Monte Carlo Studies on Equilibrium Globular Protein Folding. III. The Four Helix Bundle

ANDRZEJ SIKORSKI* and JEFFREY SKOLNICK,† *Institute of Macromolecular Chemistry, Department of Chemistry, Washington University, St. Louis, Missouri 63130*

Synopsis

The nature of the equilibrium conformational transition from the denatured state to a four-member α-helical bundle was studied employing a dynamic Monte Carlo algorithm in which the model protein chain was confined to a tetrahedral lattice. The model chain was allowed to hunt over all phase space, the target native state was not assumed a priori, and no site-specific interactions were introduced. The exterior vs the interior part of the protein is distinguished by the pattern of hydrophilic and hydrophobic interactions encoded into the primary sequence. The importance of a statistical preference for forming bends, as a function of bend location in the primary sequence, and helical wheel type cooperative interactions were examined, and the necessary conditions for collapse of the chain to the unique native structure were investigated. It was found that an amphipathic pattern of hydrophobic/hydrophilic interactions along with a statistical preference of the central residues for bend formation are sufficient to obtain the four-helix bundle. The transition to the native state has an all-or-none character.

INTRODUCTION

Proteins and polypeptides have long been the object of extensive study.[1-5] In spite of this, there are still a large number of unsolved problems concerning the nature of the transition from the random coil to the globular native conformation. The role of short- vs long-range interactions is not well understood, and the ability to predict the tertiary structure of a folded protein from its primary structure does not yet exist. Even in the case of small globular proteins, the causes of the all-or-none character of the transition from the denaturated to the native state are not fully elucidated. The solution to the above poses a severe theoretical challenge.

Computer simulations are a very useful tool for investigating the physical properties of proteins. Unlike in real systems, it is possible using Monte Carlo (MC) simulations to examine the contributions of the various kinds of interactions separately, i.e., one can do a computer experiment. Unfortunately, exact methods like molecular dynamics[6,7] or Brownian dynamics[8] are limited to studying local rearrangements because of the prohibitive amount of computer time required for the transition from a denatured to a native state. Monte Carlo simulations which employ realistic potentials, experience similar problems.[9-13] Because of the large separation in configurational space between the native and the denaturated state, and the large number of intervening local

*Permanent address: Department of Chemistry, University of Warsaw, 02-093 Warsaw, Poland.
†To whom correspondence should be addressed.

Biopolymers, Vol. 28, 1097–1113 (1989)

CCC 0006-3525/89/061097-17$04.00

minima on the free energy surface, it is still impossible, employing realistic potential energy surfaces, to reach a native state without the a priori introduction of some features of the desired native state. Thus, simplifications in the model are required. One possible approach is to use a lattice model of the protein to reduce the number of degrees of freedom and to assume a simplified model for the interactions, thereby leading to a finite (and reasonable) amount of computer time required for the simulation.

Recently a very simple model was proposed for the qualitative description of the folding of β-barrel globular proteins.[14,15] The model protein chain is confined to a tetrahedral lattice and an α-carbon representation of the protein is used. A very efficient dynamic MC algorithm that allows the system under consideration to hunt over the entire configurational space is employed. There is no assumed pathway to the folded native state; all configurations and all energetic contacts are allowed. This model was applied to β-proteins. The necessary conditions for the formation of a unique four-member β-barrel are found to be the general pattern of hydrophobic and hydrophilic residues encoded in the primary sequence and the presence in the primary sequence of regions that have a statistical preference to form bends. The analysis of this model also suggested a mechanism of protein evolution via random mutation. Similar considerations allow for a unique all-or-none transition to the Greek key topology quite close to that found in plastocyanin.[5,17]

In the present paper, a very similar model is applied to study the conditions required for the folding of model α-helical globular proteins. Because of lattice constraints, interacting α-helical stretches on a diamond lattice can only be parallel or antiparallel, whereas in real proteins there are a variety of mutual orientations of the helices, especially for short stretches. The desired native structure under consideration consists of a left-handed four-helix bundle having three short bends and without long loops.[5] This model can be viewed as a crude approximation of cytochrome c' [5,18] or myohemerythrin,[5,19] and is in fact rather similar to the synthetic four-helix bundle synthesized by DeGrado and co-workers.[20] In order to find the necessary conditions to fold the four-helix bundle, models having different primary sequences and with different kinds of allowed long-range interactions were studied.

DESCRIPTION OF THE MODEL

The macromolecule under consideration consists of $n = 48$ beads ($n - 1$ "bonds") confined to the tetrahedral lattice. Each bead is an α-carbon representation of a particular amino acid residue. Every segment of the chain has a bond length equal to $3^{1/2}$ and can adopt one of the orientations described by the bond vectors $[\pm 1, \pm 1, \pm 1]$, consistent with the tetrahedral valence angle of $\theta = 109°$. Such a chain has $n - 3$ internal rotational degrees of freedom. Every sequence of three bonds can exist in one of three distinguishable rotational conformational states: the planar *trans* (t), and the out-of-plane *gauche* minus (g^-) and *gauche* plus (g^+) conformations. Excluded volume is implemented by forbidding the multiple occupancy of the lattice sites. Because the object of this work is to investigate models of α-helical globular proteins, various possible intramolecular interactions were introduced into the model. Short-range interactions (when employed) are implemented as the

preference for the g^- state. Thus, even in the absence of long-range interchain contacts, an intrinsic preference to form right-handed α-helical turns can be accommodated. The statistical weights of the three allowed conformations are unity for a g^- state, and for g^+ or t, the statistical weight is $\omega = \exp\{-\epsilon_g/k_B T\}$, where $\epsilon_g > 0$, k_B is Boltzmann's constant, and T is the absolute temperature. The parameter $T^* = k_B T/\epsilon_g$ defines the reduced temperature scale.

In order to include the possibility of helical wheel effects[21] resulting from interactions between residues i and $i + 4$ in a helical turn, cooperative interactions denoted here as type I were included in some of the models. If a consecutive sequence of two rotational states are all in the g^- conformation, they form one α-helical turn. Thus, if residues i and $i + 4$ are a distance 4 apart, they interact with an attractive energy ϵ_c' [see Fig. 1(A)].

The question next arises as to how these parameters are related to traditional helix–coil transition theory.[22]

Strictly speaking, to obtain the Zimm–Bragg[22] helix initiation parameter (σ) and propagation parameter s, one should simulate the α-helix–coil transition without long-range interactions. From the helix content vs temperature profiles, σ and s can be extracted. Such a fitting procedure would be necessary because the dynamic MC method does not provide a direct means of obtaining the entropic contribution to these quantities. A very crude estimate of σ and s for the two models may be obtained by asserting that a helical conformation must consist of at least two consecutive g^- conformations, i.e., one helical turn. For the first model, the statistical weight of a *chhc* conformation (with c denoting a random coil and h a helical state), relative to the random coil state is

$$\sigma s^2 = \frac{1}{4\omega(1 + \omega)} \tag{1}$$

and the statistical weight of a *chhhc* sequence, relative to the random coil state is

$$\sigma s^3 = (2\omega + 12\omega^2 + 8\omega^3)^{-1} \tag{2}$$

Taking the ratio of Eq. (2) to Eq. (1) gives

$$s = (2 + 2\omega)/(1 + 6\omega + 4\omega^2) \tag{3}$$

and

$$\sigma = \frac{(1 + 6\omega + 4\omega^2)^2}{16\omega(1 + \omega)^3} \tag{4}$$

At $T^* = 1$ ($\omega = e^{-1}$), $\sigma = 0.932$. Observe, however, that unlike the traditional Zimm–Bragg model, estimates of σ and s are in fact sequence-length dependent.

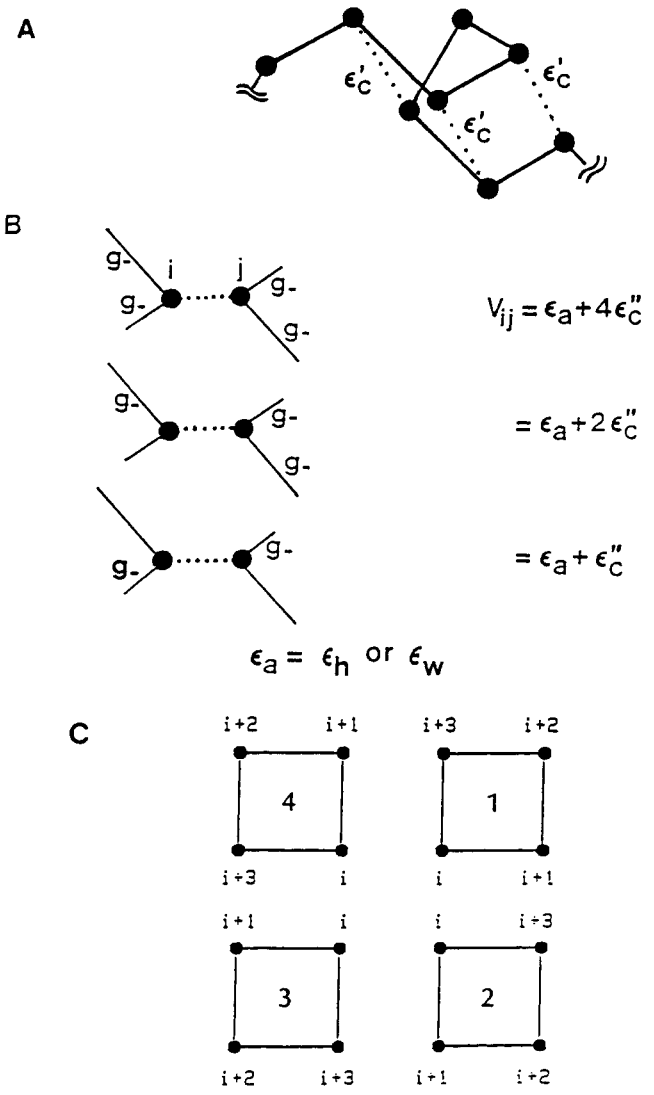

Fig. 1. Schematic representation of the possible kinds of interactions introduced into the model.

The second model employing helical wheel type cooperative interactions gives, for the statistical weight of a *chhc* sequence,

$$\sigma s^2 = \frac{1}{8} e^{-\epsilon'_c/k_B T} \tag{5}$$

and for the *chhhc* sequence relative to the *cccc* sequence,

$$\sigma s^3 = \frac{e^{-2\epsilon'_c/k_B T}}{22} \tag{6}$$

Hence,

$$s = \frac{4}{11} e^{-\epsilon_c'/k_B T} \tag{7}$$

and the Zimm–Bragg σ parameter is

$$\sigma = \frac{121}{128} e^{\epsilon_c'/k_B T} \tag{8}$$

At $|\epsilon_c/k_B T| = 1$, $\sigma = 0.35$.

Thus, in both models, if tertiary interactions are not introduced, a broad population having small lengths of fluctuating α-helices would be expected. This particular choice of parameters is employed to demonstrate, as shown below, that even in the very worst case of systems having rather small cooperativity based on local interactions, due to the fact that tertiary interactions stabilize the native conformation, an all-or-none type of transition emerges.

There is another cooperative effect between nonbonded nearest neighbor conformational states that may stabilize the tertiary structure of a globular protein,[15] shown in Fig. 1(B). This effect can be modeled by cooperative type II interactions, which occur when any nonbonded nearest neighbor pair of residues are in the g^- state giving the contribution to the energy of ϵ_c'', $2\epsilon_c''$, or $4\epsilon_c''$, depending on the mutual orientations of the parts of the chain in contact.

Attractive interactions between nearest neighbor pairs that mimic hydrophobic interactions between side groups and salt bridges[2] are included as a potential of mean force ϵ_h. Nearest neighbor contacts between a pair of hydrophilic residues or a hydrophilic–hydrophobic pair result in a repulsive potential of mean force ϵ_w.[16] The amphipathic pattern in the primary sequence is introduced as follows: Every α-helical stretch that might result in an α-helix in the native structure is represented as $\mathbf{H}_i(k)$ in the primary sequence —i denotes the number of the stretch (numbering down the chain) which consists of k residues. Consider now the ith helical turn, which contains 4 residues: i, $i + 1$, $i + 2$, and $i + 3$. Figure 1(C) presents the four-member α-helix bundle native structure viewed from the top of the molecule where every square corresponds to an α-helical turn. Residues located in the inner part of native structure (i, $i + 1$, and $i + 3$ from every sequence) should form the hydrophobic core. The following set of hydrophobic contact pairs were assumed between stretches 1–2, 2–3, 3–4 and 1–4: (i, i), $(i, i + 3)$, $(i + 1, i)$, and $(i + 2, i + 3)$. Observe that all residues in each of the four stretches will interact with the same pattern as those of the fully in-register molecule, e.g., residue i of sequence 1 interacts with residue i' of sequence 2, etc. All the remaining contacts are assumed to be hydrophilic. This set of interactions can of course occur not only in the native α-helical state but also for every configuration of the chain. In some of the later simulations, all interior hydrophobic residues i are allowed to have a favorable interaction free energy. Results identical to those with the somewhat more restrictive set of interactions described above are obtained.

In the expected native structure, there are three regions along the chain where bends might be formed. Each bend consists of three conformational states (two bonds). The preference for the formation of a bend is introduced by modifying the local intrinsic preference for the rotational states, i.e., ϵ_g of the three bend-type residues might be either zero or negative. In all cases, $\epsilon_h = \epsilon_c'' = \epsilon_c' = \epsilon_w = 0$ in the putative bend region. The turn regions in the native four-helix bundle involve residues 11–13, 23–25, and 35–37. These region are denoted by $\mathbf{b}_i$ in the primary sequence.

The MC sampling procedure is based on an asymmetric Metropolis sampling scheme.[23-25] Five kinds of micromodifications of the chain configuration were introduced to sample configuration space; these exhibit good sampling efficiency for both the native and denaturated states. A single MC cycle is composed of a number of sequences of the following local moves, each attempted on a randomly chosen piece of the chain.[16,23-25]

(i) Three-bond kink motion that changes the conformation g^+ (g^-) into the conformation g^- (g^+); see Fig. 2(A).[23,24]

(ii) Four-bond kink motion that changes the sequence $g^{\pm}g^{\mp}$ into $g^{\mp}g^{\pm}$. This kind of modification creates new conformations in the interior of the chain; see Fig. 2(B).[23]

(iii) Modification of the chain ends. At high temperatures, the two end bonds are randomly rotated, as in Fig. 2(C). At lower temperatures, as shown in Fig. 2(D), the entire end part (end stretch) of the molecule is randomly chosen, clipped off, and rebuilt at random into a new conformation. The maximum length of such a modification is not longer than the length of the conjectured α-helical stretch in the native state. The latter kind of modification helps the system surmount deep local free energy minima.

(iv) Four-bond wave motion in which a sequence of four bonds containing the conformation g^+g^- or g^-g^+ is interchanged with two consecutive bonds elsewhere in the molecule; see Fig. 2(E).

(v) Five-bond wave motion in which a conformation having five of six bonds forming a closed cyclohexane-like ring is interchanged with one bond in another part of the chain; see Fig. 2(F).

After moves iv and v, the beads are renumbered to preserve the primary sequence. Modifications iv and v shift local conformations along the chain and are very important in temperature regions where ordered structures (α-helices) appear. They can melt partially ordered (nonnative) structures, thereby avoiding the locking of the system in a deep local free energy minimum.

After a small number of micromodifications, a new configuration is accepted according to the standard Metropolis criterion.[24] To achieve thermodynamic equilibrium, the number of MC steps was in range of 10^7 cycles, with the particular number of iterations employed depending strongly on temperature. That is, the transition region is sampled for longer times than those temperatures where only denatured or folded conformations dominate. To ensure that the algorithm is ergodic and that the important regions of configuration space are sampled, simulations were performed for various high-temperature initial configurations followed by cooling–heating sequences. The properties of both the native and denaturated states are well characterized. Difficulties arise in the vicinity of native $\rightleftarrows$ denatured transition where both native (N) and

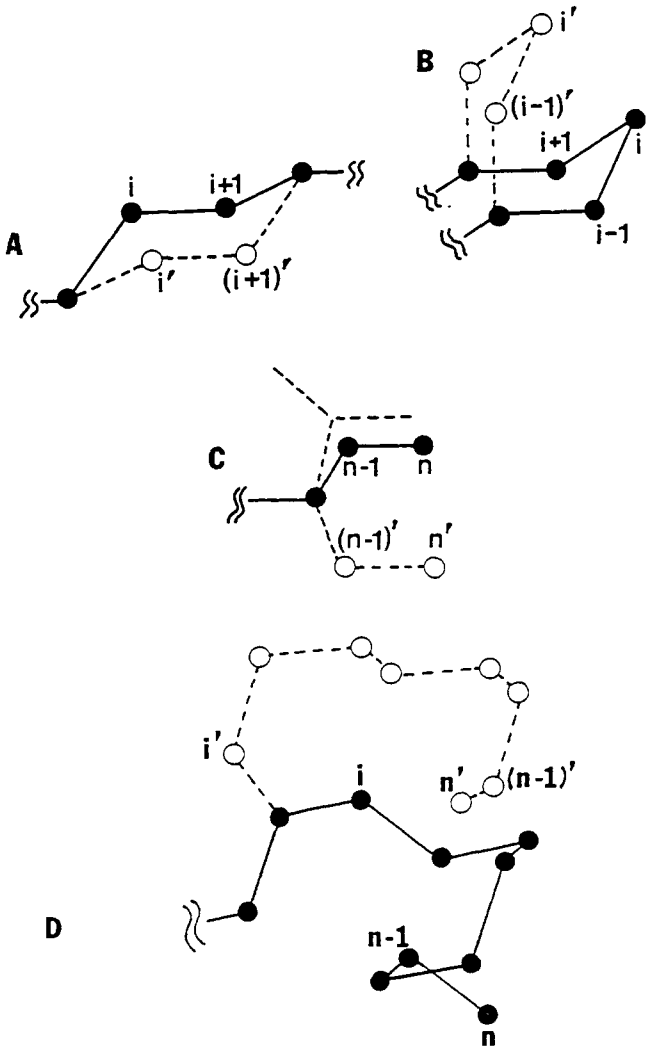

Fig. 2. Set of MC micromodifications used in the sampling algorithm.

denatured (D) states are populated, and only a small number of jumps between these states are observed because of the finite simulation time, i.e., the equilibrium constant between D and N is not well characterized in the transition region.

RESULTS

Six different models of α-helical globular proteins were considered to examine the requirements for formation from the denatured state of a unique four-member α-helical bundle. Representative conformations of the denatured state and the native state are shown in Fig. 3, structures 1 and 2 respectively. A very useful quantity to characterize the protein dimensions is the mean

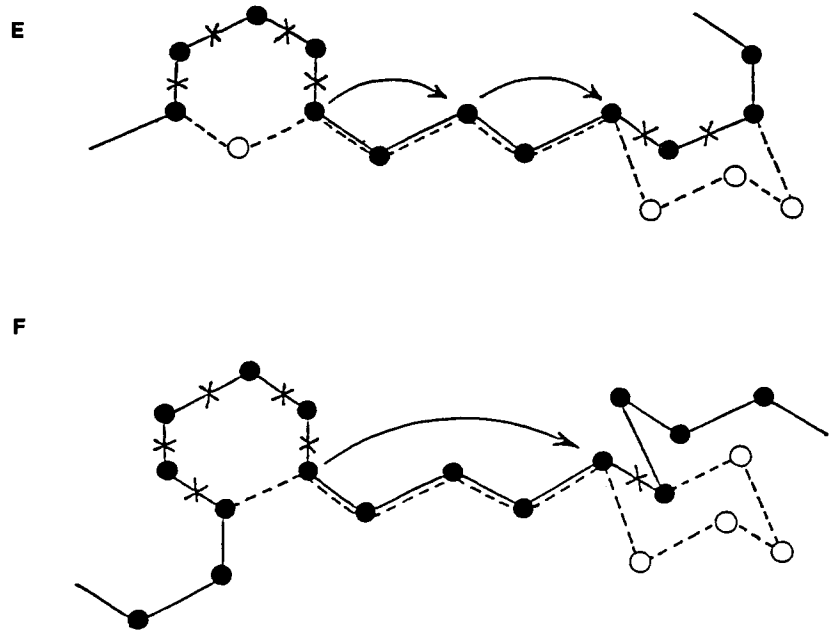

Fig. 2. (*Continued from the previous page.*)

square radius of gyration defined by

$$\langle S^2 \rangle = \frac{1}{M} \sum_{k=1}^{M} \left(\frac{1}{n^2} \sum\sum_{j>i=1}^{n} r_{ij}^2(k) \right) \tag{9}$$

M is number of MC steps and $r_{ij}(k)$ is the distance between the ith and jth residues in a given configuration of the kth MC cycle. Another useful parameter describing the model chain under consideration is the normalized fraction of α-helix (helix content) θ_h:

$$\theta_h(T^*) = \frac{f_h(T^*) - f_h^{\text{coil}}}{f_h^{\text{nat}} - f_h^{\text{coil}}} \tag{10}$$

where $f_h(T^*)$, f_h^{coil}, and f_h^{nat} are fractions of g^- states at T^*, in the random coil conformation and in the native structure, respectively. f_h^{coil} is taken to be 1/3 for convenience (actually it is slightly larger); $f_h^{\text{nat}} = 0.8666$. Table I summarizes the various models developed below.

Model A has primary sequence pattern $\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{H}_3(12)\mathbf{H}_4(12)$ with cooperative type II interactions given by $\epsilon_c'' = -\epsilon_g/2$ and with hydrophobic/hydrophilic interactions $\epsilon_h = -\epsilon_g/4$ and $\epsilon_w = 2\epsilon_g$. There is no intrinsic preference for bend formation. All the residues interact with ϵ_g, ϵ_h, and ϵ_w appropriate to the particular pair, and every bead has the same intrinsic preference for helical states (i.e., a nonzero ϵ_g).

In Fig. 4, the curve denoted by the open squares presents a plot of $\langle S^2 \rangle$ vs T^*. The analogous change in helix content, θ_h, with temperature T^* is given

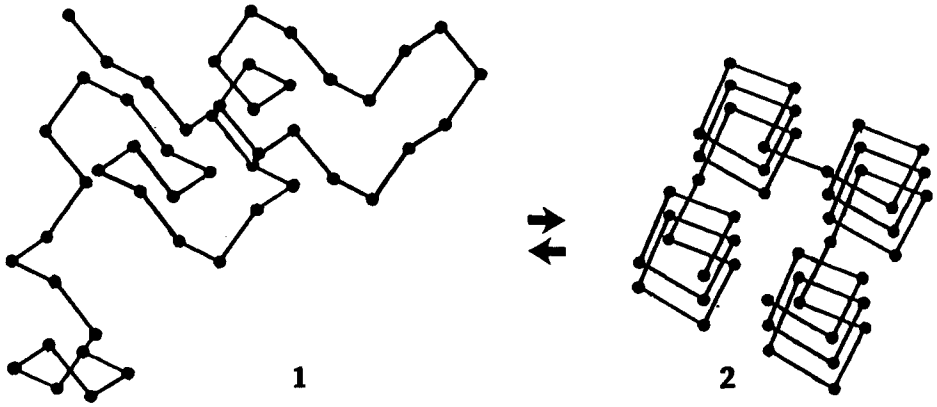

Fig. 3. Representative configurations of the model chain in a denaturated conformation (1) and in the native four-member α-helical bundle (2) states.

in Fig. 5. At low temperature, the fraction of α-helical states is close to that expected for the native structure. However, while stable ordered structures, many times resembling a pair of joined α-helical hairpins in an "S" shape as shown in Fig. 6, were obtained, the native four-member α-helical bundle 2 of Fig. 3 was never observed. At intermediate temperatures, we obtain a hairpin consisting of two α-helical stretches (stretches 1 and 2 with residues 11–13 forming a tight bend or stretches 3 and 4 with a bend located at residues 35–37) plus a long coiled tail. This is only half of the desired native structure. This is suggestive of the importance of the middle bend (located at residues 23–25) in forming the native globular protein. Thus, the hydrophobic/hydrophilic pattern of interactions (ϵ_h and ϵ_w), along with the local intrinsic preference for α-helices (ϵ_g) and cooperative type II intractions (ϵ_c''), seem to be insufficient to produce a conformational transition to the desired native state. Perhaps this is partially due to the inability of the algorithm to effectively move around large pieces of α-helical structure; i.e., the failure to obtain the four-helix bundle may be a kinetic effect.

Model B differs from the previous one by the introduction of a statistical preference of residues 23–25 for bend formation and has the primary sequence pattern $\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{b}_2\mathbf{H}_3(12)\mathbf{H}_4(12)$. The *trans* conformation of residue 23 has an intrinsic energetic preference of $-2\ \epsilon_g$. Any *gauche* state of residue 24 has an intrinsic energetic preference of $-\epsilon_g$, and any *gauche* state of residue 25 has an intrinsic energetic preference of $-2\ \epsilon_g$. The two lowest energy conformations, the native bend conformation tg^-g^+ and the nonnative bend conformation tg^+g^- are equally preferred, and the system is free to adopt all of the 27 possible rotational states. In curve B (solid squares) of Fig. 4, $\langle S^2 \rangle$ vs T^* obtained as an average over four cooling sequences is presented. For this case, the unique native four-helix bundle 2 of Fig. 3 was always obtained at lower temperatures. This is confirmed by the low-temperature value of θ_h, curve B in Fig. 5. The exterior turns involving residues 11–13 and 35–37 are induced. Because of the arrangement of nearest neighbor pairs per unit of secondary structure in the α-helix as compared to the four-member β-barrel, in the helix bundle the presence of the central statistical turn region alone is

TABLE I
Compilation of Properties of Models

Model	Primary Sequence[a]	ϵ_g[b]	ϵ'_c[c]	ϵ''_c[d]	ϵ_h[e]	ϵ_w[f]
A	$\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{H}_3(12)\mathbf{H}_4(12)$	>0	0	$-\epsilon_g/2$	$-\epsilon_g/4$	$2\epsilon_g$
B	$\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{b}_2\mathbf{H}_3(12)\mathbf{H}_4(12)$	>0	0	$-\epsilon_g/2$	$-\epsilon_g/4$	$2\epsilon_g$
C	$\mathbf{H}_1(12)\mathbf{b}_1\mathbf{H}_2(12)\mathbf{b}_2\mathbf{H}_3(12)\mathbf{b}_3\mathbf{H}_4(12)$	>0	0	$-\epsilon_g/2$	$-\epsilon_g/4$	$2\epsilon_g$
D	$\mathbf{H}_1(12)\mathbf{b}_1\mathbf{H}_2(12)\mathbf{b}_2\mathbf{H}_3(12)\mathbf{b}_3\mathbf{H}_4(12)$	>0	0	$-\epsilon_g/2$	$-\epsilon_g/4$	$2\epsilon_g$
E	$\mathbf{H}_1(12)\mathbf{b}_1^0\mathbf{H}_2(12)\mathbf{b}_2^0\mathbf{H}_3(12)\mathbf{b}_3^0\mathbf{H}_4(12)$	0	$1/2\ \epsilon_h$	0	<0	$-2\epsilon_h$
F	$\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{b}_2^0\mathbf{H}_3(12)\mathbf{H}_4(12)$	0	$1/2\ \epsilon_h$	0	<0	$-2\epsilon_h$

[a] $H(k)$ indicates an amphipathic amino acid pattern consisting of k consecutive residues in sequence i. $\mathbf{b}_i$ ($i = 1$–3 corresponds to residues 11–13 ($i = 1$), 23–25 ($i = 2$), and 35–37, respectively) indicates that the native conformation of the ith putative turn is among the energetically favored states. If $\mathbf{b}_i$ is not explicitly displayed there is no intrinsic preference for bend formation. $\mathbf{b}_i^0$ indicates that all the conformations of putative turn i based on short-range interactions are isoenergetic.

[b] Intrinsic energy (>0) of a g^+ or t state relative to a g^- state.

[c] Helical wheel type of interaction, which reflects enhanced stability of a g^-g^- sequence.

[d] Cooperative interaction arising when any two nearest neighbor nonbonded residues are in a g^- conformation.

[e] Nonbonded nearest neighbor interaction free energy of a hydrophobic/hydrophobic pair of residues.

[f] Nonbonded nearest neighbor interaction free energy of a hydrophilic/hydrophobic or hydrophilic/hydrophilic pair of residues.

sufficient to produce a unique, in-register native state, whereas in the β-barrel it is not.[16]

For the assumed set of interactions, the native protein is very stable at low temperatures, with very little end fluctuations. The fraction of helicity (at $T^* = 1.191$), before the denatured → native transition, θ_h equals 0.384, i.e., the model chain possesses some secondary structure before renaturation occurs.

Model C possesses the primary sequence pattern $\mathbf{H}_1(12)\mathbf{b}_1\mathbf{H}_2(12)\mathbf{b}_2\mathbf{H}_3(12)\mathbf{b}_3\mathbf{H}_4(12)$, and has a statistical preference for the three bends at residues 11–13, 23–25, and 35–37 in the same sense as above. The averages over four cooling sequences of $\langle S^2 \rangle$ vs T^* and θ_h vs T^*, curve C (open circles) are given in Figs. 4 and 5, respectively. The unique native state 2 of Fig. 3 was obtained in all simulations; however, the transition occurs at slightly higher temperatures than in model B, where only a central turn is preferred. This is not surprising because the introduction of an additional statistical preference for the other bends involving residues 11–13 and 35–37 stabilizes the native state. The preference of two additional bends does not change the magnitude or location of the conformational fluctuations, which involve the tails. No nonnative structures at low temperature are observed. All configurational properties of the native state such as $\langle S^2 \rangle$ and nearest neighbor contacts are the same as in model B. The fraction of helicity of the denatured state near the transition (at $T^* = 1.282$) is $\theta_h = 0.298$. Again, we point out that the introduction of a statistical preference for the two bends in the primary sequence at residues 11–13 and 35–37 is not necessary to obtain the native state; only the existence

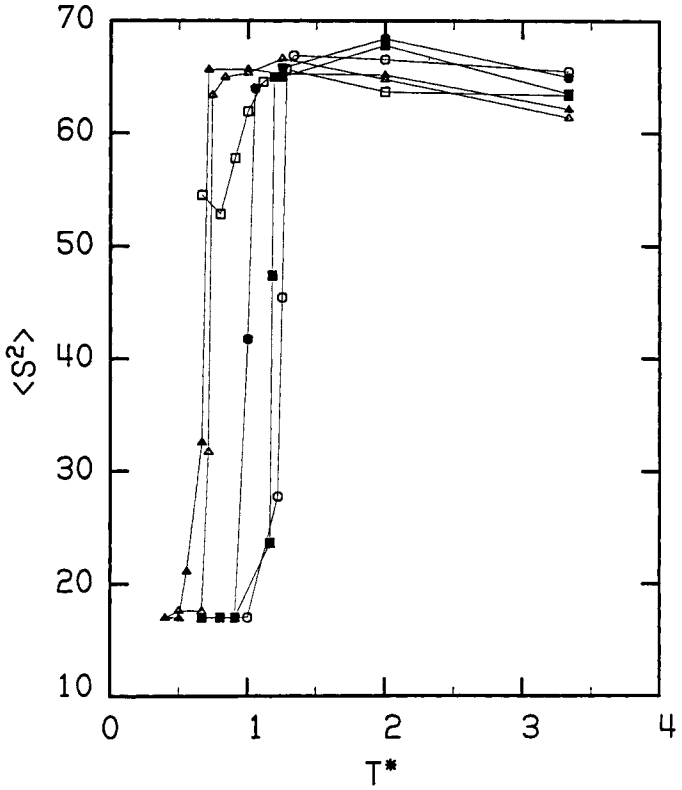

Fig. 4. Plot of the mean square radius of gyration $\langle S^2 \rangle$ vs reduced temperature T^* for renaturation (cooling) sequences (see text): model A (open squares), model B (solid squares), model C (open circles), model D (solid circles), model E (open triangles), and model F (solid triangles).

of the central region that has a statistical preference to adopt a bend conformation is crucial for the collapse to the unique native state.

Comparing the cooling and heating sequences of the simulation, one finds that there is a difference in the location of renaturation and denaturation transitions. In Fig. 7, $\langle S^2 \rangle$ is plotted vs T^* for cooling (open circles) and heating (solid circles) sequences. The heating and cooling curves should be identical in the limit of an infinite number of MC steps. However, because of the finite length of the simulations and due to the fact that the native conformation is in a very deep free energy well, the transition occurs at higher temperatures for denaturation (heating) than on renaturation (cooling).

In model D, cooperative type II interactions were eliminated, i.e., $\epsilon_c'' = 0$, in order to check if their existence is essential to reproduce the main features of the α-helical protein folding. The pattern of hydrophobic/hydrophilic interactions, the statistical preference of the three bends, and ϵ_g were the same as in model C, that is, model D has primary sequence $\mathbf{H}_1(12)\mathbf{b}_1\mathbf{H}_2(12)\mathbf{b}_2\mathbf{H}_3(12)\mathbf{b}_3\mathbf{H}_4(12)$. The results for $\langle S^2 \rangle$ vs T^* are shown in Fig. 4, curve D (solid circles), and for θ_h vs T^* in curve D (solid circles) of Fig. 5. The unique four-member α-helical native structure was obtained in three cooling

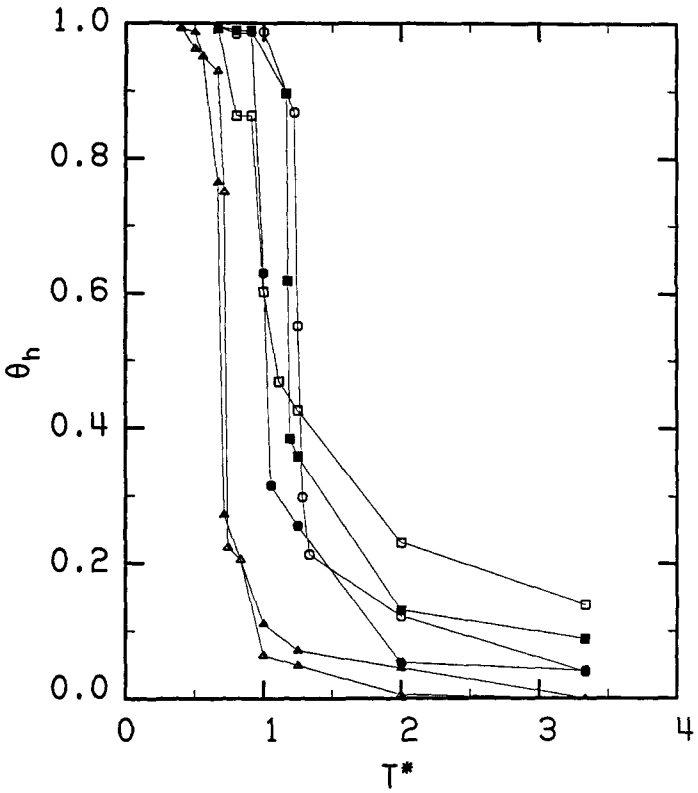

Fig. 5. Plot of normalized helix content, θ_h, vs reduced temperature T^* for renaturation (cooling) sequences. See text for more details. Model A (open squares), model B (solid squares), model C (open circles), model D (solid circles), model E (open triangles), and model F (solid triangles).

sequences. This demonstrates that cooperative type II interactions between residues are not necessary to fold the tertiary structure. Similar conclusions were reached in studies on a β-sheet globular protein model.[16] The interactions between α-helical stretches are considerably weaker in model D than in models B and C, and thus the examination of the nature of the native $\rightleftarrows$ denatured transition is easier (see below). The amount of secondary structure in the denatured state (at the temperature $T^* = 1.053$) in this model is

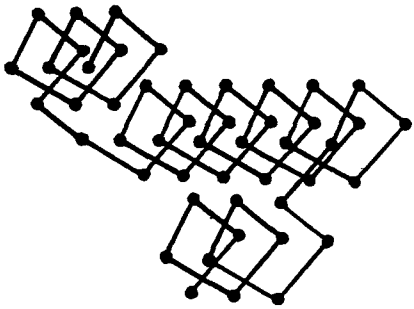

Fig. 6. Representative conformation of model A obtained under strongly renaturing conditions.

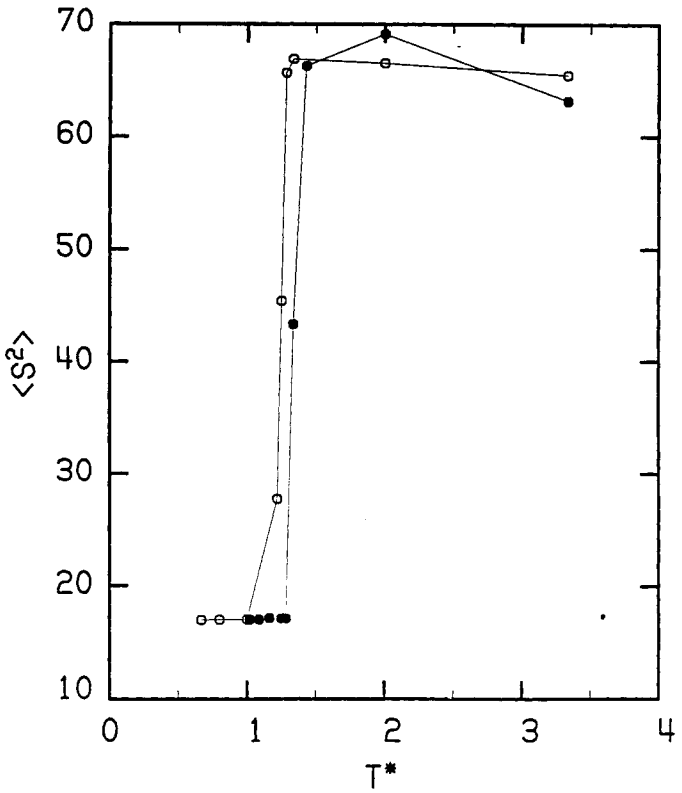

Fig. 7. Plot of the mean square radius of gyration $\langle S^2 \rangle$ vs reduced temperature for model C: A (open circles) is a cooling sequence and B (solid circles) is a heating sequence.

$\theta_h = 0.314$. Weaker tertiary interactions between stretches shift the renaturation transition to lower temperatures.

In model E, in addition to the previous pattern of hydrophobic/hydrophilic interactions and the statistical preference for tight turn regions (involving residues 11–13, 23–25, and 35–37), cooperative type I interactions were introduced and type II interactions were not included. In this model, no particular conformation of the bends is preferred (such a bend region will be called "neutral"). The bend region has $\epsilon_c' = \epsilon_h = \epsilon_w = \epsilon_g = 0$. Because type I cooperative interactions favor the formation of α-helical turns, the intrinsic local rotational potential parameter ϵ_g is not required. For all residues not involved in forming bends, the interactions are assumed to be $\epsilon_c' = 1/2\ \epsilon_h = -1/4\ \epsilon_w$. In this model, the parameter ϵ_h forms the basis of the reduced temperature scale ($T^* = kT/\epsilon_h$). The primary sequence of model E is $\mathbf{H}_1(12)\mathbf{b}_1^0\mathbf{H}_2(12)\mathbf{b}_2^0\mathbf{H}_3(12)\mathbf{b}_3^0\mathbf{H}_4(12)$, where a superscript zero indicates the location of a neutral bend. Thus, results in Fig. 4, curve E (open triangles), for $\langle S^2 \rangle$ vs T^* and in Fig. 5, curve E, θ_h vs T^* (open triangles) cannot be directly compared with all the previous models. Collapse to the unique native state 2 of Fig. 3 was observed for all three cooling sequences performed. The amount of secondary structure in the denatured state at $T^* = 0.741$ is considerably lower than in all previous models having a value $\theta_h = 0.223$.

Model F differs from model E only in the bend preference. Only a central bend neutral region located at residues 23–25 is present in the primary sequence $\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{b}_2^0\mathbf{H}_3(12)\mathbf{H}_4(12)$. As in model E, based on local short-range interactions, there is no intrinsic preference for bend formation. That is, all conformations of residues 23–25 are equally probable, and there are no long-range interactions ($\epsilon_g = \epsilon_c' = \epsilon_h = \epsilon_w = 0$). Each time in a series of three cooling sequences, the unique four-helix bundle **2** of Fig. 3 was obtained. In Fig. 4, curve F (solid triangles), the dependence of $\langle S^2 \rangle$ on reduced temperature is plotted. Compared to model E, the renaturation occurs at a lower temperature. The fraction of helicity, $\theta_h = 0.271$ at $T^* = 0.714$, is higher than in the previous case. This kind of behavior of model system was expected. The existence of a weak preference for central bend formation seems to be the sufficient condition to fold the four-helix bundle tertiary structure.

Having established the sufficient conditions for the formation of a four-helix bundle, we next examine the character of the conformational transition. In Fig. 8(A–C), a flow chart of the number of pairs of native contacts, ν_{nat}, vs "time" is plotted for model D. Results at three distinct temperatures are presented. Each unit time equals 25,000, 150,000, and 50,000 micromodifications for Fig. 8(A–C), respectively. The number of pairs of native contacts is a good unique measure of whether or not the system is in the native state. At $T^* = 1.25$, the chain exists in a purely random coil state. The number of

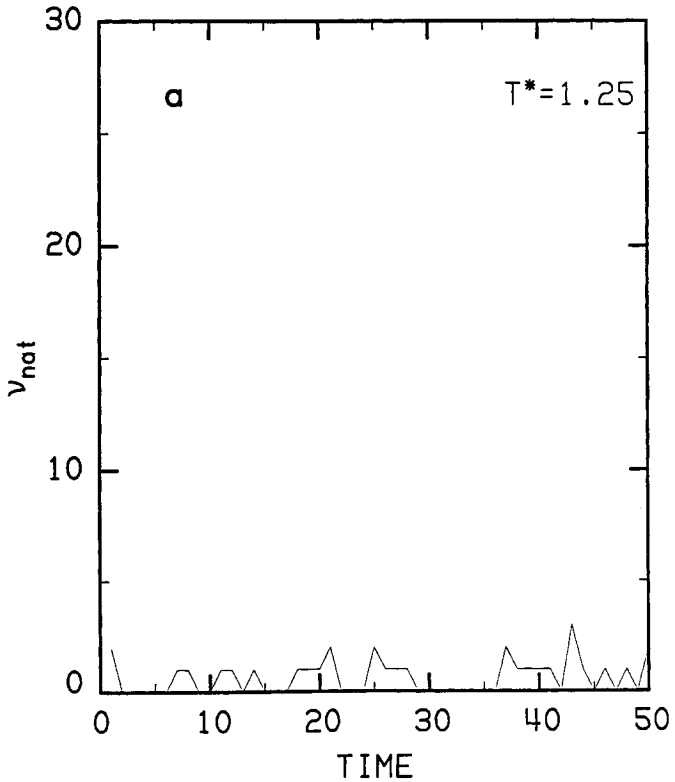

Fig. 8. Single-run flow charts for the number of pairs of native contacts ν_{nat} vs "time" in model D (see text), at reduced temperatures $T^* = 1.25$ (a), 1.0 (b), and 0.667 (c), respectively.

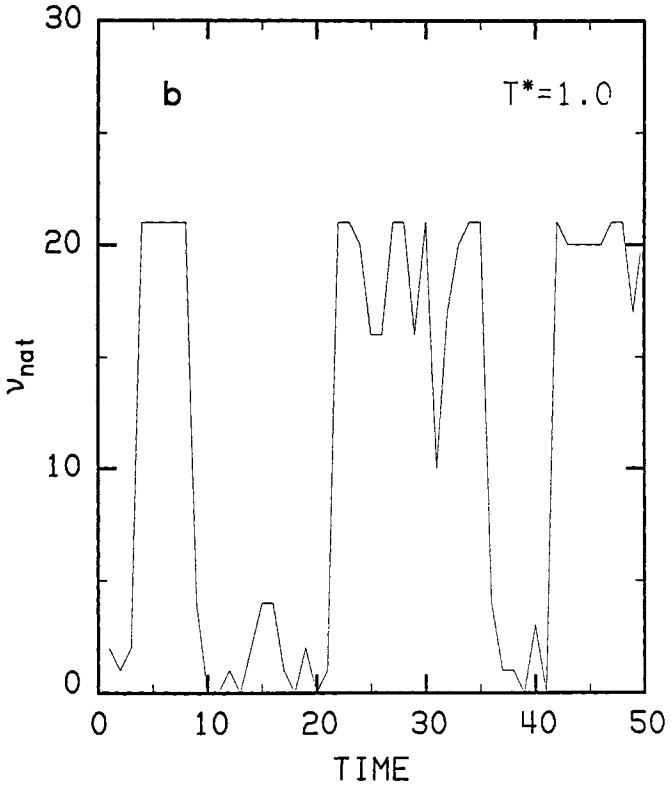

Fig. 8. *(Continued from the previous page.)*

native contacts is usually equal to zero with only a single, occasional native-like contact. In the low-temperature region, at $T^* = 0.667$, the system exists only in the unique four-member α-helix bundle native state. The number of native nearest neighbor pairs in the four-helix bundle equals 21, with a fluctuation from this value usually of no more than one contact pair. These fluctuations involve a few of the end bonds in stretches 1 or 4. In the transition region at $T^* = 1.00$, at a given moment only the entirely ordered native structure or a random coil state chain is populated. During a single simulation in the transition temperature region, the native structure is created and destroyed many times. This behavior strongly suggests that an all-or-none transition is present, as is observed in real proteins. Unfortunately, the finite time of simulations does not allow us to compute the equilibrium constant between the native and denatured states. For a given amount of CPU time, the frequency of occurrence of the native state depends on the particular choice of interaction parameters employed.

DISCUSSION

In the context of an α-carbon representation of a protein on a tetrahedral lattice, the sufficient conditions required for the folding of a model α-helical globular protein whose conformational transition from the denaturated to native state is well approximated by an all-or-none model were explored.

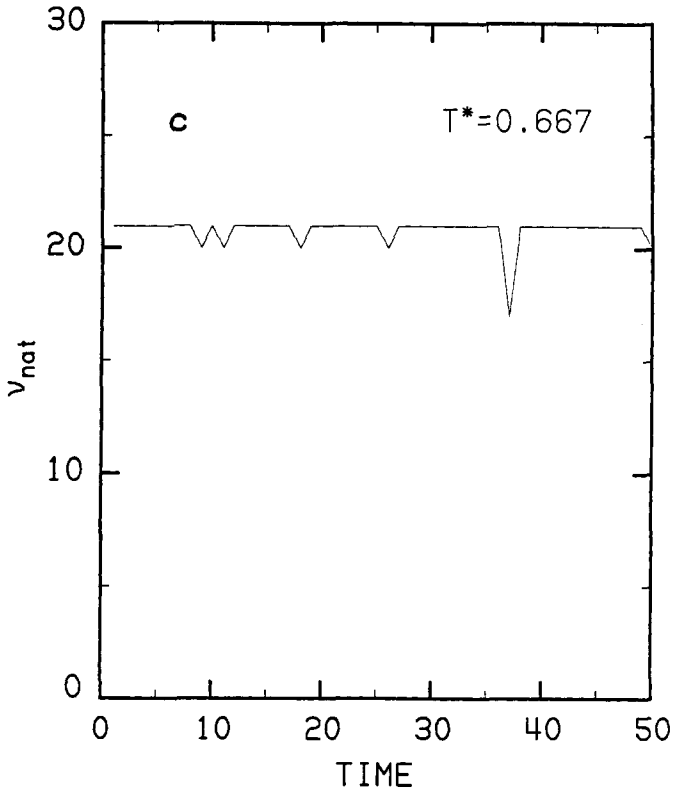

Fig. 8. (*Continued from the previous page.*)

Similar to previous studies on model β-barrel-like proteins,[14-16] it is shown that site-specific interactions of amino acid residues are not necessary to obtain a unique native α-helical globular protein at low temperatures. The conditions required for the formation of a unique four-member α-helical left-handed bundle (which can be viewed as a crude approximation of the tertiary structure in cytochrome c'[18] or myohemerythrin[19] or the synthetic four-helix bundle of DeGrado et al.[20]) are the general pattern of hydrophobic and hydrophilic interactions, the occurrence of some secondary structure in the denaturated state near the transition, and a weak statistical tendency to form essential bends. In the four-helix bundle case, the essential bend is the central turn of the four-helix bundle. This is unlike the case in four-member β-barrel models, which require in the primary sequence the two outer turns to have a statistical preference for their formation as well.[16] This presumably reflects the greater cost of forming out-of-register structures in helices than β-sheets due to the large number of contacts per repeat unit in the four-helix bundle. Overall, though, the rules for folding are the same as those found for lattice β-barrel-like globular proteins.[16]

The role of bends in the folding of tertiary structures and their influence on folding pathways requires further study.[26] The introduction of bend regions helps to fold α-helical stretches, according to hydrophobic/hydrophilic register. Elimination of the central statistical bend-forming region leads to col-

lapsed but nonnative globular states, thereby indicating that the central bend is essential for formation of the four-helix bundle. Bends probably also serve as nucleation centers for the folding of tertiary structure. Moreover, after folding they lock in the native-like conformation. Basically, it costs free energy to move the hydrophobic residues from an amphipathic helical region into a bend.

It has been shown that a very simple lattice model of globular proteins gives a qualitatively good description of the folding transition. However, further refinements in the model are required, for example, the introduction of side chains and the use of a more flexible lattice, before this model can be applied to the more complicated mixed α/β globular proteins. Such work is now underway.

Stimulating discussions with Professors Andrzej Kolinski and Robert Yaris are gratefully acknowledged. This research was supported in part by NIH grant GM-37408 from the Division of General Medical Science, United States Public Health Service. The assistance of the Washington University Business School Computing Facility is gratefully acknowledged.

References

1. Jaenicke, R., Ed. (1980) *Protein Folding, Proceedings of the 28th Conference of the German Biochemical Society*, Elsevier/North Holland Biomedical Press, Amsterdam.
2. Ptitsyn, O. B. & Finkelstein, A. V. (1980) *Quart. Rev. Biophys.* **13**, 339–386.
3. Tanford, C. (1968) *Adv. Protein Chem.* **23**, 121–282.
4. Anfinsen, C. B. (1973) *Science* **181**, 223–230.
5. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167–339.
6. Levy, R. M. Karplus, M., Kushick, J. & Perahio, D. (1984) *Macromolecules* **17**, 1370–1374.
7. Karplus, M. (1986) *Ann. NY Acad. Sci.* **482**, 255–266.
8. Lee, S., Karplus, M., Bashford, D. & Weaver, D. (1987) *Biopolymers* **26**, 481–506.
9. Levitt, M. (1982) *Ann. Rev. Biophys. Bioeng.* **11**, 251–271.
10. Ueda, Y., Taketomi, H. & Go, N. (1978) *Biopolymers* **17**, 1531–1548.
11. Krigbaum, W. R. & Lin, S. F. (1982) *Macromolecules* **15**, 1135–1145.
12. Segawa, S.-I. & Kawai, T. (1986) *Biopolymers* **25**, 1815–1835.
13. Miyazawa, S. & Jernigan, R. L. (1982) *Biopolymers* **21**, 1333–1363.
14. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7267–7271.
15. Kolinski, A., Skolnick, J. & Yaris, R. (1987) *Biopolymers* **26**, 937–962.
16. Skolnick, J., Kolinski, A. & Yaris, R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5057–5061.
17. Skolnick, J., Kolinski, A. & Yaris, R. *Proc. Natl. Acad. Sci. USA*, in press.
18. Weber, P. C., Bartsch, R. G., Cusanovich, M. A., Hamlin, R. C., Howard, A., Jordan, S. R., Kamen, M. D., Meyer, T. E., Weatherford, D. W., Xuong, N. H. & Salemme, F. R. (1980) *Nature* (London) **286**, 302–304.
19. Hendrickson, W. A. & Ward, K. B. (1977) *J. Biol. Chem.* **252**, 3012–3018.
20. Ho, S. P. & DeGrado, W. F. (1987) *J. Am. Chem. Soc.* **109**, 6751–6758.
21. Schiffer, M. D. & Edmunson, A. (1967) *Biophys. J.* **1**, 121–135.
22. Zimm, B. & Bragg, J. (1959) *J. Chem. Phys.* **31**, 526–535.
23. Valeur, B., Jarry, J. P., Geny, F. & Monnerie, L. (1975) *J. Polym. Sci. Polym. Phys. Ed.* **13**, 667–674.
24. Binder, K., Ed. (1984) *Application of the Monte Carlo Method in Statistical Physics*, Springer, Berlin, chap. 5.
25. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *J. Chem. Phys.* **85**, 3585–3597.
26. Rose, G., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1–109.

Received April 18, 1988
Accepted August 8, 1988