

Monte Carlo simulations of the folding of β -barrel globular proteins

(protein conformational transition/protein evolution/all-or-none transition/unique native state)

JEFFREY SKOLNICK^{†‡}, ANDRZEJ KOLINSKI[§], AND ROBERT YARIS[†]

[†]Institute of Macromolecular Chemistry, Department of Chemistry, Washington University, Saint Louis, MO 63130; and [§]Department of Chemistry, University of Warsaw, 02-093 Warsaw, Poland

Communicated by S. I. Weissman, March 11, 1988

ABSTRACT With the use of dynamic Monte Carlo simulations, the necessary conditions for the collapse from a random-coil denatured state to a structurally unique four-member β -barrel native state of a model globular protein have been investigated. These systems are free to roam through all of configuration space—both native and nonnative interactions are allowed. The relative importance of hydrophobic and hydrophilic interactions and the presence or absence of statistical bend-forming regions for the formation of a unique native state are examined, and the conditions necessary for a denatured-to-native (and vice versa) conformational transition that is thermodynamically all-or-none and which always results in collapse to the same, four-member β -barrel are explored. These conditions are found to be a general pattern of hydrophobic/hydrophilic residues that allows the native state to differentiate the interior from the exterior of the protein and the presence of regions that are, at the very least, neutral toward turn formation. The former set of interactions seems to define the mean length of the β -stretch, and the latter set serves to lock the native state into the lowest free energy state, the native conformation. These folding simulations strongly suggest that the general rules of protein folding are rather robust and that site-specific tertiary interactions are only involved in structural fine tuning. The conditions required for the formation of a structurally unique native state from a manifold of collapsed conformations that are originally quite close in energy is highly suggestive of a mechanism of protein evolution by means of random mutations. The implications of these folding studies for such a mechanism are qualitatively explored.

Although the equilibrium conformational transition of globular proteins has long been studied, many fundamental questions about the process remain unanswered (1–7). Is there any native-like secondary structure (α -helix, β -pleated sheet, native-like turns) present in the random-coil denatured state (D) (8)? What is the role of short- vs. long-range interactions in determining the stability of the native state (9–12)? What is the role played by site-specific interactions in determining the uniqueness of the native conformation? Does the general pattern of hydrophobic and hydrophilic residues determine the overall topology, with site-specific interactions “fine tuning” the native state, or must such interactions be explicitly enumerated, with the overall tertiary structure being extremely sensitive to local details? What is the role of the loops and bends (9, 11, 12)? Finally, for small globular proteins, what makes the conformational transition all-or-none, even though the free energy of stabilization per residue is small (7)?

Basically, two complementary theoretical approaches have emerged to attack the protein-folding problem. In the first, one attempts to use a very realistic potential energy

surface, in many cases including the solvent explicitly (13, 14). One then stimulates (typically using the technique of molecular or Brownian dynamics) the equilibrium and dynamic properties of the protein. Unfortunately, the characteristic time scale of these simulations is on the order of tens to hundreds of picoseconds, and they typically require the crystal structure of the folded state to serve as the initial configuration. Similarly, hybrid-folding approaches based on local energy minimization are intrinsically limited to the very small polypeptides (15). The complete folding or unfolding of a protein would require computers at least a million times faster than those that are presently available. Moreover, even if it were possible to fold a protein with all the myriad of interactions included, one would be no closer to elucidating the general rules of protein folding—i.e., given a very complicated many-body potential, dissecting out the essential elements that direct the folding from those that merely contribute to the structural fine tuning is difficult.

Thus, we are attempting to find a minimal set of interactions that faithfully reproduce in a qualitative sense the globular protein conformational transition (16–18). In particular, the model protein must be able to freely hunt through all of configuration space to find the native state (not specified in advance) that it prefers. Interactions between all residues are allowed. For the small model proteins discussed below, the conformational transition must be all-or-none. A given molecule must either be a member of the denatured state population or it must fold to a structurally unique native state; of course, small structural fluctuations in the latter occur, as they do in real proteins.

Clearly, for the above approach to be viable, algorithms that surmount the multiple minima problem must be used. We use a Monte Carlo (MC) technique that is efficient at sampling the broad expanse of the denatured state as well as the rather steep confines of the native state. In a previous series of publications (16–18) we demonstrated for β -barrel-like proteins that the folding to a thermodynamically unique native state requires a small amount of secondary structure in the denatured state. This acts to prepartition configuration space and guide the folding process. Moreover, in these models all interactions are the same for all the residues. While the resulting “homopolymeric” β -barrel model protein exhibited a number of qualitative aspects of globular protein folding, it did not fold to a structurally unique state; rather, a spectrum of folded structures resulted.

How then can the basic model be improved to more accurately model the globular protein-folding processes? The introduction into the primary sequence of regions that have a statistical tendency to form turns is one possibility (9, 12, 19–21). These regions will surely serve as nucleation sites of protein folding and should help the system find the native state. Once it is found, they would tend to reduce the global

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: D, random-coil denatured state; N, four-member β -barrel native state; MC, Monte Carlo; f_i , fraction of β -sheet that is in planar trans state; g , gauche state; k , no. of consecutive residues. [‡]To whom reprint requests should be addressed.

fluctuations (e.g., shifts in registration of the residues). If a tight bend forms in a region that is not particularly hydrophobic, then the shift of hydrophobic residues previously found in the interior of the β -sheet into the turn region would result in a free energy increase. Thus, bends could act to reduce conformational fluctuations from the native state. However, the presence of loops or bends cannot in and of itself be the whole story. Consider, for example, the formation of a β -hairpin (22). There are two mirror image isomers—one exposes the hydrophobic-type residues to the solvent and one buries them in the interior. Clearly, then, the presence of hydrophobic and hydrophilic type residues must be introduced. This is done in a very naive way described below.

In the present paper, we investigate the folding transition of a hypothetical four-member β -barrel protein, which is a major structural element in the Greek key and chymotrypsin-like motifs of β -barrel proteins (22). Moreover, such an idealized model turns out to provide a qualitative guide to what we believe are a number of essential features of globular protein folding.

The Model

Consider a consecutive sequence of 46 beads (each of which represents an amino acid residue) confined to a tetrahedral lattice. Thus, we use an α -carbon representation of a globular protein whose virtual bond angles of 105° have been slightly distorted to fit onto the tetrahedral lattice. To implement the effect of excluded volume, the multiple occupancy of any lattice site is prohibited. Each of the 43 interior bonds has three accessible rotational states—the lowest energy, planar trans state (t), or either of the two out-of-plane gauche (g^+ or g^-) states. The energy of a gauche state relative to a trans state is specified by ϵ_g . In what follows, the reduced temperature of the system $T^* = k_B T / \epsilon_g$ with k_B Boltzmann's constant and T the absolute temperature.

We next introduce hydrophobic and hydrophilic interactions. Consider a pair of nonbonded nearest-neighbor beads. If both beads are hydrophobic, then ϵ_h (negative) is the attractive interaction energy between the two beads and represents the hydrophobic interaction. (ϵ_h might also mimic the attractive interaction due to formation of a salt bridge.) Similarly, if both residues are hydrophilic, or if one is hydrophobic and the other hydrophilic, then their effective repulsive interaction is modeled by a positive ϵ_w . Hence ϵ_h and ϵ_w are effective potentials of mean force that qualitatively model the well-known hydrophobic effect. In what follows, all hydrophobic (hydrophilic)-type interactions are taken to have the same ϵ_h (ϵ_w). Finally, we introduce a cooperativity parameter, ϵ_c , which allows for the explicit conformational coupling between nearest-neighbor, nonbonded trans states. ϵ_c is a preaveraged conformational-interaction parameter. This parameter is used both to aid the conformational search and to include a pattern of cooperativity that is, in principle, longer ranged than between nearest neighbors down the chain. In this case, it extends to second-nearest neighbors. Qualitatively identical results to those described below have been found on setting $\epsilon_c = 0$. A more detailed discussion of ϵ_c is found in ref. 18 (see in particular figure 1 of ref. 18).

As mentioned above, a successful algorithm must be able to efficiently sample configuration space for both the denatured and native states. We employ a dynamic MC (18, 23, 24) sampling technique based on an asymmetric Metropolis scheme. A given MC sampling cycle consists of the following steps, each of which is performed on randomly chosen beads: (i) three-bond kink motion consisting of the interchange $g^\pm \rightarrow g^\mp$; (ii) four-bond kink motion involving $g^+g^- \rightarrow g^-g^+$ or vice versa; (iii) four-bond wave motion in which a given randomly chosen four-bond stretch having the conformation g^+g^- or g^-g^+ is interchanged with a randomly chosen

two-bond stretch located somewhere else in the molecule; (iv) five-bond wave motion in which a conformation consisting of five of the six bonds that form a closed cyclohexane-like ring are interchanged with a single bond that has been clipped out from somewhere else in the chain (in steps iii and iv, the beads are renumbered so that the sequence of residues remains the same); (v) end-bond rotations in which an end is randomly chosen and the rotation of one or two end bonds is attempted.

For every set of parameters, at least three independent cooling and heating sequences of at least 3×10^6 MC cycles at each temperature were run. Great care was taken to ensure equilibrated systems. However, in the transition region, the number of jumps from the N \rightarrow D state (Fig. 1) or vice versa is small (typically on the order of 10–20). Thus, while each of the N and D states are well characterized, the relative equilibrium populations of the two states are not. Better computation of the equilibrium fraction of states in the transition region is presently beyond our computational capabilities.

The following conventions specify the primary sequence. $B_i(k)$ specifies a stretch consisting of k consecutive residues. The odd elements are hydrophobic-type residues. The even elements are either hydrophilic- or hydrophobic-type residues; these will be specified. Note the given stretch need not necessarily form a β -sheet. All the ϵ_g are the same for all the beads. Regions that have a tendency to form tight bends are denoted b_i and consist of the last two beads of region B_i and the first bead of region B_{i+1} . If b_i is unspecified, then the interaction pattern of the turn is that of B_i and B_{i+1} —i.e., turns are not preferred. Otherwise, the energy of the three conformational states in the putative bend region relative to the trans state must be specified. For simplicity, for the bends, we set ϵ_c , ϵ_h , and ϵ_w equal to zero, but ϵ_g need not necessarily be zero. Finally, although we report results generated from a particular choice of parameters, qualitatively similar results are seen for a wide range of parameters; in the interest of brevity, only representative cases are displayed.

Results

In this section we shall attempt to identify those features essential for the conformational transition from the random coil (Fig. 1, structure D) to the structurally unique four-member β -barrel (Fig. 1, structure N) having bends involving residues 10–12, 21–23, and 33–35.

We begin by investigating those features that are absolutely necessary for the formation of a structurally unique four-member β -barrel. Consider the primary sequence of case A: $B_1(11)B_2(11)B_3(12)B_4(12)$, where all the residues in B_1 and B_3 and all the odd residues in B_2 and B_4 are hydrophobic type

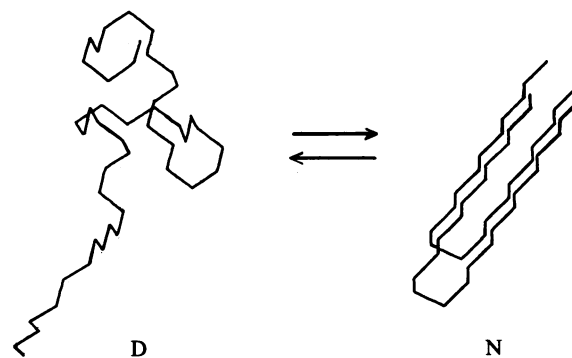


FIG. 1. Structure D depicts a representative random coil conformation in equilibrium with structure N, the "native" state four-member β -barrel.

with $\epsilon_c = -\epsilon_g/2$ and $\epsilon_h = -\epsilon_g/4$. The even residues in B_2 and B_4 are exterior/hydrophilic type with $\epsilon_c = -\epsilon_g/2$ and $\epsilon_w = 2\epsilon_g$. This system has no *a priori* preference whatsoever for the formation of bends; rather all the residues prefer to be in the trans state based on the pattern of both short- and long-range interactions. In Fig. 2, curve A (open squares), we plot the mean squared radius of gyration (S^2), obtained as an average over three cooling runs, vs. T^* . On cooling, among the observed collapsed states is the desired lowest energy structure (Fig. 1, structure N), whose bends involve residues 10–12, 21–23, and 33–35. However, we have also observed conformations that are nonnative. Various combinations of bends involving residues 8–10, 23–25, and 35–37 are also seen. Due to the odd–even pattern of the hydrophobic/hydrophilic interactions, for nonnative collapsed states there are shifts in registration by two residues from the location of the bends in the native conformation. Although the desired “native state” is appreciably populated, this state is by no means the only collapsed conformation seen. However, as T^* is lowered, the relative population of lowest energy, native β -barrel structure N increases. Consequently, whereas the presence of regions having a tendency to form turns is not a requirement for the formation of the native state, the general pattern of hydrophobic/hydrophilic residues alone is *not sufficient* to ensure that native state is *uniquely populated*. Moreover, in the denatured state, the fraction of β -sheet (trans), $f_t = 0.590$, as compared with the native state value of $f_t = 0.786$ in equilibrium with it. Parenthetically, we note that the f_t of the pure-native structure N is 0.7907; the deviations of f_t from this value are due to local fluctuations in the conformation of the ends. Hence, this system does not reproduce two essential features of the globular protein-folding transition (1–7). The collapse is to a nonunique state, and the denatured state has far too much secondary structure. However, even here tertiary interactions have served to induce additional secondary structure.

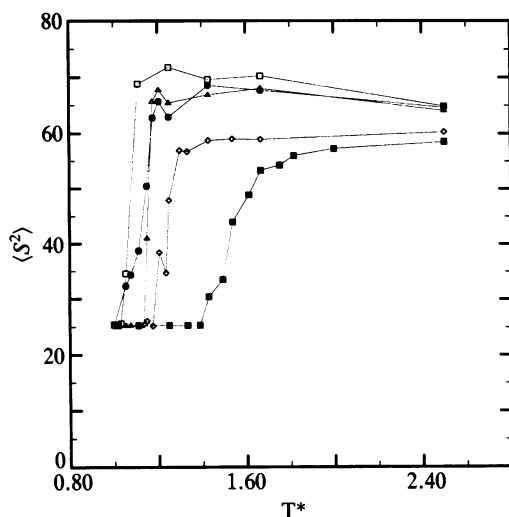


FIG. 2. Plot of the mean squared radius of gyration (S^2) vs. the reduced temperature T^* . \square , Curve A; renaturation curve of the molecule having the primary sequence of case A: $B_1(11)B_2(11)B_3(12)B_4(12)$. \bullet , Curve B; renaturation curve of the molecule having the primary sequence of case B: $B_1(12)B_2(11)b_2B_3(12)B_4(12)$. Δ , Curve C; renaturation curve of molecule having the primary sequence of case C: $B_1(11)b_1B_2(11)b_2B_3(12)b_3B_4(12)$. \diamond , Curve D; renaturation curve of the molecule having the primary sequence of case D: $B_1(11)b_1B_2(11)-b_2(11)B_3(12)b_3B_4(12)$, with a strong intrinsic preference to form turns. \blacksquare , Curve E; renaturation curve of the molecule having the primary sequence of case E: $B_1(11)b_1B_2(11)b_2B_3(12)b_3B_4(12)$, but where strand four has a cooperativity parameter of $-\epsilon_g$ and b_3 has a strong bend preference.

To examine whether or not the general pattern of hydrophobic/hydrophilic residues is responsible for the formation of the N structure and the two-residue out-of-register nonnative conformers, we set $\epsilon_w = 0$ in sequences B_2 and B_4 . These systems do not collapse to a unique N; rather, an even broader manifold of collapsed conformational states is seen. All the structures are four-member barrels whose strands differ in length—i.e., there is hypervariability of bend location. The various conformational states do not differ in free energy by an amount sufficient to ensure the unique formation of the lowest free energy N conformation. Examples of this kind of nonunique collapse transition may be found elsewhere (16–18) and need not concern us further.

We next investigate what features can provide a faithful representation of the globular protein-folding transition. One possibility is to introduce into the primary sequence the presence of regions where bend formation might be preferred. We begin with the primary sequence of case B: $B_1(11)B_2(11)b_2B_3(12)B_4(12)$ where the tertiary interactions are the same as in case A above except for those involving residues 21–23 (i.e., the central bend region) where $\epsilon_g = \epsilon_c = \epsilon_h = \epsilon_w = 0$. These residues are neutral toward formation of the central bend. In the previous case, central bend formation costs $3\epsilon_g$, which is about $3k_B T$ at the transition midpoint. In curve B (solid circles) of Fig. 2 we plot (S^2), averaged over four cooling runs, vs. T^* . Because the system does not have to pay an energetic price to form the central bend, the collapse occurs at a somewhat higher T^* . The fraction of trans states increases from about 0.589 in the denatured state to 0.782 in the native state. As in case A, a number of nonnative collapsed conformations in the transition region are seen before the N structure dominates. The central bend 21–23 now remains immobile, but the location of bends between β -strands B_1 and B_2 and between β -strands B_3 and B_4 are seen to fluctuate, although not as frequently as in case A. Clearly the lack of bend-forming regions between strands 1–2 and strands 3–4 increases the structural polymorphism of the low-temperature-ordered conformations. Thus, we next examine those features that might produce the desired N structure as well as to extend its range of stability to higher temperatures.

Formation of the Unique β -Barrel Structure

The most obvious step is to substitute bend neutral regions for those hydrophobic regions involving residues 10–12, 21–23, and 33–35 (i.e., all of the desired bend regions). Thus, we examine the primary sequence pattern of case C: $B_1(11)-b_1B_2(11)b_2B_3(12)b_3B_4(12)$. In B_1 and B_3 (B_2 and B_4), all the (odd) residues are hydrophobic type with $\epsilon_c = -\epsilon_g/2$; $\epsilon_h = -\epsilon_g/4$. In B_2 and B_4 the even residues are hydrophilic type with $\epsilon_c = -\epsilon_g/2$ and $\epsilon_w = 2\epsilon_g$. For all b type residues, $\epsilon_c = \epsilon_h = \epsilon_g = \epsilon_w = 0$. Three cooling and heating runs were performed; the same unique native state was found 19 times. No structures involving incorrect (i.e., nonnative) bend location were observed. In the transition region, being an all-or-none transition, one would expect multiple fluctuations between the native and denatured states (2, 4, 7, 25). Consistent with this prediction during the course of these simulations 19 transitions from a denatured state to the same exact native state were observed (see below for additional examples, in particular Fig. 3). In curve C (open triangles) of Fig. 2, a plot of (S^2) vs. T^* averaged over three cooling runs is presented. In comparison to curves A and B, the transition is seen to occur at a higher temperature. Moreover, f_t jumps from 0.531 before the collapse to $f_t = 0.780$, characteristic of the native state. Furthermore, the mean energy of the native state where renaturation takes place is essentially the same for cases A–C. Comparison of the average values for (S^2), f_t , the mean energy, and heat capacity obtained from three

cooling and heating sequences reveals that the cooling and heating curves are quite close, with denaturation occurring at a slightly higher temperature. [Note that the two curves should coincide in the limit of infinitely long runs (23); but, due to the limitations of our computational resources, we cannot run the simulations long enough to achieve coalescence.]

The major effect of the introduction of the additional pair of bend neutral regions, b_1 and b_3 , has been to eliminate the presence of conformers containing nonnative bends between strands 1–2 and strands 3–4. Hence, the sought-after-conformational transition to one particular collapsed state, structure N of Fig. 1, has been achieved, but this has not enhanced the range of thermal stability of the β -barrel-like structures very much. Perhaps by augmenting the bend-forming tendency of regions b_1 , b_2 , and b_3 , thermal stability of the native state can be enhanced. Thus, in case D, the energy of any gauche conformation (not necessarily the native sequence $g^+g^-g^+$) is set equal to $-2|\epsilon_g|$ for residues 10–12, 21–23, and 33–35 (the bend regions). The pattern of interactions in each of the B regions remains the same as in case C. For this series of runs comprising three cooling and two heating cycles, the unique β -barrel native structure was obtained a total of 17 times. Again, no nonnative collapsed structures were seen. In curve D (open diamonds) of Fig. 3, we plot the average $\langle S^2 \rangle$ vs. T^* obtained on cooling from an essentially infinite-temperature pure-random-coil state. Renaturation occurs at a substantially higher temperature than cases A–C; $\langle S^2 \rangle$ of the random-coil state is seen to be essentially temperature independent (with somewhat more collapsed conformations of the random-coil now favored). The collapse results in an augmentation of f_t from 0.47 to 0.78. Observe that now one has a rather faithful globular protein model.

Fig. 3 *a–c* presents a flow chart of the reduced energy (U) vs. “time” for three temperatures: at $T^* = 1.667$ (*a*), a high-temperature denaturing condition where the chain is a random coil with an $f_t = 0.44$; at $T^* = 1.205$ (*b*) in the transition region with an $f_t = 0.65$; and at $T^* = 1.111$ (*c*), a low-temperature, renaturing condition where the native state is strongly preferred and with an $f_t = 0.780$. Each “time” step represents 75,000 cycles. It is immediately apparent that the conformational transition is indeed all-or-none (2, 4, 7, 25). At any temperature, U_N , the reduced energy of N equals $-53 \epsilon_g/k_B T$ for the set of parameters chosen. At high temperature (Fig. 3*a*), all the populated states are pure-random coil. In the transition region (see Fig. 3*b*), the system is either a random-coil or a four-member β -barrel structure, ($U_N = -43.99$) with the small fluctuations in U corresponding to the wiggling of the ends. Characteristic of an all-or-none transition where intermediate states are sparsely populated (2–4, 7, 25), the conformational transition between the denatured and native states is extremely rapid. Further cooling reduces the frequency of the jumps from native to denatured and vice versa, until by $T^* = 1.111$ (Fig. 3*c*) the system is entirely in the native state ($U_N = -47.70$). Fig. 3*a–c* is representative of the kind of energy (or any other parameter)-vs.-time curves seen for all the simulations.

A further step toward stabilizing these model β -barrels can be achieved by increasing the magnitude of the attractive interactions between the strands. In case E the primary sequence is of the form $B_1(11)b_1B_2(11)b_2B_3(12)b_3B_4(12)$. The bend region b_3 in addition to having a gauche energy of $-2\epsilon_g$ for residues 33–35 has a gauche energy equal to $2\epsilon_g$ for residues $i = 36–38$ [observe, once again, that the particular gauche state (g^+ or g^-) is not specified]. Thus the tendency of the b_3 region to form some kind of bend (but not necessarily the native) is very strong even in the denatured state. Furthermore, the cooperativity parameter of residues 36–45 is twice that of B_1 for $i = 1–3$, that is $\epsilon_c = -\epsilon_g$. Curve E (solid

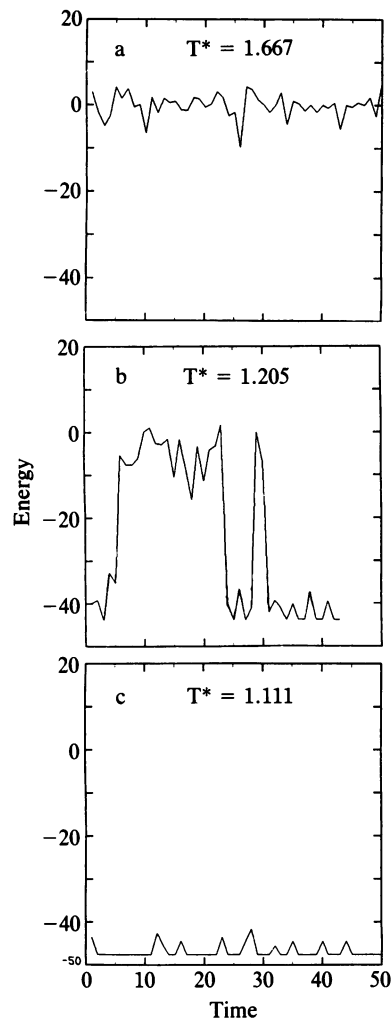


FIG. 3. Plot of the reduced energy obtained from trajectories of case D as a function of “time” at $T^* = 1.667$ (*a*), 1.205 (*b*), and 1.111 (*c*); each “time” step represents 75,000 cycles. See text for details concerning mean energy.

squares) of Fig. 2 presents a plot for the average over eight cooling runs of $\langle S^2 \rangle$ vs. T^* . The addition of an extra sticky strand is seen to substantially elevate the transition temperature. Moreover, the frequency of the jumps between the native and denatured states in the transition region relative to case D is reduced. The mean fraction of trans states increases from about 0.46 in the denatured state to 0.78 in the native state.

Possible Implications for Protein Evolution

These studies on the necessary conditions for the formation of the four-member β -barrel model protein suggest a schematic mechanism of protein evolution. We realize, of course, that this is highly speculative and cannot be proven. Moreover, these simulations say nothing about which came first—proteins, nucleic acids, or something else. Let us start with the premise that somehow “protoproteins” form. For sufficiently short chains, α - or β -hairpin-like structures or hydrophobic pockets might fold (3, 26, 27). Unfortunately, the “interior” residues are probably poorly shielded from solvent, and these structures would be expected to be marginally stable. As the size of the protoprotein increases and if a general pattern of hydrophobic and hydrophilic interactions consistent with the formation of a β -barrel (or an α -helical bundle for that matter) is synthesized by random selection,

then based on the folding criteria seen here they will spontaneously form. The collapsed conformations would occupy a manifold of conformational states that are quite close in energy, might only be stable over a relatively narrow temperature range, and therefore would be miserable proteins by today's standards. Further, suppose that random mutations subsequently (or concurrently) introduce into the primary sequence amino acids having a tendency to form turns. If the bend-forming region occurs at a place inconsistent with the general pattern of interactions (e.g., in the middle of a β -stretch), then free energy considerations that dictate the mean length of a β -stretch will override the formation of the bend in the collapsed state. Such systems would have less thermal stability than those having no bend residues at all, and therefore they would not be able to compete as well with those organisms lacking the incorrect bend-type residues. If, however, the mutation occurs at a place where an essential bend for the formation of the native state is located, then a partially unique native structure with a broader range of thermal stability will result. Organisms possessing these mutations will compete more successfully than those having protoproteins that lack the structural specificity on folding; for the latter, presumably a smaller fraction of the accessible folded isomers are biologically active. Subsequent random mutations would act to introduce additional bend-forming regions until a structurally unique native state resulted. Additional random mutations might produce amino acid substitutions that extend the environmental range over which the protein is stable. In other words, this mechanism allows natural selection to drive the development of proteins to ever more stable and complex tertiary structures (28). The alternative viewpoint that the uniqueness of the native structure depends on the presence of a large number of site-specific interactions before the protein can form has statistics working against it. The present assertion that the folding rules are robust is conceptually plausible and allows for a straightforward rationalization of protein evolution.

Discussion

The conclusions of the present series of simulations on the folding of model four-member β -barrel proteins are as follows. First, it appears that the necessary conditions for the formation of the unique N structure are (i) the presence of a small, fluctuating amount of secondary structure in the denatured state, (ii) the general pattern of hydrophobic and hydrophilic interactions, and (iii) a predilection of certain regions of the molecule to form bends. This turn-forming tendency, in fact, can be rather weak. In the simplest case considered, which folded to a unique native state, no intrinsic preference for turn formation whatsoever was present. Bends stabilize the native state by eliminating conformers that have shifts in registration consistent with the hydrophilic/hydrophobic pattern in the amino acid sequence, and thereby they localize the collapsed conformation to its lowest free energy native state. Basically, the removal of a hydrophobic-type residue from the interior into a bend region costs free energy, whereas formation of a turn in a bend-neutral region does not. Thus, bends tend to "lock" the protein into the lowest free energy state, and thereby bends produce a conformational transition to a well-defined, essentially specific native state. Furthermore, the range of thermal stability of the native state can be enhanced by augmenting the attractive interactions

between β -sheets and by the introduction of strong bend-forming regions.

In the sense that the conformational transition described here is all-or-none and folds to a unique structural state (with only small fluctuations about the global free energy minimum), we have identified systems that are faithful mimics of β -barrel proteins. Of course, due to the α -carbon lattice representation a large number of finer details have been omitted. Thus, these are models of real proteins viewed at very low resolution. The topology and thermodynamics of the transition appear to be in the same class as real globular proteins, but further refinements of the method are clearly required. Nevertheless, these simple models point the way to the folding of more complicated systems, such as the Greek key motif as well as the predominantly α -helical and mixed α/β proteins (22).

This research was supported by National Institutes of Health Grant GM-37408 from the Division of General Medical Science, U.S. Public Health Service. The assistance of the Washington University Business School Computing Facility is gratefully acknowledged.

1. Anfinsen, C. B. (1973) *Science* **181**, 223–230.
2. Jaenicke, R., ed. (1980) *Protein Folding: Proceedings of the Twenty-Eighth Conference of the German Chemical Society*, (Elsevier/North Holland Biochemical, Amsterdam).
3. Ptitsyn, O. B. & Finkelstein, V. A. (1980) *Q. Rev. Biophys.* **13**, 339–386.
4. Ghelis, C. & Yon, J. (1982) *Protein Folding* (Academic, New York).
5. Go, N. (1983) *Annu. Rev. Biophys. Bioeng.* **12**, 183–210.
6. Wetlaufer, D., ed. (1984) *The Protein Folding Problem* (Westview, Boulder, CO).
7. Creighton, T. E. (1985) *J. Phys. Chem.* **89**, 2452–2459.
8. Tanford, C. (1968) *Adv. Protein Chem.* **23**, 121–282.
9. Lewis, P. N., Go, N., Go, M., Kotelchuck, D. & Scheraga, H. A. (1970) *Proc. Natl. Acad. Sci. USA* **65**, 810–815.
10. Skolnick, J. (1985) *Macromolecules* **28**, 1073–1083.
11. Fasman, G. (1987) *Biopolymers* **26**, S59–S79.
12. Rose, G. D., Winters, R. H. & Wetlaufer, D. B. (1976) *FEBS Lett.* **63**, 10–16.
13. McCammon, J. A. & Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids* (Cambridge Univ., Cambridge).
14. McCammon, J. A. (1987) *Science* **238**, 486–491.
15. Li, Z. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6611–6615.
16. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7267–7271.
17. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *J. Chem. Phys.* **85**, 3585–3597.
18. Kolinski, A., Skolnick, J. & Yaris, R. (1987) *Biopolymers* **26**, 937–962.
19. McCammon, J. A. (1984) *Rep. Prog. Phys.* **47**, 1–46.
20. Karplus, M. & McCammon, J. A. (1981) *CRC Crit. Rev. Biochem.* **9**, 293–315.
21. Van Gunsteren, W. F. & Berendsen, H. J. C. (1982) *Biochem. Soc. Trans.* **10**, 301–309.
22. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167–339.
23. Kremer, K., Baumgartner, A. & Binder, K. (1981) *J. Phys. A* **15**, 2879–2883.
24. Binder, K., ed. (1984) *Application of the Monte Carlo Method in Statistical Physics* (Springer, Berlin), pp. 195–223.
25. Privalov, P. (1979) *Adv. Protein Chem.* **33**, 167–241.
26. Matheson, R. R. & Scheraga, H. A. (1978) *Macromolecules* **11**, 819–829.
27. Auer, H. E. & Miller-Auer, H. (1986) *Biopolymers* **25**, 1607–1613.
28. Eigen, M. & Winkler, R. (1981) *Laws of the Game* (Harper & Row, New York).