Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 2668–2672, April 1989
Biophysics

Monte Carlo simulation of equilibrium globular protein folding: α-Helical bundles with long loops

(left-handed four-helix bundle/apoferritin/somatotropin)

ANDRZEJ SIKORSKI† AND JEFFREY SKOLNICK‡

Institute of Macromolecular Chemistry, Department of Chemistry, Washington University, Saint Louis, MO 63130

Communicated by Marshall Fixman, January 6, 1989

ABSTRACT **To help elucidate the general rules of globular protein folding, computer simulations of the conformational transition in model proteins having the left-handed, four-helix bundle motif in which the helices are joined by one or two long loops, as in apoferritin and somatotropin, respectively, have been undertaken. In the context of simple tetrahedral lattice protein models, these unique native helix bundle motifs can be obtained by a set of interactions similar to those found in previous simulations of the folding of four-member α-helical bundles with tight bends and β-barrel proteins including the Greek key motif. The essential features sufficient to produce the four-helix bundle motif with long loops are as follows: (*i*) a general pattern of hydrophobic and hydrophilic type residues which differentiate the interior from the exterior of the molecule; (*ii*) the existence of hydrophilic regions in the amino acid sequence that, on the basis of short-range interactions, are indifferent to loop formation but that interact favorably with all the exterior residues of the helix bundle. Thus, these simulations indicate that, to reproduce all varieties of the left-handed four-helix bundle motif, site-specific interactions are not required.**

The problem of predicting the tertiary structure of a globular protein, given the primary sequence of amino acid residues that constitute it, is one of the most important and most exciting problems in biochemistry, and many attempts recently have been made to solve this problem (1–7). While a wealth of information on the three-dimensional structure of globular proteins exists (5) remarkably little is known about the essential factors governing the folding of proteins to their native conformation. The main questions one confronts are: What is the relative role of long-range vs. short-range interactions in determining the native state (8–10)? Are specific site–site interactions required to achieve a unique, folded native structure (11)? If not, what level of detail must be employed to predict tertiary structure (12)? What is the role of some secondary structure in the denatured state? What is the role of short bends and longer loops in forming tertiary structure (8, 13–15)?

A number of theoretical approaches have been developed to address the above questions, including molecular dynamics or Brownian dynamics techniques (7, 16). Unfortunately these techniques, when applied with a realistic potential energy surface, are limited to very short time scales. An alternative approach, espoused by Scheraga and coworkers (17, 18), employs a set of realistic potentials typically obtained from a build-up procedure, and the folded structure is obtained by energy minimization. A major problem encountered here is the presence of many local minima on the free energy surface. Up to the present, these methods have been successfully applied only for very short polypeptide chains. Finally, it should be mentioned there are some Monte Carlo simulations of lattice models of globular proteins (19–22), but they suffer from the limitation that the target native state structures are assumed in advance and the possibility of nonnative interactions is ignored.

An alternative method of studying protein folding has recently been proposed in the context of simple lattice models having a minimal set of interactions (12, 23–27). By employing an efficient Monte Carlo algorithm that allows the model chain to hunt over all configurational space, interactions between all spatially close residues are allowed, and the native structure is not assumed *a priori*. Furthermore, no site–site specific interactions are introduced. That is, all residues of a given type (e.g., hydrophobic) are taken to be the same. From the simulation results on the folding of β-barrels (12, 25, 26) and α-helical protein models (27), some general rules of folding have begun to emerge. It seems that the minimal sufficient set of conditions required to obtain a unique tertiary structure are as follows: (*i*) a general hydrophobic/hydrophilic pattern of interactions; (*ii*) short-range interactions that marginally favor secondary structure formation (a slight statistical preference for trans conformations in β-sheet proteins and the existence of helical-wheel type interactions in helical proteins, respectively); (*iii*) the presence of some regions that, based on tertiary interactions, exhibit a statistical preference to form bends. These are called "bend neutral" because without tertiary interactions the native bend conformation is but one of a large number of equally likely rotational states.

In the case of proteins whose topology is complicated by the presence of long loops in the native conformation, the effect of such loops must be addressed. Recently the role of loops was examined (25) for the case of a Greek key, β-barrel structure. Not surprisingly, loops were found to provide an entropic barrier that must be surmounted to produce a fully folded state, and, depending on the stability of the loop conformation associated with the native state, a single- or multiple-domain protein could result. In fact, if the loop marginally interacts with the folded conformation of the remainder of the molecule, a nonunique native state can result. In the present paper, the effect of long loops on the structure and the thermodynamics of the conformational transition in the left-handed four-helix bundle motif is examined.

Previously, we considered the equilibrium conformational transition to a four-helix bundle having three tight bends and no long loops (27)—i.e., the topology of cytochrome *c'* (28) and myohemerythrin (29). This is denoted by model 0 in what follows (zero for a native structure devoid of long loops) and is represented schematically by structure **0** of Fig. 1. In all cases discussed below, a thick (thin) line indicates a junction between helices at the top (bottom) of the helix bundle. In the

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.

†Present address: Department of Chemistry, University of Warsaw, 02-093 Warsaw, Poland.
‡To whom reprint requests should be addressed.

present paper, we examine some sufficient conditions required to fold a helix bundle having one long loop, model 1, as occurs in apoferritin (30); a schematic representation of the native state is shown in **1** of Fig. 1. Similarly, in model 2, and shown in **2** of Fig. 1, the conformational transition to and from a four-helix bundle having two long loops, and representative of the topology of methionyl porcine somatotropin (31), is investigated. These three models include all the possible connections between helices in an antiparallel, left-handed, four-helix bundle. Thus, it is of importance to demonstrate whether in fact our approach is capable of reproducing this class of native state structures.

MODEL

The model represents every amino acid residue by a single tetrahedral lattice point. Hence, the protein consists of N consecutive points, connected by $N-1$ bonds, each of length $l = 3^{1/2}$. The set of allowed orientations of the bonds are given by the permutations of $(\pm 1, \pm 1, \pm 1)$, which maintain tetrahedral bond angles. The short-range repulsion that prohibits chain crossing is implemented by forbidding the multiple occupancy of all lattice vertices. Every vertex represents an entire residue with side chains included; i.e., an α-carbon representation of the protein is employed. Thus, these models of α-helical proteins are obviously very crude in their representation of local details. Furthermore, in real proteins, the number of residues per α-helical turn is 3.6. Due to lattice restrictions, there are 4 residues per turn in the models considered here. Moreover, in real proteins, the helices are slightly tilted with respect to each other, and the four-helix bundles have a left-handed supertwist (3, 32). On the diamond lattice, the helices are perfectly parallel and a supertwist cannot be accommodated. Nevertheless, if one views these as models of proteins at low resolution, but with the correct backbone topology, a number of important qualitative insights can be obtained.

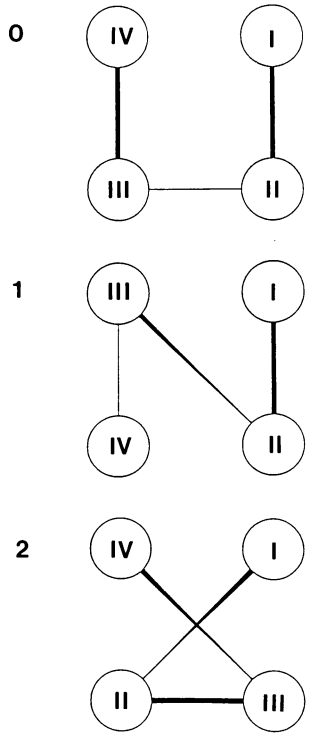

FIG. 1. Schematic representation of the topology of globular protein models 0 (see ref. 27), 1 (see ref. 30), and 2 (see ref. 31). Every native state contains a left-handed, antiparallel, four-helix bundle. The thick (thin) lines represent connections on the top (bottom) of the structure.

We next turn to the nature of the interactions that are included in the model. Each residue has three accessible rotational states: the planar trans state (t) and the two out-of-plane states, gauche plus (g^+) and gauche minus (g^-). A consecutive sequence of g^- states produces a right-handed α-helix. Elsewhere, we have shown that while an intrinsic preference for g^- states can be included in the model, it is not obligatory (27). The local preference for forming α-helical states is implemented here by including helical-wheel type interactions (33) (i.e., cooperative interactions) between residues i and $i+4$. More specifically, if all residues from the ith to the $(i+4)$th are in a g^- conformational state, then these two residues interact with an attractive potential of mean force ε_c. Helical-wheel type interactions are allowed between every residue in the chain, except for those in the putative loop and bend regions (see below).

An amphipathic sequence of residues (34, 35) is introduced to allow for differentiation of the exterior from the interior surfaces of the α-helices. All hydrophobic or hydrophilic interactions are assumed to occur only between nonbonded nearest neighbor pairs of residues and are independent of their conformation. A hydrophilic–hydrophobic pair of residues interacts with a repulsive potential of mean force ε_w. Hydrophilic/hydrophilic residues may be attractive, indifferent, or repulsive; for convenience and unless otherwise specified, we assign them an $\varepsilon_w > 0$ as well. Hydrophobic pairs of residues interact with an attractive potential of mean force ε_h. The value of ε_h or ε_w depends only on the kinds of residues that are nonbonded nearest neighbors. ε_h defines the reduced temperature scale by $T^* = k_B T/\varepsilon_h$, in which k_B, is the Boltzmann constant and T is the absolute temperature. For a more detailed description of hydrophobic and hydrophilic interactions see ref. 27.

There are some regions in the model chain which form tight bends or long loops in the desired native structure. Such regions in the primary sequence might be bend neutral. That is, for these residues, there is no statistical preference for any particular configuration, and $\varepsilon_c = 0$ as well. Values of ε_h and ε_w in a bend neutral region can be equal to zero or not equal to zero; qualitatively identical behavior is observed. In other models (type h; see below), these residues might also be part of the amphipathic sequence (with ε_h and $\varepsilon_w \neq 0$), and, based on local helical-wheel type interactions ($\varepsilon_c \neq 0$), these regions favor helix formation.

Residues belonging to those portions of the chain which are expected to form the long loop(s) in the native state might have different properties compared to those in regions that produce helices and bends. First, loop(s) might be relatively more flexible (36), and thus originally no cooperative interactions are introduced ($\varepsilon_c = 0$). Subsequently, ε_c in the loop regions was set equal to ⅛ and ¼ of the value in the remainder of the chain. Folding behavior qualitatively identical to that when $\varepsilon_c = 0$ is observed. We therefore restrict our attention to reporting results when $\varepsilon_c = 0$ in the putative loop regions. Second, loops should be located outside of the core of the helical bundle, but they should also be attracted to the hydrophilic residues that form the external surface (31). Thus, an attractive potential of mean force ε_a is introduced between hydrophilic residues that might be on the exterior of the helix bundle and those residues located in possible loop(s) regions. The potential ε_a is an extension of the hydrophilic pattern to loop regions and corresponds to an ε_h type of interaction. A separate symbol is introduced for notational convenience to make the distinction between the α-helical core and the possible loop-forming parts of the protein so as to underline the similarities of all the models of the four-helix bundle under consideration. For the special case of model 2 having two such loops in the native state, as a first approx-

imation, it is assumed that other than a hard core repulsion there is no interaction between the loops. Again, it is emphasized that the assumed set of interactions is valid not only in the native state but also for every configuration of the chain in which the pair of interacting residues are nearest neighbors. Thus the Monte Carlo algorithm has *only* information about primary structure (the values of ε_c, ε_h, ε_w, and ε_a for every residue) without any built-in bias towards the native conformation. Of course, the primary sequence introduced by us is one which we guessed would form a four-helix bundle with the appropriate topology.

As in previous work (27), the primary sequence is specified by the following shorthand notation. $\mathbf{H}_i(k)$ corresponds to putative helical stretch number i consisting of k residues and having a typical amphipathic amino acid sequence. Short bend neutral regions are denoted by $\mathbf{b}_i^o$, and the ith putative loop consisting of k residues is denoted as $\mathbf{L}_i(k)$. The number of the model (e.g., 0, 1, or 2) indicates the number of putative loop type regions encoded into the amino acid sequence. An n indicates that the bend regions are neutral in the sense defined above, and an h indicates that the regions have an amphipathic amino acid sequence and, based on local interactions, would prefer to adopt an α-helical configuration.

A number of cases having the native state topologies of structures **1** and **2** of Fig. 1 were studied. Both of them have an α-helical core that is almost the same as in previous studies on the four-helix bundle with tight bends (27).

The first model under consideration, model 1, consists of $n = 61$ residues with a short bend (every turn involves 3 residues) between helices I and II and between helices III and IV and a long loop (15 residues) connecting helices II and III (see structure **1** of Fig. 1 and Fig. 2*A*). Model 1n has the primary sequence $\mathbf{H}_1(9)\mathbf{b}_1^o\mathbf{H}_2(11)\mathbf{L}_1(15)\mathbf{H}_3(10)\mathbf{b}_2^o\mathbf{H}_4(10)$. Model 1h, having the primary sequence $\mathbf{H}_1(11)\mathbf{H}_2(12)\mathbf{L}_1(15)$-$\mathbf{H}_3(12)\mathbf{H}_4(11)$, differs from model 1n by the absence of a turn neutral sequence between putative helices I and II and helices III and IV. Model 2n (see structure **2** of Fig. 1) consists of $N = 78$ residues with one short bend between helices II and III (consisting of 3 residues) and two long loops. One loop, involving 18 residues, is located between helices I and II and the second loop, containing 13 residues, lies between helices III and IV (Fig. 1 structure **2** and Fig. 2*B*). The primary sequence of model 2n is $\mathbf{H}_1(12)\mathbf{L}_1(18)\mathbf{H}_2(10)$-$\mathbf{b}_1^o\mathbf{H}_3(11)\mathbf{L}_2(13)$-$\mathbf{H}_4(11)$. Finally, model 2h has the primary sequence $\mathbf{H}_1(12)\mathbf{L}_1(18)\mathbf{H}_2(12)\mathbf{H}_3(12)\mathbf{L}_2(13)\mathbf{H}_4(11)$.

MONTE CARLO ALGORITHM

The Monte Carlo sampling algorithm (37) is basically the same as that employed elsewhere (12, 25–27, 37). The main idea is to allow the model system to sample all the relevant regions of configuration space, especially those configurations close to the native state and those in the randomly coiled state. Each Monte Carlo sampling cycle consists of an attempt to make the following five kinds of moves (local changes of configuration) on a randomly chosen section of the chain. (*i*) Three-bond kink motion, which changes $g^{\pm}$ into $g^{\mp}$. This move serves to diffuse local orientations down the chain. (*ii*) Four-bond kink motion, which changes $g^{\pm}g^{\mp}$ into $g^{\mp}g^{\pm}$, thereby introducing new local orientations into the interior part of the chain. (*iii*) Chain end modifications. To effect more efficient sampling, different modifications of the chain ends are used depending on the temperature. At high temperatures where random coil conformations dominate, only the two end segments are rotated. At lower temperatures, where native-like conformations contribute substantially to the population, an entire end section of the chain of length l_s is cut off and then randomly reconstructed. The magnitude of l_s is determined randomly, but the maximum length of such a piece does not exceed the length of an α-helical stretch in the guessed native state. (*iv*) Four-bond wave motion, where four consecutive bonds having the conformational sequence g^+g^- or g^-g^+ are interchanged with two consecutive bonds located elsewhere in the chain. (*v*) Five-bond wave motion, where five of the six bonds forming a closed cyclohexane-like ring are interchanged with one bond located elsewhere in the chain. Of course, the residues are renumbered after modifications *iv* and *v*. Modifications *iv* and *v*, and the low-temperature version of *iii* change and shift pieces of local conformations along the chain, thereby permitting the rapid dissolution of partially folded secondary structures which might be nonnative, and/or out of register. Hence, these moves help to surmount deep local free energy minima in a fairly efficient manner.

The periodic boundary conditions normally applied to Monte Carlo simulations were not used because of the cost in computer time associated with projecting the coordinates back into the original Monte Carlo box. With the advent of computers having large amounts of memory, an alternative method exists that is particularly efficient for single chains. The model protein chain is placed in a Monte Carlo box large enough to ensure that after a small number of Monte Carlo steps it cannot leave the box (in our particular case, the length of an edge of the box equals 80). After every 100 Monte Carlo cycles, the center of mass of the chain is shifted back to the center of the Monte Carlo box. This produces a substantial speed up (up to a factor of 3) in the algorithm.

Every local change of chain configuration is tentatively accepted subject to excluded volume restrictions. After a few modifications, a new configuration is accepted according to the standard Metropolis criterion (37). The number of modifications required to sample the equilibrium properties depends strongly on the temperature and on the length of the chain, but it is typically on the order of 10^7 Monte Carlo cycles for the cases studied below. To ensure that the algorithm is ergodic and that the model chain reaches the global free energy minimum, different initial configurations were used and cooling–heating sequences (thermal renaturation and denaturation cycles) were performed. The native and denatured states are both well characterized, but the relative populations of these states in the transition region are not. This is because an insufficient number of jumps between the native and denatured state are observed to accurately calculate the equilibrium constant between the unfolded and folded forms of the model protein.

RESULTS AND DISCUSSION

The ratio of the potentials of mean force are the same as those employed in previous work on the four-helix bundle having three tight bends; namely, $\varepsilon_a = \varepsilon_c = \frac{1}{2}\varepsilon_h = -\frac{1}{2}\varepsilon_w$. For all the models under consideration, this set of interactions leads to the unique native structure. The ratio between potentials can vary, but in general both contributions to the total energy (short-range interactions involving ε_c and long-range interactions involving ε_h) should have comparable values. Thus, a broad range of parameters lead to qualitatively similar behavior. Furthermore, since hydrophilic interactions of the ε_w type do not occur in the pure native state, their variation only influences the simulation time required to fold to the native state by the suppression (or lack thereof) of incorrectly folded species (e.g., the mirror image of a native hairpin). When an inappropriate ratio of potential parameters is chosen this gives either a collapsed, nonunique state (ε_h too strong) or too expanded a molecule, containing long isolated α-helical pieces (ε_c too strong).

In Fig. 2, representative native configurations obtained from the simulations on models 1 and 2 are shown. For each and every simulation involving a given primary sequence type, the same native state is found. Contrary to the conclu-

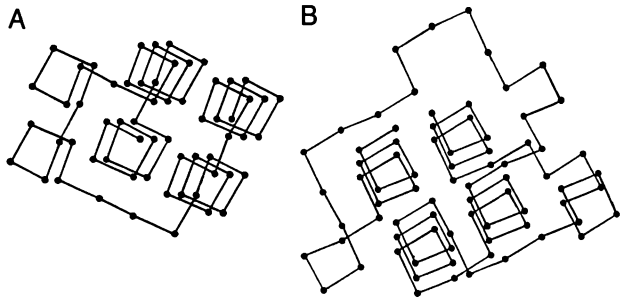

FIG. 2. Representative conformations of the native states obtained from the simulations of the four-helix bundle with one (*A*; model 1) and two (*B*; model 2) long loops.

sions from both models 0n and in four-member β-barrel structures (12, 26), where the presence of at least a central turn neutral region is required to obtain the unique native state, no tight bend preference is necessary to produce structures **1** and **2**. That is, models 1h and 2h uniquely fold up to **1** and **2** as well. We return to this point later.

Fig. 3 shows the mean-square radius of gyration $\langle S^2 \rangle$ for models 1n, 1h, 2n, 2h, and 0n (the latter was obtained from ref. 27, where it was called model F) as a function of T^*. Model 1n is observed to have a higher transition temperature than model 1h (within the error of the simulation the relative ordering of models 2n and 2h cannot be determined). The origin of this is as follows. In an h type model, a helical configuration of the bends is preferred (this is reflected in the lower value of $\langle S^2 \rangle$ at high T^*); this results in a lower free energy denatured state as compared to the bend neutral case. Thus, the n series should always melt at a higher temperature than the corresponding h series for a given topology, an effect seen in model 0 as well (27).

Due to the greater configurational entropy of the denatured state arising with an increase in the number of loops, the transition temperature increases when the number of long loops is decreased. However, in all cases, in the high-temperature limit a chain has virtually the same statistics as that of a random coil. Decreasing the temperature leads to more frequent but unstable locally ordered structures. In addition to the values of $\langle S^2 \rangle$ reported in Fig. 3, this is verified by, for example, the calculation of the mean helix content, θ_h, defined as

$$\theta_h(T^*) = \frac{f(T^*) - f_{coil}}{f_{nat} - f_{coil}}, \qquad [1]$$

with $f(T^*)$, f_{coil}, and f_{nat} the fraction of g^- (helical) states at T^*, in the denatured state, and in the native conformation, respectively. To facilitate a comparison between the models, we report average values of $\theta_h(T^*)$ calculated for those residues that are part of the four-helix bundle and not involved in loops. Just prior to renaturation there is a small amount of fluctuating secondary structure in the chain: θ_h = 0.223 at T^* = 0.741 for model 0n; θ_h = 0.229 at T^* = 0.667 for model 1n; θ_h = 0.391 at T^* = 0.588 for model 1h; θ_h = 0.417 at T^* = 0.556 for model 2n; and θ_h = 0.406 at T^* = 0.556 for model 2h.

Under native conditions, θ_h assumes values close to unity rather than precisely unity because the native conformation experiences small local fluctuations. After the collapse to the native state, $\langle S^2 \rangle$, as observed in Fig. 3, decreases slightly as the temperature is lowered in models 1 and 2, whereas $\langle S^2 \rangle$ is constant in model 0. First of all, there is an all-or-none (see below) renaturation transition of the protein chain, which forms the four-member α-helical core with the loop(s) arranged on the outside. On further decrease of the temperature, the loop(s) undergo(es) a local rearrangement to increase the number of contacts with the helical bundle, whose structure remains unchanged.

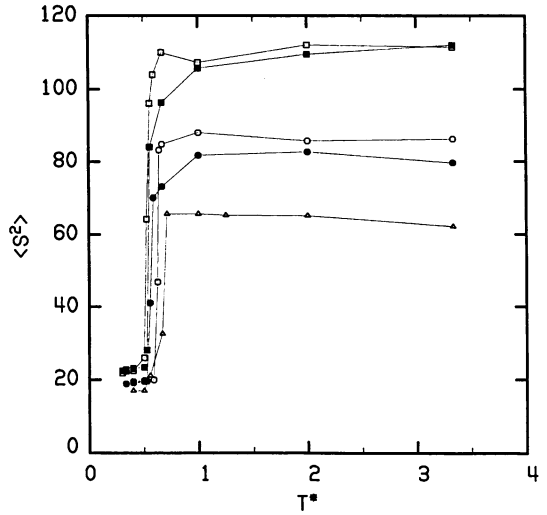

FIG. 3. Plot of the mean-square radius of gyration vs. reduced temperature T^* averaged over a number of cooling (renaturation) sequences. ○, Model 1n; ●, model 1h; □, model 2n; ■, model 2h; and △, model 0n (see ref. 27, where model 0n is referred to as model F).

The nature of the conformational transition emerges from an analysis of the average number of native helical contacts, ν_{nat}, as a function of time. We focus explicitly on the results of model 2h. The behavior of the other models is qualitatively the same. In Fig. 4, the number of native contacts, ν_{nat}, at three different temperatures is plotted vs. "time." ν_{nat} = 21 for the fully native, four-helix bundle **2** of Fig. 1. The time unit corresponds to 2×10^5, 6×10^5, and 10^5 Monte Carlo cycles, respectively, for Fig. 4 *A*, *B*, and *C*. In Fig. 4*A*, at a temperature T^* = 0.556, the model chain is randomly coiled and has some fluctuating secondary structure, among which are a few transient native contacts. In the transition region, at T^* = 0.526, the system undergoes a series of renaturation and denaturation transitions. The transition is more complicated than the transitions for model 0 having 3 tight bends and closely resembles that found for the Greek-key, β-barrel structure (25). At any given moment, the protein chain exists in one of three states: a randomly coiled chain, a pure folded helical core plus loop(s), or a folding intermediate with 10 contacts that is marginally (<8%) populated. Thus, to a good approximation the model protein undergoes the all-or-none transition. This is the same kind of transition as is observed in real proteins (5, 38, 39). Finally, at a temperature T^* = 0.5, the chain is folded into the unique native state and has ν_{nat} = 21. This number is slightly altered by fluctuations.

On the basis of the above results, we conclude that a minimal set of sufficient conditions required to obtain the folded α-helical structure having the complicated topology of **1** and **2** of Fig. 1 is almost the same as the set required for formation of an α-helical bundle with short bends and for β-barrels, including the Greek key motif (3). This strongly suggests that the folding rules are insensitive to the particular topology and require the following: (*i*) A hydrophobic/hydrophilic pattern of residues which if suitably arranged in an amphipathic pattern will produce an α-helical protein and if arranged in an odd/even pattern will result in a β-protein (35). (*ii*) Short-range (local) interactions which impart marginal stability to secondary structures such as α-helices, turns, or β-strands. In the case of four-helix bundles having tight bends, it is found that there must be at least a single central turn neutral region to produce the unique, in-register native state. Here, for helical bundles having long loops we have found that tight turn neutral regions are not obligatory.

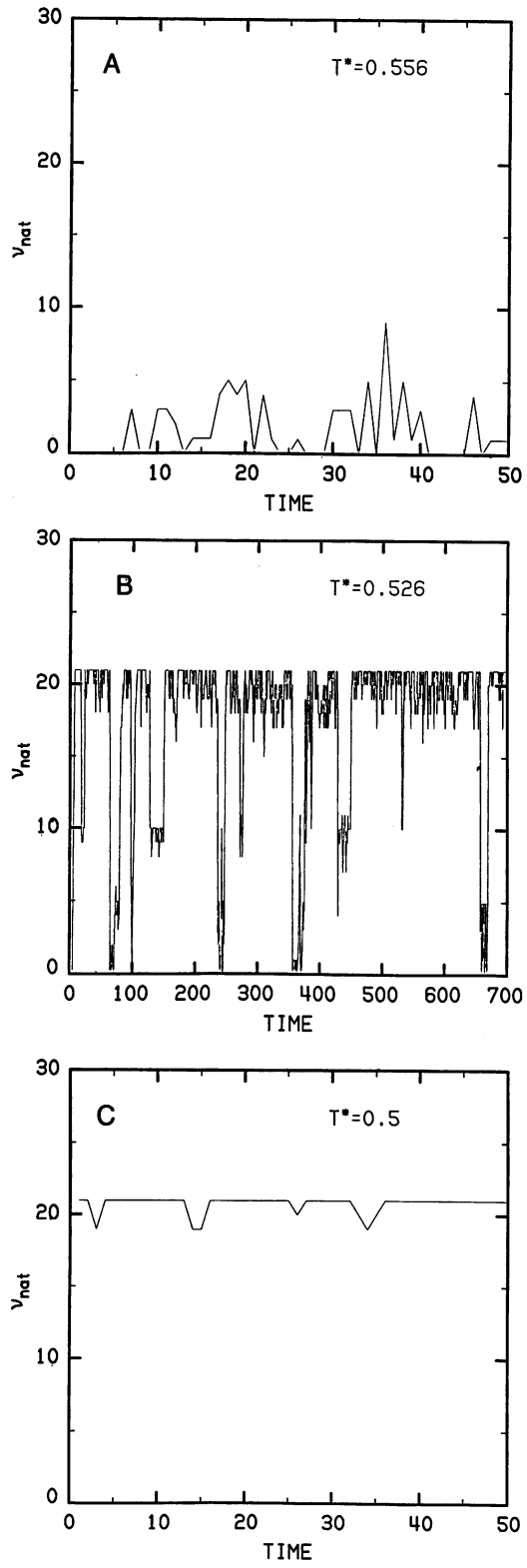

FIG. 4. Plot of the number of native contacts ν_{nat} (in the α-helical core) in a single Monte Carlo run vs. "time" (see text) for model 2h at $T^* = 0.556$ (*A*), 0.526 (*B*), and 0.5 (*C*).

This can be rationalized as follows. Basically, the loop region has no intrinsic preference for helical states and therefore effectively acts as a "stop" to helix propagation. This acts to reduce the extent of helix length fluctuations [a recent analysis of a possible role of specific residues in this regard has been presented by Presta and Rose (9)]. Furthermore, because of loop entropy, the loops tend to be as tight as possible; thus, in-register helix states are further favored. Finally, the loop interacts with the exterior faces of the helices; this further stabilizes the native in-register helix conformations. Hence, due to the interplay of energetic and entropic factors, turn neutral regions are not obligatory in helix bundles with long loops. That is, the loops reduce the fluctuations in registration inherent in the four-helix bundle with tight bends.

In conclusion, this series of simulations has again demonstrated that a simple lattice model without any specific site–site interactions can mimic the conformational transition of globular proteins in that both the overall topology and thermodynamics are reproduced. Of course, further refinements of the model are required to study mixed α/β proteins and to reproduce the finer details.

Stimulating conversations with Dr. Sherin S. Abdel-Meguid of Monsanto are gratefully acknowledged. This research was supported in part by Grant GM-37408 from the National Institute of General Medical Sciences, U.S. Public Health Service. The assistance of the Washington University Business School Computing Facility is also appreciated.

1. Jaenicke, R., ed. (1980) *Protein Folding*, Proceedings of the 28th Conference of the German Biochemical Society (Elsevier/North-Holland, Amsterdam).
2. Ptitsyn, O. B. & Finkelstein, A. V. (1980) *Q. Rev. Biophys.* **13**, 339–386.
3. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167–339.
4. Creighton, T. E. (1985) *J. Phys. Chem.* **89**, 2452–2459.
5. Mutter, M. (1985) *Angew. Chem. Int. Ed. Engl.* **24**, 639–653.
6. Karplus, M. (1986) *Ann. N.Y. Acad. Sci.* **482**, 255–266.
7. McCammon, J. A. (1987) *Science* **238**, 486–491.
8. Fasman, G. D. (1987) *Biopolymers* **26**, S59–S79.
9. Presta, L. D. & Rose, G. D. (1988) *Science* **240**, 1632–1641.
10. Skolnick, J. (1985) *Macromolecules* **28**, 1073–1083.
11. McCammon, J. A. (1984) *Rep. Prog. Phys.* **47**, 1–46.
12. Skolnick, J., Kolinski, A. & Yaris, R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5057–5061.
13. Lewis, P. N., Gō, N., Gō, M., Kotelchuck, D. & Scheraga, H. A. (1970) *Proc. Natl. Acad. Sci. USA* **65**, 810–815.
14. Rose, G. D., Winters, R. H. & Wetlaufer, D. B. (1976) *FEBS Lett.* **63**, 10–16.
15. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1985) *J. Mol. Biol.* **201**, 201–217.
16. McCammon, J. A. & Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids* (Cambridge Univ. Press, Cambridge, U.K.).
17. Scheraga, H. A. & Paine, G. H. (1986) *Ann. N.Y. Acad. Sci.* **482**, 60–68.
18. Li, Z. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6611–6615.
19. Kingbaum, W. R. & Lin, S. F. (1982) *Macromolecules* **15**, 1135–1145.
20. Levitt, M. (1982) *Ann. Rev. Biophys. Bioeng.* **11**, 251–271.
21. Segawa, S.-I. & Kawai, T. (1986) *Biopolymers* **25**, 1815–1835.
22. Taketomi, H., Kano, F. & Gō, N. (1988) *Biopolymers* **27**, 527–559.
23. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7267–7271.
24. Kolinski, A., Skolnick, J. & Yaris, R. (1987) *Biopolymers* **26**, 937–962.
25. Skolnick, J., Kolinski, A. & Yaris, R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1229–1233.
26. Skolnick, J., Kolinski, A. & Yaris, R. (1989) *Biopolymers*, in press.
27. Sikorski, A. & Skolnick, J. (1989) *Biopolymers*, in press.
28. Weber, R. C., Bartsch, R. G., Cusanovich, M. A., Hamlin, R. C., Howard, A., Jordan, S. R., Kamen, M. D., Meyer, T. E., Weatherford, D. W., Xuong, N. M. & Salemme, F. R. (1980) *Nature (London)* **286**, 302–304.
29. Hendrickson, W. A. & Ward, K. B. (1977) *J. Biol. Chem.* **252**, 3012–3018.
30. Ford, G. C., Harrison, P. M., Rice, D. W., Smith, J. M. A., Treffry, A., White, J. L. & Yariv, J. (1984) *Phil. Trans. R. Soc. London Ser. B.* **304**, 551–565.
31. Abdel-Meguid, S. S., Shieh, H.-S., Smith, W. W., Dayringer, H. E., Violand, N. V. & Bentle, L. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6434–6437.
32. Weber, P. C. & Salemme, F. R. (1980) *Nature (London)* **287**, 82–84.
33. Schiffer, M. R. & Edmundson, A. (1967) *Biophys. J.* **1**, 121–135.
34. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13**, 669–678.
35. Lim, V. I. (1974) *J. Mol. Biol.* **88**, 857–872.
36. Rose, G., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1–109.
37. Baumgartner, A. (1984) in *Application of the Monte Carlo Method in Statistical Physics*, ed. Binder, K. (Springer, Berlin), pp. 137–192.
38. Tanford, C. (1964) *J. Am. Chem. Soc.* **84**, 4240–4247.
39. Brandts, J. & Lumry, R. (1963) *J. Phys. Chem.* **67**, 1484–1494.