and J. Lee. 2004. Folding of small proteins using a single continuous potential. *J. Chem. Phys.* 120: 8271–8276.

[54] Bursulaya, B.D., and C.L. Brooks III. 1999. Folding free energy surface of a three-stranded β-sheet protein. *J. Am. Chem. Soc.* 121: 9947–9951.

[55] Colombo, G., D. Roccatano, and A.E. Mark. 2002. Folding and stability of the three-stranded β-sheet peptide betanova: insights from molecular dynamics simulations. *Proteins* 46: 380–392.

[56] Favrin, G., A. Irbäck, B. Samuelsson, and S. Wallin. 2003. Two-state folding over a weak free-energy barrier. *Biophys. J.* 85: 1457–1465.

[57] Ferrenberg, A.M., and R.H. Swendsen. 1988. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61: 2635–2638.

[58] Cochran, A.G., N.J. Skelton, and M.A. Starovasnik. 2001. Tryptophan zippers: stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. USA* 98: 5578–5583.