Three-helix-bundle protein in a Ramachandran model

Anders Irbäck*, Fredrik Sjunnesson, and Stefan Wallin

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

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved September 21, 2000 (received for review May 26, 2000)

We study the thermodynamic behavior of a model protein with 54 amino acids that forms a three-helix bundle in its native state. The model contains three types of amino acids and five to six atoms per amino acid and has the Ramachandran torsional angles ϕ_i, ψ_i as its degrees of freedom. The force field is based on hydrogen bonds and effective hydrophobicity forces. For a suitable choice of the relative strength of these interactions, we find that the three-helix-bundle protein undergoes an abrupt folding transition from an expanded state to the native state. Also shown is that the corresponding one- and two-helix segments are less stable than the three-helix sequence.

It is not yet possible to simulate the formation of proteins' native structures on the computer in a controlled way. This goal has been achieved in the context of simple lattice and off-lattice models, where typically each amino acid is represented by a single interaction site corresponding to the C_α atom, and such studies have provided valuable insights into the physical principles of protein folding (1–5) and the statistical properties of functional protein sequences (6, 7). However, these models have their obvious limitations. Therefore, the search for computationally feasible models with a more realistic chain geometry remains a highly relevant task.

In this paper, we discuss a model based on the well-known fact that the main degrees of freedom of the protein backbone are the Ramachandran torsional angles ϕ_i, ψ_i (8). Each amino acid is represented by five or six atoms, which makes this model computationally slightly more demanding than C_α models. On the other hand, it also makes interactions such as hydrogen bonds easier to define. The formation of native structure is, in this model, driven by hydrogen-bond formation and effective hydrophobicity forces; hydrophobicity is widely held as the most important stability factor in proteins (9, 10), and hydrogen bonds are essential to properly model the formation of secondary structure.

In this model, we study in particular a three-helix-bundle protein with 54 amino acids, which represents a truncated and simplified version of the four-helix-bundle protein *de novo* designed by Regan and DeGrado (11). This example was chosen partly because there have been earlier studies of similar-sized helical proteins using models at comparable levels of resolution (12–18). The behavior of small fast-folding proteins is a current topic in both theoretical and experimental research, and a three-helix-bundle protein that has been extensively studied both experimentally (19, 20) and theoretically (14, 17, 21, 22) is fragment B of staphylococcal protein A.

In addition to the three-helix protein, to study size dependence, we also look at the behavior of the corresponding one- and two-helix segments. By using the method of simulated tempering (23–25), a careful study of the thermodynamic properties of these different chains is performed.

Not unexpectedly, it turns out that the behavior of the model strongly depends on the relative strength of the hydrogen-bond and hydrophobicity terms. In fact, the situation is somewhat reminiscent of what has been found for homopolymers with stiffness (26–29), with hydrogen bonds playing the role of the stiffness term. Throughout this paper, we focus on one specific empirical choice of these parameters.

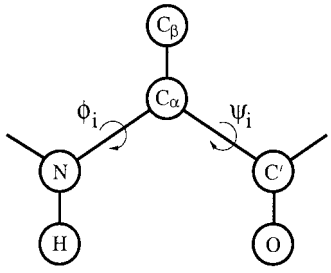

Fig 1. Schematic figure showing the representation of one amino acid.

For this choice of parameters, we find that the three-helix-bundle protein has the following three properties. First, it does form a stable three-helix bundle (except for a 2-fold topological degeneracy). Second, its folding transition is abrupt, from an expanded state to the native three-helix-bundle state. Third, compared to the one- and two-helix segments, it forms a more stable secondary structure. It should be stressed that these properties are found without resorting to the popular Gō approximation (30), in which interactions that do not favor the desired structure are ignored.

The Model

The model we study is a reduced off-lattice model. The chain representation is illustrated in Fig. 1. As mentioned in the introduction, each amino acid is represented by five or six atoms. The three backbone atoms N, C_α, and C′ are all included. Also included are the H and O atoms shown in Fig. 1, which we use to define hydrogen bonds. Finally, the side chain is represented by a single atom, C_β, which can be hydrophobic, polar, or absent. This gives us the following three types of amino acids: A with hydrophobic C_β, B with polar C_β, and G (glycine) without C_β.

The H, O, and C_β atoms are all attached to the backbone in a rigid way. Furthermore, in the backbone, all bond lengths, bond angles, and peptide torsional angles (180°) are held fixed. This leaves us with two degrees of freedom per amino acid, the Ramachandran torsional angles ϕ_i and ψ_i (see Fig. 1). The parameters held fixed can be found in Table 1.

Our energy function

$$E = E_{\rm loc} + E_{\rm sa} + E_{\rm hb} + E_{\rm AA} \qquad [1]$$

is composed of four terms. The local potential $E_{\rm loc}$ has a standard form with 3-fold symmetry,

$$E_{\rm loc} = \frac{\varepsilon_\phi}{2}\sum_i(1+\cos 3\phi_i) + \frac{\varepsilon_\psi}{2}\sum_i(1+\cos 3\psi_i). \qquad [2]$$

This paper was submitted directly (Track II) to the PNAS office.

*To whom reprint requests should be addressed. E-mail: irback@thep.lu.se.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.240245297. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.240245297

Table 1. Geometry parameters

Bond lengths, Å		Bond angles, °	
NC_α	1.46	$C'NC_\alpha$	121.7
$C_\alpha C'$	1.52	$NC_\alpha C'$	111.0
$C'N$	1.33	$C_\alpha C'N$	116.6
NH	1.03	$NC_\alpha C_\beta$	110.0
$C_\alpha C_\beta$	1.53	$C'C_\alpha C_\beta$	110.0
$C'O$	1.23		

The self-avoidance term E_{sa} is given by a hard-sphere potential of the form

$$E_{sa} = \varepsilon_{sa} \sum_{i<j}{}' \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}, \qquad [3]$$

where the sum runs over all possible atom pairs except those consisting of two hydrophobic C_β. The hydrogen-bond term E_{hb} is given by

$$E_{hb} = \varepsilon_{hb} \sum_{ij} u(r_{ij}) v(\alpha_{ij}, \beta_{ij}), \qquad [4]$$

where

$$u(r_{ij}) = 5\left(\frac{\sigma_{hb}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{hb}}{r_{ij}}\right)^{10} \qquad [5]$$

$$v(\alpha_{ij}, \beta_{ij}) = \begin{cases} \cos^2 \alpha_{ij} \cos^2 \beta_{ij} & \alpha_{ij}, \beta_{ij} > 90^\circ \\ 0 & \text{otherwise} \end{cases} \qquad [6]$$

In Eq. **4**, i and j represent H and O atoms, respectively, and r_{ij} denotes the HO distance, α_{ij} the NHO angle, and β_{ij} the HOC′ angle. Any HO pair can form a hydrogen bond. The last term in Eq. **1**, the hydrophobicity term E_{AA}, has the form

$$E_{AA} = \varepsilon_{AA} \sum_{i<j} \left[\left(\frac{\sigma_{AA}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{AA}}{r_{ij}}\right)^{6} \right], \qquad [7]$$

where both i and j represent hydrophobic C_β. To speed up the simulations, a cutoff radius r_c is used,† which is 4.5 Å for E_{sa} and E_{hb} and 8 Å for E_{AA}.

In this energy function, roughly speaking, the first two terms, E_{loc} and E_{sa}, enforce steric constraints, whereas the last two terms, E_{hb} and E_{AA}, are the ones responsible for stability. Force fields similar in spirit, emphasizing hydrogen bonding and hydrophobicity, have been used with some success to predict structures of peptides (31) and small helical proteins (15).

The parameters of our energy function were determined largely by trial and error. The final parameters are listed in Table 2. The parameters σ_{ij} of Eq. **3** are given by

$$\sigma_{ij} = \sigma_i + \sigma_j + \Delta\sigma_{ij},$$

Table 2. Parameters of the energy function

					σ_i, Å							
ε_ϕ	ε_ψ	ε_{sa}	ε_{hb}	ε_{AA}	N	C_α	C′	H	C_β	O	σ_{hb}, Å	σ_{AA}, Å
1	1	0.0034	2.8	2.2	1.65	1.85	1.85	1.0	2.5	1.65	2.0	5.0

Energies are in dimensionless units, in which the folding transition occurs at $kT \approx 0.65$ for the three-helix-bundle protein (see below).

†The cutoff procedure is $f(r) \mapsto \tilde{f}(r)$, where $\tilde{f}(r) = f(r) - f(r_c) - (r - r_c) f'(r_c)$, if $r < r_c$ and $\tilde{f}(r) = 0$ otherwise.

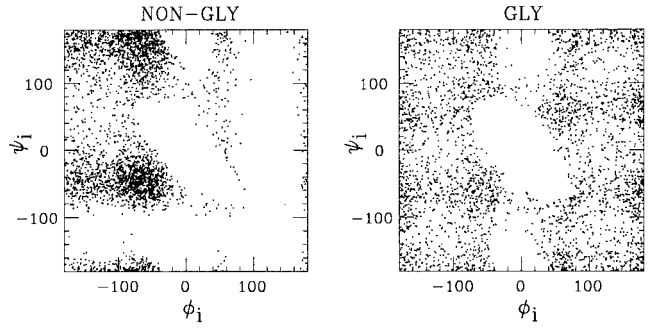

Fig 2. ϕ_i, ψ_i scatter plots for nonglycine and glycine, as obtained by simulations of the chains GXG for X = A/B and X = G, respectively, at $kT = 0.625$ (shown is ϕ_i, ψ_i for X).

where σ_i, σ_j can be found in Table 2, and $\Delta\sigma_{ij}$ is zero except for $C_\beta C'$, $C_\beta N$, and $C_\beta O$ pairs that are connected by three covalent bonds. In these three cases, we put $\Delta\sigma_{ij} = 0.625$ Å. This could equivalently be described as a change of the local ϕ_i and ψ_i potentials. In Fig. 2, we show ϕ_i, ψ_i scatter plots for nonglycine (A and B) and glycine for our final parameters, which are in good qualitative agreement with the ϕ_i, ψ_i distributions of real proteins (8, 32).

Finally, we determined the strengths of the hydrogen-bond and hydrophobicity terms on the basis of the resulting overall thermodynamic behavior of the three-helix sequence. For this purpose, we performed a set of trial runs for fixed values of the other parameters. An alternative would have been to use the method of Shea *et al.* (33). The result of our empirical determination of ε_{hb} and ε_{AA} does not seem unreasonable; at the folding temperature of the three-helix sequence (see below), we get $\varepsilon_{hb}/kT \approx 4.3$ and $\varepsilon_{AA}/kT \approx 3.4$.

In this model, we study the three sequences shown in Table 3, which contain 16, 35, and 54 amino acids, respectively. Following the strategy of Regan and DeGrado (11), the A and B amino acids are distributed along the sequence 1H in such a way that this segment can form a helix with all hydrophobic amino acids on the same side. The sequence 3H, consisting of three such stretches of As and Bs plus two GGG segments, is meant to form a three-helix bundle. This particular sequence was recently studied by Takada *et al.* (18), who used a more elaborate model with nonadditive forces.

Results

To study the thermodynamic behavior of the chains described in the previous section, we use the method of simulated tempering. This means that we first select a set of allowed temperatures and then perform simulations in which the temperature is a dynamical variable. This is done to speed up low-temperature simulations. In addition, it provides a convenient method for calculating free energies.

An example of a simulated-tempering run is given in Fig. 3, which shows the Monte Carlo evolution of the energy E and radius of gyration R_g (calculated over all backbone atoms) in a simulation of the three-helix sequence. Also shown (Fig. 3 *Bottom*) is how the system jumps between the different temperatures. Two distinct types of behavior can be seen. In one case, E is high, fluctuations in size are large, and the temper-

Table 3. The sequences studied

1H:	BBABBAABBABBAABB
2H:	1H-GGG-1H
3H:	1H-GGG-1H-GGG-1H

BIOPHYSICS

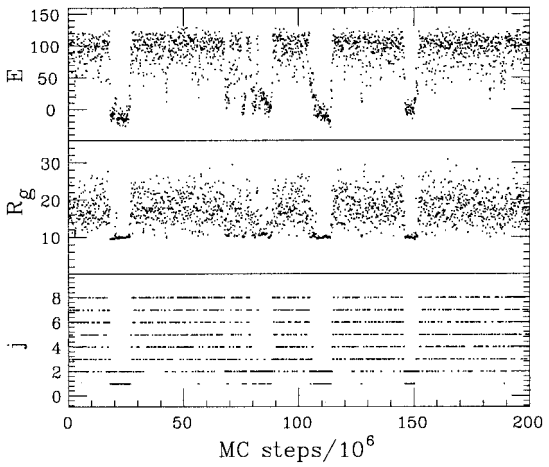

Fig 3. Monte Carlo evolution of the energy and radius of gyration in a typical simulation of the three-helix sequence. *Bottom* shows how the system jumps between the allowed temperatures T_j, which are given by $T_j = T_{\min}(T_{\max}/T_{\min})^{(j-1)/(J-1)}$ (34) with $kT_{\min} = 0.625$, $kT_{\max} = 0.9$, and $J = 8$. The temperature $T_{\min}$ is chosen to lie just below the collapse transition, whereas $T_{\max}$ is well into the coil phase (see Fig. 4).

atures visited are high. In the other case, E is low, the size is small and almost frozen, and the temperatures visited are low. Interesting to note is that there is one temperature, the next-lowest one, which is visited in both cases. Apparently, both types of behavior are possible at this temperature.

In Fig. 4*a*, we show the specific heat as a function of temperature for the one-, two-, and three-helix sequences. A pronounced peak can be seen that gets stronger with increasing chain length. In fact, the increase in height is not inconsistent with a linear dependence on chain length, which is what one would have expected if it had been a conventional first-order phase transition with a latent heat.

Our results for the radius of gyration (not displayed) show that the specific heat maximum can be viewed as the collapse temperature. The specific heat maximum is also where hydrogen-bond formation occurs, as can be seen from Fig. 4*b*. Important to note in this figure is that the decrease in hydrogen-bond energy *per amino acid* with decreasing temperature is most rapid for the three-helix sequence, which implies that, compared to the shorter ones, this sequence forms more stable secondary structure. The results for the chain entropy shown in Fig. 4*c* provide further support for this; the entropy loss per amino acid with decreasing temperature is largest for the three-helix sequence.

It should be stressed that the character of the collapse transition depends strongly on the relative strength of the hydrogen-bond and hydrophobicity terms. Fig. 4 shows that the transition is very abrupt or "first-order-like" for our choice $(\varepsilon_{hb}, \varepsilon_{AA}) = (2.8, 2.2)$. A fairly small decrease of $\varepsilon_{hb}/\varepsilon_{AA}$ is sufficient to get a very different behavior with, for example, a much weaker peak in the specific heat. In this case, the chain collapses to a molten globule without specific structure rather than to a three-helix bundle. A substantially weakened transition was observed for $\varepsilon_{hb} = \varepsilon_{AA} = 2.5$. If, on the other hand, $\varepsilon_{hb}/\varepsilon_{AA}$ is too large, then it is evident that the chain will form one long helix instead of a helical bundle.

We now turn to the three-dimensional structure of the three-helix sequence in the collapsed phase. It turns out that it does form a three-helix bundle. This bundle can have two distinct topologies: if we let the first two helices form a U, then the third helix can be either in front of or behind that U. The model is, not unexpectedly, unable to discriminate between these two possibilities. To characterize low-temperature conformations, we therefore determined two representative structures, one for each topology, which, following ref. 18, are referred to as FU and BU, respectively. These structures are shown in Fig. 5. They were generated by quenching a large number of low-T structures to zero temperature, and we feel convinced that they provide good approximations of the energy minima for the respective topologies. Given an arbitrary conformation, we then measure the root-mean-square distances δ_i (i = FU, BU) to these two structures (calculated over all backbone atoms). These distances are converted into similarity parameters Q_i by using

$$Q_i = \exp(-\delta_i^2/100\ \text{Å}^2). \qquad [8]$$

At temperatures above the specific heat maximum, both Q_i tend to be small. At temperatures below this point, the system is found to spend most of its time close to one or the other of the representative structures; either Q_{FU} or Q_{BU} is close to 1. Finally, at the peak, all three of these regions in the Q_{FU}, Q_{BU} plane are populated, as can be seen from Fig. 6*a*. In particular, this implies that the folding transition coincides with the specific heat maximum.

The folding transition can be described in terms of a single "order parameter" by taking $Q = \max(Q_{FU}, Q_{BU})$ as a measure of nativeness. Correspondingly, we put $\delta = \min(\delta_{FU}, \delta_{BU})$. In Fig. 6*b*, we show the free-energy profile $F(Q)$ at the folding temperature. The free energy has a relatively sharp minimum at $Q \approx 0.9$, corresponding to $\delta \approx 3$ Å. This is followed by a weak barrier around $Q = 0.7$, corresponding to $\delta \approx 6$ Å. Finally, there is a broad minimum at small Q, where $Q = 0.2$ corresponds to $\delta \approx 13$ Å.

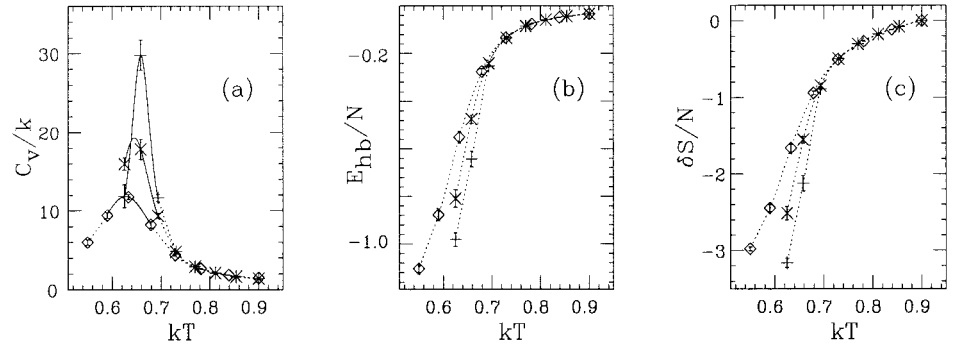

Fig 4. Thermodynamic functions against temperature for the sequences 1H (◇), 2H (×), and 3H (+) in Table 3. (*a*) Specific heat $C_v = (\langle E^2\rangle - \langle E\rangle^2)/NkT^2$, N being the number of amino acids. (*b*) Hydrogen-bond energy per amino acid, E_{hb}/N. (*c*) Chain entropy per amino acid, $\delta S/N = [S - S(kT = 0.9)]/N$. The full lines in *a* represent single-histogram extrapolations (35). Dotted lines are drawn to guide the eye.

Fig 5. Representative low-temperature structures, FU and BU, respectively. Drawn with RASMOL (36).

What does the nonnative population at the folding temperature correspond to in terms of R_g and E_{hb}? This can be seen from the Q, R_g, and Q, E_{hb} scatter plots in Fig. 7. These plots show that the low-Q minimum of $F(Q)$ corresponds to expanded structures with a varying but not high secondary-structure content. Although a detailed kinetic study is beyond the scope of this paper, we furthermore note that the free-energy surfaces corresponding to the distributions in Fig. 7 are relatively smooth. Consistent with that, we found that standard fixed-temperature Monte Carlo simulations were able to reach the native state, starting from random coils.

Let us finally mention that we also performed simulations of some random sequences with the same length and composition as the three-helix sequence. The random sequences did not form stable structures and collapsed more slowly with decreasing temperature than the designed three-helix sequence.

Summary and Outlook

We have studied a reduced protein model where the formation of native structure is driven by a competition between hydrogen bonds and effective hydrophobicity forces. Using this force field, we find that the three-helix-bundle protein studied has the following properties:

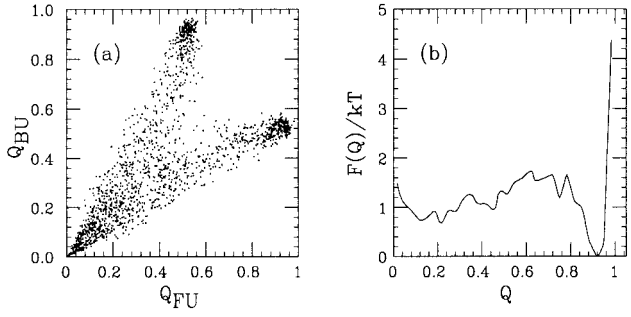

Fig 6. (*a*) Q_{FU}, Q_{BU} (see Eq. **8**) scatter plot at the specific heat maximum (kT = 0.658). (*b*) Free energy $F(Q)$ as a function of Q = max(Q_{FU}, Q_{BU}) at the same temperature.

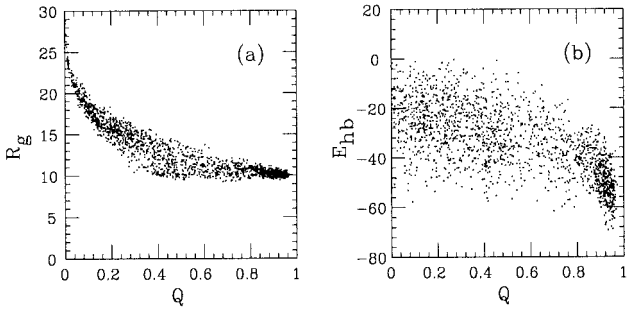

Fig 7. (*a*) Q, R_g, and (*b*) Q, E_{hb} scatter plots at the folding temperature (kT = 0.658).

- It does form a stable three-helix-bundle state, except for a 2-fold topological degeneracy.
- It undergoes an abrupt folding transition from an expanded state to the native state.
- It forms more stable secondary structure than the corresponding one- and two-helix segments.

An obvious question that remains to be addressed is what is needed to lift the topological degeneracy. Not obvious, however, is whether this question should be addressed at the present level of modeling, before including full side chains.

A first-order-like folding transition that takes the system directly from the unfolded state to the native one is what one expects for small fast-folding proteins. For the model to show this behavior, careful tuning of the relative strength of the hydrogen-bond and hydrophobicity terms, $\varepsilon_{hb}/\varepsilon_{AA}$, is required. This $\varepsilon_{hb}/\varepsilon_{AA}$ dependence may at first glance seem unwanted but is not physically unreasonable; ε_{hb} can be thought of partly as a stiffness parameter, and chain stiffness has important implications for the phase structure, as shown by recent work on homopolymers (26–29). Note also that incorporation of full side chains makes the chains intrinsically stiffer, which might lead to a weaker $\varepsilon_{hb}/\varepsilon_{AA}$ dependence.

Our three-helix sequence has been studied previously by Takada *et al.* (18), who used a more elaborate force field. It was suggested that it is essential to use context-dependent hydrogen bonds for the three-helix-bundle protein to make more stable secondary structure than its one-helix fragments. Our model shows this behavior, although its hydrogen bonds are context-independent.

Let us finally stress that we find a first-order-like folding transition without using the Gō approximation. Evidence for first-order-like folding transitions has been found for proteins with similar lengths in some C_α models (5, 14, 17, 33), but these studies use this approximation.

This work was in part supported by the Swedish Foundation for Strategic Research.

1. Săli, A., Shakhnovich, E. & Karplus, M. (1994) *J. Mol. Biol.* **235,** 1614–1636.
2. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins Struct. Funct. Genet.* **21,** 167–195.
3. Dill, K. A. & Chan, H. S. (1997) *Nat. Struct. Biol.* **4,** 10–19.
4. Klimov, D. K. & Thirumalai, D. (1998) *J. Chem. Phys.* **109,** 4119–4125.
5. Nymeyer, H., García, A. E. & Onuchic, J. N. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 5921–5928.
6. Pande, V. S., Grosberg, A. Y. & Tanaka, T. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12976–12979.
7. Irbäck, A., Peterson, C. & Potthast, F. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 9533–9538.
8. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23,** 283–437.
9. Dill, K. A. (1990) *Biochemistry* **29,** 7133–7155.
10. Privalov, P. L. (1992) in *Protein Folding*, ed. Creighton, T. E. (Freeman, New York), pp. 83–126.
11. Regan, L. & DeGrado, W. F. (1988) *Science* **241,** 976–978.
12. Rey, A. & Skolnick, J. (1993) *Proteins Struct. Funct. Genet.* **16,** 8–28.
13. Guo, Z. & Thirumalai, D. (1996) *J. Mol. Biol.* **263,** 323–343.
14. Zhou, Z. & Karplus, M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 14429–14432.
15. Koretke, K. K., Luthey-Schulten, Z. & Wolynes, P. G. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2932–2937.

16. Hardin, C., Luthey-Schulten, Z. & Wolynes, P. G. (1999) *Proteins Struct. Funct. Genet.* **34,** 281–294.
17. Shea, J.-E., Onuchic, J. N. & Brooks, C. L., III (1999) *Proc. Natl. Acad. Sci. USA* **96,** 12512–12517.
18. Takada, S., Luthey-Schulten, Z. & Wolynes, P. G. (1999) *J. Chem. Phys.* **110,** 11616–11629.
19. Bottomley, S. P., Popplewell, A. G., Scawen, M., Wan, T., Sutton, B. J. & Gore, M. G. (1994) *Protein Eng.* **7,** 1463–1470.
20. Bai, Y., Karimi, A., Dyson, H. J. & Wright, P. E. (1997) *Protein Sci.* **6,** 1449–1457.
21. Guo, Z., Brooks, C. L., III, & Boczko, E. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10161–10166.
22. Kolinski, A., Galazka, W. & Skolnick, J. (1998) *J. Chem. Phys.* **108,** 2608–2617.
23. Lyubartsev, A. P., Martsinovski, A. A., Shevkunov, S. V. & Vorontsov-Velyaminov, P. V. (1992) *J. Chem. Phys.* **96,** 1776–1783.
24. Marinari, E. & Parisi, G. (1992) *Europhys. Lett.* **19,** 451–458.
25. Irbäck, A. & Potthast, F. (1995) *J. Chem. Phys.* **103,** 10298–10305.
26. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7267–7271.
27. Doniach, S., Garel, T. & Orland, H. (1996) *J. Chem. Phys.* **105,** 1601–1608.
28. Bastolla, U. & Grassberger, P. (1997) *J. Stat. Phys.* **89,** 1061–1078.
29. Doye, J. P. K., Sear, R. P. & Frenkel, D. (1998) *J. Chem. Phys.* **108,** 2134–2142.
30. Gō, N. & Taketomi, H. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 559–563.
31. Ishikawa, K., Yue, K. & Dill, K. A. (1999) *Protein Sci.* **8,** 716–721.
32. Zimmerman, S. S., Pottle, M. S., Némethy, G. & Scheraga, H. A. (1977) *Macromolecules* **10,** 1–9.
33. Shea, J.-E., Nochomovitz, Y. D., Guo, Z. & Brooks, C. L., III (1998) *J. Chem. Phys.* **109,** 2895–2903.
34. Hansmann, U. H. E. & Okamoto, Y. (1997) *J. Comput. Chem.* **18,** 920–933.
35. Ferrenberg, A. M. & Swendsen, R. H. (1988) *Phys. Rev. Lett.* **61,** 2635–2638, and erratum (1989) **63,** 1658, and references given in the erratum.
36. Sayle, R. & Milner-White, E. J. (1995) *Trends Biochem. Sci.* **20,** 374–376.