

Peptide Folding and Aggregation Studied Using a Simplified Atomic Model

Anders Irbäck¹

Complex Systems Division, Department of Theoretical Physics
Lund University, Sölvegatan 14A, SE-223 62 Lund, Sweden

Submitted to *J. Phys.: Condens. Matter*

Abstract:

Using an atomic model with a simplified sequence-based potential, the folding properties of several different peptides are studied. Both α -helical (Trp cage, F_s) and β -sheet (GB1p, GB1m2, GB1m3, Betanova, LLM) peptides are considered. The model is able to fold these different peptides for one and the same choice of parameters, and the melting behaviour of the peptides (folded population against temperature) is in very good agreement with experimental data. Furthermore, using the same model with unchanged parameters, the aggregation behaviour of a fibril-forming fragment of the Alzheimer's A β peptide is studied, with very promising results.

PACS: 87.15.Aa

¹E-mail: anders@thep.lu.se

1 Introduction

Atomic simulations are becoming an increasingly useful tool in protein folding studies, as witnessed by recent studies that combined simulations and experiments in a fruitful manner (for a review, see [1]). Many of these calculations focused on unfolding rather than folding, or incorporated sequence-specific structural information into the energy function. Unbiased simulations of folding are different and impose much sharper constraints on the interaction potential. Such simulations are becoming feasible for short chains (for a review, see [2]), which is an exciting development, but there are fundamental questions about the potentials used in the simulations that remain to be fully understood. One difficulty is that different potentials give very different relative weights to the α -helix and β -strand regions of the Ramachandran space [3]. 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. It seems that most current models need further calibration in order to give a temperature dependence that is not too weak [4]; as a result, calculated melting temperatures tend to be unrealistically high. To be able to address these thermodynamic questions in a systematic manner, extensive conformational sampling is required.

Here I discuss folding [5–7] and aggregation [8] studies of peptides, which were carried out using an atomic model with a simplified sequence-based interaction potential. For computational efficiency, the model contains no explicit water molecules and only torsional degrees of freedom. The interaction potential has three major terms representing excluded-volume effects, hydrogen bonding and effective hydrophobic attraction. 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.

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 behaviour 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 behaviour 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.

2 Model and methods

Our model contains all atoms of the polypeptide chains, including hydrogen atoms. It 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 [5].

The interaction potential

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

is composed of four terms. Next, a brief description of these different terms is given; further details can be found in [7].

The first term in Eq. 1 represents excluded-volume effects and has the form

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

where the summation is over pairs of atoms (i, j) , and $\sigma_i = 1.77, 1.75, 1.55, 1.42$ and 1.00 \AA for S, C, N, O and H atoms, respectively. The scale factor λ_{ij} in Eq. 2 has the value 0.75 for all pairs except those connected by three covalent bonds, for which $\lambda_{ij} = 1$. When the two atoms belong to different chains, the smaller value $\lambda_{ij} = 0.75$ is always used. To speed up the calculations, Eq. 2 is evaluated using a cutoff of $r_{ij}^c = 4.3\lambda_{ij} \text{ \AA}$.

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

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

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 Gly and Pro amino acids which have very different ϕ , ψ distributions. Neither is it used for the two end amino acids, unless these are protected by capping groups.

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

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

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

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

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

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 function $u(r)$ is calculated using a cutoff of $r^c = 4.5 \text{ \AA}$. 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. Intrachain hydrogen bonds between parts that are very close in sequence are 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 . For interchain hydrogen bonds, we make no such restrictions. A reduced strength is assigned to hydrogen bonds involving chain ends; 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

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

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. C_{IJ} is given by

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

where A_I and A_J denote predetermined sets of atoms in side chains I and J , and 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 $A < x < B$ [$A = (3.5 \text{ \AA})^2$ and $B = (4.5 \text{ \AA})^2$]. 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. The strength of the hydrophobic attraction is reduced for side-chain pairs that are nearest or next-nearest neighbours along the sequence; M_{IJ} is reduced by a factor of 2 for next-nearest neighbours, and taken to be 0 for nearest neighbours.

The parameters of this potential were largely determined by a trial and error procedure. 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. 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.

To study the thermodynamic behaviour of this model, Monte Carlo methods were used. Monte Carlo details can be found in [7, 8]. All our simulations were started from random configurations. All statistical errors quoted are 1σ errors obtained from the variation between independent runs. For each system, about 10 independent runs were performed.

For a peptide with N amino acids, we define the α -helix content H and β -strand content S as the fractions of the $N - 2$ inner amino acids with their (ϕ, ψ) pair in the α -helix and β -strand regions of the Ramachandran space. α -helix is assumed to correspond to $-90^\circ < \phi < -30^\circ$, $-77^\circ < \psi < -17^\circ$, and β -strand to $-150^\circ < \phi < -90^\circ$, $90^\circ < \psi < 150^\circ$. To distinguish between parallel and antiparallel β -strands in our multichain systems, we examine the orientation of end-to-end vectors. For a given multichain configuration, we first determine all pairs of chains such that (i) their interchain hydrogen bond energy is less than $-1.5\epsilon_{\text{hb}}^{(1)}$ (roughly corresponding to 2–3 hydrogen bonds), and (ii) both chains have a β -strand content higher than 0.5. For each such pair, the scalar product of the normalised end-to-end unit vectors is calculated. If this scalar product is greater than 0.7 (less than -0.7), we say that the two chains are parallel (antiparallel). The numbers of parallel and antiparallel pairs are denoted by n_+ and n_- , respectively.

3 Results

3.1 Folding

Using the model described above, folding studies were performed for a set of sequences with about 20 amino acids each [5–7]. This set contains α -helical peptides (Trp cage, F_s), β -hairpins (GB1p, GB1m2, GB1m3) and three-stranded β -sheets (Betanova, LLM). The different geometries of the peptides studied are schematically illustrated in Fig. 1. It turns out that the model has native-like free-energy minima for all these different sequences, for one and the same choice of model parameters. To achieve that goal, a considerable amount of fine-tuning was required. Once this had been accomplished, the melting behaviour of the peptides was investigated without making any further changes of the potential. In this section, a brief summary of these thermodynamic studies is given.

Let us begin with the α -helical sequences. The optimised 20-residue Trp cage (NLYIQWLKDG-GPSSGRPPPS) is a “miniprotein” with a compact folded state and a melting temperature of 315 K [10]. Its NMR-derived native structure [10] 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). In our calculations, the melting temperature of the Trp cage is used to set the overall energy scale of the model. Figure 2a shows the temperature dependence of the helix content H . A simple two-state fit provides an excellent description of the data. The midpoint temperature from this fit, T_m , is set to 315 K, the experimental melting temperature. Having done that, there is no free parameter left in the model. The fitted energy difference ΔE is, in contrast to T_m , not used for calibration, but is rather a prediction of the model. This prediction can be tested by comparing the resulting native population, $1/\{1 + \exp[-(1/kT - 1/kT_m)\Delta E]\}$, with experimental data. From Fig. 2b it can be seen that the native population obtained using $T_m = 315$ K our fitted value of $\Delta E = 11.5 \pm 0.2$ kcal/mol is in good agreement with

experimental data over the entire temperature range, with a maximum deviation of $\sim 5\%$ at the lowest temperatures. Native-like structures have been obtained in previous simulations of this sequence [11–15], but a similar temperature dependence has not been reported.

The 21-residue F_s peptide is given by Suc-A₅(AAARA)₃A-NH₂, (where Suc is succinyl acid) [16, 17]. F_s forms an α -helix and was studied using circular dichroism (CD) as well as infrared (IR) spectroscopy. The melting temperature measured by IR was 334 K [18], whereas the CD-based studies obtained $T_m = 308$ K [17] and $T_m = 303$ K [19]. Computational studies of F_s were also reported [20–22]. A calculation with explicit water [21] gave $T_m = 345$ K, which is in reasonable agreement with the IR-based result. A two-state fit of our data for the helix content, shown in Fig. 3, gives $T_m = 304 \pm 1$ K, which is significantly lower than the IR-based result but in perfect agreement with the CD studies, especially that of Thompson et al. [19]. For the energy difference, we obtain $\Delta E = 11.9 \pm 0.3$ kcal/mol, which also agrees well with what Thompson et al. found, namely $\Delta E = 12 \pm 2$ kcal/mol.

Let us now turn to the β -sheet peptides. The 41–56-residue hairpin GB1p (GEWTYDDATK-TFTVTE) [23] from the protein G B1 domain is probably the most extensively studied β -sheet peptide, at least computationally [4, 24–32]. Very recently, two mutants of GB1p with enhanced stability were designed [33], GB1m2 and GB1m3, by replacing the turn segment DDATKT by NPATGK. The mutant GB1m2 (GEWTYNPATGKFTVTE) is identical to GB1p except for this change, while GB1m3 (KKWTYNPATGKFTVQE) differs from GB1p at the chain ends as well. At 298 K, GB1m3, GB1m2 and GB1p were estimated by CD and NMR to be $86 \pm 3\%$, $74 \pm 5\%$ and $\sim 30\%$ folded, respectively [33]. A Trp fluorescence study [34] obtained a significantly higher folded population for GB1p ($T_m = 297$ K).

Figure 4a shows our results for the hydrophobicity energy E_{hp} for GB1p and GB1m3. This quantity should be strongly correlated with Trp fluorescence, as Trp43 forms a hydrophobic cluster together with Tyr45, Phe52 and Val54. A two-state fit to the data for GB1p gives $T_m = 297 \pm 1$ K and $\Delta E = 14.2 \pm 0.2$ kcal/mol, which indeed is in good agreement with the Trp fluorescence study ($T_m = 297$ K and $\Delta E = 11.6$ kcal/mol [34]). The same fit gives $T_m = 321 \pm 1$ K and $T_m = 322 \pm 2$ K and for GB1m3 and GB1m2, respectively. These values lie close to the melting temperature measured by CD and NMR for GB1m2 (320 ± 2 K [33]), and somewhat below the corresponding result for GB1m3 (333 ± 2 K [33]).

An independent and more direct estimate of the folded populations can be obtained by counting native backbone hydrogen bonds. Figure 4b shows the probability distribution of the number of such bonds, N_{hb}^{nat} , for GB1p and GB1m3 at 299 K. The distribution has a clear bimodal shape for both peptides, with one native and one unfolded peak. The native peak is larger for the mutant GB1m3 than for GB1p, as it should. 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 above mentioned CD and NMR results (at 298 K) is very good, although the model slightly overestimates the folded fraction for GB1m2.

At 299 K, our E_{hp} - and N_{hb}^{nat} -based folded populations for GB1p are 46% and 27%, respec-

tively. The magnitude of this difference is similar to that between the different experiments. In fact, the $N_{\text{hb}}^{\text{nat}}$ -based value agrees well with CD and NMR data [33], whereas the E_{hp} -based value agrees well with Trp fluorescence data [34].

Betanova is a designed antiparallel three-stranded β -sheet peptide with 20 residues (RGWSVQNGKYTNNGKTTEGR) [35]. It is only marginally stable [39], but mutants with higher stability were recently developed [39], such as the triple mutant LLM (Val5Leu, Asn12Leu, Thr17Met). Computational studies were reported for the original Betanova sequence [36–38]. The NMR-based folded populations of LLM and Betanova are 36 % and 9 %, respectively, at 283 K [39].

Figure 5 shows results obtained using our model on Betanova and LLM. A two-state fit of the E_{hp} data in Fig. 5a gives $T_{\text{m}} = 314 \pm 1$ for Betanova and $T_{\text{m}} = 302 \pm 1$ K for LLM. These melting temperatures are high compared with the NMR results, especially for Betanova. However, these peptides do not show ideal two-state behaviour in our model, and the apparent folded populations depend strongly on the observable used. Figure 5b shows the probability distribution of the number of native backbone hydrogen bonds at 287 K. This distribution has three peaks for both Betanova and LLM; in addition to the folded and unfolded peaks at high and low $N_{\text{hb}}^{\text{nat}}$, there is also a peak at $N_{\text{hb}}^{\text{nat}} = 4$. The typical conformation at this peak contains the first (N-terminal) β -hairpin but not the second (C-terminal) one. Taking conformations with $N_{\text{hb}}^{\text{nat}} \geq 6$ as native, the folded populations are 6 ± 1 % and 38 ± 2 % for Betanova and LLM, respectively. These values are small compared to our E_{hp} analysis above, but in good agreement with the NMR results. That Betanova and LLM do not show ideal two-state behaviour is not surprising given their small size and high flexibility. It is striking, however, how hard it can be to detect deviations from the two-state picture from the temperature dependence of a single quantity [40], as in Fig. 5a.

3.2 Aggregation

Using exactly the same model, the oligomerisation properties of the $A\beta_{16-22}$ peptide (acetyl-KLVFFAE-NH₂) were studied [8], by unbiased thermodynamic simulations of systems of one, three and six $A\beta_{16-22}$ peptides. The three- and six-chain systems were contained in periodic boxes of sizes $(35\text{\AA})^3$ and $(44\text{\AA})^3$, respectively.

Small fibril-forming peptides like this seven-residue fragment of the β -amyloid peptide associated with the Alzheimer’s disease, are well suited as model systems for probing the mechanisms of aggregation and fibril formation. While the structure of amyloid fibrils is not known in atomic detail, it is well established that the core of the typical amyloid fibril is composed of β -sheets whose strands run perpendicular to the fibril axis [41]. For $A\beta_{16-22}$ fibrils, there is evidence from solid-state NMR that the β -strands have an antiparallel organisation [42, 43].

Figure 6 shows our results for the α -helix and β -strand contents H and S against temperature

for the different numbers of chains, N_c . For $N_c = 1$, H and S are both small, showing that the $A\beta_{16-22}$ monomer is mainly a random coil at all temperatures. The $N_c = 3$ and $N_c = 6$ systems show a qualitatively different behaviour; S increases sharply with decreasing temperature, to values of $S = 0.6$ and higher, whereas H is very small. These results clearly show that unless the temperature is too high, the three- and six-chain systems self-assemble into ordered structures with a high β -strand content.

Computer simulations of $A\beta_{16-22}$ systems have been performed before [44,45], and our $N_c = 1$ and $N_c = 3$ results can be compared with results from molecular dynamics simulations with explicit water by Klimov and Thirumalai [45]. Our results are in reasonable agreement with theirs for $N_c = 1$, but disagree with theirs for $N_c = 3$. For $N_c = 3$, they obtained a smaller β -strand content and a larger α -helix content compared to their own $N_c = 1$ results; whereas we observe a much larger β -strand content for $N_c = 3$ compared to $N_c = 1$. For the $N_c = 3$ system, Klimov and Thirumalai [45] furthermore found evidence for an obligatory α -helical intermediate. No sign of such an intermediate was found in our calculations.

As mentioned above, there is evidence that the β -strands in full $A\beta_{16-22}$ fibrils have an antiparallel organisation. To study the β -strand organisation in our model, we consider the joint probability distribution $P(n_+, n_-)$, where n_+ and n_- count the numbers of interacting chain pairs with high β -strand contents that are parallel and antiparallel, respectively (see Section 2). Table 1 shows this distribution for $N_c = 3$ at 275 K. For this system, the most probable combination of (n_+, n_-) is $(1, 1)$, corresponding to a mixed β -sheet. At the same time, the distribution shows a clear asymmetry. The frequency of occurrence for antiparallel β -sheets with $(n_+, n_-) = (0, 2)$ is a factor of 7 higher than that for parallel β -sheets with $(n_+, n_-) = (2, 0)$. The corresponding results for $N_c = 6$, at 287 K, are shown in Table 2. As in the $N_c = 3$ case, a majority of the configurations contain mixed β -sheet structure, n_+ and n_- both being nonzero. The asymmetry of the (n_+, n_-) distribution is even more pronounced for $N_c = 6$ than for $N_c = 3$. In particular, it can be seen that large n_- values are much more probable than large n_+ values; the combination $(n_+, n_-) = (4, 0)$ is, e.g., very unlikely to occur, whereas $(n_+, n_-) = (0, 4)$ does occur with a significant frequency. Compared to purely antiparallel β -sheet structures, it is possible that mixed β -sheet structures are more difficult to extend to large stable structures. To be able to check whether or not this is the case, simulations of larger systems are required.

Why are antiparallel β -sheets favoured over parallel ones? Klimov and Thirumalai [45] concluded that $A\beta_{16-22}$ peptides make antiparallel β -sheets because of Coulomb interactions between charged side chains; the two end side chains of the $A\beta_{16-22}$ peptide carry opposite charges, which indeed should make the antiparallel orientation electrostatically favourable. However, our model completely ignores Coulomb interactions between side-chain charges and still strongly favours the antiparallel organisation. Other mechanisms than Coulomb interactions between side-chain charges might therefore play a significant role, such as the geometry of backbone-backbone hydrogen bonds, steric effects, and the precise distribution of hydrophobicity along the chains.

In addition to $A\beta_{16-22}$, we also studied a few control sequences. Some of these, including

the polar sequence studied in [46], had a low overall hydrophobicity. These peptides showed a much lower propensity to aggregate than $A\beta_{16-22}$, and a higher peptide concentration was required to promote aggregation. As an example of a peptide with a significant hydrophobicity but an uneven distribution of it, the peptide acetyl-KFFAAAE-NH₂ was studied, in which the two hydrophobic Phe amino acids are asymmetrically placed. For this sequence, aggregated β -sheet structures were obtained, but with a predominantly parallel β -strand organisation.

In our model, the six-chain $A\beta_{16-22}$ system does not exhibit a single dominating free-energy minimum, but rather a number of more or less degenerate local minima. Figure 7 shows two snapshots of such minima. In the simplest class of typical structures observed in our simulations, five of the chains form a relatively flat β -sheet, whereas the remaining chain is a random coil and held in contact with the β -sheet by hydrophobic attraction. Six-stranded β -sheets also occur in the simulations, but with a low frequency. Further, for the six-chain system, new structures emerge with no analogs in the three-chain simulations. The second structure in Fig. 7 illustrates this. Here stability is achieved by stacking two different, three-stranded, β -sheets together, which brings hydrophobic side chains from the two β -sheets in close contact. Such “sandwiches” occur with a non-negligible frequency in our simulations.

4 Summary

A simplified sequence-based interaction potential for protein folding studies was discussed. To what extent it will be possible to extend this approach to larger chains remains to be seen. In its present form, the model is able to give a good description of the folding behaviour of several different peptides with about 20 amino acids, as shown by our calculations for the Trp cage, F_s, GB1p, GB1m2, GB1m3, Betanova and LLM. The model is not only capable of folding these sequences to structures similar to their experimental structures, but the melting behaviour of the peptides is also realistic; both melting temperatures and the rates at which the folded populations change with temperature are in good agreement with experimental data. The model correctly predicts some of the β -sheet peptides studied to be quite unstable. A comparison of calculated and experimental folded populations for our β -sheet peptides can be found in Table 3.

In the same model, $A\beta_{16-22}$ peptides show a high propensity to self-assemble into aggregated structures with a high β -strand content, while the isolated $A\beta_{16-22}$ peptide is mainly a random coil. Both parallel and antiparallel arrangements of the β -strands occur in the model, with a definite preference for the antiparallel arrangement. It is important to note that this preference for the antiparallel β -strand orientation exists despite that the model ignores the Coulomb interactions between the two charged side chains at the ends of the peptide. In fact, it has been suggested [45] that such Coulomb interactions are the main determinant for the antiparallel orientation. While these Coulomb interactions might enhance the tendency for $A\beta_{16-22}$ peptides to form β -sheets with an antiparallel organisation, our results strongly

suggest that other factors play a significant role, too.

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.

Acknowledgement

This work was in part supported by the Swedish Research Council.

References

- [1] Vendruscolo M and Paci E 2003 *Curr. Opin. Struct. Biol.* **13** 82.
- [2] Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu K Y and García A E 2003 *Curr. Opin. Struct. Biol.* **13** 168.
- [3] Zaman M H, Shen M Y, Berry R S, Freed K F and Sosnick T R 2003 *J. Mol. Biol.* **331** 693.
- [4] Zhou R, Berne B J and Germain R 2001 *Proc. Natl. Acad. Sci. USA* **98** 14931.
- [5] Irbäck A, Samuelsson B, Sjunnesson F and Wallin S 2003 *Biophys. J.* **85** 1466.
- [6] Irbäck A and Sjunnesson F 2004 *Proteins* **56** 110.
- [7] Irbäck A and Mohanty S 2004 Lund preprint LU TP 04-28 (submitted to *Biophys. J.*).
- [8] Favrin G, Irbäck A and Mohanty S 2004 Lund preprint LU TP 04-18 (to appear in *Biophys. J.*).
- [9] Sayle R and Milner-White E J 1995 *Trends Biochem. Sci.* **20** 374.
- [10] Neidigh J W, Fesinmeyer R M and Andersen N H 2002 *Nat. Struct. Biol.* **9** 425.
- [11] Snow C D, Zagrovic B and Pande V S 2002 *J. Am. Chem. Soc.* **124** 14548.
- [12] Simmerling C, Strockbine B and Roitberg A E 2002 *J. Am. Chem. Soc.* **124** 11258.
- [13] Schug A, Herges T and Wenzel W 2003 *Phys. Rev. Lett.* **91** 158102.
- [14] Pitera J W and Swope W 2003 *Proc. Natl. Acad. Sci. USA* **100** 7587.
- [15] Zhou R 2003 *Proc. Natl. Acad. Sci. USA* **100** 13280.
- [16] Lockhart D J and Kim P S 1992 *Science* **257** 947.
- [17] Lockhart D J and Kim P S 1993 *Science* **260** 198.
- [18] Williams S, Causgrove T P, Gilmanshin R, Fang K S, Callender R H, Woodruff W H and Dyer R B 1996 *Biochemistry* **35** 691.
- [19] Thompson P A, Eaton W A and Hofrichter J 1997 *Biochemistry* **36** 9200.
- [20] Vila J A, Ripoll D R and Scheraga H A 2000 *Proc. Natl. Acad. Sci. USA* **97** 13075.
- [21] García A E and Sanbonmatsu K Y 2002 *Proc. Natl. Acad. Sci. USA* **99** 2782.
- [22] Nymeyer H and García A E 2003 *Proc. Natl. Acad. Sci. USA* **100** 13934.
- [23] Blanco F J, Rivas G and Serrano L 1994 *Nat. Struct. Biol.* **1** 584.

- [24] Roccatano D, Amadei A, Di Nola A and Berendsen H J C 1999 *Protein Sci.* **8** 2130.
- [25] Pande V S and Rokhsar D S 1999 *Proc. Natl. Acad. Sci. USA* **96** 9062.
- [26] Dinner A R, Lazaridis T and Karplus M 1999 *Proc. Natl. Acad. Sci. USA* **96** 9068.
- [27] García A E and Sanbonmatsu K Y 2001 *Proteins* **42** 345.
- [28] Zagrovic B, Sorin E J and Pande V 2001 *J. Mol. Biol.* **313** 151.
- [29] Kussell E, Shimada J and Shakhnovich E I 2002 *Proc. Natl. Acad. Sci. USA* **99** 5343.
- [30] Zhou R 2003 *Proteins* **53** 148.
- [31] Bolhuis P G 2003 *Proc. Natl. Acad. Sci. USA* **100** 12129.
- [32] Wei G, Mousseau N and Derreumaux P 2004 *Proteins* **56** 464.
- [33] Fesinmeyer R M, Hudson F M and Andersen N H 2004 *J. Am. Chem. Soc.* **126** 7238.
- [34] Muñoz V, Thompson P A, Hofrichter J and Eaton W A 1997 *Nature* **390** 196.
- [35] Kortemme T, Ramírez-Alvarado M and Serrano L 1998 *Science* **281** 253.
- [36] Bursulaya B D and Brooks III C L 1999 *J. Am. Chem. Soc.* **121** 9947.
- [37] Colombo G, Roccatano D and Mark A E 2002 *Proteins* **46** 380.
- [38] Kim S Y, Lee J and Lee J 2004 *J. Chem. Phys.* **120** 8271.
- [39] López de la Paz M, Lacroix E, Ramírez-Alvarado M and Serrano L 2001 *J. Mol. Biol.* **312** 229.
- [40] Favrin G, Irbäck A, Samuelsson B and Wallin S 2003 *Biophys. J.* **85** 1457.
- [41] Sunde M and Blake C 1997 *Adv. Protein Chem.* **50** 123.
- [42] Balbach J J, Ishii Y, Antzutkin O N, Leapman R D, Rizzo N W, Dyda F, Reed J and Tycko R 2000 *Biochemistry* **39** 13748.
- [43] Gordon D J, Balbach J J, Tycko R and Meredith S C 2004 *Biophys. J.* **86** 428.
- [44] Ma B and Nussinov R 2002 *Proc. Natl. Acad. Sci. USA* **99** 14126.
- [45] Klimov D K and Thirumalai D 2003 *Structure* **11** 295.
- [46] Diaz-Avalos R, Long C, Fontano E, Balbirnie M, Grothe R, Eisenberg D and Caspar D L D 2003 *J. Mol. Biol.* **330** 1165.

Tables

Table 1. The probability distribution $P(n_+, n_-)$ for $N_c = 3$ A β_{16-22} peptides at 275 K. $P(n_+, n_-)$ values smaller than 10^{-3} are omitted. The numbers in parentheses are the statistical errors in the last digits.

n_+	n_-		
	0	1	2
0	0.17 (2)	0.22 (3)	0.14 (3)
1	0.13 (2)	0.32 (6)	
2	0.020 (7)		

Table 2. Same as Table 1 for $N_c = 6$ A β_{16-22} peptides at 287 K.

n_+	n_-				
	0	1	2	3	4
0	0.028 (5)	0.059 (11)	0.08 (2)	0.06 (2)	0.030 (15)
1	0.038 (6)	0.12 (2)	0.16 (3)	0.10 (3)	0.006 (3)
2	0.026 (11)	0.11 (5)	0.14 (5)	0.004 (2)	
3	0.008 (5)	0.013 (9)	0.015 (12)		

Table 3. 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 [33], whereas those on Betanova and LLM are from [39].

	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)

Figure captions

Figure 1. Schematic illustration of the different geometries of the peptides studied. Shown from left to right are the Trp cage, F_s, GB1m3 and Betanova. Drawn with RasMol [9].

Figure 2. The Trp cage. (a) Helix content against temperature. The line is a fit to the two-state expression $H(T) = (H_u + H_n K(T))/(1 + K(T))$. 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 $\Delta E = E_u - E_n$ is the energy difference between the unfolded and native states. (b) Native population against temperature in the model (line), as obtained from the fit of the helix content. Plot symbols show experimental results [10] based on CD (○) and NMR (●), respectively.

Figure 3. Same as Fig. 2a for the F_s peptide.

Figure 4. (a) The hydrophobicity energy E_{hp} against temperature for GB1p (○) and GB1m3 (●). The lines are two-state fits (see Fig. 2a). 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. (b) Probability distribution of the number of native backbone 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).

Figure 5. (a) The hydrophobicity energy E_{hp} against temperature for Betanova (○) and LLM (●). The lines are two-state fits (see Fig. 2a). (b) Probability distribution of the number of native backbone hydrogen bonds, N_{hb}^{nat} , for LLM (full line) and Betanova (dotted line) at 287 K. In Betanova notation, the native hydrogen bonds are Ser4(N)-Thr11(O), Ser4(O)-Thr11(N), Gln6(N)-Lys9(O), Gln6(O)-Lys9(N), Tyr10(N)-Thr17(O), Tyr10(O)-Thr17(N), Asn12(N)-Lys15(O) and Asn12(O)-Lys15(N).

Figure 6. (a) The α -helix content H against temperature T for A β_{16-22} for $N_c = 1$ (○), $N_c = 3$ (●) and $N_c = 6$ (▲). Lines joining data points are only a guide for the eye. (b) Same for the β -strand content S . Note that the scales in (a) and (b) are different.

Figure 7. Two typical low-energy structures from our simulations of six A β_{16-22} peptides: a five-stranded β -sheet (left), and two three-stranded β -sheets “sandwiching” several of their hydrophobic side-chains between them (right). Drawn with RasMol [9].



Fig. 1

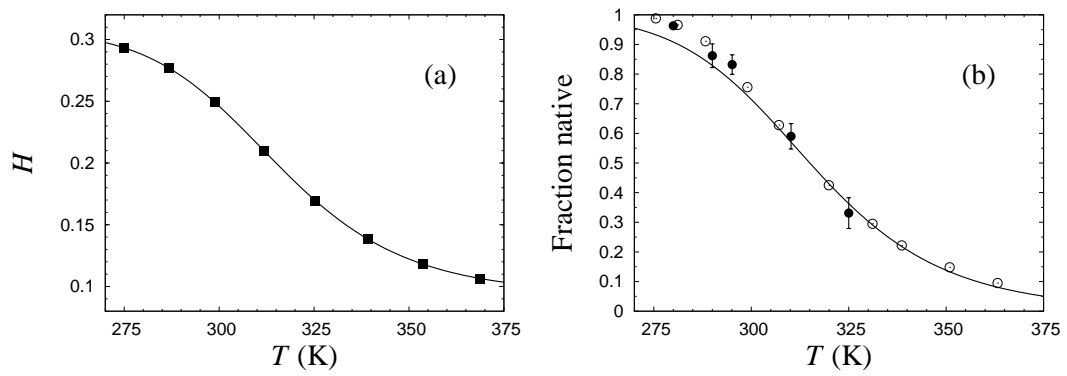


Fig. 2

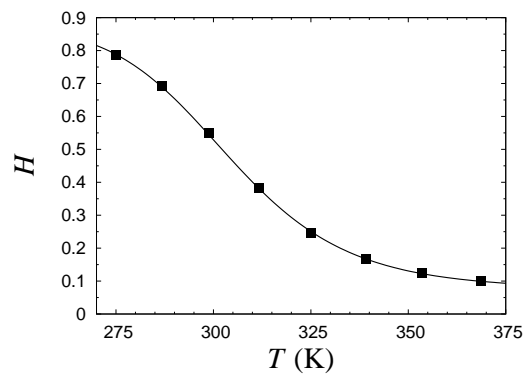


Fig. 3

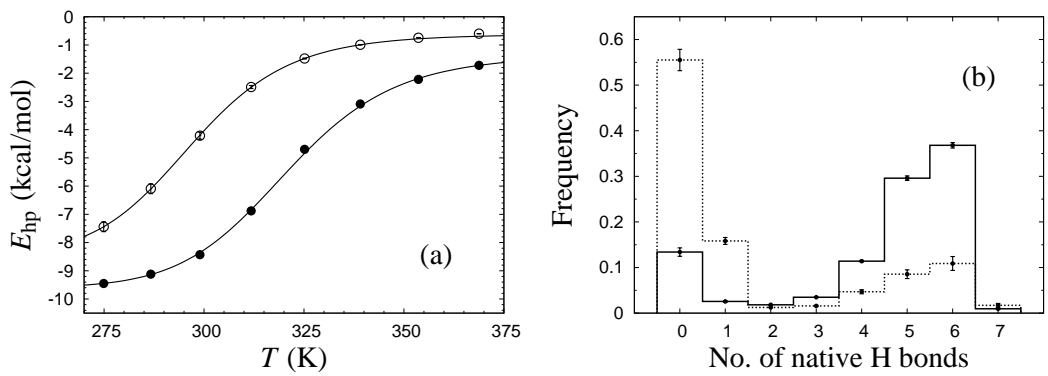


Fig. 4

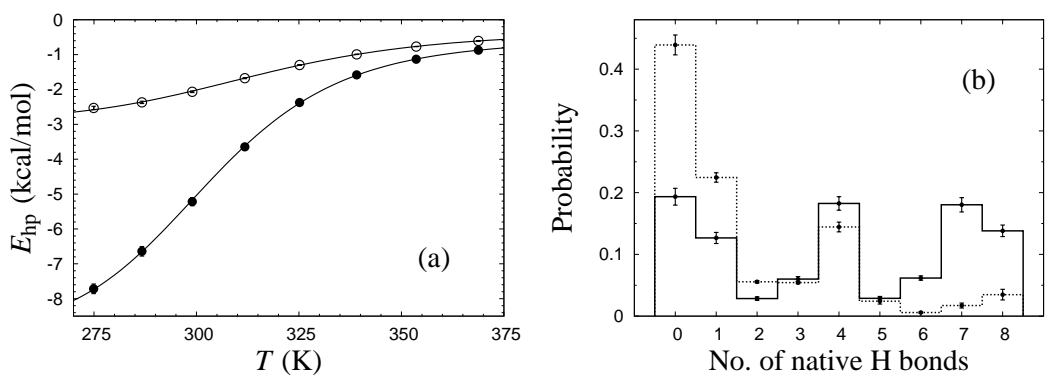


Fig. 5

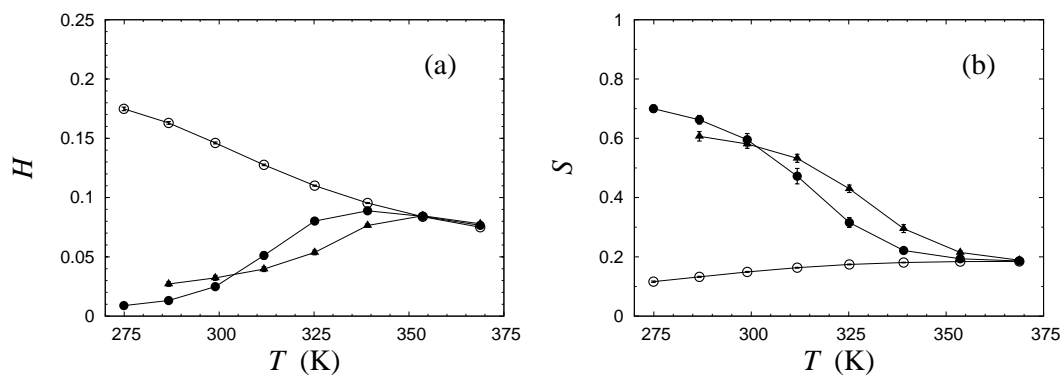


Fig. 6



Fig. 7