PROTEINS: Structure, Function, and Genetics 47:99–105 (2002)

Folding of a Small Helical Protein Using Hydrogen Bonds and Hydrophobicity Forces

Giorgio Favrin, Anders Irbäck,* and Stefan Wallin
Department of Theoretical Physics, Complex Systems Division, Lund University, Lund, Sweden

ABSTRACT **A reduced protein model with five to six atoms per amino acid and five amino acid types is developed and tested on a three-helix-bundle protein, a 46-amino acid fragment from staphylococcal protein A. The model does not rely on the widely used Gō approximation, which ignores non-native interactions. We find that the collapse transition is considerably more abrupt for the protein A sequence than for random sequences with the same composition. The chain collapse is found to be at least as fast as helix formation. Energy minimization restricted to the thermodynamically favored topology gives a structure that has a root-mean-square deviation of 1.8 Å from the native structure. The sequence-dependent part of our potential is pairwise additive. Our calculations suggest that fine-tuning this potential by parameter optimization is of limited use. Proteins 2002;47:99–105.**

Key words: protein folding; folding thermodynamics; folding kinetics; Monte Carlo simulation; energy minimization

INTRODUCTION

In recent years, several important insights have been gained into the physical principles of protein folding.[1–6] Still, in terms of quantitative predictions, it is clear that it would be extremely useful to be able to perform more realistic folding simulations than are currently possible. In fact, most models used so far for statistical–mechanical simulations of folding rely on one or both of two quite drastic approximations, the lattice and Gō[7] approximations.

The reason that lattice models have been used to study basics of protein folding is partly computational, but also physical—on the lattice, the potential to use for stable and fast-folding sequences to exist is known (a simple contact potential is sufficient). By contrast, how to satisfy these criteria for off-lattice chains is largely unknown, and therefore many current off-lattice models[5,8–15] use Gō-type potentials[7] where non-native interactions are ignored. The use of the Gō approximation has some support from the finding that the native structure is a determinant for folding kinetics.[16,17] However, it is an uncontrolled approximation, and it is useless when it comes to structure prediction because it requires prior knowledge of the native structure.

This article discusses an off-lattice model that does not follow the Gō prescription. Using this model, we perform extensive folding simulations for a small helical protein. The force field of the model is simple and is based on hydrogen bonds and effective hydrophobicity forces (no explicit water). Other non Gō-like models with more elaborate force fields have been used for structure prediction with some success.[18–20] However, the dynamic properties of these models are unclear.

The original version of our model was presented earlier[21] and includes three types of amino acids: hydrophobic, polar, and glycine. This version was applied to a designed three-helix-bundle protein with 54 amino acids.[21] For suitable relative strength of the hydrogen bonds and hydrophobicity forces, this sequence was found to form a stable three-helix bundle, except for a twofold topologic degeneracy. In addition, its folding transition is first-order-like and coincides with the collapse transition (parameter σ of ref. 4 is zero).

In the present study, we extend this model from three to five amino acid types, by taking alanine to be intermediate in hydrophobicity between the previous two hydrophobic and polar classes, and by introducing a special geometric representation for proline, which is needed to be able to mimic the helix-breaking property of this amino acid. Otherwise, the model is the same as the earlier model. The modified model is tested on a real three-helix-bundle protein, the 10–55-amino acid fragment of the B domain of staphylococcal protein A. The structure of this protein has been determined by nuclear magnetic resonance (NMR),[22] and an energy-based structure prediction method has been tested on the sequence.[18] The folding properties have been studied too, both experimentally[23,24] and theoretically.[8,11,12,25–27] In particular, this means that we can compare the behavior of previous Gō-like models with that of our more realistic model.

MATERIALS AND METHODS

Geometry

Our model is an extension of that introduced previously.[21] It uses three different amino acid representations: one for glycine, one for proline, and one for the rest. The nonglycine, nonproline representation is illustrated in Figure 1(a) and is identical to that of hydrophobic and

Grant sponsor: Swedish Foundation for Strategic Research.

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

Received 9 August 2001; Accepted 12 November 2001

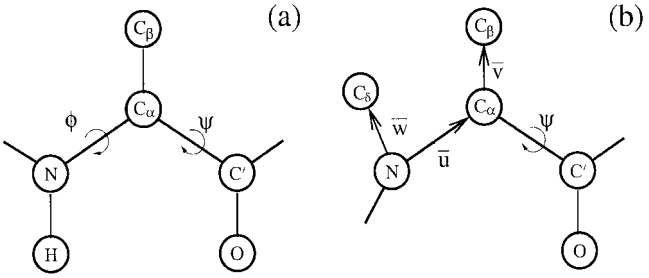

Fig. 1. **a:** Schematic representation of the common geometric representation for all amino acids except glycine and proline. **b:** The representation of proline. The C_δ atom is assumed to lie in the plane of the N, C_α and C_β atoms. The N—C_δ bond vector **w** is given by $\mathbf{w} = -0.596\mathbf{u} + 0.910\mathbf{v}$, where the vectors **u** and **v** are defined in the figure. The numerical factors were obtained by an analysis of structures from the Protein Data Bank (PDB).[28]

polar amino acids in the original model. The three backbone atoms N, C_α, and C′ are all included, whereas the side-chain is represented by a single atom, a large C_β. The remaining two atoms, H and O, are used to define hydrogen bonds. The representation of glycine is the same, except that C_β is missing.

The representation of proline is new compared with the original model. The side-chain of proline is attached to the backbone not only at C_α, but at N as well. A well-known consequence is that proline can act as a helix breaker. For the model to be able to capture this important property, we introduce a special representation for proline, illustrated in Figure 1(b). It differs from that represented in Figure 1(a) in two ways: first, the Ramachandran angle ϕ is held constant, at −65°; and second, the H atom is replaced by a side-chain atom, C_δ. This more realistic representation of proline is needed when studying the protein A fragment, which has one proline at each of the two turns.

All amino acids except proline have the Ramachandran torsion angles ϕ and ψ [Fig. 1(a)] as their degrees of freedom, whereas ψ is the only degree of freedom for proline. All bond lengths, bond angles, and peptide torsion angles (180°) are held fixed. Numerical values of the bond lengths and bond angles can be found in the earlier study[21] and in Figure 1(b).

The helix-breaking property of proline is clearly manifested in the shape of the ψ distribution for amino acids that are followed by a proline in the sequence (with the proline on their C′ side). Helical values of ψ are suppressed for such amino acids. This is illustrated in Figure 2(a), where the peak on the left corresponds to the α-helix. Figure 2(b) shows that the model exhibits a qualitatively similar behavior.

Force Field

Our energy function

$$E = E_{\mathrm{loc}} + E_{\mathrm{sa}} + E_{\mathrm{hb}} + E_{\mathrm{col}} \tag{1}$$

is composed of four terms. The first two terms E_{loc} and E_{sa} are local ϕ, ψ and self-avoidance potentials, respectively.[21] The third term is the hydrogen-bond energy E_{hb}, which is given by

$$E_{\mathrm{hb}} = \epsilon_{\mathrm{hb}} \sum_{ij} \left[5\left(\frac{\sigma_{\mathrm{hb}}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{\mathrm{hb}}}{r_{ij}}\right)^{10} \right] v(\alpha_{ij}, \beta_{ij}) \tag{2}$$

$$v(\alpha_{ij}, \beta_{ij}) = \begin{cases} \cos^2 \alpha_{ij} \cos^2 \beta_{ij} & \alpha_{ij}, \beta_{ij} > 90^\circ \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where i and j represent H and O atoms, respectively, and where r_{ij} denotes the HO distance, α_{ij} the NHO angle, and β_{ij} the HOC′ angle.

The last term in equation (1), the hydrophobicity or collapse energy E_{col}, has the form

$$E_{\mathrm{col}} = \epsilon_{\mathrm{col}} \sum_{i<j} \Delta(s_i, s_j) \left[\left(\frac{\sigma_{\mathrm{col}}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{\mathrm{col}}}{r_{ij}}\right)^{6} \right] \tag{4}$$

where the sum runs over all possible $C_\beta C_\beta$ pairs, and s_i denotes amino acid type. To define $\Delta(s_i, s_j)$, we divide the amino acids into three classes: hydrophobic (H; Leu, Ile, Phe), alanine (A; Ala), and polar (P; Arg, Asn, Asp, Gln, Glu, His, Lys, Pro, Ser, Tyr).* There are then six kinds of $C_\beta C_\beta$ pairs, and the corresponding $\Delta(s_i, s_j)$ values are taken to be

$$\Delta(s_i, s_j) = \begin{cases} 1 \text{ for HH and HA pairs} \\ 0 \text{ for HP, AA, AP, and PP pairs} \end{cases} \tag{5}$$

The main change in the force field compared with that described previously[21] is that alanine forms its own hydrophobicity class in addition to the previous two hydrophobic and polar classes. Alanine is taken as intermediate in hydrophobicity, meaning that there is a hydrophobic interaction between HA pairs, but not between AA pairs. In addition, the interaction strength ϵ_{col} is increased slightly, from 2.2 to 2.3.† Finally, in the self-avoidance potential, the C_δ atom of proline is assigned the same size as C_β atoms. Otherwise, the entire force field, including parameter values, is exactly the same as described previously.[21]

With these changes in geometry and force field, we end up with five different amino acid types in the new model. First, we have hydrophobic, alanine, and polar, which share the same geometric representation but differ in hydrophobicity, and then glycine and proline with their special geometries.

In this study, we test this model on the 10–55-amino acid fragment of the B domain of staphylococcal protein A. Calculated structures are compared with the minimized average NMR structure[22] with Protein Data Base (PDB) code 1bdd. Throughout the present study, this structure is referred to as the native structure.

As a first test of our model, two different fits to the native structure were made. The first fit is purely geometric. Here, we simply minimized the root-mean-square deviation (RMSD) from the native structure, δ (calculated over all backbone atoms). This was done by using simulated annealing, and the best result was δ = 0.14 Å. In the

*Cys, Met, Thr, Trp, and Val do not occur in the sequence studied.

†The energy unit is dimensionless and such that $kT_c = 0.62$, with T_c the collapse temperature (see Results and Discussion).

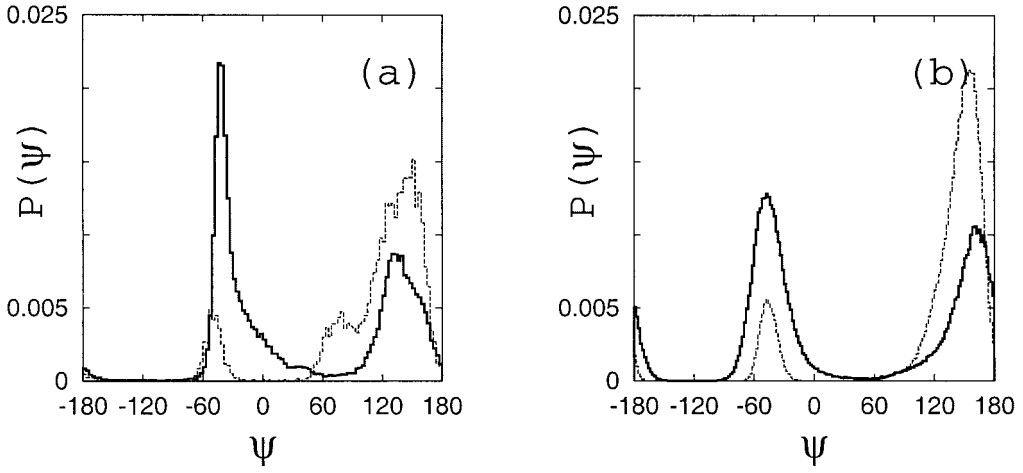

Fig. 2. **a:** Distributions of the Ramachandran angle ψ, based on Protein Data Bank (PDB) data. The full (dashed) line represents nonglycine, nonproline amino acids that are followed by a nonproline (proline) in the sequence. **b:** The corresponding histograms for the model, as obtained by simulations of Gly-X-X (full line) and Gly-X-Pro (dashed line) at $kT = 0.55$, where *X* denotes polar amino acids (the curves show the ψ distribution for the middle of the three amino acids).

second fit, we took into account the limitations imposed by the first three terms of the potential, by minimizing the function

$$\tilde{E} = E_{\text{loc}} + E_{\text{sa}} + E_{\text{hb}} + \kappa \sum_i (\mathbf{r}_i - \mathbf{r}_i^0)^2 \qquad (6)$$

where $\kappa = 1$ Å^{-2} and $\{\mathbf{r}_i^0\}$ denotes the structure obtained from the first fit. The minimum $\tilde{E}$ structure had $\delta = 0.32$ Å. These results show that, in spite of relatively few degrees of freedom, our model permits quite an accurate description of the real structure.

Numerical Methods

To simulate the thermodynamic behavior of this model, we use simulated tempering,[29–31] which means that the temperature is a dynamic variable (for details, see refs. 29–31). The temperature update is a standard Metropolis step. Our conformation updates are of two different types: the simple nonlocal pivot move, in which a single torsion angle is turned, and a semilocal move.[32] The latter method works with the Ramachandran angles of four adjacent amino acids. These are turned with a bias toward local rearrangements of the chain. The degree of bias is governed by a parameter b. In our thermodynamic simulations, we take $b = 10$ (rad/Å)^2, which gives a strong bias toward deformations that are approximately local.[32]

Figure 3 shows the evolution of the energy in a simulated-tempering run that took about two weeks on an 800-MHz processor. Data corresponding to all the different temperatures are shown (eight temperatures, ranging from $kT = 0.54$ to $kT = 0.90$). We see that there are many independent visits to low-energy states, which is necessary in order to get a reliable estimate of the relative populations of the folded and unfolded states. To test the value of the semilocal update, we repeated the same calculation using pivot moves only. The difference in performance was not quantified, but it was clear that the sampling of low energies was less efficient in the run that relied solely on pivot moves.

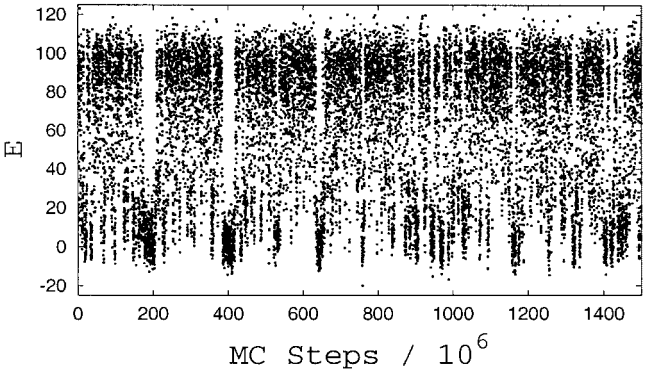

Fig. 3. Monte Carlo evolution of the energy in a simulated-tempering run.

For our kinetic simulations, we do not use the pivot update, but only the semilocal method. Parameter b is taken to be 1 (rad/Å)^2 in the kinetic runs, which turned out to give an average change in the end-to-end vector squared of $\sim$0.5 Å^2.

RESULTS AND DISCUSSION

Thermodynamics

We begin our study of the model defined in Materials and Methods by locating the collapse transition. Figure 4 shows the radius of gyration (calculated over all backbone atoms) against temperature for both the protein A sequence and three random sequences with the same length and composition. The random sequences were generated keeping the two prolines of the protein A sequence fixed at their positions, one at each turn. The remaining 44 amino acids were randomly reshuffled.

Naively, one may expect these sequences to show similar collapse behaviors, since the composition is the same.

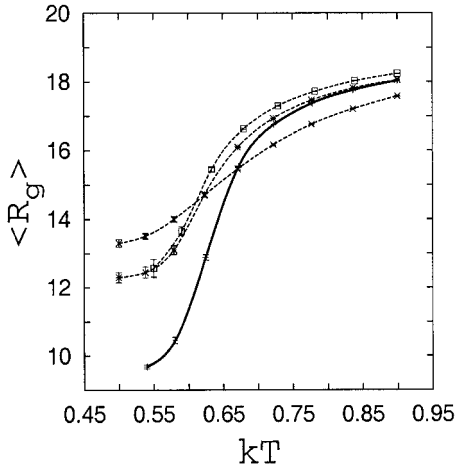

Fig. 4. The radius of gyration (in Å) against temperature. Full and dashed lines represent the protein A sequence and the three random sequences (see text), respectively.

However, the protein A sequence turns out to collapse much more efficiently than the random sequences (Fig. 4). The native structure has a radius of gyration of 9.25 Å, which is significantly smaller than one finds for the random sequences in this temperature range. The specific heat (data not shown) has a pronounced peak in the region where the collapse occurs. Taking the maximum as the collapse temperature T_c, we obtain $kT_c = 0.62$ for the protein A sequence.

The chain collapse is not as abrupt for the protein A sequence as for the designed sequence studied in ref. 21. This is not surprising, as that sequence has a hydrophobicity pattern that fits its native structure perfectly. Although the protein A sequence does not have a fully perfect hydrophobicity pattern, the collapse behavior is highly cooperative, as can be seen from the comparison with the random sequences.

Next, we turn to the structure of the collapsed state. As a measure of similarity with the native structure, we use

$$Q = \exp(-\delta^2/100\ \text{Å}^2) \tag{7}$$

where δ, as before, denotes RMSD. An alternative would be to base the similarity measure on the number of native contacts present, rather than RMSD. The problem with such a definition is that it does not provide an efficient discrimination between the two possible topologies of a three-helix bundle[33]—the third helix can be either in front of, or behind, the U formed by the first two helices. This problem is avoided by using RMSD.

Figure 5(a) shows the free-energy profile $F(Q)$ in the collapsed phase at $kT = 0.54$. We see that there is a broad minimum at $Q \approx 0.8$–0.9, with two distinct local minima at $Q = 0.78$ and $Q = 0.90$, respectively. Both minima correspond to the native overall topology. There is also a minimum at $Q = 0.50$, which corresponds to the wrong topology. The $Q = 0.50$ minimum is more narrow and slightly higher, so the native topology is the favored one. It should be stressed, however, that it is difficult to discriminate between the two topologies using a pairwise additive potential (see the section Fine-tuning?). To be able to do that properly, it is likely that one has to include multibody terms and/or more side-chain atoms in the model.

The main difference between the two minima at $Q = 0.78$ and $Q = 0.90$ lies in the shape and orientation of helix III, which comprises amino acids 41–55 in the native structure. At the $Q = 0.78$ minimum, there tends to be a sharp bend in this segment, and the amino acids before the bend, 41–44, are disordered rather than helical. The remaining amino acids, 45–55, tend to make a helix, but its orientation differs from that in the native structure. Relative to the $Q = 0.90$ minimum, where helix III is much more native-like, we find that the $Q = 0.78$ minimum is entropically favored but energetically disfavored. The separation in energy between these minima is probably underestimated by our model. There is, for example, a stabilizing electrostatic interaction between helices I and III in the native structure (Glu16—Lys50), which should favor the $Q = 0.90$ minimum but is missing in our model.

Also shown in Figure 5(a) is the result for one of the random sequences. The probability of finding this sequence in the vicinity of the native structure is, not unexpectedly, very low. The same holds true for the other two random sequences too (data not shown).

To extract representative conformations for the collapsed state, we used simulated annealing followed by a conjugate–gradient minimization. Using this procedure, a large set of low-temperature Monte Carlo conformations were quenched to zero temperature. Figure 5(b) shows the quenched conformations with lowest energy in a Q, E scatter plot. Our minimum-energy structure is found at $Q = 0.44$, corresponding to $\delta = 9.1$ Å. However, our thermodynamic calculations show that this conformation is not very relevant, in spite of its low energy. If we restrict ourselves to conformations with the native-like and thermodynamically most relevant topology, the lowest energy is at $Q = 0.97$, corresponding to $\delta = 1.8$ Å. This conformation is shown in Figure 6 along with the native structure. It is worth noting that both the $Q = 0.44$ and $Q = 0.97$ minima were revisited in independent runs.

These results can be compared with those of Scheraga and coworkers,[18] who tested an energy-based structure prediction method on the same sequence. With their energy function, the global minimum was found to have an RMSD of 3.8 Å from the native structure (calculated over C_α atoms).

Helix Stability

Having discussed the overall thermodynamic behavior, we now take a closer look at the stability of the secondary structure and at how it varies along the chain. To this end, we monitored the hydrogen-bond energy between the CO group of amino acid i and the NH group of amino acid $i + 4$ [see equations (2) and (3)], $e_{hb}(i)$, as a function of i. This was done not only for the protein A sequence, but also for the corresponding three one-helix segments, which are listed in Table I. An experimental study[24] of essentially

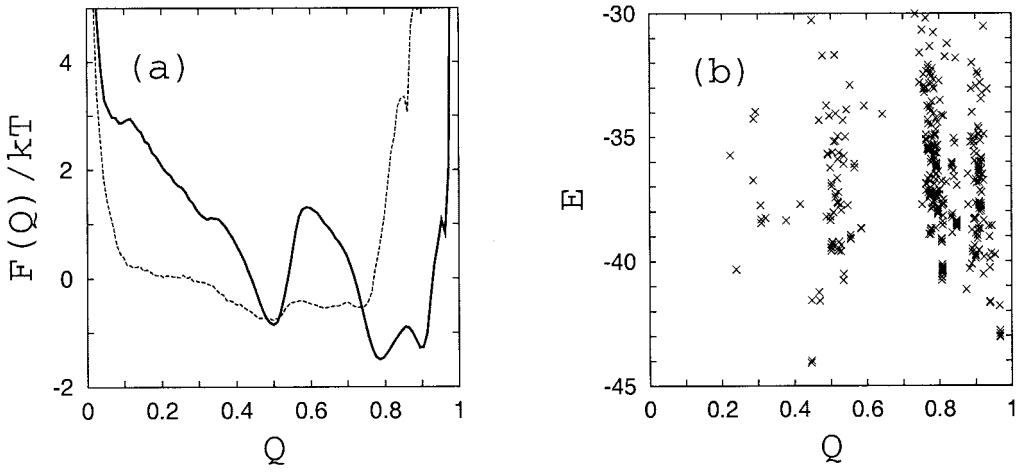

Fig. 5. **a:** Free-energy profile $F(Q) = -kT \ln P(Q)$ at $kT = 0.54$ (full line), where $P(Q)$ is the probability distribution of Q. Also shown (dashed line) is the result for one of the random sequences at $kT = 0.50$. **b:** Q, E scatter plot for quenched conformations with low energy.

Fig. 6. Schematic illustrations of the native structure (left) and our minimum-energy structure for the native topology (right). Drawn with RasMol.[34]

TABLE I. The One-Helix Fragments Studied

Segment	Sequence	Amino acids
I	QQNAFYEILHL	10–20
II	NEEQRNGFIQSLKDD	24–38
III	QSANLLAEAKKLNDA	41–55

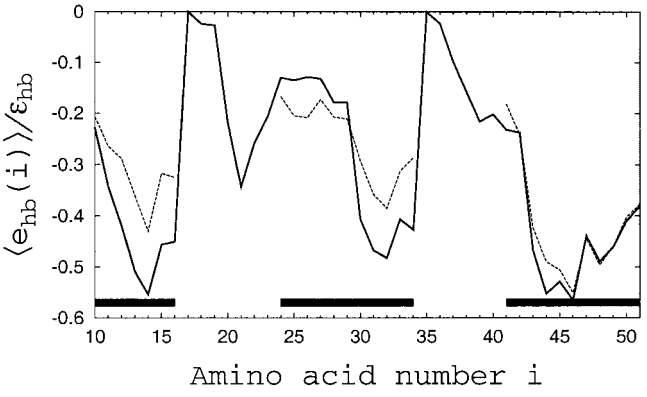

Fig. 7. Hydrogen-bond profile showing the normalized average energy of α-helical hydrogen bonds, $\langle e_{hb}(i)\rangle/\epsilon_{hb}$, against amino acid number i, at $kT = 0.58$. The full line represents the protein A sequence, whereas the dashed lines represent the corresponding three one-helix segments (Table I). Thick horizontal lines indicate hydrogen bonds present in the native structure.

the same three segments found segment III to be the only one that shows some stability on its own.

The results of our calculations are shown in Figure 7, from which we see that the difference between the full sequence and the one-helix segments is not large in the model. However, segments I and II definitely make less stable helices on their own than as interacting parts of the full system; they are stabilized by interhelical interactions. Furthermore, among the three one-helix segments, the model correctly predicts segment III to be the most stable. That this segment does not become more stable as part of the full system is probably related to the earlier observation that helix III is distorted at the $Q = 0.78$ minimum.

A striking detail shown in Figure 7 is that the beginning of segment II is quite unstable. This can be easily understood. This segment has a flexible glycine at position 30, and the amino acids before the glycine, 24–29, are all polar, so there are no hydrophobic interactions that can help stabilize this part.

Kinetics

Using the semilocal update,[32] we performed a set of 30 kinetic simulations at $kT = 0.54$. The runs were started from random coils. There are big differences between these runs, partly because the system, as it should, sometimes spends a significant amount of time in the wrong topology. Nevertheless, the data show one stable and interesting trend, namely, that the formation of helices was never faster than the collapse. This is illustrated in Figure 8, which shows the evolution of the similarity parameter Q_0, the hydrogen-bond energy E_{hb}, and the radius of gyration, R_g, in one of the runs. Q_0 is defined as Q in equation (7), except that it measures similarity to the optimized model structure in Figure 6, rather than the native structure. In Figure 8, we see that E_{hb} converges slowly, whereas the collapse occurs relatively early.

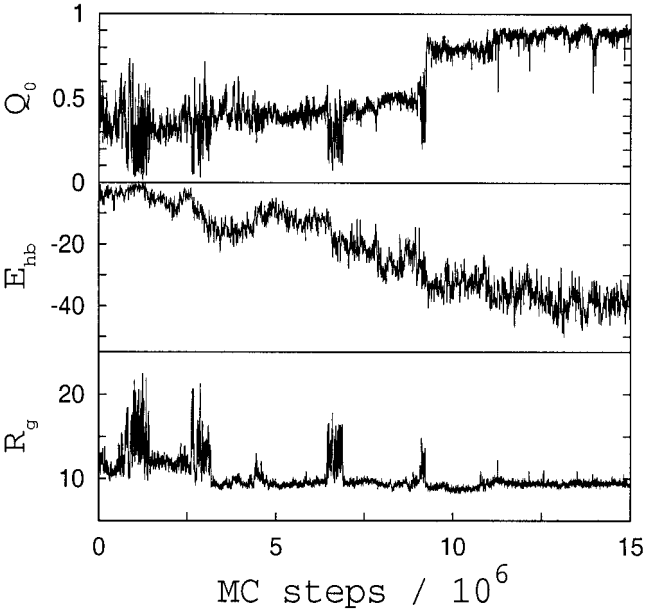

Fig. 8. Monte Carlo evolution of the similarity parameter Q_0 (top), the hydrogen-bond energy E_{hb} (middle) and the radius of gyration R_g (bottom) in a kinetic simulation at $kT = 0.54$.

Now, at a first glance, it may seem easy to make the helix formation faster by simply increasing the strength of the hydrogen bonds. Therefore, it is important to note that the hydrogen bonds cannot be made much stronger without making the ground state noncompact, thus destroying the three-helix bundle.[35] This means that the conclusion that the collapse is at least as fast as helix formation holds for any reasonable choice of parameters in this model.

It is interesting to compare these results with those of Zhou and Karplus,[11] who studied the same protein using a Gō-type potential and observed fast folding when the Gō forces were strong. Under these conditions, the helix formation was found to be fast, whereas the collapse was the rate-limiting step.

However, a Gō-like model ignores a large fraction of the interactions that drive the collapse, which can make the collapse artificially slow. In a recent Gō model study,[15] this problem was addressed by eliminating backbone terms from the potential until a reasonable helix stability was achieved. No such calibration was carried out by Zhou and Karplus.[11] This may explain why these investigators find a behavior that our model cannot reproduce.

Finally, let us mention that we also performed the same type of kinetic simulations for the designed sequence discussed earlier,[21] which has a very abrupt collapse transition. It turns out that E_{hb} and R_g evolve in a strongly correlated manner in this case. Thus, the helix formation and collapse occur simultaneously for this sequence.

Fine-tuning?

In the section Thermodynamics, we discussed the relative weights of the two possible overall topologies, which is a delicate issue. What changes are needed in order for the model to suppress the wrong topology more strongly? Is it necessary to change the form of the energy function, or would it be sufficient to fine-tune the interaction matrix $\Delta(s_i, s_j)$ in equation (4)?

One way to accomplish such a fine-tuning of $\Delta(s_i, s_j)$ would be to maximize $\langle Q \rangle'$, where Q is the similarity parameter and $\langle \cdot \rangle'$ denotes a thermodynamic average restricted to compact conformations (e.g., $R_g < 10$ Å). This is essentially the overlap method described in ref. 36. The gradient of the quantity $\langle Q \rangle'$ can be written as

$$\frac{\partial \langle Q \rangle'}{\partial \Delta(s_i, s_j)} = -\frac{\epsilon_{col}}{kT} (\langle QX \rangle' - \langle Q \rangle' \langle X \rangle') \qquad (8)$$

where X is a sum of Lennard–Jones terms $(\sigma_{col}/r_{ij})^{12} - 2(\sigma_{col}/r_{ij})^6$, over all possible $C_\beta C_\beta$ pairs of type s_i, s_j.

We calculated the Q, X correlation in equation (8) for all pairs s_i, s_j with $\Delta(s_i, s_j) = 1$ at $kT = 0.54$ and found that $|\partial \langle Q \rangle' / \partial \Delta(s_i, s_j)|$ was small (≤ 0.15) for all these pairs. Hence, there is no sign that a significant increase in $\langle Q \rangle'$ can be achieved by fine-tuning $\Delta(s_i, s_j)$; the contact patterns appear to be too similar in the two topologies. To include more side-chain atoms and/or multibody terms in the model is likely to be a more fruitful approach.

CONCLUSION

We have explored a five-letter protein model with five to six atoms per amino acid, in which the formation of native structure is driven by hydrogen bonding and effective hydrophobicity forces. This model, which does not follow the Gō prescription, was tested on a small but real sequence, a three-helix-bundle fragment from protein A.

With this model, the protein A sequence was found to collapse much more efficiently than do random sequences with the same composition. In the collapsed phase, we found that the native topology dominates, although the suppression of the wrong three-helix-bundle topology is not strong. Energy minimization constrained to the thermodynamically favored topology gave a structure with an RMSD of 1.8 Å from the native structure.

In our kinetic simulations, the collapse was always at least as fast as helix formation, in sharp contrast with previous results for the same protein that were obtained using a Gō-like C_α model.[11] A possible explanation for the conflicting conclusions is that the Gō approximation makes the collapse artificially slow by ignoring a large fraction of the interactions driving the collapse. In our model, the conclusion that the helix formation is not faster than collapse seems unavoidable; if one tries to speed up the helix formation by increasing the strength of the hydrogen bonds, the chain does not fold into a compact helical bundle.

The force field of our model was deliberately kept simple. In particular, the hydrophobicity potential was taken to be pairwise additive, with a simple structure for the interaction matrix $\Delta(s_i, s_j)$ [see equation (5)]. It would be very interesting to look into the behavior of the model in the presence of multibody terms. A simpler alternative is to stick to the pairwise additive potential and fine-tune the parameters $\Delta(s_i, s_j)$. However, the calculations in this

article give no indication that there is much to be gained from such fine-tuning.

REFERENCES

1. Săli A, Shakhnovich E, Karplus M. Kinetics of protein folding: a lattice model study of the requirements for folding to the native state. J Mol Biol 1994;235:1614–1636.
2. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167–195.
3. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–19.
4. Klimov DK, Thirumalai D. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. J Chem Phys 1998;109:4119–4125.
5. Nymeyer H, García AE, Onuchic JN. Folding funnels and frustration in off-lattice minimalist protein landscapes. Proc Natl Acad Sci USA 1998;95:5921–5928.
6. Hao M-H, Scheraga HA. Theory of two-state cooperative folding of proteins. Acc Chem Res 1998;31:433–440.
7. Gō N, Taketomi H. Respective roles of short- and long-range interactions in protein folding. Proc Natl Acad Sci USA 1978;75: 559–563.
8. Zhou Y, Karplus M. Folding thermodynamics of a model three-helix-bundle protein. Proc Natl Acad Sci USA 1997;94:14429–14432.
9. Shea J-E, Nochomovitz YD, Guo Z, Brooks CL III. Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist β-barrel model. J Chem Phys 1998;109:2895–2903.
10. Micheletti C, Banavor JR, Maritan A, Seno F. Protein structures and optimal folding from a geometrical variational principle. Phys Rev Lett 1999;82:3372–3375.
11. Zhou Y, Karplus M. Interpreting the folding kinetics of helical proteins. Nature 1999;401:400–403.
12. Shea J-E, Onuchic JN, Brooks CL III. Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. Proc Natl Acad Sci USA 1999;96:12512–12517.
13. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000;298: 937–953.
14. Clementi C, Jennings PA, Onuchic JN. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1β. Proc Natl Acad Sci USA 2000;97:5871–5876.
15. Shimada J, Kussell EL, Shakhnovich EI. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. J Mol Biol 2001;308:79–95.
16. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–994.
17. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
18. Lee J, Liwo A, Scheraga HA. Energy-based de novo protein folding by conformational space annealing and an off-lattice united-residue force field: application to the 10–55 fragment of staphylococcal protein A and to apo calbindin D9K. Proc Natl Acad Sci USA 1999;96:2025–2030.
19. Pillardy J, Czaplewski C, Liwo A, Lee J, Ripoll DR, Kaźmierkiewicz R, Ołdziej S, Wedemeyer WJ, Gibson KD, Arnautova YA, Saunders J, Ye Y-J, Scheraga HA. Recent improvements in prediction of protein structure by global optimization of a potential energy function. Proc Natl Acad Sci USA 2000;98:2329–2333.
20. Hardin C, Eastwood MP, Luthey-Schulten Z, Wolynes PG. Associative memory Hamiltonians for structure prediction without homology: alpha-helical proteins. Proc Natl Acad Sci USA 2000;97:14235–14240.
21. Irbäck A, Sjunnesson F, Wallin S. Three-helix-bundle protein in a Ramachandran model. Proc Natl Acad Sci USA 2000;97:13614–13618.
22. Gouda H, Torigoe H, Saito A, Sato M, Arata Y, Shimada I. Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures. Biochemistry 1992;31:9665–9672.
23. Bottomley SP, Popplewell AG, Scawen M, Wan T, Sutton BJ, Gore MG. The stability and unfolding of an IgG binding protein based upon the B domain of protein A from *Staphylococcus aureus* probed by tryptophan substitution and fluorescence spectroscopy. Protein Eng 1994;7:1463–1470.
24. Bai Y, Karimi A, Dyson HJ, Wright PE. Absence of a stable intermediate on the folding pathway of protein A. Protein Sci 1997;6:1449–1457.
25. Boczko EM, Brooks CL III. First-principles calculation of the folding free energy of a three-helix bundle protein. Science 1995; 269:393–396.
26. Guo Z, Brooks CL III, Boczko EM. Exploring the folding free energy surface of a three-helix bundle protein. Proc Natl Acad Sci USA 1997;94:10161–10166.
27. Kolinski A, Galazka W, Skolnick J. Monte Carlo studies of the thermodynamics and kinetics of reduced protein models: application to small helical, β and α/β proteins. J Chem Phys 1998;108: 2608–2617.
28. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
29. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsov-Velyaminov PV. New approach to Monte Carlo calculation of the free energy: method of expanded ensembles. J Chem Phys 1992;96: 1776–1783.
30. Marinari E, Parisi G. Simulated tempering: a new Monte Carlo scheme. Europhys Lett 1992;19:451–458.
31. Irbäck A, Potthast F. Studies of an off-lattice model for protein folding: sequence dependence and improved sampling at finite temperature. J Chem Phys 1995;103:10298–10305.
32. Favrin G, Irbäck A, Sjunnesson F. Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. J Chem Phys 2001;114:8154–8158.
33. Wallin S, Farwer J, Bastolla U. Testing distance measures for protein structures. In preparation.
34. Sayle R, Milner-White EJ. RasMol: biomolecular graphics for all. Trends Biochem Sci 1995;20:374–376.
35. Irbäck A, Sjunnesson F, Wallin S. Hydrogen bonds, hydrophobicity forces and the character of the collapse transition. J Biol Phys 2001;27:169–179.
36. Bastolla U, Vendruscolo M, Knapp E-W. A statistical mechanical method to optimize energy functions for protein folding. Proc Natl Acad Sci USA 2000;97:3977–3981.