Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 146–150, January 1995
Biophysics

Forces of tertiary structural organization in globular proteins

KAIZHI YUE AND KEN A. DILL

Department of Pharmaceutical Chemistry, Box 1204, University of California at San Francisco, CA 94143

Communicated by Bruno H. Zimm, University of California, San Diego, CA, July 25, 1994 (received for review February 14, 1994)

ABSTRACT **The tertiary structures of globular proteins have remarkable and complex symmetries. What forces cause them? We find that a very simple model reproduces some of those symmetries. Proteins are modeled as copolymers of specific sequences of hydrophobic (H) and polar (P) monomers (HP model) configured as self-avoiding flights on simple three-dimensional cubic lattices. The model has no parameters; we just seek the conformations that have the global maximum number of HH contacts for any given sequence. Finding global optima for chains in this model has not been computationally possible before for chains longer than 36-mers. We report here a procedure that can find all the globally optimal conformations, the number of which defines the *degeneracy* of a sequence, for chains up to 88 monomers long. It is about 37 orders of magnitude faster than previous exact methods. We find that degeneracy is an important aspect of sequence design. So far, we have found that four-helix bundles, α/β-barrels, and parallel β-helices are globally optimal conformations of polar/nonpolar sequences that have minimal degeneracy.**

The tertiary structures of globular proteins are remarkable in their great variety and often high symmetries (1–3). But ever since the atomic structures of the first globular proteins appeared (4), it has not been clear how the tertiary structures of proteins are encoded within their amino acid sequences. Tertiary structures are characterized by hydrogen-bonded secondary structures, helices, and sheets and by their compactness, which appears to be driven by the burial of nonpolar amino acids. Based on these and other energetic interactions, some computer models produce native-like conformations (5). But because those models have many parameters and cannot predict the full set of protein structures, it is not yet known what physical interactions are predominant in real proteins or what is the minimal model that can predict tertiary architectures.

Here we seek such a minimal model. Motivated by earlier results showing that helices and sheets can be stabilized by chain compactness (6, 7) and that the uniqueness of native structures can be largely encoded within the sequences of nonpolar and polar amino acids (8, 9), we ask whether the sequence of polar and nonpolar monomers might be sufficient to encode tertiary structures. Until now it has not been possible to address such questions because to do so requires knowledge of the native states (i.e., conformations of lowest free energy) for chains longer than about 40 monomers in three dimensions. We have recently developed a method called "constrained hydrophobic core construction (CHCC)," which is based on discrete geometry (10), with which we found globally optimal conformations for chain lengths up to about 36 monomers (11).

Here we report new developments with CHCC that allow us to find globally optimal lattice conformations for chain lengths up to 88 monomers in minutes to hours on computer workstations. It finds lowest energy states about 37 orders of magnitude faster than any other method that is currently able to guarantee finding the global minimum. Thus, we can explore model sequence–structure relationships for chains long enough to resemble real proteins. Since the CHCC method finds all globally optimal solutions, it is unique among existing conformational search methods in its abilities (*i*) to establish with certainty the conformations that are true global optima of this model, and (*ii*) to compute the degeneracies of the global optima. The conclusions that emerge here are crucially dependent on both of these capabilities.

The Model

Protein chains are modeled as copolymers of specific sequences of hydrophobic (H) and polar/charged (P) monomers (HP model) (12). An example is as follows: P1 H2 P3 H4 H5 P6 H7 H8 P9 H10 P11 P12 H13 H14 H15 P16 P17 P18 H19 P20 H21 H22 P23 H24 H25 P26 H27 P28 P29 H30 H31 H32 P33 P34 P35 H36 P37 H38 H39 P40 H41 H42 P43 H44 P45 P46 H47 H48 H49 P50 P51 P52 H53 P54 H55 H56 P57 H58 H59 P60 H61 P62 P63 H64 H65 H66 P67. Within a sequence, a run of monomers of a single type is called a segment. For example, residues 13, 14, and 15 in the above sequence constitute an H-segment of length 3. We call a P-segment of length 1 a P-singlet (see P6 and P9).

Chains are configured as self-avoiding flights on three-dimensional simple cubic lattices. A monomer can make at most $z = 6$ contacts with nearest neighbor sites on the simple cubic lattice. A contact between two H monomers has a favorable free energy, and the contact free energy for all other types of contacts is 0. Conformations are native if they have the lowest free energy—i.e., the maximum number of HH contacts—over all possible conformations. While the HP lattice model is crude, it has the following characteristics of real proteins: when the HH attraction is strong, chains fold to a few native states with nonpolar cores and with secondary structures [defined by topological measures (13, 14)]. HP lattice proteins also resemble real proteins in some mutational (15) and kinetic properties (16). Most importantly for our present purposes, the lattice model has the same conformational search problem that proteins have. That is, both proteins and the HP lattice model have conformational spaces too large to be enumerated by computer. Because proteins fold to "single" native states, there are few global minima. Hence, folding represents finding few global minima on a large landscape. It is this problem we call the Levinthal paradox. The HP lattice model has very few global minima conformations, as we will show. In these respects, the model resembles proteins. The model does not accurately represent microscopic details of proteins.

Finding Native Structures of the HP Model by CHCC

We aim to find native structures of any HP lattice model sequence. The CHCC strategy is as follows (11). We focus on collections of H-monomers, which we call the "H-core." These collections determine the total interaction energy. It has been

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.

Abbreviation: CHCC, constrained hydrophobic core construction.

proven that if a conformation has an H-core with the minimal possible surface area, it is native (11). To find such conformations, the CHCC method systematically introduces geometric constraints to prune branches of a conformational search tree. When the method finds a branch that provably cannot lead to the minimal surface of H-core, the branch is pruned. The CHCC method first constructs an optimal geometry for an H-core and then systematically attempts to lay the chain sequence into it, subject to geometric and energetic constraints. Then CHCC tries a different core geometry and proceeds again.

To compute a "best possible" size and shape for the H-core, CHCC counts the number of H monomers in the sequence and constructs the most nearly spherical lattice container for them. If we first suppose that there were no chain connectivity and we "pour" all the H monomers into this container, then the resultant surface of H-core must certainly be a lower bound for any realizable chain fold of the sequence, since chain connectivity is a constraint that could only lead to a "worse" core than this (i.e., having larger H-core surface). This gives the first estimate for the H-core surface area. Second, in optimal conformations P-singlets must be located at the surface of the H-core because placing P-singlets at any other locations will have larger surface (i.e., higher energy) because of burial of P residues in the core. Third, to avoid burial of P monomers in the H-core, the H monomers at the center of the core must be contained in long H-segments of a sequence. Optimal configurations will have such "tether" segments configured to begin at the surface of the H-core, pass through the center, and return to some other point on the surface. H residues in shorter H-segments should not occupy these positions because it would force P monomers to be buried. To form the H-core in Fig. 1 requires H-segments of length greater than 2. Furthermore, only the H residues in the middle of such segments can occupy the positions of H14, H31, H48, and H65 monomers, which are deepest in the core. We now describe a theoretical treatment that extends earlier work (11) and applies to chains of lengths approaching those of real proteins.

For illustration here, we limit our discussion to conformations with no buried P residues in the hydrophobic cores. We slice the H-core into residue-thick layers (see Fig. 2) and compute the total surface area by summing up the lateral surface areas of the layers. Assume we slice the H-core in the direction perpendicular to the z axis, as shown in Fig. 3. Let the contribution of the layer to the surface area be p_z. The subscript z indicates the location of the layer along the z axis. Under the assumption of no burial of P residues in H-cores, we have:

$$p_z = 2[(a_{2,z} - a_{1,z}) + (b_{2,z} - b_{1,z})], \quad [1]$$

where $a_{2,z}$ and $a_{1,z}$ delineate the boundaries of the layer on the x axis. Similarly, $b_{2,z}$ and $b_{1,z}$ define boundaries on the y axis. Note that p_z is independent of the exact shape of the layer—e.g., g_z and f_z in Fig. 3. Let a, b, and c be the sizes of the H-core in x, y, and z directions. Summing over x, y, and z axes,

$$\sum_{z=1}^{c} p_z + \sum_{x=1}^{a} p_x + \sum_{y=1}^{b} p_y = 2S, \quad [2]$$

where the factor of two corrects for double-counting each side of the H-core surface S. The volume of the H-core, V, is

$$V = \sum_{z=1}^{c} \sum_{x=a_{1,z}}^{a_{2,z}} [g_z(x) - f_z(x)]. \quad [3]$$

To avoid burying P monomers, a P-singlet cannot be colinear with its two flanking H residues but must be accommodated by a "staircase" configuration—for example, H2 and H4

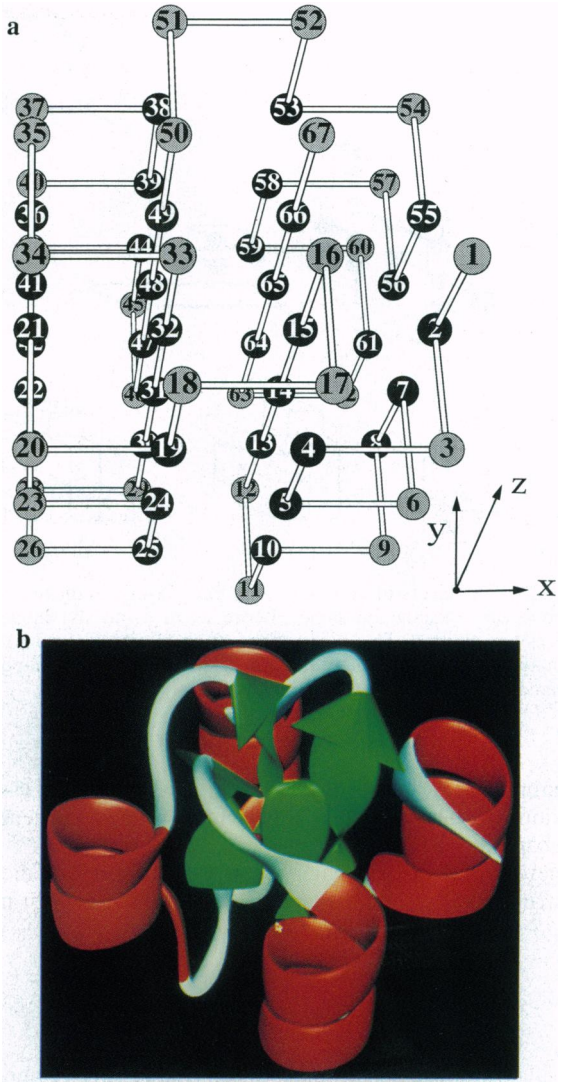

FIG. 1. (*a*) A HP lattice model conformation resembling an α/β-barrel in real proteins. Residues 2–10 are a lattice helix, and residues 12–16 are a lattice strand. As in real α/β-barrels, the β-strands are hydrophobic, protected by the surrounding helices. (*b*) The ribbon diagram of the conformation.

in Fig. 1. The number of stairs, J_z, in layer z in the z direction is (See Fig. 3)

$$J_z = \sum_{x=a_{1,z}}^{a_{2,z}-1} \{[1 - \delta_{g_z(x+1),g_z(x)}] + [1 - \delta_{f_z(x+1),f_z(x)}]\}, \quad [4]$$

where for integers k, m, $\delta_{km} = 1$ if $k = m$ and $\delta_{km} = 0$ otherwise. That is, upon proceeding along g_z (or f_z), one stair increment is added if g_z changes; no increment is added otherwise. J_x and J_y are defined as in Eq. **4** but for x and y dimensions. The total number of stairs J_t in the H-core is the sum of J_x, J_y, and J_z, corrected for double and triple counting when there are overlapping stairs.

Now we give the generalization to the tethering constraint (11). Let the sequence lengths from an H monomer to the two terminating P residues in its segment be r_1 and r_2. Let the lengths of the two shortest distinct paths from a position within the H-core to a position on its surface be d_1 and d_2. To assign that H residue to occupy that position requires, for an optimal conformation,

$$\min(r_1, r_2) \geq \min(d_1, d_2), \quad [5]$$

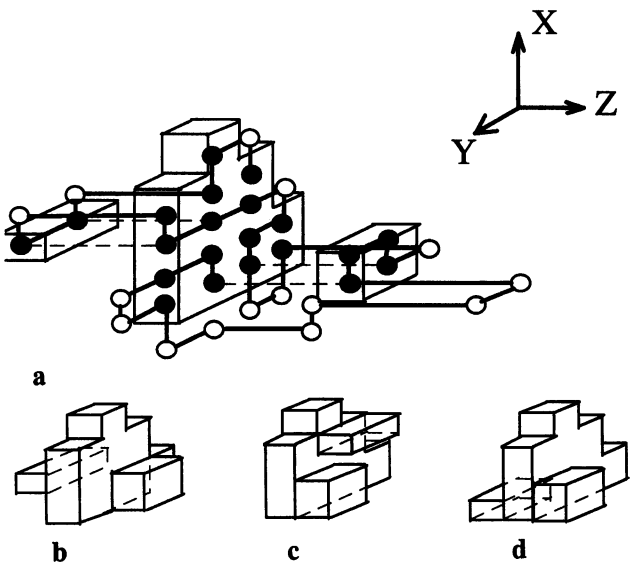

FIG. 2. The layers of an H-core. (*a*) Each layer is indicated by an envelope. (*b*) Exactly the same H-core as in *a* but the individual residues are omitted. This is the shape of an H-core. (*c* and *d*) Two alternative shapes of the same H-core. The volumes of the shapes are the same, and the numbers of stairs are nearly equal.

$$\max(r_1, r_2) \geq \max(d_1, d_2). \quad [6]$$

Defining g_z and f_z for $z = 1, 2, 3, \ldots, c$, all r_1 and r_2 of all H-core positions are defined. Thus, Eqs. **5** and **6** relate a sequence to the shape of its H-core.

Therefore, if the total number of P-singlets is n_{1P}, then to construct a native conformation requires that we must minimize S subject to Eqs. **5** and **6**, while

$$J_t \geq n_{1P}, \quad [7]$$

$$V = n_H. \quad [8]$$

Using the above constraints, we can find H-core geometries for large subclasses of sequences. For example, it has been shown (11) that the H-cores with minimal surface area are close to a cube. Furthermore, if an H-core of minimal surface area is enclosed by the tightest possible imaginary box, then all of the P-singlets must also reside inside this box. If the dimensions of the enclosing box are a, b, and c, then

$$abc \geq n_H + n_{1P}. \quad [9]$$

The P-singlets often stack on top of one another, forming a P-edge (for example, see the conformation of P3, P6, and P9 in Fig. 1). The more P-edges there are in a given H-core, the smaller the surface area. Since each P-singlet has two flanking H residues, these H residues will form two columns of H residues flanking the P-edge. No other residues can occupy those positions. The P-singlets within a pattern of sequence may form a P-edge among themselves—for example, the pattern "HPHHPHHPH" (see monomers H2–H10 in Fig. 1). More details are given in ref. 11.

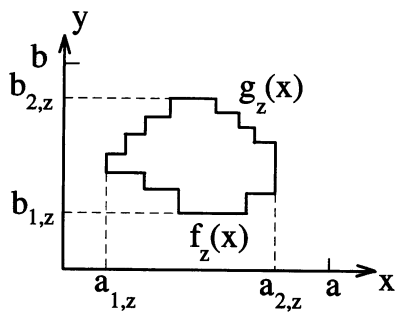

FIG. 3. A layer that is perpendicular to the z axis. Assume the maximum dimensions of such layers are a and b. In general, a particular layer may be confined inside $(a_{1,z}, a_{2,z})$ and $(b_{1,z}, b_{2,z})$; $g_z(x)$ and $f_z(x)$ give the upper boundary and lower boundary of the layer, respectively.

Demonstrating the CHCC Method: Folding a Small α/β-Barrel Sequence

The constraints in the CHCC method are powerful. We demonstrate this by the example HP sequence given in *The Model* section. While the CHCC method can find globally optimal conformations for any HP sequence, certain sequences can be folded without computer enumeration. Here is an example.

This sequence has 36 H residues and 12 P-singlets. Using Eq. **9**, we find that 48 such monomers can be contained exactly within an enclosing box of dimensions $4 \times 4 \times 3$. This box potentially describes a minimal surface H-core. Any other enclosing box can be shown to entail a larger surface area of H-core (e.g., either $4 \times 4 \times 4$ or $5 \times 5 \times 3$). Since we have exactly 12 P-singlets, an optimal core, if we can achieve it, must have four P-edges, and each edge must contain 3 P-singlets. Furthermore, the 2 flanking H monomers of each P-singlet must form two surface columns adjacent and parallel to its P-edge. These eight columns must occupy all of the H-core surface area in the direction parallel to the P-edges. This is the z direction in Fig. 1. With this step, we have now preliminarily assigned 36 residues.

All of the remaining H monomers in the sequence occur as four H-segments, each of length 3. They must fill the remaining positions in the H-core. By the tethering rule, the central H monomers in these segments must fill the center core sites. But this requirement can be satisfied by more than one configuration. For example an H-segment can pass straight by H14 as it is configured in Fig. 1. Or the H-segment could turn sideways at position H14 and return to the surface at H7 or H5. But the latter is impossible because all of the side exits are occupied by the flanking H residues that have already been placed. Consequently, all four H-segments must lie parallel to one another, and their ending P residues, e.g., P16, must be placed directly above or below them in the z direction. This now preliminarily assigns 12 additional monomers.

Next, we introduce the remaining chain connectivity and consider the exact positions of individual P-singlets. Each P-singlet occurs in one of four HPHHPHHPH patterns in the sequence. Each such pattern can form the lattice equivalent of an amphipathic helix (17) as in H2 to H10 in Fig. 1. The intervening P residues can readily be placed, producing a connected chain consistent with all of the prior placements. Therefore, the conformation has a minimal surface H-core and thus is native.

Now we show how this process can count global optima. We show that this sequence has only three native conformations: (*i*) the one in Fig. 1, (*ii*) one in which P1 in Fig. 1 flips to the right, and (*iii*) one in which the four α/β units go counterclockwise (i.e., H19 goes to the position of H55). We show that the other potential possibilities, placing H5 instead at the position occupied by H19 (in Fig. 1) and P6 at the position occupied by P20, would have been unsuccessful. All of the ending P residues of the long H-segments—e.g., P12, P16, P46, and P50—are on P-segments of lengths 2 or 3 and are positioned in the center of the top (or bottom) of the H-core. They are too short to connect the top and bottom of the H-core; thus each segment must reside entirely at the top or bottom. Thus, since all of the HPHHPHHPH patterns and the H-segments of length 3 are alternately connected in the sequence, all of the ending H

residues of the HPHHPHHPH patterns must be on the top or bottom layer. There are two possibilities for each HPHHPHHPH pattern. Either both of its ends can be in the same layer, or one end will be at the top and the other at the bottom. The positions of the P-singlets are fixed. Having both ends in the same layer entails having every monomer in the pattern in the same layer. This is clearly unacceptable because the pattern is too short for sidesteps—e.g., for moving H5 to the position of H19 in Fig. 1—because the ending H residues then would not reach their required positions on the layer at the opposite end. Thus, the only possible configuration of the HPHHPHHP pattern is an amphipathic helix. Finally, the intervening P-segments are short and must make small loops at each end. Hence, the only possible connectivity involves the α/β-barrel conformations shown.

Finding Native Conformations by Computer

We have automated the CHCC method in a computer algorithm. The search is divided into three steps. First, we identify the possible boxes that can envelop the eventual minimal-surface-area H-cores. We try all of the possible ways of distributing the n_H H residues to layers. The distributions with the minimal surface area of the H-core and with a sufficient number of stairs determine the dimensions of the boxes.

Second, for each such box, we enumerate the possible detailed shapes of the native H-cores. A given set of layers can be stacked in different ways. For example, in Fig. 2 *b, c,* and *d,* the layers are identical but stacked differently. We examine all possible combinations of placing an H or P monomer at each position in the box. The number of computer operations in this search is far fewer than 2^{abc} because of many constraints.

Third, we place residues one-at-a-time to attempt to construct viable conformations. We first assign the P-singlets, because they should be placed at stairs and are few in number. Each P-singlet starts a new "thread." We construct the conformation by expanding the threads at positions where there are the fewest choices. The choices for placing other residues are limited. For example, an H can only be placed at a depth in the H-core allowed by the tether constraint. We backtrack when a dead end is reached. A native conformation is found when all residues are placed.

The run time for the program, t, is dependent on the HP sequence, but it scales with chain length, L, roughly as

$$t \approx 0.001 D^{0.9} 1.125^{L}, \qquad [10]$$

where D is the degeneracy of the sequence. On an Apollo workstation, the program can find native states for any arbitrary sequence of up to $L = 70$ in hours and for sequences of low degeneracy of up to $L = 88$ in minutes.

Origin of Tertiary Symmetries

To explore the relationship between sequence and structure in the model, we have used the CHCC method to fold up HP sequences for chain lengths from 40 to 88 monomers. We also apply the CHCC method in reverse [to do "inverse" folding (18)], where we are given a target structure, and CHCC designs sequences that fold to it. We now briefly describe an approach we call "tinkering," which involves both folding and inverse folding—i.e., changing both the sequence and the structure. In tinkering, we (*i*) begin with a target structure and tentatively design a sequence; (*ii*) fold the sequence to all of its native conformations; (*iii*) modify both the target structure and the sequence to reduce the degeneracy, if any; and then (*iv*) repeat this cycle until it converges on a "good" sequence/structure pair—i.e., a structure that is encoded in a sequence of lowest possible degeneracy. In the HP model, most arbitrarily chosen sequences will fold to many conformations—e.g., for some 48-mers, we found degeneracies ranging from thousands to millions (19). This may not be unrepresentative of real amino acid sequences, however, since "designed" proteins also appear to have broad conformational distributions (20). Even so, by tinkering we have found a few good structures that are encoded in sequences that fold to <100 other native conformations.

Remarkably, these few good structures appear to have some of the tertiary symmetries found in globular proteins. The example described above is the closest lattice equivalent to an α/β-barrel. The sequence shown has a degeneracy of 3. That is, to within certain internal mirror symmetries, there are only three globally optimal conformations for this sequence of 67 monomers. Fig. 4*b* shows that a second class of good structures is the class of four-helix bundles of different chain lengths; the one shown has a degeneracy of 67. Fig. 5 shows a third tertiary architecture we have found, with a degeneracy of 4, to within mirror symmetries. It bears some resemblance to the parallel β-helix fold recently observed (21).

Hence, these results offer a prediction about the origins of tertiary structures in proteins. Tertiary structures would seem to require hydrophobic collapse and hydrogen bonding. But the present results offer a different conjecture—of a balance of two tendencies. On the one hand, these model chains are driven to collapse by hydrophobic interactions to compact conformations in which steric constraints are severe—to globally optimal conformations. But almost any sequence can do this, and collectively such sequences would fold to a very broad array of possible compact states. On the other hand, according to the present results another factor contributes to high-symmetry tertiary architectures. That factor is the "encodability" of the structure. For most (poor) structures, the "best" HP sequences that can encode them will also encode many other incorrect conformations. But for the few (good) structures, the HP sequences that encode them also encode very few other incorrect alternatives. In the HP model, these good structures have some of the high symmetries observed in globular proteins. The latter implies that proteins will be marginally stable because to design uniqueness requires that a sequence have a minimal number of H monomers (18). The observations that (*i*) real proteins are marginally stable and (*ii*) conformationally diverse "designed" proteins are extremely stable (22, 23) would support this hypothesis. The elegant experiments of Kamtekar *et al.* (24) support the hypothesis that the HP sequence alone is sufficient to encode tertiary architectures.

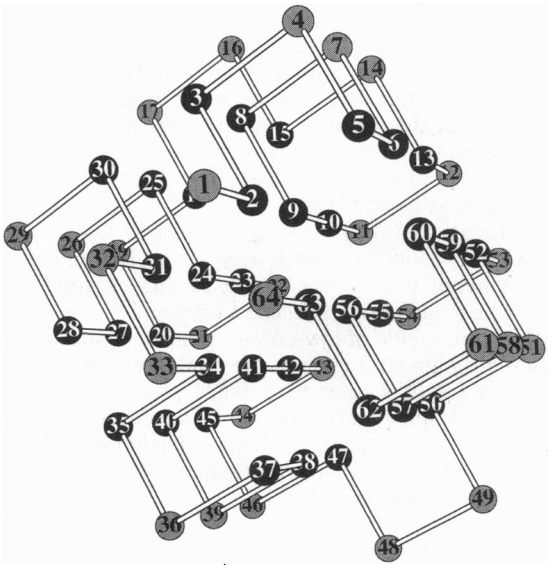

FIG. 4. A four-helix bundle.

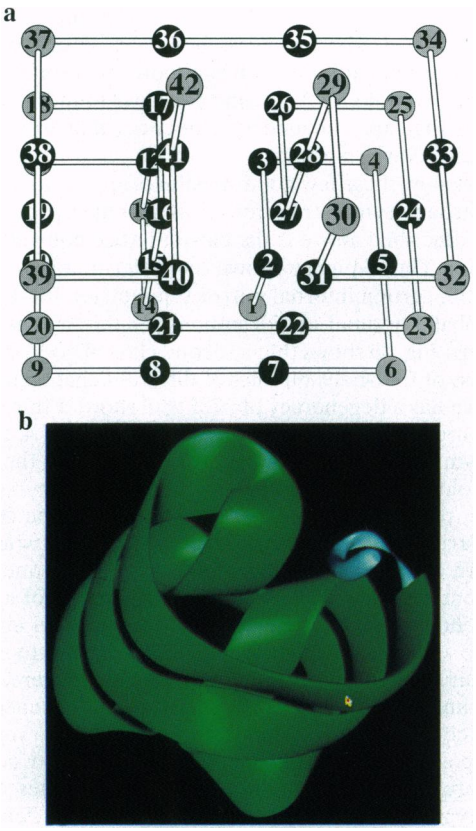

FIG. 5. (*a* and *b*) A β-helix.

Of course real proteins are more complex, and their structures undoubtedly are also determined by the other types of intramolecular interactions, neglected here. In addition, we know that some tertiary structures, mainly in which β-sheets predominate, cannot arise on the simple cubic lattice because the spatial representation of the chain partly dictates the sequence–structure relationships. But we believe the methods developed here can be generalized to off-lattice models. Many constraints discussed in this paper can be easily generalized. For example, Eqs. **5** and **6** for tether lengths will not change much in an off-lattice implementation of the present method.

It is not yet possible to design amino acid sequences to fold uniquely and without conformational diversity in the absence of prosthetic groups. One explanation is that the side chains haven't been suitably designed to fit together or that more monomer diversity is required than can be supplied in a two-letter (HP) code. This might suggest that an HP model is too simple. But the present results suggest another explanation. It may be that even the hydrophobic and polar sequence patterns in designed peptides are poor. That is, designing a sequence to have a good H core is not sufficient; it is necessary to also "design out" (18) [sometimes called "negative design" (25)] incorrect folds, even with only H and P interactions. A homopolymer of all H monomers will also have a good H-core, but it will be a bad design because a homopolymer will fold to a very large ensemble of low-energy compact states. Hence, it may be the design strategy, not the model, that is too simple. The present results suggest that simple protein design strategies—just put H monomers in the core—are too simple. Perhaps sequences can be designed from a simple HP binary code to fold to unique native states. But conformational degeneracies must become better understood first.

The present results also bear on the Levinthal paradox, the issue of how a chain with such a large conformational space can find the one native conformation on the biological (short) time scale. It has been argued that the folding problem is NP-complete. That is, according to a widely accepted assumption, the computer running time grows exponentially with the size of the problem. In the case of the three-dimensional simple cubic lattice, the number of accessible conformations for a chain of length n is approximately 4.7^n. However, Eq. **10** shows that the exponential growth is not an insurmountable problem as long as the base of the exponential is sufficiently small (1.125 here). The CHCC method is an example of a practical algorithm that can find all of the small number of global minima in minutes to hours on workstations based only on sequence information and no additional arbitrary constraints. It implies that conformational searching is not the limiting problem in folding proteins, at least for this model. The problem is adequate potential functions and chain representations.

Proteins have remarkable symmetries. We have asked here what is a minimal physical model that could account for the tertiary symmetries in proteins. We have found protein-like tertiary structures arise in a simple model in which chain molecules have specific sequences of H (solvent-averse) and P (neutral) monomers. Such symmetries are found in those HP sequences designed to fold to the fewest possible stable states.

We thank the Office of Naval Research for financial support.

1. Levitt, M. & Chothia, C. (1976) *Nature (London)* **261,** 552.
2. Richardson, J. & Richardson, D. C. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. Fasman, G. (Plenum, New York), pp. 1–97.
3. Branden, C. & Tooze, J. (1991) *Introduction to Protein Structure* (Garland, New York).
4. Kendrew, J. C. (1958) *Nature (London)* **181,** 662.
5. Skolnick, J. & Kolinski, A. (1991) *J. Mol. Biol.* **221,** 499–531.
6. Chan, H. S. & Dill, K. A. (1989) *Macromolecules* **22,** 4559–4573.
7. Yee, D. & Dill, K. A. (1993) *Protein Sci.* **2,** 884–899.
8. Lau, K. F. & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 638–642.
9. Chan, H. S. & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6388–6392.
10. Gruber, P. M. & Lekkerkerker, C. G. (1987) *Geometry of Numbers* (Elsevier, New York).
11. Yue, K. & Dill, K. A. (1993) *Phys. Rev. E* **48,** 2267–2278.
12. Dill, K. A. (1990) *Biochemistry* **29,** 7133–7155.
13. Lau, K. F. & Dill, K. A. (1989) *Macromolecules* **22,** 3986–3997.
14. Chan, H. S. & Dill, K. A. (1991) *J. Chem. Phys.* **95,** 3775–3787.
15. Shortle, D., Chan, H. S. & Dill, K. A. (1992) *Protein Sci.* **1,** 201–215.
16. Gronenborn, A. M. & Clore, G. M. (1994) *Science* **263,** 536–536.
17. Eisenberg, D., Weiss, R. M. & Terwillinger, T. C. (1984) *Proc. Natl. Acad. Sci. USA* **82,** 140–144.
18. Yue, K. & Dill, K. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4163–4167.
19. Yue, K., Fiebig, K., Chan, H. S., Thomas, P., Shakhnovich, E. & Dill, K. A. *Proc. Natl. Acad. Sci. USA* **92,** 325–329.
20. Handel, T. & Degrado, W. F. (1992) *Science* **112,** 6710–6711.
21. Yoder, M. D., Keen, N. T. & Jurnak, F. (1993) *Science* **260,** 1503–1507.
22. Pakula, .A. A. & Sauer, R. T. (1990) *Nature (London)* **344,** 363–364.
23. DeGrado, W. F., Wasserman, Z. R. & Lear, J. D. (1989) *Science* **243,** 622–628.
24. Kamtekar, S., Xiong, H., Babik, J. M. & Hecht, M. H. (1993) *Science* **262,** 1680–1685.
25. Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. (1990) *Science* **249,** 884.