Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6388–6392, August 1990
Chemistry

Origins of structure in globular proteins

(protein folding/compact polymers/steric forces/secondary structures/conformational entropy)

HUE SUN CHAN AND KEN A. DILL

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

Communicated by Bruno H. Zimm, May 7, 1990

ABSTRACT **The principal forces of protein folding—hydrophobicity and conformational entropy—are nonspecific. A long-standing puzzle has, therefore, been: What forces drive the formation of the specific internal architectures in globular proteins? We find that any self-avoiding flexible polymer molecule will develop large amounts of secondary structure, helices and parallel and antiparallel sheets, as it is driven to increasing compactness by any force of attraction among the chain monomers. Thus structure formation arises from the severity of steric constraints in compact polymers. This steric principle of organization can account for why short helices are stable in globular proteins, why there are parallel and antiparallel sheets in proteins, and why weakly unfolded proteins have some secondary structure. On this basis, it should be possible to construct copolymers, not necessarily using amino acids, that can collapse to maximum compactness in incompatible solvents and that should then have structural organization resembling that of proteins.**

What force causes the internal architectures in globular proteins? Although the dominant force of folding is the hydrophobic interaction (1), that type of interaction would seem to be too nonspecific to account for the considerable amounts of helices and sheets in globular proteins. Secondary structures are hydrogen-bonded. Therefore for many years, it was assumed that globular proteins would have simple regular crystal-like internal architectures (2), all helix or all sheet (3–9). It was expected that all proteins would be the same, irrespective of the amino acid sequence. For example, the discovery by Pauling *et al.* (3–7) of the α-helix and parallel and antiparallel sheets was premised on the assumption of "equivalence": that suitable bonding patterns should be the same for each amino acid. However, the appearance of protein structure offered several surprises. First, architectures in globular proteins are not simple and regular; about 53% of all residues are not in helices, sheets, or turns (10). Second, studies of model compounds in water show that hydrogen bonding is not a strong force for folding (11–14). Third, since different sequences code for different native structures, the internal architectures of globular proteins cannot be accounted for by backbone interactions; they must arise instead from *side-chain* interactions. It is only in the side chains that one amino acid differs from another and, therefore, by which one amino acid sequence differs from another. However, most hydrogen bonds in proteins (about 85%) are among peptide backbone groups rather than side chains (10). Fourth, the following evidence indicates that "intrinsic propensities" (i.e., the local interactions among neighboring residues along the chain that are responsible for helix-coil behavior) do not account for internal architecture in proteins. (*i*) For peptides in solution, helix stability increases with chain length (15–17) but, for the helices in globular proteins, the stabilities (observed in the distribution; see below) decrease with length (18–20). (*ii*) The most stable helices (21, 22) are typically longer than 15 residues and are not completely helical; in contrast in globular proteins, the average helix length is 12 residues, the most common helix length is less than 6 residues (18–20), and these are 100% helical up to nearly the denaturation temperature of the protein. (*iii*) "Local" sequence information has been insufficient to predict protein structures (23–26). What then is the origin of the regular and irregular conformations that comprise globular proteins?

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.

Lattice Approach

The purpose of the present work is to study patterns of spatial adjacencies of monomers in flexible polymer molecules having various degrees of compactness. We do so by exhaustive computer enumeration of every possible sterically allowed conformation of a chain molecule on a three-dimensional simple cubic lattice. First, it is helpful to define two terms to distinguish between two different types of spatial nearest neighboring monomers. We refer to two monomers that are adjacent in the chain sequence $(i, i + 1)$ as "connected" neighbors. We refer to monomers (i, j), $i \neq j$, that are spatially adjacent but not connected $(i \neq j - 1, j + 1)$ as "topological" neighbors. The secondary structures of proteins are simple patterns of intrachain topological neighbor pairs. A helix is defined by its series of contacts; $(i, i + 3)$, $(i + 1, i + 4)$, $(i + 2, i + 5)$, ... A parallel sheet is defined by the series: (i, j), $(i + 1, j + 1)$, $(i + 2, j + 2)$, ... An antiparallel sheet is defined by the series: (i, j), $(i + 1, j - 1)$, $(i + 2, j - 2)$, ...

Corresponding to the two classes of neighbor types, there are two classes of folding interactions: forces among connected neighbors and forces among topological neighbors. There is some justification for believing that the free energies of these two types of interaction may be approximately additive in polymers and proteins (27–29). To the extent that additivity holds, it implies that folding need not be described within a single model: a geometrically accurate model is required to account for the interactions among connected neighbors (i.e., of the bond angles ϕ, ψ, χ, for example); a topologically accurate model is required to account for the interactions among topological neighbors. Our focus in this work is on this topological problem: what is the nature of the monomer pairings (i, j) in the full spectrum of conformations of a chain molecule? We, therefore, require a model with which we can explore the full conformational space and with which we can accurately represent the "neighborness" of residue pairs, but it follows from the arguments above that the accuracy of representation of geometric detail is less important for this problem. This purpose is well-served by the use of a lattice model of a chain. The lattice serves as a device for representing all the possible conformations of the chain backbone at low resolution by a finite countable set of configurations using fixed bond angles (27, 28, 30).

First, we consider what errors are incurred by using a lattice model to represent secondary structures. The Pauling–

Corey–Branson α-helix cannot be represented with perfect accuracy on any lattice of low coordination number, but the helices observed in crystallography data bases of real proteins are likewise not perfect α-helices. The identification of helices varies depending on different definitions (18, 19, 31–37); in addition, some helices of different topology also occur in proteins (with low frequency), namely the 3_{10}- and π-helices (18, 38–40). Our four representations of helices on the simple cubic lattice are described elsewhere (41).

The quality of model helices can be assessed by comparing the set of spatial distances from residue i to $(i + 2, i + 3, i + 4, \ldots)$ in the model with the corresponding distances in the true α-helix. The distance between the (i, j) residue pair in an α-helix is denoted by $d_\alpha(i, j)$. Because of helical symmetry, this distance depends only on $|j - i|$ and is independent of the starting position i. On the other hand, the corresponding distance $d_m(i, j)$ in the four model helices does depend on i. The average distance $\bar{d}_m(i, j)$ is obtained by averaging over all possible inequivalent starting positions. The average distances $\bar{d}_m(i, j)$ for model helices and the corresponding distances $d_\alpha(i, j)$ for real α-helices are plotted in Fig. 1 for comparison.

We define a simple measure of the overall quality of representation of α-helices by models, in terms of the error $\delta(i, j) \equiv |d_\alpha(i, j) - \bar{d}_m(i, j)|/d_\alpha(i, j)$, the "energy-weighted" average,

$$\langle\delta\rangle_E \equiv \lim_{N\to\infty} \frac{\sum_{k=2}^{N} [\delta(i, i + k)/d_\alpha(i, i + k)^6]}{\sum_{k=2}^{N} [1/d_\alpha(i, i + k)^6]}. \quad [1]$$

According to this criterion, the models used here and shown in figure 9 of ref. 41 are the closest topological representations of a three-dimensional α-helix on the simple cubic lattice. Since no one of these four lattice helix models is clearly superior to the others or uniquely "correct" by several different criteria, and since, together, they broadly characterize the range of variation among observed helices in proteins, we make no further attempt to select from among them a single "best" model. All the cubic lattice models represent helix topology more accurately than the tetrahedral lattice model (see Fig. 1), despite the more accurate representation of a single bond angle on the tetrahedral lattice.

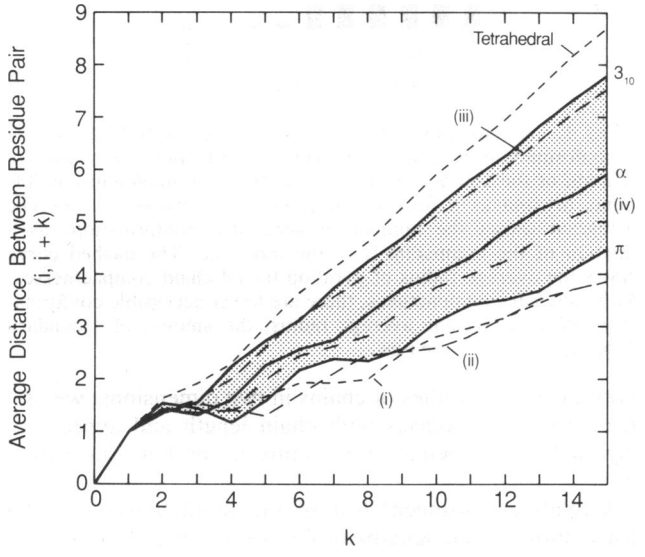

FIG. 1. Average distance between residues i and $i + k$ in helices as a function of separation, k, along the chain, in units of (virtual) bond lengths. This measure compares helical topologies, of α-, 3_{10}-, and π-helices (solid lines), of cubic lattice model helices (*i–iv*) in figure 9 of ref. 41, and of the tetrahedral lattice model helix (dashed lines). The shaded region between the 3_{10}- and π-helices indicates the variation allowed by observers of protein data bases in helix identification. Data compiled by Kabsch and Sander (18) show that the majority of helices are identified as α-helices, π-helices are very rare, but 3_{10}-helices are not uncommon—they make up approximately 10% of residues observed in helices.

The compactness ρ of a chain molecule is defined as the ratio of the number of intrachain topological contacts t (i.e., the number of dots on the contact map) to the maximum possible number of such contacts, t_{max},

$$\rho \equiv \frac{t}{t_{max}}, \qquad 0 \le \rho \le 1. \quad [2]$$

Maximally compact chains are characterized by $\rho = 1$. The value of t_{max} is determined by simple geometric packing considerations (for three dimensions, see ref. 41; for two dimensions, see ref. 42). Although the compactness is inversely related to the radius of gyration, these measures are not otherwise identical (41–43); the former is most convenient here.

Our purpose below is to explore the conformations of polymer chains of varying compactness, from low-density open conformations, $\rho \approx 0$, to high-density maximally compact conformations, $\rho = 1$. Chains with a single self-contact have been studied in detail elsewhere (41–44). We consider here conformations with any two specified intrachain contacts. This case is interesting because two contacts are the minimum required to specify the smallest building blocks of secondary structure. If a chain molecule has one given topological contact (i, j), then which contacts (i', j') are the most probable second ones to form? We count all the chain conformations that do not violate excluded volume constraints and that have both contacts, (i, j) and (i', j'); this number is $Q(i, j; i', j')$. If the total number of all possible conformations is Q_0, then the restriction of conformational space due to the two constraints is $Q(i, j; i', j')/Q_0$. The entropy loss for the formation of these two contacts, converted to a free energy, is

$$-kT \ln\left[\frac{Q(i, j; i', j')}{Q_0}\right]. \quad [3]$$

For a given first contact, we perform this enumeration for every possible second contact pair (i', j').

We have found that a useful way to display the results of this type of calculation is to insert these free energies at their respective positions on the contact map (41–43) and connect similar values with contour lines. This gives a "topological" free energy surface (see Fig. 2). The lowest free energies are shown as the darkest regions; the dot on that figure indicates the position of the first presumed contact (i, j). It is clear from Fig. 2 that out of approximately $n^2/2$ possible second contacts, only two are strongly preferred: those of the helix and antiparallel sheet. A chain prefers to simply "zip up" a helix or antiparallel sheet from a small loop because these are the configurations that cause the least further reduction in configurational entropy [i.e., given a small loop $(i, i + 3)$ as the initial "nucleus," the most favored second additional contacts are $(i + 2, i + 5)$ or $(i - 2, i + 1)$ or $(i - 1, i + 4)$]. The same feature is observed for chains on two-dimensional square lattices (43), an observation that was the motivation for the present work and suggests that this feature is not simply a function of dimensionality or geometry of the lattice. The present simulations take into account only conformational freedom and excluded volume, which are properties of

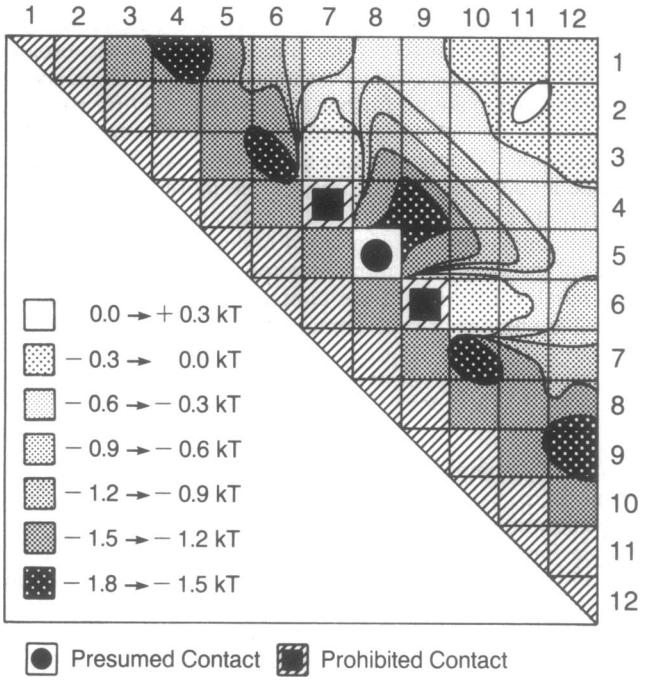

FIG. 2. Topological free energy surface. Given a contact between residue pair (5,8) in a 12-segment three-dimensional chain, these contours represent the relative free energies for formation of any second contact pair (i', j'). The antiparallel sheet (4,9) and/or the helical contacts (3,6), (7,10) are most favored relative to all other configurations, because they impose the least additional restriction on conformational freedom.

all chain molecules, and do not consider the effects of any specific interactions.

Now consider chain molecules of increasingly greater compactness. The probability that an arbitrarily chosen monomer is in a particular form of secondary structure is shown in Fig. 3, as a function of chain compactness and as a function of the radius of gyration of the chain. These distributions are of all positions of the monomer within the chain and are taken over the full ensemble of conformations of a given number of intrachain contacts (compactness) or of a given range of radius of gyration. The number of accessible conformations diminishes rapidly as the compactness increases. Within this ensemble of accessible conformations, the fraction of conformations containing secondary structures increases sharply as compactness increases. For example, about 50% of the monomers in a chain are in some form of secondary structure in the ensemble of conformations that are maximally compact. Thus any force that causes a chain to have many intrachain contacts (i.e., to become compact) will consequently cause formation of much secondary structure. For proteins, this force is presumably the hydrophobic interaction.

The result shown in Fig. 3 also implies that chain molecules which are relatively, but not maximally, compact should also have some degree of structure. This offers an explanation for why some denatured proteins have secondary structure (45–49): the ensemble of unfolded configurations of any weakly denatured protein will be of relatively high density (50) and thus should contain much secondary structure. On the other hand, proteins that are strongly denatured (for example, in high concentrations of urea or guanidinium hydrochloride) will be of much lower compactness and should, therefore, have much less secondary structure.

The fraction of residues participating in secondary structure is not much affected by chain length (over the narrow range tested) in these three-dimensional studies (Table 1). In

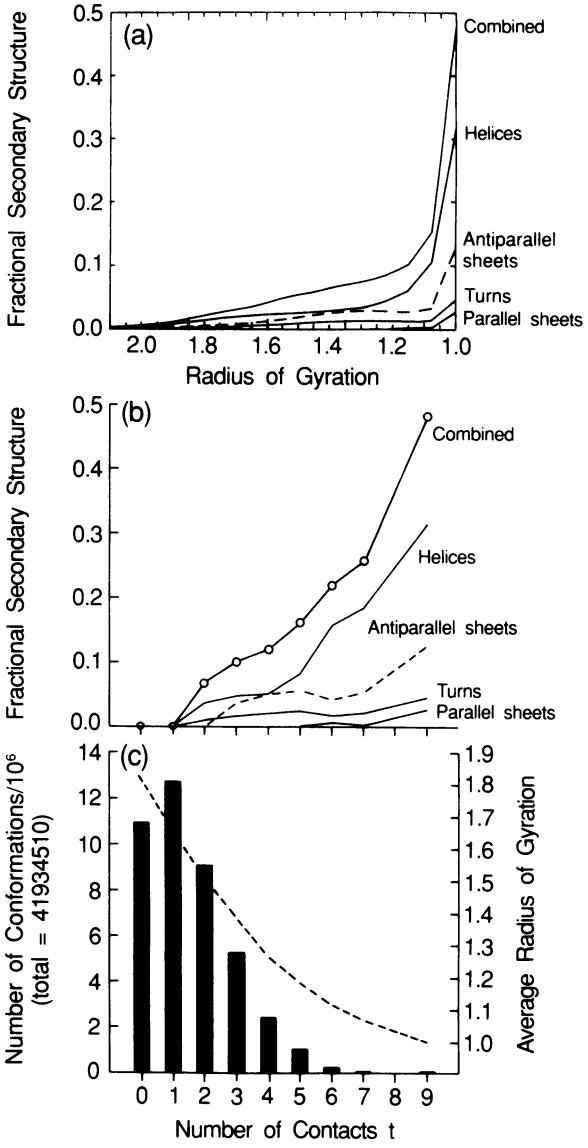

FIG. 3. Amount of secondary structure over the full ensemble of all conformations of 12-segment chains: as a function of the radius of gyration of the chains measured in units of the minimum possible radius (*a*) and as a function of chain compactness (*b*). (*c*) The histogram shows the number of accessible conformations as a function of the compactness of the molecule. The dashed curve shows the average radius of gyration for all chain compactnesses. With increasing compactness, there are fewer accessible conformations. With decreasing average radius, the amount of secondary structure strongly increases.

more extensive studies of chains in two dimensions, we find that structure increases with chain length and appears to approach 100% secondary structure in the long chain limit (42).

A relatively stringent test of the predictions is of the distributions of the lengths of helices and parallel and antiparallel sheets. The simulations are compared with the Kabsch and Sander observations (18) of protein crystal structures in Fig. 4; the agreement is good. For both the theory and data-base observations, the relative amounts of the shortest helices and parallel sheets are about the same, and both are about 25% as probable as the shortest antiparallel sheets. The prediction that antiparallel sheets are strongly favored is in agreement with the observations of Richardson (51, 52) that parallel sheets are less stable, always buried, and seldom occur in small protein domains. Also,

Table 1. Percentage secondary structure participation in compact molecules in simple cubic lattices

Structure(s)	% secondary structure		
	12 residues	18 residues	27 residues
Helix i	10.5	5.7	2.9
Helix ii	16.0	7.2	4.2
Helix iii	9.6	15.8	16.5
Helix iv	0.9	0.7	0.6
All helices	**31.5**	**26.0**	**22.5**
Antiparallel sheet	12.6	18.6	19.8
Parallel sheet	2.7	4.7	6.9
Turn	4.6	6.1	6.6
All sheets and turns	**18.0**	**26.8**	**29.1**
All structures	**48.2**	**51.2**	**49.8**

Definitions of various secondary structure types are given in ref. 41. The combined participation rates (in all structures) are slightly smaller than the sum of the participation rates in all helices and all sheets, because it is possible for residues to participate in both an antiparallel sheet and a type ii helix. Similarly, because it is possible for a residue to participate in more than one secondary structure, subtotals of all helices and all sheets and turns are less than the sum of the respective types of helices and sheets. Totals and subtotals are highlighted by bold numbers.

both the theory and the data-base observations agree that, for all structures, there is a monotonic decrease in probability with increasing length of the structure, the rate of decline increasing in the following order: helices < parallel sheets < antiparallel sheets. The minor differences between theory and experiment are due to the limited sizes of the chains in the simulations and otherwise have no significant bearing on the test of principle.

The total amount of secondary structure of each type is found as the integral under each of these distributions. Theory and experiment are in qualitative agreement. For example, the simulations predict 22–32% helix and 18–29%

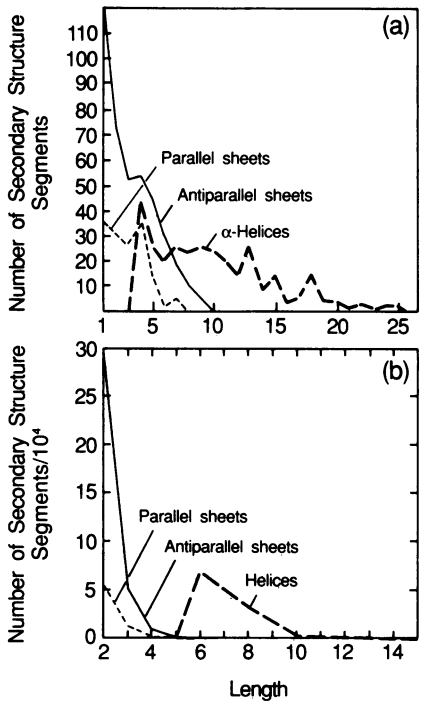

FIG. 4. Length distributions of secondary structures. (*a*) Data-base observations of Kabsch and Sander (18) on 62 proteins of different chain lengths. (*b*) Exhaustive simulations of maximally compact chains of 26 residues on a two-dimensional square lattice. "Length" refers to the number of residues in a secondary structural element.

sheet overall, in agreement with studies of protein crystal data bases showing 21–40% helix and 14–28% sheet (18, 19, 31–37). For example, the study of Chou and Fasman (31, 32) shows 890/2473 = 36% helix and 424/2473 = 17.1% sheet in 15 globular proteins; the study of Levitt and Greer (19) shows 3627/9213 = 39.4% helix and 2136/9213 = 23.2% sheet in 43 globular proteins; and the study of Kabsch and Sander (18) shows 21.2% helix, 13.9% antiparallel sheet, and 4.5% parallel sheet in 62 globular proteins. More quantitative comparison is not presently warranted for several reasons. (*i*) The data-base observations depend on the specific details of the definitions of structures and on data-base size. (*ii*) By neglecting zigzags, the simulations are conservative in estimating the number of sheets. (*iii*) These results may depend on chain length for longer chains (see above), and since lattice under- and overestimates due to chirality and other geometric details are unknown, more quantitative comparison is difficult.

The steric principle of internal architecture could also be tested by construction of copolymers, not necessarily comprised of amino acids, that could collapse to maximum compactness in poor solvents (i.e., those that are incompatible with a sufficient number of monomers in the chain) and that should thus be driven to the formation of internal structural organization. The experiments of Rao *et al.* (53) on random terpolymers of lysine, alanine, and glutamic acid are consistent with this expectation. They found that within a solution containing a very large number of different random sequences, a significant fraction of the chains is highly compact, and they observe 46% helix by circular dichroism measurements. Such copolymers might also provide the basis for the design of new materials with microscopic organization resembling that of proteins, but whose construction strategies would follow from well-established polymer chemistry.

This conclusion that internal architecture in proteins arises from steric packing would seem to leave no role for hydrogen bonds, a principal signature of secondary structures. We believe the role of hydrogen bonds is more subtle. Neither amino acids nor hydrogen bonds are required for helical conformations of chain molecules. Of the 176 crystal structures of synthetic polymers known to be reliable, 49 are planar zigzags (resembling sheets), and 79 are helices of 22 different types (54). Several crystalline polymers have helices closely resembling the α-helix, with three to four monomers per turn, including polybutadiene, polybutene, polyvinylnapthalene, polypropylene, and even fibrous sulfur, none of which form hydrogen bonds. In our view, the packing forces in proteins can drive, for example, the formation of periodic repeating stretches of (i, i + 3) contacts. What then "decides" that the geometry will be specifically that of the α-helix, as opposed to any of the dozens of other types of closely related helices, will be the local factors, including hydrogen bonding, that distinguish amino acids as monomers from other types of chemical units. Thus in our view hydrogen bonding and intrinsic propensities do contribute to internal organization but at the level of "fine-tuning" rather than as dominant forces.

The packing force is significantly stronger than intrinsic connected-neighbor interactions for the formation of short helices in globular proteins. (The "packing" free energy is defined as the free energy of a structure in the ensemble of compact conformations relative to the same structure within the full ensemble of all conformations.) The intrinsic helix/coil equilibrium constant of 1.05 (55) corresponds to $0.20kT$, for a 6-residue segment of four rotatable bonds. In contrast, our simulations show that the packing free energy favors the helix relative to the coil by approximately $1.7kT$ for a 6-residue segment in three dimensions.

The steric "force" for internal organization can, therefore, be viewed as follows. In the ensemble of all chain configu-

rations, helices and sheets are a very small fraction of the ensemble and are, therefore, entropically costly to form. However, in the much smaller ensemble of only compact configurations, helices and sheets are a relatively larger fraction of all possible conformations and, therefore, are entropically less costly to form. Therefore, any force toward compactness also decreases the entropic cost of forming secondary structures.

Conclusions

We have explored every possible conformation accessible to short chain molecules in three-dimensional space on a simple cubic lattice by computer enumeration. We conclude that internal architecture is a natural consequence of compactness in chain molecules. For proteins, compactness is a result of hydrophobic interactions. This view resolves several puzzles. (*i*) It shows how helical peptides that are not stable isolated in solution can be stable in globular proteins. (*ii*) It provides a single framework to account also for stabilities of sheets and irregular structures that are equally important components of protein architecture but that are intrinsically "nonlocal" in character. (*iii*) It explains why some unfolded proteins are observed to have secondary structure. It follows that other polymer molecules, not comprised of amino acids, when driven to the maximally compact state by solvent aversion of some of the residues, might also have protein-like internal structures.

We thank Fred Cohen, Kit Fun Lau, Greg Petsko, Jolanda Schreurs, and Dave Yee for helpful discussions and the National Institutes of Health, the University Research Initiative Program of Defense Advanced Research Projects Agency, and the Pew Scholars Program in the Biomedical Sciences for research support.

1. Kauzmann, W. (1959) in *Advances in Protein Chemistry*, eds. Anfinsen, C. B., Jr., Anson, M. L., Bailey, K. & Edsall, J. T. (Academic, New York), Vol. 14, pp. 1–63.
2. Wu, H. (1931) *Chin. J. Physiol.* **5**, 321–344.
3. Pauling, L., Corey, R. B. & Branson, H. R. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 205–211.
4. Pauling, L. & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 235–240.
5. Pauling, L. & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 251–256.
6. Pauling, L. & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 272–281.
7. Pauling, L. & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 729–740.
8. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958) *Nature (London)* **181**, 662–666.
9. Kendrew, J. C. (1961) *Sci. Am.* **205(6)**, 96–110.
10. Baker, E. N. & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol.* **44**, 97–179.
11. Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg Ser. Chim.* **29**, 223–229.
12. Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg Ser. Chim.* **29**, 230–259.
13. Klotz, I. M. & Franzen, J. S. (1962) *J. Am. Chem. Soc.* **84**, 3461–3466.
14. Klotz, I. M. & Farnham, S. B. (1968) *Biochemistry* **7**, 3879–3882.
15. Zimm, B. H. & Bragg, J. K. (1959) *J. Chem. Phys.* **31**, 526–535.
16. Goodman, M., Verdini, A. S., Toniolo, C., Phillips, W. D. & Bovey, F. A. (1969) *Proc. Natl. Acad. Sci. USA* **64**, 444–450.
17. Poland, D. C. & Scheraga, H. A. (1970) *Theory of the Helix-Coil Transition* (Academic, New York).
18. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
19. Levitt, M. & Greer, J. (1977) *J. Mol. Biol.* **114**, 181–293.
20. Srinivasan, R. (1976) *Indian J. Biochem. Biophys.* **13**, 192–193.
21. Marqusee, S. & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8898–8902.
22. Bradley, E. K., Thomason, J. F., Cohen, F. E., Kosen, P. A. & Kuntz, I. D. (1990) *J. Mol. Biol.*, in press.
23. Kabsch, W. & Sander, C. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1075–1078.
24. Argos, P. (1987) *J. Mol. Biol.* **197**, 331–348.
25. Thornton, J. (1988) *Nature (London)* **335**, 10–11.
26. Qian, N. & Sejnowski, T. J. (1988) *J. Mol. Biol.* **202**, 865–884.
27. Flory, P. J. (1953) *Principles of Polymer Chemistry* (Cornell Univ. Press, Ithaca, NY).
28. de Gennes, P.-G. (1979) *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, NY).
29. Dill, K. A. (1985) *Biochemistry* **24**, 1501–1509.
30. Hill, T. L. (1960) *An Introduction to Statistical Thermodynamics* (Addison-Wesley, Reading, MA).
31. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–222.
32. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245.
33. Kolaskar, A. S., Ramabrahmam, V. & Soman, K. V. (1980) *Int. J. Pept. Protein Res.* **16**, 1–11.
34. Ramakrishnan, C. & Soman, K. V. (1982) *Int. J. Pept. Protein Res.* **20**, 218–237.
35. Cantor, C. R. & Schimmel, P. R. (1980) *Biophysical Chemistry* (Freeman, San Francisco).
36. Creighton, T. E. (1983) *Proteins, Structures and Molecular Properties* (Freeman, New York), p. 230.
37. Richards, F. M. & Kundrot, C. E. (1988) *Proteins* **3**, 71–84.
38. Low, B. W. & Baybutt, R. B. (1952) *J. Am. Chem. Soc.* **74**, 5806–5807.
39. Donohue, J. (1953) *Proc. Natl. Acad. Sci. USA* **39**, 470–478.
40. Schulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure* (Springer, New York), p. 69.
41. Chan, H. S. & Dill, K. A. (1990) *J. Chem. Phys.* **92**, 3118–3135.
42. Chan, H. S. & Dill, K. A. (1989) *Macromolecules* **22**, 4559–4573.
43. Chan, H. S. & Dill, K. A. (1989) *J. Chem. Phys.* **90**, 492–509.
44. Jacobson, H. & Stockmayer, W. H. (1950) *J. Chem. Phys.* **18**, 1600–1606.
45. Brazhnikov, E. V., Chirgadze, Y. N., Dolgikh, D. A. & Ptitsyn, O. B. (1985) *Biopolymers* **24**, 1899–1907.
46. Ptitsyn, O. B. (1987) *J. Protein Chem.* **6**, 273–293.
47. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989) *Biochemistry* **28**, 7–13.
48. Shortle, D. & Meeker, A. K. (1989) *Biochemistry* **28**, 936–944.
49. Goto, Y. & Fink, A. L. (1989) *Biochemistry* **28**, 945–952.
50. Dill, K. A., Alonso, D. O. V. & Hutchinson, K. (1989) *Biochemistry* **28**, 5439–5449.
51. Richardson, J. S. (1977) *Nature (London)* **268**, 495–500.
52. Richardson, J. S. (1981) in *Advances in Protein Chemistry*, eds. Anfinsen, C. B., Edsall, J. T. & Richards, F. M. (Academic, New York), Vol. 34, pp. 167–339.
53. Rao, S. P., Carlstrom, D. E. & Miller, W. G. (1974) *Biochemistry* **13**, 943–952.
54. Tadokoro, H. (1979) *Structure of Crystalline Polymers* (Wiley, New York).
55. Lewis, P. N., Gō, N., Gō, M., Kotelchuck, D. & Scheraga, H. A. (1970) *Proc. Natl. Acad. Sci. USA* **65**, 810–815.