

# Monte Carlo Studies on Equilibrium Globular Protein Folding. II. $\beta$ -Barrel Globular Protein Models

JEFFREY SKOLNICK,\* *Institute of Macromolecular Chemistry, Department of Chemistry, Washington University, St. Louis, Missouri 63130*; ANDRZEJ KOLINSKI, *Department of Chemistry, University of Warsaw, Pasteura-1, 02-093, Warsaw, Poland*; and ROBERT YARIS, *Institute of Macromolecular Chemistry, Department of Chemistry, Washington University, St. Louis, Missouri 63130*

## Synopsis

In the context of dynamic Monte Carlo simulations on a model protein confined to a tetrahedral lattice, the interplay of protein size and tertiary structure, and the requirements for an all-or-none transition to a unique native state, are investigated. Small model proteins having a primary sequence consisting of a central bend neutral region flanked by two tails having an alternating hydrophobic/hydrophilic pattern of residues are seen to undergo a continuous transition to a  $\beta$ -hairpin collapsed state. On increasing the length of the tails, the  $\beta$ -hairpin structural motif is found to be in equilibrium with a four-member  $\beta$ -barrel. Further increase of the tail length results in the shift of the structural equilibrium to the four-member  $\beta$ -barrel. The random coil to  $\beta$ -barrel transition is of an all-or-none character, but while the central turn is always the desired native bend, the location of the turns involving the two external strands is variable. That is,  $\beta$ -barrels having the external strands that are two residues out of register are also observed in the transition region. Introduction into the primary sequence of two additional regions that are at the very least neutral toward turn formation produces an all-or-none transition to the unique, native, four-member  $\beta$ -barrel. Various factors that can augment the stability of the native conformation are explored. Overall, these folding simulations strongly indicate that the general rules of globular protein folding are rather robust—namely, one requires a general pattern of hydrophobic/hydrophilic residues that allow the protein to have a well-defined interior and exterior and the presence of regions in the amino acid sequence that at the very least are locally indifferent to turn formation. Since no site-specific interactions between hydrophobic and hydrophilic residues are required to produce a unique four-member  $\beta$ -barrel, these simulations strongly suggest that site specificity is involved in structural fine-tuning.

## INTRODUCTION

Since the time Kendrew obtained the first low-resolution x-ray crystal structure of myoglobin,<sup>1</sup> the elucidation of the factors responsible for the folding of globular proteins has been an objective of biophysical chemistry.<sup>2-8</sup> In particular, the role of short-range vs long-range interactions has long been debated.<sup>9</sup> Simply stated, can one predict the tertiary structure based primarily on the down-chain, short-range interactions,<sup>10,11</sup> or are tertiary interactions themselves responsible for the creation as well as the stabilization of the native conformation, or is it some combination of the two?<sup>12</sup> If tertiary

\*To whom correspondence should be addressed.

interactions are in fact very important (and since almost all proteins do not adopt appreciable amounts of  $\alpha$ -helical or  $\beta$ -sheet structure in the denatured state), it is not clear at what level they must be treated. Must one know all the intimate details of the interactions? Or since homologous proteins such as the globins possess quite small sequence homology and yet have very similar tertiary structure,<sup>13</sup> perhaps the general rules of protein folding are very robust and, provided the rules are not violated, local differences in amino acids merely produce minor structural fine-tuning. It is the latter viewpoint we have adopted in a series of Monte Carlo (MC) simulations on globular protein folding.<sup>14-17</sup> In the current work on model protein systems, we shall examine the factors responsible for the formation from a random-coil state of a structurally unique four-member  $\beta$ -barrel native conformation as well as the features that make the conformational transition all-or-none.<sup>8</sup>

Two basically complementary theoretical/computational approaches have emerged to address the factors determining protein conformation and stability. The first approach employs as realistic a potential energy surface for the protein as is currently available.<sup>18-21</sup> One then starts with an initial folded state typically obtained from the x-ray structure, and employs molecular or Brownian dynamics to obtain information about the equilibrium and dynamic properties of the protein. This is an extremely powerful technique for obtaining local, relatively short-time (on the order of hundreds of picoseconds) information. Unfortunately, proteins neither fold nor globally unfold on this time scale. To obtain somewhat longer time information, an umbrella sampling technique is employed.<sup>18</sup> Here, one has to guess at the reaction coordinate for the moiety undergoing the transformation. Again this technique involves relatively minor conformational adjustments, e.g., distortion of the lattice accompanying the rotation of a tyrosine ring. If one could really specify with a large degree of certainty the unfolding or folding pathway of the protein, a substantial part of the problem would already be solved; it is precisely this information that is lacking.

An alternative approach espoused by Scheraga et al.<sup>22,23</sup> and again using a "realistic set" of interactions, involves the buildup of a polypeptide chain starting from its constituent amino acids. In its most sophisticated variant to help alleviate the multiple minima problem, a Monte Carlo/energy minimization approach is used.<sup>23</sup> So far the method is useful for the study of small polypeptides (e.g., pentapeptides), but so far has not been shown capable of treating the number of amino acids present in even a small protein.

Such detailed studies of local protein conformations are extremely important in providing insight into aspects of local globular protein properties. However, due to the very large number of degrees of freedom involved, this approach would require a computer many orders of magnitude faster than any currently available to fold a protein from the random coil to the native state.<sup>8</sup> Moreover, given the complicated potentials employed, even if such folding of a protein were successful, one would be no closer to understanding the general rules of protein folding—i.e., one would then have to dissect the free energy surface to distinguish the essential from the nonessential features. In other words, one would pare down the potential until the point when the protein failed to fold; then one would have established the necessary conditions for

the folding of the particular protein under consideration. The folding general rules would have to emerge by induction.

We have opted for an alternative approach and are attempting to develop a series of model proteins where the minimal set of interactions necessary to fold a given topology are introduced, and the folding/unfolding transition is studied using highly efficient MC techniques.<sup>14-17</sup> Refinements in the allowed interactions are introduced only when absolutely required to reproduce a desired characteristic present in real globular proteins.

When using a schematic model, care must be taken to ensure that the conformational transition of the model and real globular proteins have the same qualitative features. First and foremost, the native (N) state cannot be a priori specified in advance. Just as in real proteins, interactions between all residues must be allowed. If only interactions that occur in the native state are permitted, this would serve as a check of the algorithm, but it cannot provide any guarantee that the assumed minimal set of interactions would fold the protein, once such a stringent restriction is eliminated. Allowing nonnative interactions also permits an examination of the factors required to produce an all-or-none transition. One of the most remarkable things about the conformational transition in small globular proteins is that the transition is well approximated by the two-state model.<sup>8,24-27</sup> Namely, a given molecule is either completely denatured (D) or it folds to a thermodynamically unique native state. When we say that the  $D \leftrightarrow N$  transition is all-or-none we do not, of course, mean that in the transition between the two states the molecule does not pass through a large number of intermediate states. We only mean that the time spent in these intermediate states is negligibly small compared to the time spent in the D and the N states. Or said another way, the probability that the protein is in a configuration that is not identifiable as either a D or an N configuration is very small. Yet the free energy of stabilization per residue is small, on the order of a few hundred calories per mole,<sup>28,29</sup> and based on naive statistical mechanical considerations, one might expect a whole manifold of states to be populated. Hence, while we do not in advance construct a model that has an all-or-none conformational transition, such a transition must emerge from the simulation. The necessary conditions that change an intrinsically continuous transition to one of the all-or-none type are investigated below. If one can in fact produce a crossover between the two qualitatively different kinds of transitions, the insights gained would presumably have applicability to the folding transition in real globular proteins as well.

Possibly the most stringent requirement is that the model protein collapse not merely to a thermodynamically unique state, but also to a structurally unique state, about which only local conformational rearrangements occur.<sup>30,31</sup> Otherwise stated, every time a  $D \rightarrow N$  transition occurs, the same native state results; otherwise one has missed what is perhaps the most essential feature of globular protein folding. Ideally, the model must not only satisfy the above requirements but must also have a three-dimensional, backbone structural motif that occurs in nature; or at the very least, it should reproduce a structural building block or domain of a naturally occurring protein tertiary structure.

So that the above approach be practical, efficient MC algorithms that surmount the multiple minima problem must be employed. We use a dynamic MC method<sup>14-17,32</sup> that works well at sampling both the vast range of configuration space accessible to the denatured state as well as the rather steep confines of the native conformation. Moreover, since we are doing a series of computer experiments, in which we guess that a particular kind of interaction is important and then examine the consequences of incorporating it into the model, the algorithms must be sufficiently fast so that an entire denaturation or renaturation run can be done in a reasonable amount of CPU time. Otherwise, the computations become prohibitively expensive, and the characterization of parameter space will of necessity be incomplete.

In work done to date, we have focused on the folding of  $\beta$ -proteins, the results of which are summarized below.<sup>14-16</sup> The first question addressed was the role (if any) played by secondary structure in the denatured state, a controversial subject.<sup>8,23</sup> What is agreed upon is that at most a small amount of secondary structure exists in the unfolded protein; whether the thermally induced denatured state is a pure random coil lacking any elements of (fluctuating) secondary structure or not is not universally agreed upon. Thus, we considered the simple case of a homopolymeric protein,<sup>16</sup> in which no elements of secondary structure at all are favored in the denatured state, yet where the local tertiary interactions favor  $\beta$ -sheet formation. Just as in the case of very flexible polymers,<sup>14,15,33</sup> collapse is to a molten globule that is globally disordered; however, here there are patches of local ordering. The transition is extremely steep, but successive heating and cooling produces a different collapsed conformation each time. Therefore, this is a miserable globular protein model. Thus, if this class of models has anything at all to say about the conformational transition in real proteins, then there appears to be a requirement that a small amount of fluctuating secondary structure is present in the denatured state.

Introduction of a small preference for  $\beta$ -pleated sheet formation, along with the same kind of tertiary interactions as in the above model, resulted in collapse to a  $\beta$ -barrel-like structure.<sup>16</sup> The mean strand length is approximately constant on successive refolding, but the number of  $\beta$ -stretches and bend locations are variable. While we were able to establish by the standard thermodynamic criterion<sup>29,34</sup> that these systems behave like two-state models, they fail to reproduce a unique  $\beta$ -barrel native state; rather a manifold of folded structures result.

It appears, then, that models having completely uniform interactions, i.e., where each amino acid residue is the same, while reproducing a number of features of the folding transition, cannot be the whole story. Two essential pieces have been omitted from the physics. One of the most obvious features of real globular proteins is that they have a well-defined interior and exterior.<sup>9</sup> The aforementioned class of models do not. Consider, for example, the formation of even the simplest of folded structures, the  $\beta$ -hairpin. If all the interactions are uniform, then there are two mirror-image conformers; the two are related by the reversal of the interior and exterior residues. Therefore, one should at the very least specify the pattern of hydrophobic and hydrophilic residues.

The introduction of hydrophobic/hydrophilic residues in an appropriate odd/even pattern characteristic of an antiparallel  $\beta$ -protein<sup>36</sup> should also serve to reduce the fluctuations in turn location and thereby the manifold of conformational states. In other words, suppose one shifts the location of the turn so that now hydrophobic and hydrophilic residues are in contact. This conformation will have a much higher free energy; thus shifts in registration by at least two residues that return the molecule to a pattern of hydrophobes interacting with other hydrophobes would be expected to occur.

Suppose one then introduces into the primary sequence regions that have a statistical tendency to form turns.<sup>9,18-20</sup> That this is physically reasonable is strongly suggested by the nmr studies of Wright and co-workers on polypeptide fragments, which find evidence of a statistical population of native-like turn conformations even in the denatured state.<sup>37-39</sup> If a tight bend forms in a region previously found in the interior of the protein, the shift of hydrophobic residues into the bend costs free energy. Thus, the presence of regions that have a statistical tendency to form bends should help reduce the multiplicity of collapsed conformations. The validity of these qualitative conjectures are examined below.

The outline of the remainder of the paper is as follows: In all cases, we focus on proteins whose collapsed structure is composed of  $\beta$ -sheets. In the next two sections, a detailed discussion of the model and the simulation algorithm are presented. The following section describes the folding of  $\beta$ -hairpins having a primary sequence pattern of a central turn neutral region flanked by two tails, each of which contains the odd/even hydrophobic/hydrophilic pattern of residues. The character of the transition from the random-coil state to a  $\beta$ -hairpin is analyzed. As the size of the protein is increased, but where the primary sequence pattern remains the same, a crossover from  $\beta$ -hairpins to the four-member, antiparallel  $\beta$ -barrel structural motif is observed. Then, the factors responsible for the collapse to a unique four-member  $\beta$ -barrel structure and the nature of the transition are discussed in detail. Finally, we conclude with a summary of the present calculations and a discussion of their implications to the folding of real proteins.

## THE MODEL

Consider a consecutive sequence of  $n$  beads occupying the vertices of a diamond lattice, and let  $l$  be the virtual bond length. Each bead represents an amino acid residue. Hence, an  $\alpha$ -carbon representation of a protein is used.<sup>40</sup> Excluded volume is incorporated into the model by prohibiting the multiple occupancy of all lattice sites. Thus, at most a single bead can occupy a lattice site. Each of the  $n - 3$  interior beads numbered from bead 2 to bead  $n - 2$  have three allowed rotational state conformations. There is the lowest energy *trans* (t) state consisting of the in-plane arrangement of three consecutive bonds and the two out-of-plane *gauche* states denoted by  $g^+$  and  $g^-$ , respectively. In the following, the reduced temperature of the system is defined by  $T^* = k_B T / \epsilon_g$ , with  $k_B$  Boltzmann's constant and  $T$  the absolute temperature.

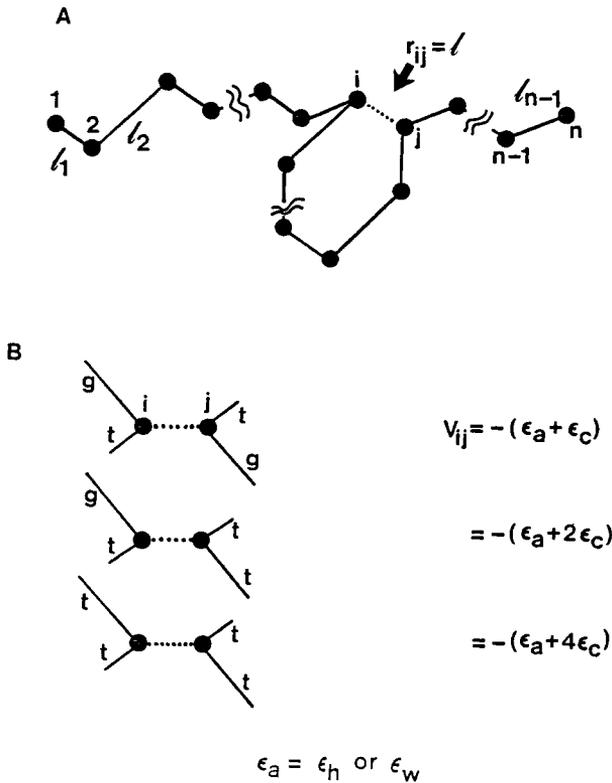


Fig. 1. (A) Schematic representation illustrating the coupling between the local conformation and the nearest neighbor interactions between the nonbonded bead pair  $i$  and  $j$ . (B)  $t$  or  $g$  denotes a *trans* or *gauche* conformation associated with the particular bond either preceding or following beads  $i$  or  $j$ .

We next discuss our simple implementation of hydrophobic and hydrophilic interactions. Consider, as in Fig. 1(A), a pair of nonbonded nearest neighbor residues. Suppose both are hydrophobic. Then,  $\epsilon_h$  (a negative quantity) is the attractive interaction free energy between the two. Thus, in a very simplistic manner  $\epsilon_h$  mimics the hydrophobic interaction. It might also mimic an attractive interaction arising from a salt bridge. If, however, both residues are hydrophilic or if they form a hydrophobic-hydrophilic pair, then a positive  $\epsilon_w$  accounts for their effective repulsive interaction. It should be pointed out that  $\epsilon_h$  and  $\epsilon_w$  are intrinsically independent of the backbone chain configuration.

To explicitly allow conformational coupling between nonbonded but nearest neighbor conformational states, an additional parameter  $\epsilon_c$  is introduced. The magnitude of the conformational coupling is depicted in Fig. 1(B).  $\epsilon_c$  is in fact a preaveraged parameter that implicitly assumes a favorable interaction pattern of amino acids in the primary sequence and allows for interactions between second nearest neighbors down a pair of adjacent sections of the protein chain. Qualitatively identical behavior has been found on neglecting  $\epsilon_c$  altogether, and an example of this is also given. We remind the reader that  $\epsilon_h$ ,  $\epsilon_c$ , and  $\epsilon_w$  are potentials of mean force, where the contribution of the solvent to the interaction has been included.

Based on the above, the total configurational energy of the model protein is

$$E = \sum_{i=2}^{n-2} (1 - \delta_{ii}) \epsilon_g - \frac{1}{2} \sum_{i=1}^n \sum_{|j-i|>1}^n [\theta(r_{ij}) - \theta(r_{ij} - l)] \epsilon_{ij} - \frac{1}{2} \sum_{i=2}^{n-1} \sum_{|i-j|>1}^n [\theta(r_{ij}) - \theta(r_{ij} - l)] [\delta_{t,i-1} + \delta_{t,i}] [\delta_{t,j-1} + \delta_{t,j}] \epsilon_c \quad (1)$$

Wherein  $\delta_{ii} = 1$  if the  $i$ th rotational state is *trans* and zero otherwise.  $\theta$  is a Heaviside step function,<sup>41</sup>  $r_{ij}$  is the distance between the beads under consideration, and  $\epsilon_{ij} = \epsilon_h$  or  $\epsilon_w$  depending on whether residues  $i$  and  $j$  are both hydrophobic or at least one of the two is hydrophilic. Observe that the zero of energy in these calculations is the pure *trans* state of the molecule.

In the simulations described below, we introduce into the primary sequence hydrophobic and hydrophilic type residues; in addition, some residues may have a statistical tendency to form turns. These turn regions are never externally constrained to adopt a native bend configuration; rather, the system must find such a state. A short hand notation described below is employed to specify the primary sequence. Let  $\mathbf{B}_i(k)$  specify the  $i$ th stretch in the primary sequence that contains  $k$  consecutive residues. The  $i$ th stretch is not constrained a priori to adopt any particular conformation, e.g., a  $\beta$ -sheet. It can choose whatever favorable conformation that it prefers based on free energy considerations. All internally odd-numbered residues (i.e.,  $k = 1, 3, 5$ , etc.) in the  $i$ th stretch are hydrophobic. The even-numbered residues may be either hydrophobic, hydrophilic, or inert; the particular residue type will be specified as needed. There are  $k - 3$  conformational states associated with each stretch of length  $k$ , and all the  $\epsilon_g$  of these states are taken to be identical unless otherwise specified. Furthermore, regions that have a tendency to form a turn are delineated by  $\mathbf{b}_i$ ; the pattern is taken to be consistent with formation of a tight turn. Hence  $\mathbf{b}_i$  consists of the last two beads in  $\mathbf{B}_i$  and the first bead in region  $\mathbf{B}_{i+1}$ . If  $\mathbf{b}_i$  is absent in the specification of the primary sequence, then the interaction pattern of residues  $k - 1$  and  $k$  in sketch  $i$  and residue 1 in stretch  $i + 1$  are taken to be that of  $\mathbf{B}_i$  and  $\mathbf{B}_{i+1}$ , respectively. When  $\mathbf{b}_i$  is explicitly displayed, then the energy of the three rotational states must be specified. In all cases  $\epsilon_h = \epsilon_c = \epsilon_w = 0$ , but  $\epsilon_g$  need not necessarily equal zero.

### MONTE CARLO ALGORITHM

One of the important features of a simulation algorithm is the ability to efficiently sample configuration space for the denatured state (a relatively simple matter to implement) and the native conformation. The sampling of configurations near the native conformation is not so straightforward because of the large free energy barriers that exist between the lowest free energy conformation (the native state) and other conformations that are also deep (but not the deepest) free energy minima. To surmount the multiple minima problem, a dynamic Monte Carlo (MC) technique based on an asymmetric Metropolis scheme is employed.<sup>32,42</sup> A MC sampling cycle is composed of the

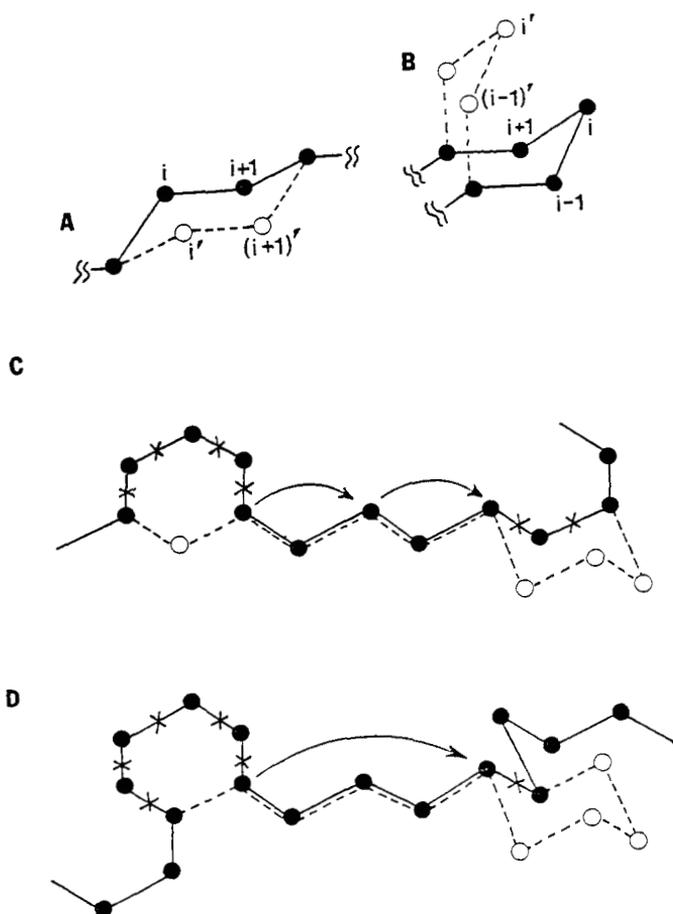


Fig. 2. Schematic representation of the kinds of MC moves employed. (A) Three-bond kink motion. (B) Four-bond kink reorientation. (C) Four-bond wave motion. (D) Five-bond wave motion. (E) High-temperature end-bond reorientation. (F) Large-scale tail rotations sometimes employed in the low-temperature state to surmount the large free energy barriers between the conformations associated with the native and native-like states.

following kinds of elementary moves, shown schematically in Fig. 2, each of which is attempted on a randomly chosen bead<sup>17</sup>:

(A) Three-bond kink motion consisting of the interchange of a  $g^{\pm}$  conformation for a  $g^{\mp}$  conformation (this is analogous to jumping from one half of the chair conformation in cyclohexane to its mirror image), see Fig. 2(A).

(B) Four-bond kink reorientation involving  $g^{\pm}g^{\mp} \rightarrow g^{\mp}g^{\pm}$ ; see Fig. 2(B).

(C) Four-bond wave motion in which a randomly selected four-bond sequence possessing the  $g^+g^-$  or  $g^-g^+$  conformation is interchanged with a randomly chosen consecutive pair of bonds located elsewhere in the chain; see Fig. 2(C).

(D) Five-bond wave motion involving the interchange of a conformation composed of five of the six bonds that form a closed cyclohexane-like ring with a single bond that has been clipped out from another part of the chain; see Fig. 2(D).

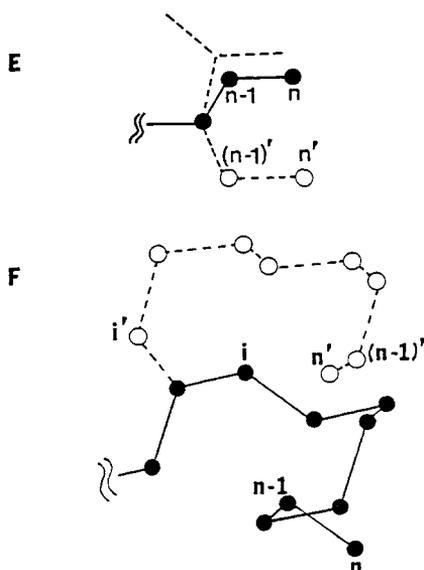


Fig. 2. (Continued from the previous page.)

In elementary motions C and D, the beads are renumbered so that the sequence of residues always remains the same.

(E) End bond rotations involving a randomly chosen end. At higher temperatures where random coil like conformations dominate, as shown in Fig. 2(E), the rotation of one or two end bonds is attempted.

(F) As the temperature decreases and the mean length of *trans* states increases, one needs the option of invoking larger tail rotations to ensure that the system does not become trapped in a deep local minimum. Thus, as shown in Fig. 2(F), if  $l_s$  denotes the guessed mean length of an end  $\beta$ -sheet in the native state, then the random rotation of an entire tail starting from a randomly chosen bead up to a distance  $l_s$  from an end is attempted. For the majority of simulations, the entire set of moves A–F are incorporated into a given MC cycle. However, in some of the later simulations, step type F has been omitted without any noticeable effect on the efficiency of the algorithm.

The set of moves described above differs from our previous simulations<sup>14–16</sup> in that the reptation move,<sup>43,44</sup> wherein a given bond is chopped off from one end of the chain and added to the other end at a random reorientation followed by bead renumbering, has been omitted. While highly efficient at sampling random coil configurations, when appreciable secondary structure that has been stabilized by tertiary interactions is present, the acceptance rate of the reptation move is very low. Thus, it is not a viable move for sampling native-like configurations. The introduction of four- and five-bond wave motions plays the role of internal reptation-like moves, and simultaneously allows the removal of defects between assembled elements of secondary structure.

After every MC cycle consisting of three attempts at moves A–F on the appropriate randomly chosen chain elements, the resulting new conformation is accepted subject to the standard Metropolis criterion. If  $E_{\text{old}}$  denotes the

energy of the configuration prior to the sequence of modifications and  $E_{\text{new}}$  is the energy of the new conformation reflecting these modifications, then the probability of acceptance of the modifications is

$$P_{\text{new/old}} = \min\{\exp[-(E_{\text{new}} - E_{\text{old}})/k_{\text{B}}T], 1\} \quad (2)$$

A uniform random number generator between  $[0, 1]$  is employed to generate a random number  $R$ . If  $R < P_{\text{new/old}}$  the conformation is accepted; otherwise the conformation is rejected. Using this criterion, in the limit of a long sequence of MC cycles, the energy tends to a Boltzmann distribution and the ensemble average of any quantity of interest can be obtained as the arithmetic mean of the quantity computed over the MC run. A discussion of the ergodicity of the present set of moves is contained in our earlier work.<sup>15,16</sup> However, there is the additional observation that the present algorithm does allow, in the transition region, a fully native molecule to denature and vice versa.

For each and every set of parameters, at least three and up to eight independent cooling and heating sequences each composed of at least  $10^6$  MC cycles per temperature were run. The systems studied were carefully equilibrated. Unfortunately, due to limitations in computational resources, we have only been able to observe between 10 and 20  $N \leftrightarrow D$  transitions for a given set of parameters. Thus, while the  $N$  and  $D$  subpopulations are well characterized, the equilibrium fraction of  $N$  vs  $D$  states are not.

### INTERPLAY OF STRUCTURAL MOTIF AND PROTEIN SIZE

Particularly important for the general understanding of the rules of protein folding is the interrelationship between the number of amino acids in the protein and the structural motif that the protein adopts.<sup>45</sup> In the following we consider a very simple case: Imagine an amino acid sequence having a central region that is neutral toward bend formation (all the  $\epsilon_g$  of these residues are zero) and where the central region is flanked by two tails composed of alternating hydrophilic and hydrophobic residues. One conformation the molecule might adopt is a  $\beta$ -hairpin.<sup>4</sup> We realize that in real  $\beta$ -hairpins the hydrophobic residues are probably still partially exposed to solvent, and as a consequence,  $\beta$ -hairpins are probably marginally stable at best. In the transition region, the random coil state 1 of Fig. 3 might be in equilibrium with a fully in register, lowest energy  $\beta$ -hairpin ( $\beta_2$ ) conformation shown in 2 of Fig. 3. On the other hand, both of the end strands could spontaneously bend over and form a four-member  $\beta$ -barrel ( $\beta_4$ ) as in 3 of Fig. 3. This doubles the number of attractive nearest neighbor interactions. However, six residues that have an intrinsic preference for *trans* ( $\beta$ ) states now reside in the two additional tight turns. The cost of forming these two turns is fixed and is independent of tail size. Thus, as the length of the tails increases, eventually the free energy decrease due to the increased number of hydrophobic interactions accompanying  $\beta$ -barrel formation will overwhelm the cost of forming the two extra turns, and the  $\beta$ -hairpin collapsed state will be replaced by the 4-member  $\beta$ -barrel conformation as the equilibrium state. This expectation is verified below where we present results on the  $\beta_2$  vs  $\beta_4$  structures as a

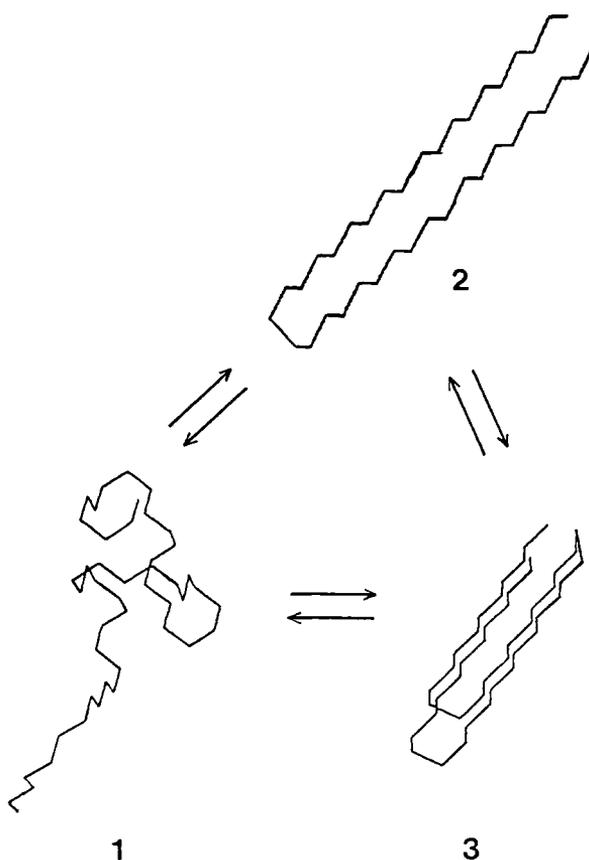


Fig. 3. Representative configurations of a random coil state 1 in equilibrium with the  $\beta$ -hairpin conformation 2 and a four-member  $\beta$ -barrel conformation 3 that a model protein of amino acid sequence  $\mathbf{B}_1(2l_s)\mathbf{b}_1\mathbf{B}_2(2l_s + 2)$  might adopt.

function of the number of residues  $n$  in the model protein. We shall examine the nature of the conformational transition appropriate to the two structural motifs.

Consider a model protein of  $n = 4l_s + 2$  residues on a diamond lattice that has the primary sequence pattern of model A  $\mathbf{B}_1(2l_s)\mathbf{b}_1^0\mathbf{B}_2(2l_s + 2)$ , wherein the conformational states associated with residues  $2l_s - 1$  to  $2l_s + 1$  are neutral toward bend formation. This is indicated by the superscript zero over  $\mathbf{b}_i$ . That is,  $\epsilon_g = \epsilon_c = \epsilon_h = \epsilon_w = 0$  in these locations. The above primary sequence when arranged in a  $\beta_4$  conformation having two  $\beta$ -strands of  $l_s$  residues followed by two  $\beta$ -strands of  $l_s + 1$  residues results in every hydrophobic residue having as its two nearest neighbors hydrophobic residues.

It is straightforward to demonstrate from Eq. (1) that the reduced energy of the  $\beta$ -hairpin configuration in units of  $k_B T$  with the above amino acid pattern is

$$E_N(\beta_2) = (k_s - 1)\epsilon_h/k_B T + (4k_s - 6)\epsilon_c/k_B T \quad (3)$$

and if  $k_s$  is even, the reduced energy of the four-member  $\beta$ -barrel configuration,

$$E_N(\beta_4) = 2(k_s - 1)\epsilon_h/k_B T + 8(k_s - 3)\epsilon_c/k_B T - \epsilon_c/k_B T + 6\epsilon_g/k_B T \quad (4a)$$

whereas if  $k_s$  is odd

$$E_N(\beta_4) = 2(k_s - 1)\epsilon_h/k_B T + 8(k_s - 3)\epsilon_c/k_B T + 6\epsilon_g/k_B T \quad (4b)$$

In what follows, the energy will always be reported in units of  $k_B T$ .

Because this is a model of a protein confined to a lattice and since  $\epsilon_h$ ,  $\epsilon_c$ , and  $\epsilon_w$  represent potentials of mean force, the free energy difference between the  $\beta$ -hairpin and the four-member  $\beta$ -barrel conformation should be very well approximated by the energy difference between  $E_N(\beta_4)$  and  $E_N(\beta_2)$ . That is, to a first approximation we ignore any additional entropic differences between the hairpin and the  $\beta$ -barrel.

Since a  $\beta$ -hairpin has only 3 residues in a *gauche* conformation, but a four-member  $\beta$ -barrel involves 9 residues in a *gauche* conformation, the fraction of *trans* states in the native (lowest energy)  $\beta$ -hairpin is

$$f_t(\beta_2) = (n - 6)/(n - 3) \quad (5a)$$

while for the four-member  $\beta$ -barrel,

$$f_t(\beta_4) = \frac{(n - 12)}{(n - 3)} \quad (5b)$$

In Table I we compile for the  $\beta$ -hairpin,  $E(\beta_2)$  and  $f_t(\beta_2)$ , and for the four-member  $\beta$ -barrel  $E(\beta_4)$  and  $f_t(\beta_4)$  as well as the location of the bends in the two structures for  $n = 22, 26, 30, 34, 38, 42$ , and  $46$ , assuming  $\epsilon_c = -\epsilon_g/2$ ,  $\epsilon_h = -\epsilon_g/4$ , and  $\epsilon_w = 2\epsilon_g$ . (Specification of  $\epsilon_w$  is not required as there are no hydrophilic/hydrophobic nearest neighbor pairs in the native state of the model system.) In the lowest energy, four-member  $\beta$ -barrel structure, the tight turns are located at residues  $l_s - 1$  to  $l_s + 1$ ,  $2l_s - 1$  to  $2l_s + 1$ , and  $3l_s$  to  $3l_s + 2$ .

By examining Table I for  $E_N(\beta_2)$  and  $E_N(\beta_4)$ , it follows for all  $n < 30$ , the dominant conformation should be the  $\beta$ -hairpin. At  $n = 30$ , the  $\beta$ -barrel should be slightly favored by  $-0.5\epsilon_g/k_B T$  over the hairpin conformation, and both motifs should be observed. Beyond  $n = 30$ , the four-member  $\beta$ -barrel should dominate. Of course, as is evident from Eqs. (3), (4a), and (4b), the range of  $n$  corresponding to the crossover from the  $\beta$ -hairpin motif to the four member  $\beta$ -barrel motif depends on the particular choice of values chosen for  $\epsilon_c$ ,  $\epsilon_g$ , and  $\epsilon_h$ . An average parameter that is sensitive to the collapse transition for both hairpins and  $\beta$ -barrels is the fraction of *trans* states  $f_t$ . In the case of small  $\beta$ -hairpins, the mean square radius of gyration defined by

$$\langle S^2 \rangle = \frac{1}{2n^2} \sum_i \sum_j \langle r_{ij}^2 \rangle \quad (6)$$

TABLE I  
Compilation of Native State Conformational Properties for  $\beta$ -hairpins and  
Four-Member  $\beta$ -Barrels<sup>a</sup>

$n$	$l_s$	$\beta$ -Hairpin			Four-Member $\beta$ -Barrel		
		$E_N(\beta_2)$	$f_t(\beta_2)$	Turn Location	$E_N(\beta_4)$	$f_t(\beta_4)$	Turn Location
22	5	$-8\epsilon_g/k_B T$	0.842	9-11	$-4\epsilon_g/k_B T$	0.526	4-6 9-11 15-17
26	6	$-10.25\epsilon_g/k_B T$	0.870	11-13	$-8\epsilon_g/k_B T$	0.609	5-7 11-13 18-20
30	7	$-12.5\epsilon_g/k_B T$	0.889	13-15	$-13\epsilon_g/k_B T$	0.667	6-8 13-15 21-23
34	8	$-14.75\epsilon_g/k_B T$	0.903	15-17	$-17.0\epsilon_g/k_B T$	0.710	7-9 15-17 24-26
38	9	$-17\epsilon_g/k_B T$	0.914	17-19	$-22\epsilon_g/k_B T$	0.743	8-10 17-19 27-29
42	10	$-19.25\epsilon_g/k_B T$	0.923	19-21	$-26\epsilon_g/k_B T$	0.769	9-11 19-21 30-32
46	11	$-21.5\epsilon_g/k_B T$	0.930	21-23	$-31\epsilon_g/k_B T$	0.791	10-12 21-23 33-35

<sup>a</sup>Calculated assuming primary sequence pattern of model A,  $\mathbf{B}_1(2l_s)\mathbf{b}_1\mathbf{B}_2(2l_s + 2)$ , with  $\epsilon_c = -\epsilon_g/2$ ,  $\epsilon_h = -\epsilon_g/4$ , and  $\epsilon_w = 2\epsilon_g$ .

(where the angle brackets denote the average over the MC run), does not change appreciably from its values in the random coil state. Basically on collapse from the random coil to the  $\beta$ -hairpin, one component of  $\langle S^2 \rangle$ , say  $\langle S_x^2 \rangle$ , grows, and the other two components shrink by an amount such that the net change in  $\langle S^2 \rangle$  is on the order of 30% for  $n = 22$  and 10% for  $n = 26$ . However, if the collapse is to a  $\beta$ -barrel from the random coil state, a substantial change in  $\langle S^2 \rangle$  is observed.

### FORMATION OF $\beta$ -HAIRPINS

In Fig. 4(A), the fraction of *trans* states vs  $T^*$  is plotted for  $n = 22$  and 26 in the curves indicated by solid squares and solid triangles, respectively. As is expected on the basis of Table I, the collapse is to a  $\beta$ -hairpin for both cases. In Fig. 4(B), we plot the mean energy per residue vs  $T^*$  for  $n = 22$  and 26 in the curves indicated by the solid squares and solid triangles. As was also evidenced in Fig. 4(A), the transition is seen to be fairly broad and occurs at rather low temperature, with the transition temperature increasing slightly with increasing  $n$ . We next dissect the character of the conformational transition and focus on the  $n = 26$  case for definiteness, but qualitatively identical behavior is observed for  $n = 22$ . In Fig. 4(C), we plot  $P_{12-K, 13+K}$  the probability that native pairs  $12 - K$  and  $13 + K$  are in contact vs the number

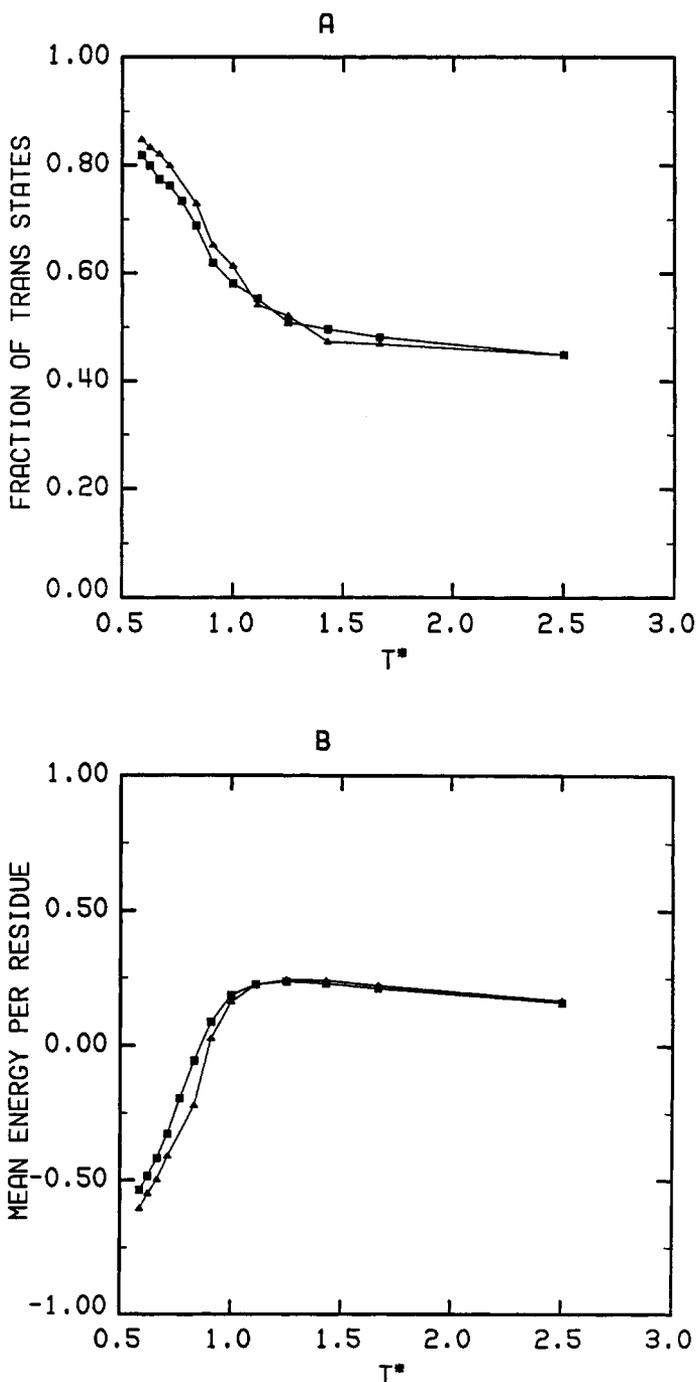


Fig. 4. (A) Fraction of *trans* states vs  $T^*$  for  $n = 22$  and 26 in the curves indicated by solid squares and triangles, respectively. (B) Mean energy per residue vs  $T^*$  for  $n = 22$  and 26 in the curves indicated by the solid squares and solid triangles, respectively. (C) For  $n = 26$ ,  $P_{12-K, 13+K}$  the probability that native pairs  $12 - K$  and  $13 + K$  are in contact vs the number of residues  $K$  from the central turn pair, residues 12 and 13, at  $T^* = 1.00$  (solid squares), 0.905 (solid circles), 0.833 (solid triangles), and 0.667 (solid diamonds) in the curves going from bottom to top. In all cases, the primary sequence is  $B_1(2l_s)b_1B_2(2l_s + 2)$ , with  $l_s = 6$ .

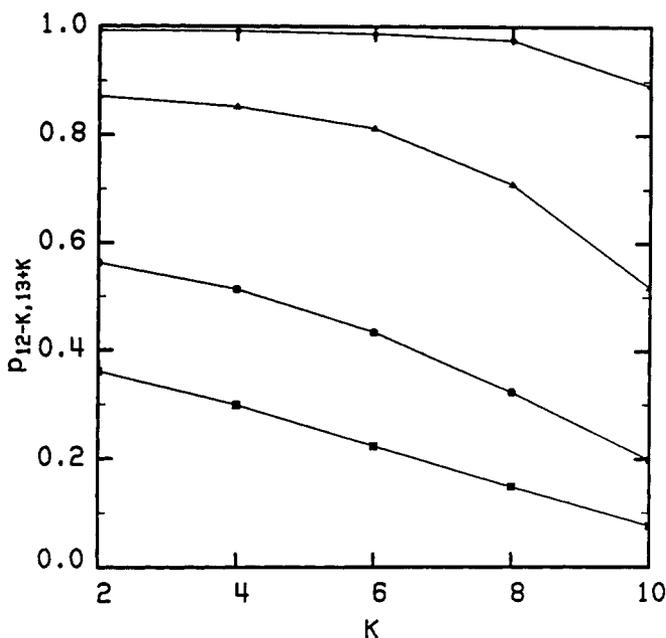


Fig. 4. (Continued from the previous page.)

of residues  $K$  from the central turn pair 12 and 13 at  $T^* = 1.00$  (solid squares), 0.909 (solid circles), 0.833 (solid triangles), and 0.667 (solid diamonds) in the curves going from bottom to top. In the  $n = 26$   $\beta$ -hairpin, proceeding out from the central turn, the nearest neighbor pairs are 10–15 ( $K = 2$ ), 8–17 ( $K = 4$ ), 6–19 ( $K = 6$ ), 4–21 ( $K = 8$ ), and 2–23 ( $K = 10$ ). Observe that  $P_{12-K, 13+K}$  monotonically decreases as one goes out from the central bend. Therefore, it follows that the  $\beta$ -hairpin zips up from the central bond, and moreover, the transition is continuous. This is verified in Fig. 5(A–C), where we plot the mean energy as a function of “time” for  $T^* = 1.25$ , 0.714, and 0.588, respectively. Each time step represents 20,000 MC cycles. In Fig. 5(A), at  $T^* = 1.25$ , the denatured state dominates. In Fig. 5(B), at  $T^* = 0.714$ , the denatured state is in equilibrium with the native state ( $E_N(\beta_2) = -14.35$ ), and in Fig. 5(C) at  $T^* = 0.588$ , the fully native  $\beta$ -hairpin dominates with a lowest energy  $\beta$ -hairpin state having a value  $-17.425$ . Figure 5(B) provides convincing evidence that a whole manifold of partially unfolded states populate the transition region, viz., the transition is continuous.

On the basis of Figs. 4 and 5, we conclude that the conformational transition in a  $\beta$ -hairpin has a number of features in common with that of single-chain polypeptide helices,<sup>46</sup> and qualitatively closely corresponds to the conformational transition expected for putative  $\alpha$ -helical hairpins.<sup>47,48</sup> Namely, due to loop entropy, the folded structure always possesses a rather tight turn and denaturation proceeds from the free ends. The free energy difference between the native  $\beta$ -hairpin and the intermediate states are insufficient to suppress formation of partially unfolded states. Thus, in a very real sense, this is an extremely poor globular protein model. A larger loop that

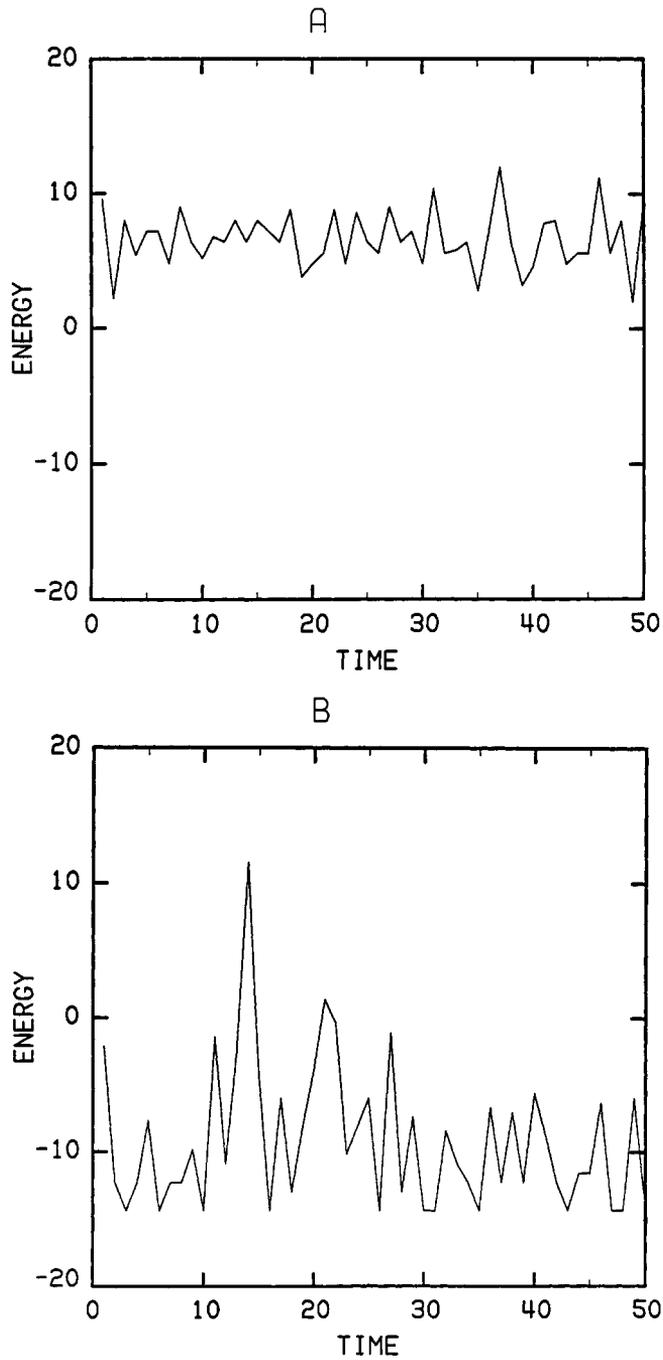


Fig. 5. For the  $n = 26$  system, plot of mean energy vs "time" for  $T^* = 1.25$  (A), 0.714 (B), and 0.588 (C), respectively. Each time point corresponds to 20,000 MC cycles.

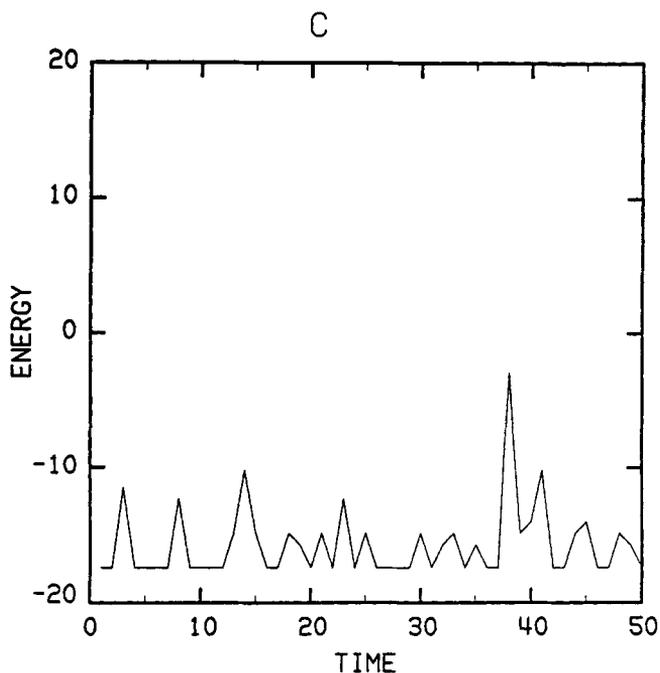


Fig. 5. (Continued from the previous page.)

makes residues 10–17 or 8–15 nearest neighbors occurs with a probability of about 3% up to  $T^* = 0.903$ . Once the  $\beta$ -hairpin states make an appreciable contribution, these larger loops occur with a probability (which decreases very rapidly with decreasing  $T^*$ ) of less than 1%. In other words, the central bend in the native structure is localized at residues 11–13 (which are intrinsically neutral toward bend formation).

That the collapse to a  $\beta$ -hairpin occurs at rather low temperature might cause the  $\beta$ -hairpin to be kinetically favored over the  $\beta$ -barrel for intermediate values of  $n$ , where the  $\beta$ -barrel should dominate the structural equilibrium. The origin of this expectation is as follows: Since  $f_t$  of the four-member  $\beta$ -barrel is fairly small for smaller values of  $n$  (see Table I, column seven), the fraction of *trans* states in the absence of any tertiary interactions,  $f_t^0$ , given by

$$f_t^0 = (1 + 2 \exp\{-1/T\})^{-1} \quad (7)$$

will be larger than the  $f_t$  of the four-member  $\beta$ -barrel. Thus, there are local free energy barriers to the formation of the bends from a *trans* state that must be surmounted to go from a  $\beta$ -hairpin to a four-member  $\beta$ -barrel. This problem should be eliminated as  $n$  increases, because both the fraction of *trans* states in the native structure and the transition temperature should increase, giving  $f_t(\beta_4) > f_t^0$  in the transition region. However, since our objective here is to study the spontaneous crossover from the  $\beta$ -hairpin to the  $\beta$ -barrel motif, we must be able to surmount this local rotational minimum problem.

A particularly straightforward solution is, on the order of 1% of the MC cycles to randomly or systematically (it does not make any difference), to allow the system to make 15 attempts at moves A-F before applying the Metropolis criterion of Eq. (2). This acts to suitably thermalize the system, and the transition from a  $\beta$ -hairpin native state to a  $\beta$ -barrel native state is observed.

In Fig. 6, we plot the average fraction of *trans* states vs  $T^*$  for the  $n = 30$  system obtained from a single cooling sequence. That is, the system starts at infinite temperature and is successively cooled;  $2 \times 10^6$  MC cycles are run at each temperature. The predominance of the  $\beta$ -hairpin conformation is responsible for the maximum in the  $f_t$  vs  $T^*$  observed in Fig. 6. Further cooling results in the exclusive dominance of the four-member  $\beta$ -barrel state. In Fig. 7A, B, we plot the mean energy vs time for  $T^* = 0.909$  and 0.833, respectively. Each time unit represents 20,000 MC cycles. At  $T^* = 0.909$  [Fig. 7(A)], the system initially starts out as a partially folded  $\beta$ -hairpin, then denatures, and refolds once more to the  $\beta$ -hairpin [ $E_N(\beta_2) = -13.75$ ]. Subsequently, it denatures and then refolds to a four-member  $\beta$ -barrel with  $E_N = -14.3$  in the flat region of the energy. The system then denatures and refolds five times to the  $\beta$ -hairpin. (This is the origin of the five dips between the time units 55-80.) Then it twice denatures, partially refolds to the  $\beta$ -hairpin before finally denaturing and then adopting the four-member  $\beta$ -barrel configuration. Further cooling to  $T^* = 0.833$  [Fig. 7(B)] and starting from the  $\beta$ -barrel

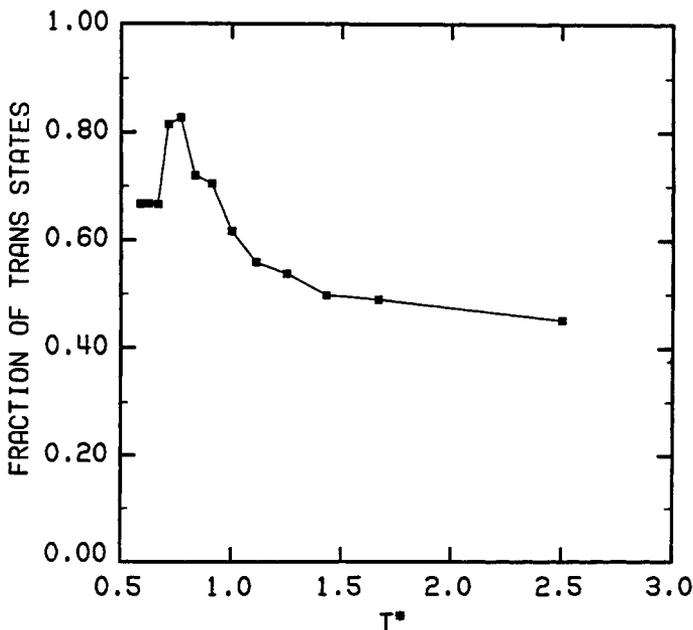


Fig. 6. Average fraction of *trans* states vs  $T^*$  for the  $n = 30$  system obtained from a single cooling sequence.

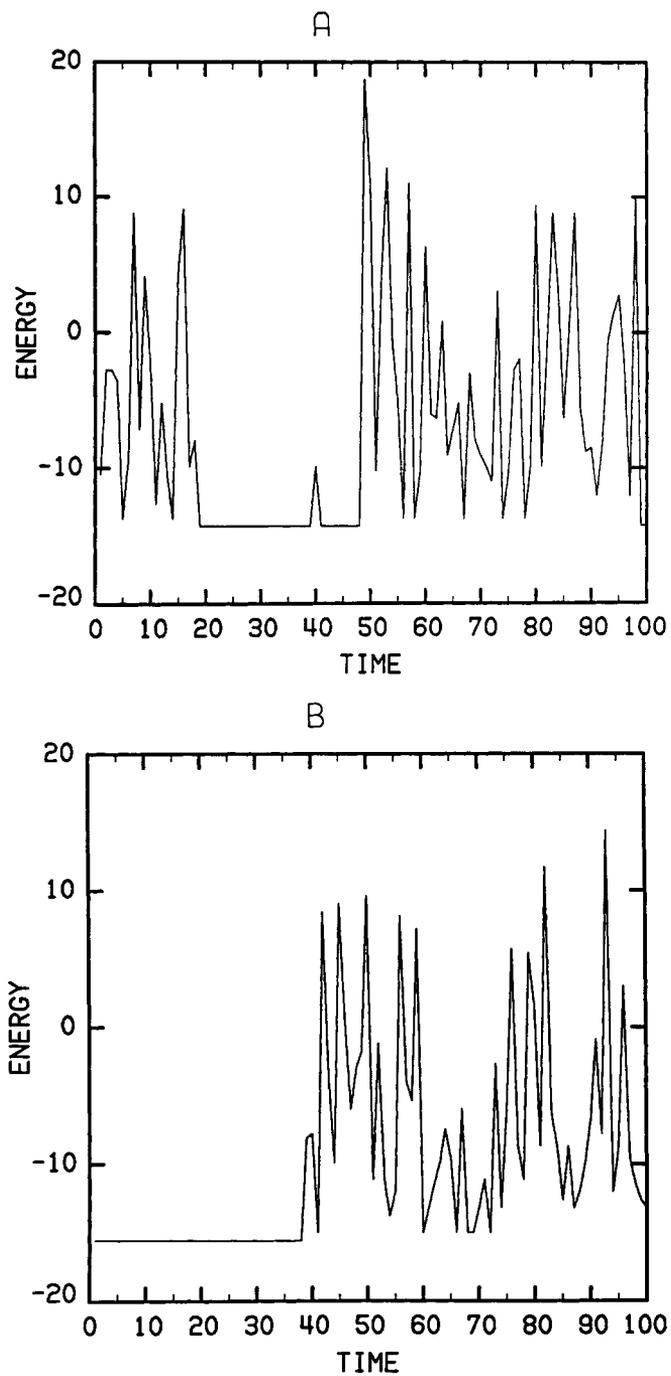


Fig. 7. For model A, plot of the mean energy vs "time" for a  $n = 30$  protein at 0.909 (A) and 0.833 (B). Each time step represents 20,000 MC cycles.

conformation ( $E_N = -15.6$ ) results in multiple transitions between the random coil state and the  $\beta$ -hairpin ( $E_N = -15.0$ ), with the final state being a partially unfolded  $\beta$ -hairpin. (Observe that the random coil to  $\beta$ -barrel transition has an all-or-none character, whereas the denatured state to  $\beta$ -hairpin is continuous.)

We next turn to the  $n = 34, 38,$  and  $42$  cases of model A where the four-member  $\beta$ -barrel should dominate the low-temperature population. In Fig. 8, we plot  $f_t$  vs  $T^*$  for  $n = 34, 38,$  and  $42$ , in the solid squares, solid triangles, and solid diamonds, respectively. In all cases, the low-temperature native state is a four-member  $\beta$ -barrel, with a decreasing frequency of occurrence of the  $\beta$ -hairpin in the transition region as  $n$  increases. In fact, for  $n = 42$ , no hairpins were observed. While the native turns between strands 1 and 2 and strands 3 and 4 dominate the population, in the transition region nonnative turns are also observed, details of which follow. For definiteness we focus on the  $n = 42$  case; similar behavior is also seen for  $n = 34$  and  $38$ . In Fig. 9(A-C), we plot the mean energy as a function of "time" at  $T^* = 1.25, 1.053,$  and  $0.769$ , respectively. Each time unit corresponds to 20,000 MC cycles. At  $T^* = 1.25$  [Fig. 9(A)], the system is in a random coil state. At  $T^* = 1.053$  [Fig. 9(B)], the system is seen to undergo an abrupt jump from the denatured to the renatured state. The energy of the pure native state at this temperature equals  $-24.7$ . On further cooling past the transition region to renaturing conditions [Fig. 9(C)], minor end fluctuations about the native state are

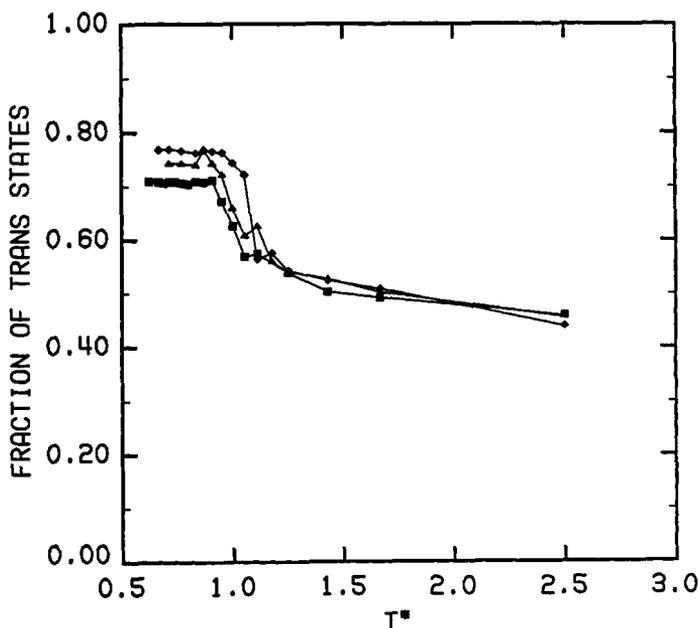


Fig. 8. Plot of the fraction of *trans* states  $f_t$  vs  $T^*$  for  $n = 34, 38,$  and  $42$  in the solid squares, triangles, and diamonds, respectively.

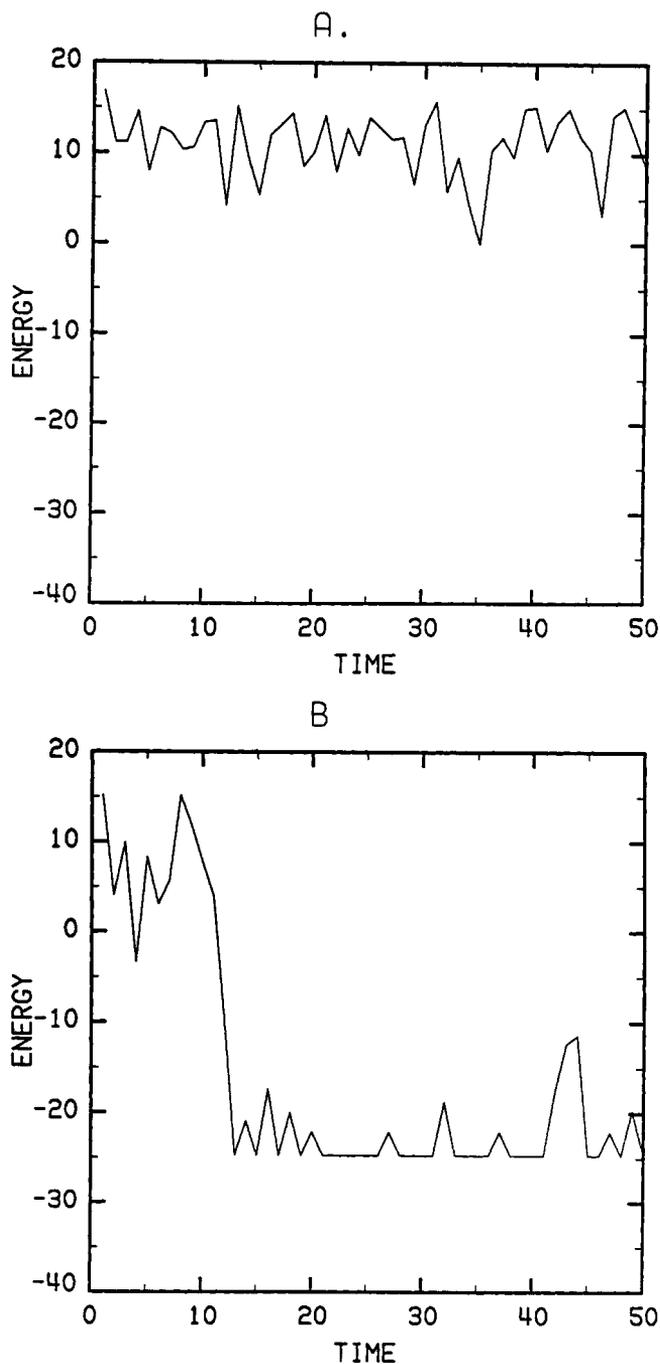


Fig. 9. For  $n = 42$  system, plot of the average energy vs "time" at  $T^* = 1.25$  (A), 1.053 (B), and 0.769 (C), respectively. Each time unit corresponds to 20,000 MC cycles.

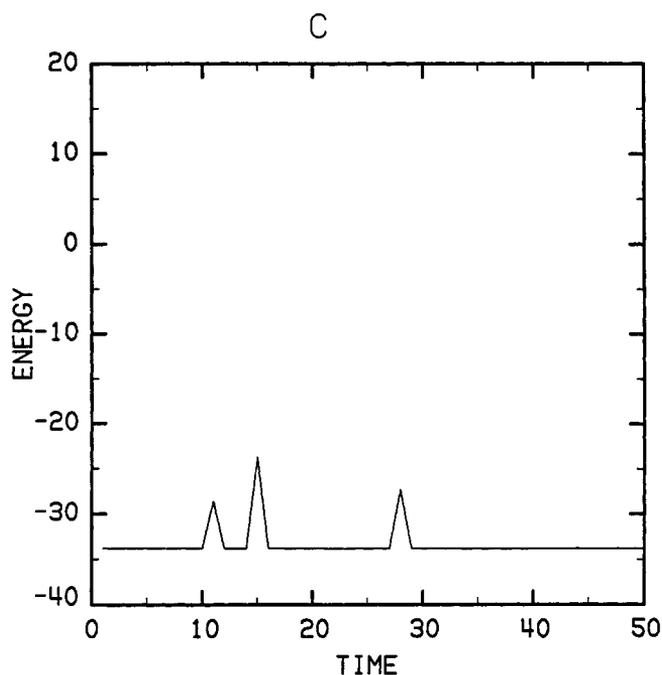


Fig. 9. (Continued from the previous page.)

observed. Thus based on Fig. 9(A–C) this conformational transition is all or none, a necessary feature if this class of models is to mimic globular proteins.

We next turn to a finer, more microscopic characterization of the character of the conformational transition. In Fig. 10(A), we plot, at  $T^* = 1.053$ , the probability of finding native state contacts between pairs of residues  $P_{I-K, I+1+K}$  as a function of the distance from the turn,  $K$ , between strands 1–2 with  $I = 10$  (solid squares) and strands 2–3 with  $I = 20$  (solid circles). In Fig. 10(B), we plot, at  $T^* = 1.053$ ,  $P_{I-K, J+K}$  for strands 3–4 with  $I = 31$  and  $J = 32$  (solid triangles) and strands 1–4 with  $I = 12$  and  $J = 31$  (solid diamonds). Observe that in comparison with Fig. 4(C), the profiles are very flat. It appears that for this particular run strands 2 and 3 zip up first, followed by strands 3 and 4, and then strands 1 and 2. Assembly, however, is extremely rapid, and to a very good approximation the transition is all or none. We also note that nonnative turns comprise about 4% of the population at  $T^* = 1.053$ . These involve a turn between strands 1 and 2 shifted outward by two residues to residues 11–13, which is about 5.1% populated, and a nonnative turn between strands 3 and 4 shifted outward and involving residues 32–34, which is about 2.6% populated. These profiles too are rather flat. Thus, while the transition is all or none, it is not to a unique native state. Observe that the population of the central native turn is always dominant, with less than 2% of the contacts in the vicinity of the central turn being nonnative. These probably reflect denatured state contributions to  $P_{I-K, J+K}$ .

On further cooling to  $T^* = 0.952$ , we find further evidence for a substantial population having a nonnative location of the outer two bends. In Fig. 10(C)

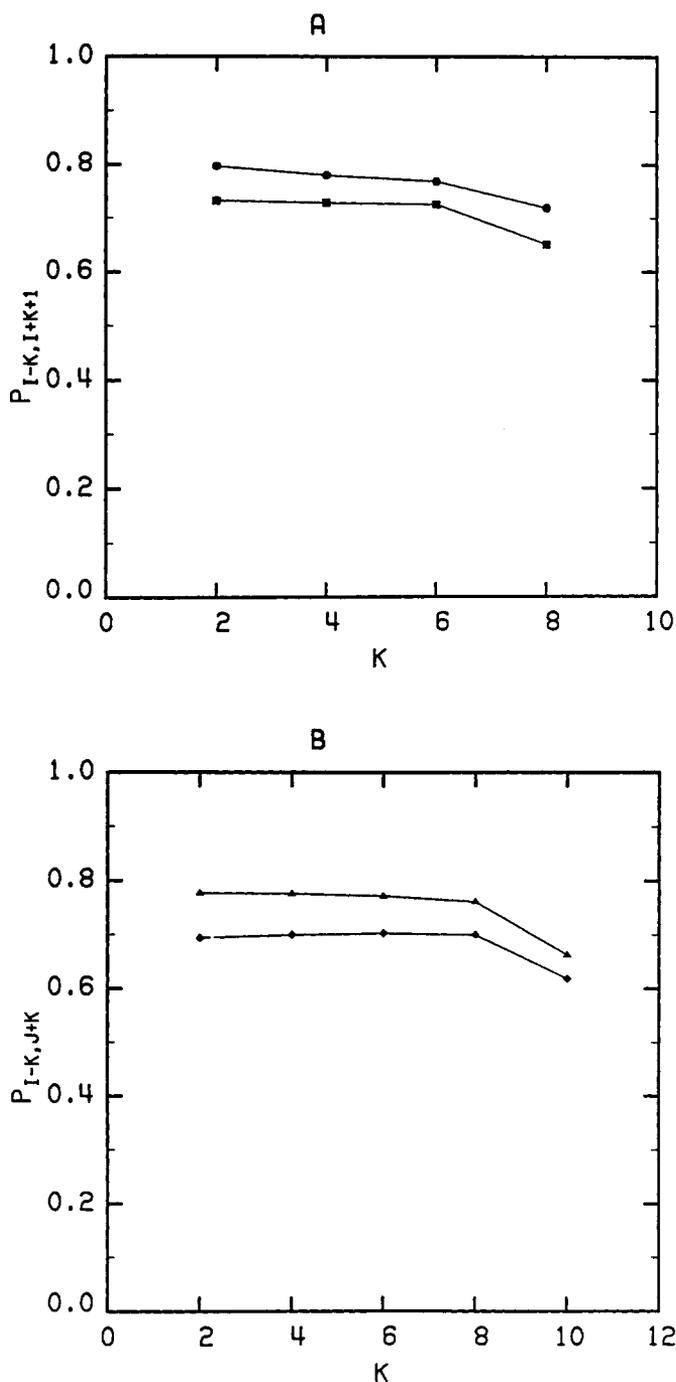


Fig. 10. (A) At  $T^* = 1.053$ , probability of finding native state contacts between pairs of residues,  $P_{I-K, I+K+1}$ , as a function of the distance from the turn for strands 1-2 with  $I = 10$  (solid squares) and strands 2-3 with  $I = 20$  (solid circles). (B) At  $T^* = 1.053$ , probability of finding native state contacts between pairs of residues  $P_{I-K, J+K}$  in strands 3-4 with  $I = 31$  and  $J = 32$  (solid triangles) and 1-4 with  $I = 12$  and  $J = 31$  (solid diamonds). (C) At  $T^* = 0.952$ , the probability of finding native state contacts between pairs of residues,  $P_{I-K, I+K+1}$ , as a function of the distance from the turn for strand 1-2 with  $I = 10$  (solid squares) and strands 2-3 with  $I = 20$  (solid circles). (D) At  $T^* = 0.952$  the probability of finding native state contacts between pairs of residues  $P_{I-K, J+K}$  in strands 3-4 with  $I = 31$  and  $J = 32$  (solid triangles) and 1-4 with  $I = 12$  and  $J = 31$  (solid diamonds).

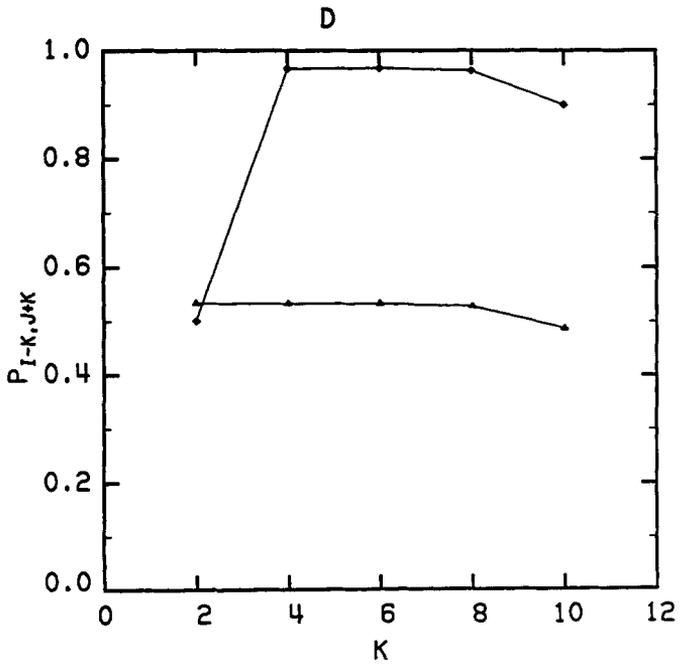
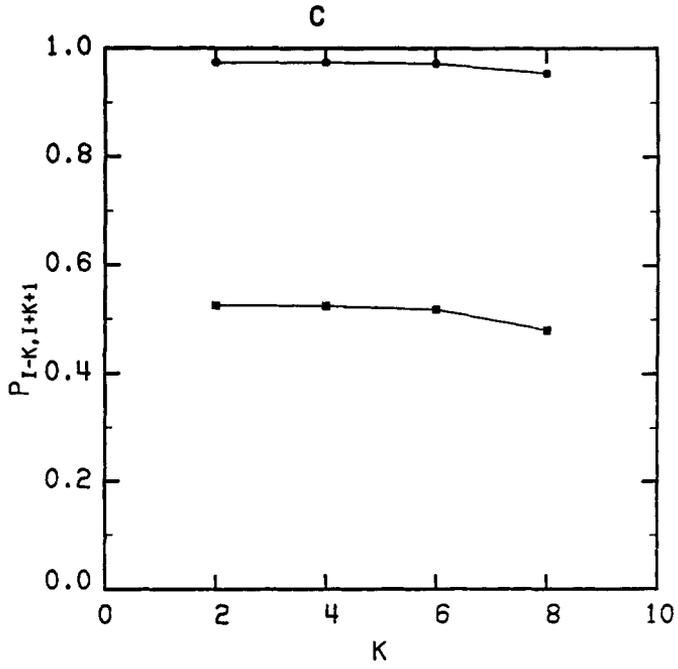


Fig. 10. (Continued from the previous page.)

we plot the probability of native contacts  $P_{I-K, I+1+K}$  for strands 1-2 with  $I = 10$  (solid squares), 2-3 with  $I = 20$  (solid circles), and in Fig. 10(D),  $P_{I-K, J+K}$  for strands 3-4 with  $I = 31$  and  $J = 32$  (solid triangles) and 1-4 with  $I = 12$  and  $J = 31$  (solid diamonds) vs the distance from the turn  $K$ . Observe that  $P_{I-K, J+K}$  for the contacts between strands 2-3 and between strands 1 and 4 is close to unity, whereas  $P_{I-K, J+K}$  between strands 1-2 and 3-4 is about 50%. The origin of this is made clear by examining the other conformations that are appreciably populated. There is also a substantial population involving a nonnative turn at residues 7-9, which produces contacts between residues 6 and 11 with  $P_{6,11} = 0.456$ , residues 4 and 13 with  $P_{4,13} = 0.456$ , and residues 2 and 15 with  $P_{2,15} = 0.446$ . There is also a nonnative turn involving residues 32-34, thereby giving contacts between residues 31 and 36 with  $P_{31,36} = 0.446$ , residues 29 and 38 with  $P_{29,38} = 0.466$ , residues 27 and 40 with  $P_{27,40} = 0.460$ , and residues 25 and 42 with  $P_{25,42} = 0.424$ . Thus, while the probability profiles are rather flat, indicative of an all-or-none type of transition to a  $\beta$ -barrel, this particular run provides unequivocal evidence that  $\beta$ -barrels having end strands that are two residues out of register as well as the fully in-register states are substantially populated. Analysis (not shown) of the mean energy vs "time" flow chart also shows the equilibrium between the fully in-register and the two residue out-of-register outer strands. The reason that the bends between strands 1-2 and 3-4 involve an equilibrium between the fully in-register and two residues out-of-register turns is due to the alternating hydrophobic/hydrophilic repeat pattern assumed in the primary sequence.

Summarizing the results of this section, we have considered the nature of the collapse transition in a homologous series of model proteins, in which the central residues are intrinsically indifferent to turn formation, i.e., for these residues the intrinsic probability of  $t$ ,  $g^-$ , and  $g^+$  states are equal. The "bend" neutral residues overwhelming form turns in the native state precisely because turn formation at that location does not cost any free energy. Shifting of the turn location moves hydrophobic residues into the turn region and thus costs free energy. Thus, due to tertiary interactions the central native-like turn dominates the population of states.

We have investigated the effect of increasing the length of the strands attached to the turn-neutral region. For sufficiently short polypeptides, a  $\beta$ -hairpin is formed at low temperature. Increasing the number of residues decreases the free energy of the native state, and thus increases the transition temperature. As one further increases the protein size, the predicted equilibrium between the  $\beta$ -hairpin and the four-member  $\beta$ -barrel is observed. Whereas the former transition from a random coil is continuous, the latter is all or none. The origin of the transition between the two structural motifs is as follows: The free energy cost of forming the additional two turns required by the  $\beta$ -barrel relative to the  $\beta$ -hairpins is a constant independent of protein size, and equals  $6\epsilon_g$  for the particular case under consideration. On the other hand, the number of nearest neighbor contacts in the  $\beta$ -barrel is double that of the  $\beta$ -hairpin. Thus, as  $n$  increases, the  $\beta$ -barrel should and indeed does dominate the population. We would eventually expect six-member (and larger)  $\beta$ -barrels as  $n$  is further increased.

These studies graphically point out that knowledge of short-range interactions alone is insufficient to specify the protein tertiary structure. These systems have the identical primary sequence pattern yet tertiary interactions dictate whether the system will fold to a  $\beta$ -hairpin or a  $\beta$ -barrel. Tertiary interactions also induce the formation of the essentially immobile central turn. Thus, these studies are highly suggestive that if one desires to predict the three-dimensional tertiary conformation in real proteins, knowledge of short-range interactions alone is unlikely to provide sufficient information.

While model A is highly suggestive of a globular protein, it still does not satisfy a major requirement. That is, the collapse is to a  $\beta$ -barrel whose central turn is always the desired native conformation, but the two turns between the central strand and the two exterior strands are not always native. The native structure does not differ by sufficient free energy from the slightly out-of-register conformers to ensure that it is the only state observed in the transition region.

### REQUIREMENTS FOR A UNIQUE FOUR-MEMBER $\beta$ -BARREL NATIVE STATE

Based on the above considerations, the primary sequence pattern  $\mathbf{B}_1(l_s)\mathbf{b}_1^0\mathbf{B}_2(l_s)\mathbf{b}_2^0\mathbf{B}_3(l_s + 1)\mathbf{b}_3^0\mathbf{B}_4(l_s + 1)$  with two additional neutral bend regions  $\mathbf{b}_1^0$  and  $\mathbf{b}_3^0$  involving residues  $l_s - 1$  to  $l_s + 1$  and  $3l_s$  to  $3l_s + 2$  should localize the turns between strands 1-2 and 3-4 to yield a desired unique "native" conformation. In what follows, we consider  $n = 46$  model proteins, and thus  $l_s = 11$ . The desired native turns involve residues 10-12, 21-23, and 33-35. A preliminary account of these models has been presented elsewhere.<sup>17</sup>

In what follows, the properties of the various models are summarized in Table II. We begin with the simplest possible case, model B: In the primary sequence of stretches  $\mathbf{B}_1$  and  $\mathbf{B}_3$  all the residues are hydrophobic, while for  $\mathbf{B}_2$

TABLE II  
Compilation of Properties of Various Models

Model	Primary Sequence <sup>a</sup>	$\epsilon_g^b$	$\epsilon_c^c$	$\epsilon_h^d$	$\epsilon_w^e$
A	$\mathbf{B}_1(2l_s)\mathbf{b}_1^0\mathbf{B}_2(2l_s + 2)$	$> 0$	$-\epsilon_g/2$	$-\epsilon_g/4$	$2\epsilon_g$
B	$\mathbf{B}_1(11)\mathbf{b}_1^0\mathbf{B}_2(11)\mathbf{b}_2^0\mathbf{B}_3(12)\mathbf{b}_3^0\mathbf{B}_4(12)$	$> 0$	$-\epsilon_g/2$	$-\epsilon_g/4$	$2\epsilon_g$
C	$\mathbf{B}_1(11)\mathbf{b}_1\mathbf{B}_2(11)\mathbf{b}_2\mathbf{B}_3(12)\mathbf{b}_3\mathbf{B}_4(12)$	$> 0$	$-\epsilon_g/2$	$-\epsilon_g/4$	$2\epsilon_g$
D	$\mathbf{B}_1(11)\mathbf{b}_1\mathbf{B}_2(11)\mathbf{b}_2\mathbf{B}_3(12)\mathbf{b}_3\mathbf{B}_4(12)^f$	$> 0$	$-\epsilon_g/2$	$-\epsilon_g/4$	$2\epsilon_g$
E	$\mathbf{B}_1(11)\mathbf{b}_1\mathbf{B}_2(11)\mathbf{b}_2\mathbf{B}_3(12)\mathbf{b}_3\mathbf{B}_4(12)$	$> 0$	0	$-2\epsilon_g$	$2\epsilon_g$

<sup>a</sup> $\mathbf{B}_i(k)$  indicates an odd/even hydrophobic/hydrophilic sequence of  $k$  residues.  $\mathbf{b}_i$  indicates that the two *gauche* states in the bonds at the junction between  $\mathbf{B}_i$  and  $\mathbf{B}_{i+1}$  are based on short-range interactions isoenergetic. See text for further details.  $\mathbf{b}_i^0$  indicates that all three conformations ( $t$ ,  $g^+$ , or  $g^-$ ) of each residue in the putative bend region are isoenergetic.

<sup>b</sup>The intrinsic energy of a  $g^+$  or  $g^-$  state relative to a  $t$  state.

<sup>c</sup>Conformational coupling parameter reflecting the enhanced stability of any pair of nearest neighbor nonbonded residues when both are in the *trans* state. See Fig. 1(B).

<sup>d</sup>Interaction free energy between a pair of nearest neighbor hydrophobic residues.

<sup>e</sup>Hydrophilic/hydrophobic or hydrophilic/hydrophilic nearest neighbor pair interaction free energy.

<sup>f</sup> $\epsilon_c$  for residues 36-45 is twice that in  $\mathbf{B}_1$ ,  $\mathbf{B}_2$ , and  $\mathbf{B}_3$ , and in addition, the energy of either *gauche* state equals  $-2\epsilon_g$  for residues 33-35 and  $2\epsilon_g$  for residues 36-38.

and  $B_4$  all the odd residues are hydrophobic. For all hydrophobic-type residues  $\epsilon_c = -\epsilon_g/2$ ,  $\epsilon_h = -\epsilon_g/4$ , and for the hydrophilic type residues,  $\epsilon_w = 2\epsilon_g$ . For the bend-type residues  $\epsilon_g = \epsilon_c = \epsilon_h = \epsilon_w = 0$ . Again there is no intrinsic preference for the bend conformation. In addition to the native turn conformation,  $g^+g^-g^+$ , the other 26 nonnative conformers are equally likely.

In Fig. 11(A) we plot the average of  $\langle S^2 \rangle$  vs  $T^*$  obtained from three cooling and three heating sequences in the solid squares and open squares, respectively. Due to the finite length of the runs, the denaturation occurs at a slightly higher temperature than renaturation. While the properties of the native and denatured state conformations are rather well described, due to the finite number of transitions between the D and N states, the equilibrium populations are not. This is the origin of the spike seen in the heating curve. Substantially better characterization of the equilibrium constant is not practical using a dynamic MC algorithm, and alternative approaches are currently being examined. Over the course of both heating and cooling sequences, the collapse from a denatured conformation to the desired unique four-member  $\beta$ -barrel structure was seen a total of 19 times.

Just as in previous cases, the transition is all or none. This is verified in Fig. 11(B), where we plot for a single run the mean energy vs time for  $T^* = 1.149$ , a temperature in the transition region. Each time step represents 25,000 MC cycles. However, unlike the previous cases,<sup>14-16</sup> now the conformational transition is to the unique native state and all the desired features of globular protein folding are reproduced.

In the curve denoted by the solid squares of Fig. 12, we plot the fraction of *trans* states obtained as an average over the three cooling cycles,  $f_t$  vs  $T^*$ .  $f_t$  prior to collapse is 0.531 and in the native state  $f_t = 0.786$ . From Table I,  $f_t$  of the native conformation is 0.791. Once again tertiary interactions are seen to induce the formation of secondary structure.

In Fig. 13, the curve denoted by the solid squares presents a plot of the mean energy per residue,  $E_b$ , vs  $T^*$ , for Model B obtained as an average over the three cooling cycles. For the particular set of parameters, the energy of the pure native state is  $-36 \epsilon_g/k_B T$ , which per residue corresponds to  $-0.7826 \epsilon_g/k_B T$ . In the vicinity of renaturation,  $E_b$  decreases steeply before it approaches the asymptotic value of the pure native state, indicated by the dashed curve.

In Fig. 14, the curves denoted by the open circles and solid squares, respectively, we compare the  $\langle S^2 \rangle$  vs  $T^*$  for the system having one statistical turn-neutral region (model A) with the present case having three statistical turn neutral regions (model B). The dimensions are comparable in the denatured state, but the transition temperature in the latter case appears to be slightly higher than in the former and is somewhat steeper. This presumably reflects the lower free energy of the native state in the case of model B, where no energetic penalty is paid for formation of the tight turns between strands 1 and 2 and strands 3 and 4, respectively.

Having established the necessary conditions for the formation of the crystallographically unique four-member  $\beta$ -barrel, we next address those features that might serve to enhance the range of thermal stability of the native state. Two things immediately come to mind. One can enhance the local statistical preference for turn conformations, and one can augment the cooperativity

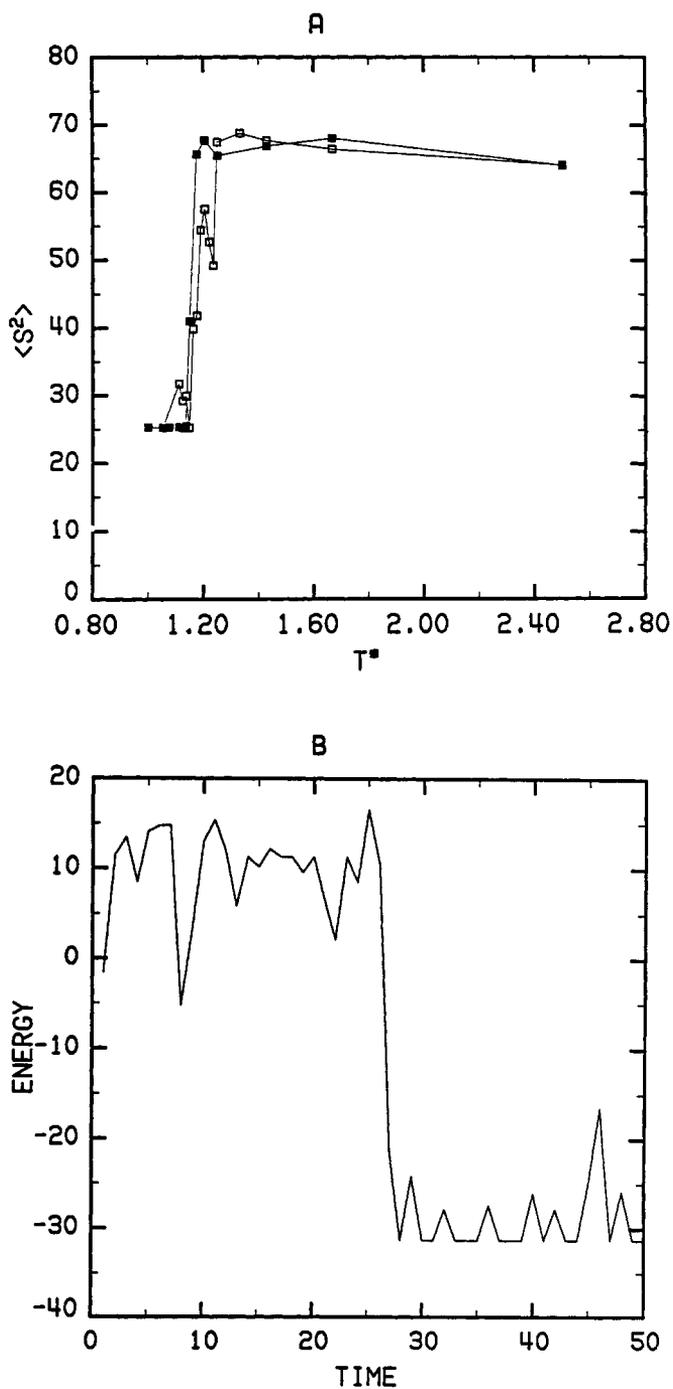


Fig. 11. (A) For model B, plot of the mean square radius of gyration  $\langle S^2 \rangle$  vs  $T^*$  for a  $n = 46$  model protein, obtained as an average over three cooling and three heating sequences in the curves indicated by the solid and open squares, respectively. (B) Plot for model B of the energy vs "time" for a single run at  $T^* = 1.149$ . Each time step represents 25,000 MC cycles.

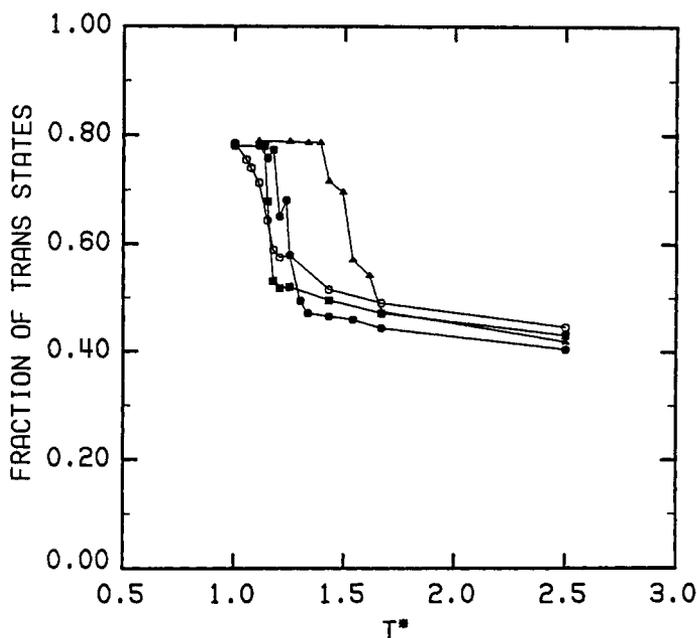


Fig. 12. Plot of the fraction of *trans* states vs  $T^*$  for model A (open circles), B (solid squares), C (solid circles), and D (solid triangles).

parameter that accounts for the stabilization of  $\beta$ -sheets when any nonbonded pair of residues are in a *trans* conformation. We consider the effect of each in turn.

Model C has the primary sequence pattern  $B_1(11)b_1B_2(11)b_2B_3(12)b_3B_4(12)$  of model B with the same magnitude of the hydrophobic, hydrophilic, and cooperative interactions. However, here the energy of any *gauche* conformation is set equal to  $-2\epsilon_g$  for those residues in the putative bend regions, i.e., for residues 10–12, 21–23, and 33–35. Again, we emphasize that in addition to the native-like turn conformation  $g^+g^-g^+$  there are seven other triplets of *gauche* states that on the basis of their short-range energetic preference are equally likely. What distinguishes the native  $g^+g^-g^+$  triplet from these others is that this conformation permits favorable tertiary interactions between the strands to occur, whereas other conformations do not.

A series of three cooling and two heating cycles for this model were run during which the unique four-member  $\beta$ -barrel native state was obtained 17 times from the denatured state. As in the turn-neutral case of model B, no non native collapsed structures were seen. In the curves denoted by the closed circles of Figs. 12–14, we plot  $f_t$ ,  $E_b$ , and  $\langle S^2 \rangle$ , respectively, vs  $T^*$  obtained from the cooling cycles. Relative to model B, the transition has shifted to higher temperatures. This is not surprising as the reduced energy of the pure native state,  $E_N(\beta_4)$ , now equals  $-53\epsilon_g/k_B T$ , whereas for model B it equals  $-36\epsilon_g/k_B T$ . The energy per residue of the native state equals  $-1.152\epsilon_g/k_B T$  and is indicated as a dashed line in Fig. 13. The most striking difference is seen in Fig. 14. The decrease in  $\langle S^2 \rangle$  for the denatured state relative to models A and B reflects the local preference for *gauche* conforma-

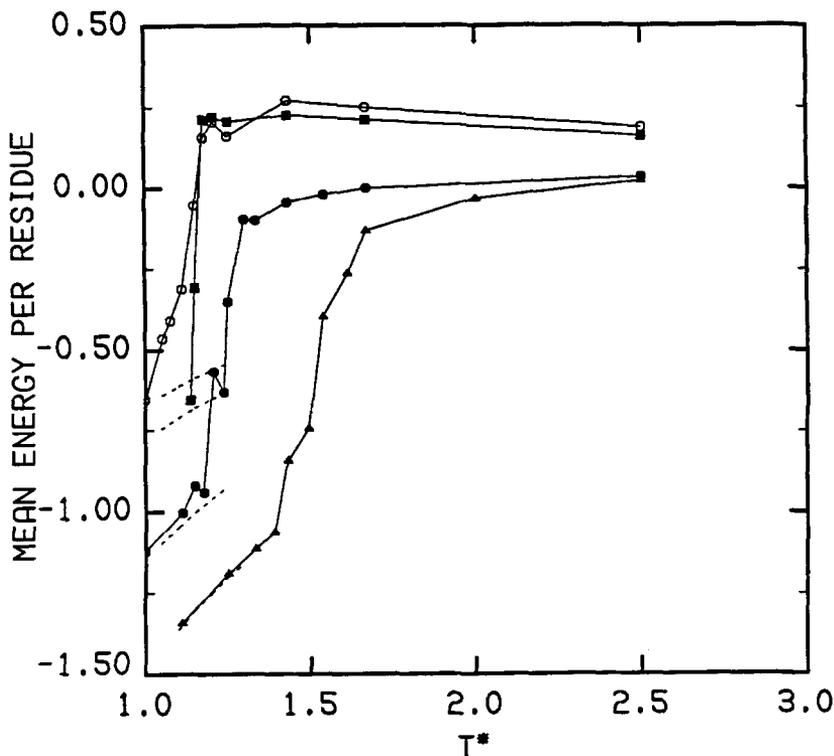


Fig. 13. Comparison of the mean energy per bond  $E_b$  vs  $T^*$  for model A (open circles), B (solid squares), C (solid circles), and D (solid triangles).

tions of the nine bend residues. From Fig. 12, one sees that  $f_i$  of the denatured state is 0.47 while  $f_i$  of the native state is 0.78. Thus in many respects, this is a rather faithful model of a globular protein.

Examination of the mean energy vs "time" curve once again indicates that the collapse is all or none. As these curves display identical qualitative behavior to those already displayed here, in the interest of brevity these are not shown. Finally, as in model B, the transition midpoint on heating from the native state is shifted by about 0.03 reduced temperature units to higher temperature from the cooling curve. Once again, rather small hysteresis effects are evident, and these should be eliminated in the limit of very long runs.

We next examine the effect of regional variations in tertiary interactions on the character of the  $\beta$ -barrel conformational transition. A particularly simple case is given in model D. Once again the primary sequence pattern is  $B_1(11)b_1B_2(11)b_2B_3(12)b_3B_4(12)$ . The stabilities of regions  $B_1$ ,  $B_2$ , and  $B_3$  and  $b_1$  and  $b_2$  are identical to those of model C. However, the cooperativity parameter  $\epsilon_c$  of residues 36-45 is twice that of regions  $B_1$ ,  $B_2$ ,  $B_3$ ; i.e., we set  $\epsilon_c$  for all the residues in strand 4 equal to  $-\epsilon_g$ . Moreover, the statistical preference for forming a turn at residues 33-35 is now no longer merely localized to the turn itself (region  $b_3$ ). In addition to setting the *gauche* energy equal to  $-2 \epsilon_g$  for residues 33-35, residues 36-38 have a strong

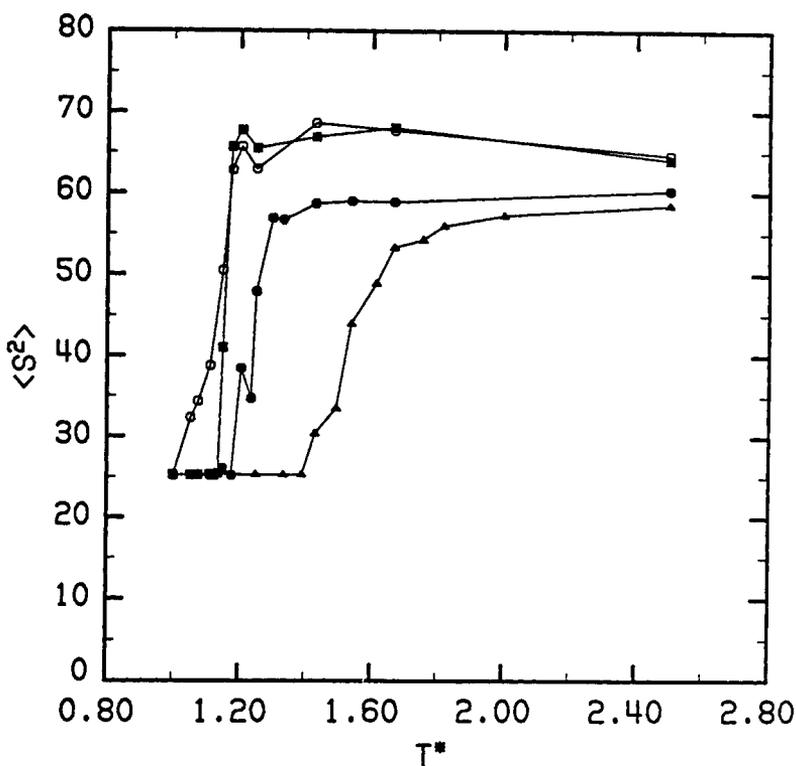


Fig. 14. Comparison of  $\langle S^2 \rangle$  vs  $T^*$  for model A (open circles), B (solid squares), C (solid circles), and D (solid triangles).

intrinsic preference for *trans* states, with the energy of either *gauche* state now equal to  $2\epsilon_g$ . Consequently, there is a very strong statistical preference to form a turn even in the denatured state. Some recent elegant nmr work by Wright and co-workers on short linear peptides provides evidence for  $\beta$ -turn populations in the denatured state, and support the plausibility of a strong turn preference<sup>38,39</sup> (even though, as seen above, turn-neutral states are sufficient to produce the desired results).

In the curves denoted by the closed triangles in Figs. 12–14, we plot  $f_t$ ,  $E_b$ , and  $\langle S^2 \rangle$ , respectively, vs  $T^*$  for model D obtained as an average over eight cooling runs. Not surprisingly, the addition of an extra sticky fourth strand substantially raises the transition temperature. As shown in Fig. 12, the  $f_t$  curve at high temperature now lies above that of model C, due to the preference of residues 36–38 to be in the *trans* state. Similarly, at high temperature the  $f_t$  curve of model B lies above both models C and D as *gauche* states in the bend regions are not favored, and  $f_t$  of model A lies above models B–D as there is only one turn neutral region, with the other two bend regions now favoring *trans* states. For model D,  $f_t$  increases from about 0.46 in the random coil state to 0.78 in the native conformation. As shown in Fig. 14, the inclusion of an even stronger tendency for *trans* states adjacent to bend 3 further depresses the equilibrium dimensions of the random coil states.

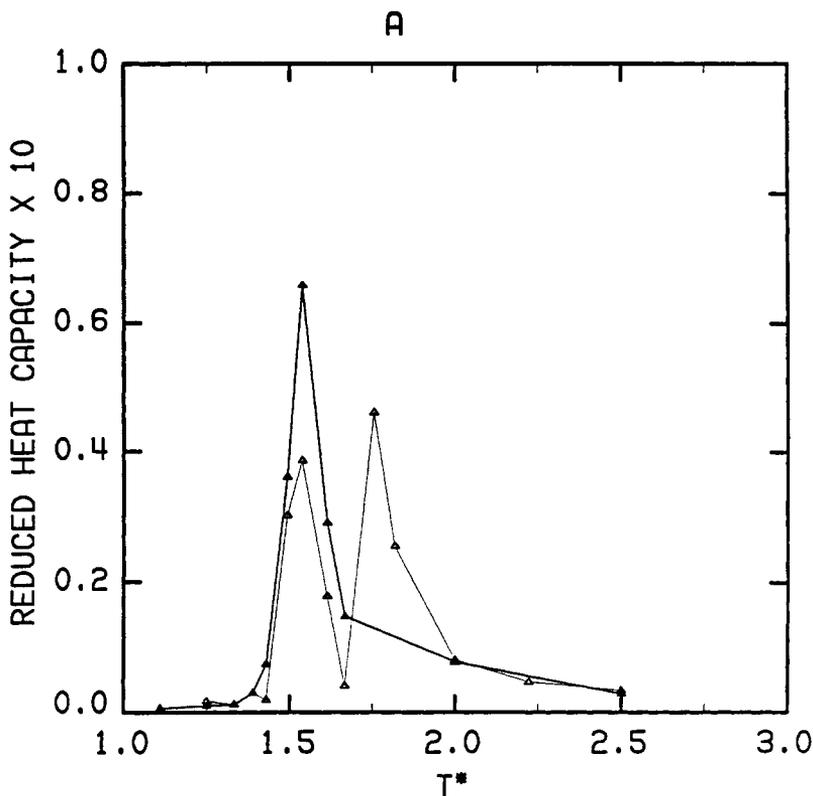


Fig. 15. (A) For model D, reduced heat capacity vs  $T^*$  obtained as an average over eight cooling (solid triangles) and three heating sequences (open triangles). (B) For model D,  $\langle S^2 \rangle$  vs  $T^*$  obtained as an average over eight cooling (solid triangles) and three heating (open triangles) sequences.

Moreover, the energy of the pure native state is  $-69\epsilon_g/k_B T$ , which translates into  $-1.500\epsilon_g/k_B T$  per residue. The energy of the pure native state is shown as a dashed line in Fig. 13.

In Fig. 15(A), we present for model D a plot of reduced heat capacity divided by the square of the number of residues,  $C_v$ , vs  $T^*$  obtained as an average over eight cooling (solid triangles) and three heating sequences (open triangles), respectively. In this particular case, unlike those examined previously, substantial hysteresis is seen. This is perhaps most graphically displayed in Fig. 15(B), where  $\langle S^2 \rangle$  vs  $T^*$  is plotted for the cooling (solid triangles) and heating (open triangles) sequences, respectively. The shift in the transition midpoint is about 0.15 in reduced temperature units and is due to the fact that the native conformation resides in a very deep local free energy minimum.

Until this point, we have examined models having a nonzero cooperativity parameter  $\epsilon_c$  between nonbonded nearest neighbor *trans* states. Physically  $\epsilon_c$  is a preaveraged parameter that implicitly assumes that an odd/even hydrophobic/hydrophilic pattern is present in the primary sequence. It allows for explicit conformational coupling between residues up to second nearest

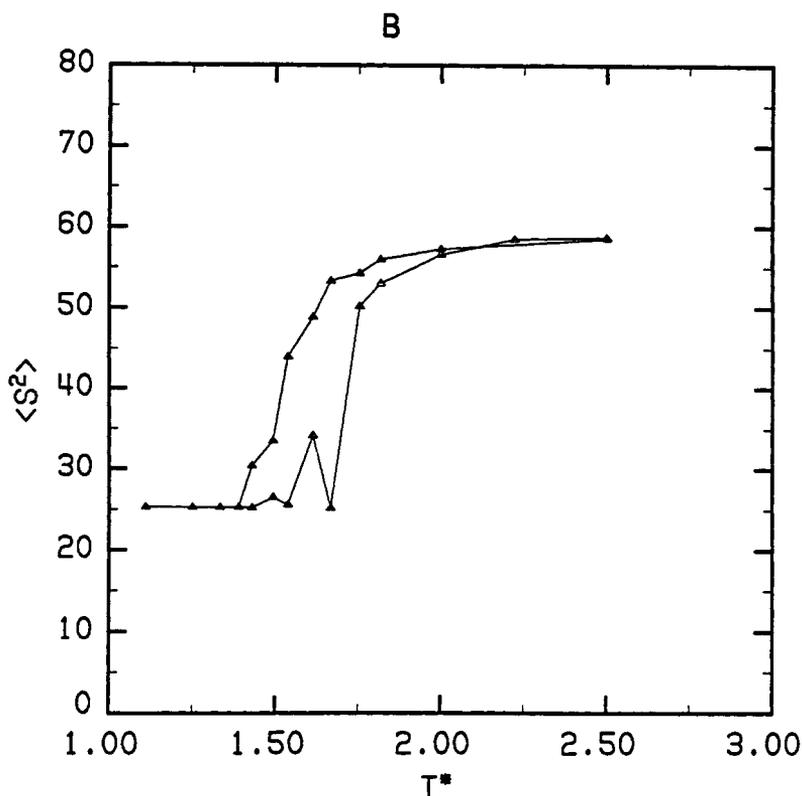


Fig. 15. (Continued from the previous page.)

neighbor down the chain. As a practical matter, it was employed because we originally believed such cooperativity would aid in the search for the native state. To investigate whether  $\epsilon_c$  is necessary, model E possesses the primary sequence  $\mathbf{B}_1(11)\mathbf{b}_1\mathbf{B}_2(11)\mathbf{b}_2\mathbf{B}_3(12)\mathbf{b}_3\mathbf{B}_4(12)$  with residues 11–12, 21–23, and 23–35 having a *gauche* energy  $-2\epsilon_g$ , and  $\epsilon_w = \epsilon_h = \epsilon_c = 0$ . All the residues in  $\mathbf{B}_1$  and  $\mathbf{B}_3$  and all the odd residues in  $\mathbf{B}_2$  and  $\mathbf{B}_4$  are of the hydrophobic type with  $\epsilon_h = -2\epsilon_g$  and  $\epsilon_c = 0$ , and all the even residues in  $\mathbf{B}_2$  and  $\mathbf{B}_4$  are of the hydrophilic type with  $\epsilon_w = 2\epsilon_g$  and  $\epsilon_c = 0$ . That is, we consider the identical pattern as in model C, but where the cooperativity parameter is turned off. For this particular case, the reduced energy of the native conformation equals  $-20\epsilon_h$ . Once again collapse is to the same unique four-member  $\beta$ -barrel. If  $\epsilon_h$  is set equal to  $-\epsilon_g$  this merely changes the location of the transition region, but leaves all the qualitative behavior the same as above. This demonstrates that a cooperativity parameter of the  $\epsilon_c$  type is not required; rather, one need merely consider the identity of the interacting pair of residues. Coupled with the presence of statistical turn-forming regions, this is sufficient to produce an all-or-none transition to a unique native conformation. Finally, this is conceptually satisfying in that  $\epsilon_c$  might be very hard to obtain experimentally, whereas  $\epsilon_h$  and  $\epsilon_w$ , being more local parameters, might be extractable from empirically measured hydrophobicity scales.<sup>49,50</sup>

## DISCUSSION

In the context of a simplified model of  $\beta$ -proteins, we have examined a number of essential features of the equilibrium globular protein folding transition. The particular general points addressed were as follows: (1) Is knowledge of the short-range interactions alone sufficient to predict the native state? (2) What makes the globular protein conformational transition all or none? (3) What are the minimum requirements for formation of a unique native state? (4) How important are site-specific interactions in determining the structure of the unique native state? The results from our simulations are summarized below.

First of all, we addressed the interplay of protein size and topology, and considered a series of models in which the primary sequence consisted of the classic  $\beta$ -protein pattern of alternating odd/even hydrophobic/hydrophilic residues plus a statistical turn-neutral region. By extending the length of the two hydrophobic/hydrophilic tails, a spontaneous transition from a  $\beta$ -hairpin to a four-member  $\beta$ -barrel was observed to occur. All the hydrophobic and hydrophilic residues were taken to be identical; i.e., the magnitude of the interaction is the same between any pair of residues of a given type. The presence of residues that are intrinsically neutral to turn formation is sufficient to localize the central turn. The size of the model protein at which the transition from a  $\beta$ -hairpin to a  $\beta$ -barrel occurs is in agreement with the predictions based on equating the energy of two motifs. This implies that knowledge of short-range interactions alone is insufficient to predict the tertiary structure. Both the hairpin and the  $\beta$ -barrel have the same intrinsic preference for  $\beta$ -sheet formation and the same down chain local set of interactions. What causes a transition between the two is, as expected, global free energy considerations. Namely, the presence of two additional turns in the  $\beta$ -barrel costs a fixed amount of free energy independent of protein size; on the other hand, the  $\beta$ -barrel has double the number of favorable hydrophobic contacts. Thus, as  $n$  increases, the lowest free energy structure shifts from the  $\beta$ -hairpin to the four-member  $\beta$ -barrel.

Due to the relatively small free energy difference between the native, fully in-register,  $\beta$ -hairpin and those conformations involving partial denaturation of the ends, the denatured to native state conformational transition is seen to be continuous and has much in common with single-chain helical polypeptides.<sup>46-48</sup> Even though the residues in the center of the molecule are neutral to turn formation, all collapsed conformations have the turn involving these central residues. Basically, singly out-of-register hairpin conformations are prohibited because this would result in repulsive hydrophobic-hydrophilic contacts. Further shift of registration by two residues now places the hydrophobic residues in the turn region and thereby raises the free energy. Hence out-of-register  $\beta$ -hairpins are eliminated. Thus, the three residues in the center of the molecule, while intrinsically neutral to turn formation, will form the central turn of the  $\beta$ -hairpin because of tertiary interactions.

Turning now to the cases having the same primary sequence pattern but of sufficient length that the  $\beta$ -barrel is favored, the nature of conformational transition is entirely different. It is no longer continuous; rather, it is all or none. Basically due to loop entropy, the only substantial fraying occurs at the

ends of the chain, but because each end now has two nearest neighbors, such fluctuations are greatly reduced. The central turn in the native state always involves the bend-neutral region, but the two turns involving the exterior strands occur at residues that produce both in-register external strands as well as two-residue out-of-register external strands. Thus, while the conformational transition is of an all-or-none character, it is not to a unique native state. The sought after faithful globular protein model can be simply achieved by introducing into the primary sequence two additional neutral turn regions. From the considerations outlined above, an all-or-none transition to the same unique lowest free energy native structure is observed. This points out that site-specific interactions are not required to produce a unique native conformation.

What is required for the formation of the unique four-member  $\beta$ -barrel is an odd/even hydrophobic/hydrophilic pattern plus the presence of statistical turn-neutral regions in the amino acid sequence. Increasing the statistical preference for any turn conformation and/or augmenting the hydrophobic interaction serves to increase the range of thermal stability of the model protein. Recent nmr measurements by Wright and co-workers on peptide fragments has found evidence of native-like  $\beta$ -turns in the denatured state, in qualitative agreement with what appears to be required here.<sup>38,39</sup> In fact, the measured turn populations are larger than those obtained from model B, which indicates that a turn-neutral region is all that is necessary and sufficient to obtain a unique native state.

The present schematic models strongly suggest that the rules of protein folding are in fact rather robust. Based on the folding simulations of four-member  $\beta$ -barrels reported here, as well as folding simulations on the  $\beta$ -barrel Greek key topology<sup>36</sup> and on the requirements for the folding of a four-helix bundle to be described elsewhere,<sup>51,52</sup> these seem to be rather simple. That this viewpoint is reasonable is suggested by the fact that homologous series of proteins possess the same topology but not the same amino acid sequence.<sup>4,13</sup> This is not to indicate that there are no regions where amino acid substitution could have drastic effects. In fact, in the present case elimination of the central turn neutral region would provide a whole manifold of nonnative  $\beta$ -barrel-like structures. In this sense certain regions (e.g., essential turns) are sensitive to the types of amino acids located there. Overall, though, site-specific interactions involving minor differences in side-chain hydrophobicity and size are likely to result in local fine-tuning about the same basic backbone topology.

In summary, we have for the first time in the present series of simulations identified systems that are rather close mimics of  $\beta$ -barrel globular proteins when viewed at low resolution. However, due to lattice restrictions, a number of the finer details of globular protein structure cannot be reproduced. Further refinements of the methodology are required, and these are currently underway. However, even given all that, the topology and thermodynamics of the transition seen here appear to be in the same class as that of real globular proteins. Overall, these simple models seem to be very useful guides for understanding the nature of the processes involved in the globular protein conformational transition.

This research was supported in part by NIH grant GM-37408, from the Division of General Medical Sciences, United States Public Health Service, by BRSG S07 RR07054-22 awarded by the Biomedical Research Support Program, Division of Research Resources, National Institutes of Health, and by Multiflow Computer, Inc., for an instrumental grant enabling the purchase of a Multiflow Trace on which these calculations were partially carried out. The assistance of the Washington University Business School Computing Facility is also gratefully acknowledged.

### References

1. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958) *Nature* (London) **81**, 662-666.
2. Afinsen, C. B. (1973) *Science* **181**, 223-230.
3. Jaenicke, R., Ed. (1980) "Protein Folding, Proceedings of the 28th Conference of the German Society," Elsevier/North Holland Biochemical Press, Amsterdam.
4. Ptitsyn, O. B. & Finkelstein, V. A. (1980) *Quart. Rev. Biophys.* **13**, 339-386.
5. Ghelis, C. & Yon, J. (1982) *Protein Folding*, Academic Press, Orlando, FL.
6. Go, N. (1983) *Ann. Rev. Biophys. Bioeng.* **12**, 183-210.
7. Wetlaufer, D., Ed. (1984) *The Protein Folding Problem*, Westview, Boulder, CO.
8. Creighton, T. E. (1985) *J. Phys. Chem.* **89**, 2452-2459.
9. Lewis, P. N., Gō, N., Gō, M., Kotelchuck, D. & Scheraga, H. A. (1970) *Proc. Natl. Acad. Sci. USA* **67**, 810-815.
10. Fasman, G. (1987) *Biopolymers* **26**, S59-S79.
11. Rose, G. D., Winters, R. H. & Wetlaufer, D. B. (1976) *FEBS Lett.* **63**, 10-16.
12. Skolnick, J. (1986) *Macromolecules* **19**, 1153-1166.
13. Bashford, D., Chothia, C., & Lesk, A. M. (1987) *J. Mol. Biol.* **196**, 199-216.
14. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7267-7271.
15. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *J. Chem. Phys.* **85**, 3585-3597.
16. Kolinski, A., Skolnick, J. & Yaris, R. (1987) *Biopolymers* **26**, 937-962.
17. Skolnick, J., Kolinski, A. & Yaris, R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5057-5061.
18. McCammon, J. A. (1984) *Rep. Prog. Phys.* **47**, 1-46.
19. Karplus, M. & McCammon, J. A. (1981) *CRC Crit. Rev. Biochem.* **9**, 293-315.
20. McCammon, J. A. & Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge.
21. McCammon, J. A. (1987) *Science* **238**, 486-491.
22. Vaáquez, M. & Scheraga, H. A. (1985) *Biopolymers* **24**, 1437-1447.
23. Li, Z. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6611-6615.
24. Schellman, J. A. (1952) *CR Trav. Lab. Carlsberg Ser. Chem.* **29**, 230-259.
25. Tanford, C. (1964) *J. Am. Chem. Soc.* **4**, 4240-4247.
26. Brandts, J. & Lumry, R. (1963) *J. Phys. Chem.* **67**, 1484-1494.
27. Wetlaufer, D. B., Malik, S. K., Stoller, L. & Coffin, R. L. (1964) *J. Am. Chem. Soc.* **86**, 508-514.
28. Tanford, C. (1968) *Adv. Protein Chem.* **23**, 121-282.
29. Privalov, P. L. (1979) *Adv. Protein. Chem.* **33**, 167-241.
30. Ansari, A., Berendzen, J., Bowne, S. F., Frauenfelder, H., Iben, I. E. T., Sauke, T. B., Shyamsunder, E. & Young, R. D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5000-5004.
31. Elber, R. & Karplus, M. (1987) *Science* **235**, 318-321.
32. Binder, K. (1986) *Application of the Monte Carlo Method in Statistical Physics*, Springer, Berlin, pp. 1-45.
33. Lifshitz, I. M., Grosberg, A. Y. & Khoklov, A. R. (1978) *Rev. Mod. Phys.* **50**, 683-695.
34. Privalov, P. L. (1982) *Adv. Protein Chem.* **35**, 1-104.
35. Chothia, C. (1975) *Nature* (London) **254**, 304-308.
36. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167-339.
37. Rose, G. D., Winters, R. H. & Wetlaufer, D. B. (1976) *FEBS Lett.* **63**, 10-16.
38. Wright, P. E., Dyson, H. J., Rance, M., Houghten, R. A. & Lerner, R. A. (1987) *Protides Biol. Fluids* **35**, 477-480.
39. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1988) *J. Mol. Biol.* **201**, 161-200, 201-217.
40. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, Wiley, New York, chap. 7.

41. Byron, F. W. & Fuller, R. W. (1970) *Mathematics of Classical and Quantum Physics, Vol. II*, Addison-Wesley, Reading, MA, pp. 388–389.
42. Kremer, K., Baumgartner, A. & Binder, K. (1981) *J. Phys.* **A15**, 2879–2883.
43. Binder, K., Ed. (1984) *Application of the Monte Carlo Method in Statistical Physics*, Springer, Berlin, chap. 5.
44. Wall, F. T. & Mandell, F. (1975) *J. Chem. Phys.* **63**, 4592–4595.
45. Dill, K. A. (1985) *Biochemistry* **24**, 1501–1509.
46. Zimm, B. & Bragg, J. (1959) *J. Chem. Phys.* **31**, 526–535.
47. Poland, D. C. & Scheraga, H. A. (1965) *Biopolymers* **3**, 305–313.
48. Skolnick, J. (1985) *Macromolecules* **18**, 1073–1083.
49. Tanford, C. (1980) *The Hydrophobic Effect*, Wiley, New York.
50. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985) *Science* **229**, 834–838.
51. Sikorski, A. & Skolnick, J. (1989) *Biopolymers*, following paper.
52. Skolnick, J., Kolinski, A. & Yaris, R. (1989) *Proc. Nat. Acad. Sci. USA* (in press).

Received April 18, 1988

Accepted August 8, 1988