PROTEINS: Structure, Function, and Genetics 47:315–322 (2002)

Geometry and Physics of Proteins

Jayanth R. Banavar,[1]* Amos Maritan,[2,3,4] Cristian Micheletti,[2,3,4] and Antonio Trovato[5]
[1]*Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania*
[2]*International School for Advanced Studies (S.I.S.S.A.), Trieste, Italy*
[3]*Istituto Nazionale Fisica della Materia (INFM), Trieste, Italy*
[4]*Abdus Salam International Center for Theoretical Physics, Trieste, Italy*
[5]*Niels Bohr Institutet, København, Denmark*

ABSTRACT **A conceptual framework for understanding the protein folding problem has remained elusive in spite of many significant advances. We show that geometrical constraints imposed by chain connectivity, compactness, and the avoidance of steric clashes can be encompassed in a natural way using a three-body potential and lead to a selection in structure space, independent of chemical details. Strikingly, secondary motifs such as hairpins, sheets, and helices, which are the building blocks of protein folds, emerge as the chosen structures for segments of the protein backbone based just on elementary geometrical considerations. Proteins 2002;47:315–322.**

INTRODUCTION

The principal theme of this paper is the study of segments of the protein backbone modeled as a tube. In spite of much work on polymers and chains, a continuum description of a chain is plagued by singularities and, until very recently, there existed no continuum description of a tube of non-zero thickness. Our recent work[1] has shown how this problem can be solved. The key idea is to discard pair-wise interactions and to use suitable three-body interactions as the basic interacting unit. (For a discrete description of a tube, pair-wise interactions can exist in addition to the three-body interactions.) The constraint of a non-zero thickness approximately captures the role of steric constraints as embodied in the Ramachandran plot.[2]

There are two features that are common to all globular proteins: their backbone and hydrophobicity. For simplicity, consider a model of the backbone with identical particles tethered to each other along a chain with the usual degrees of freedom given by the dihedral and bond angles. When one has attractive pairwise interactions that promote compactness, one finds that there are a huge number of compact states with almost all of them having no significant secondary structure.[3,4] Our objective is to uncover a general principle, which can be used to winnow out all but the folds observed in Nature. In a recent paper, Przytycka et al.[5] have explored the possibility that the pattern of all-β folds results from an underlying simple grammar.

We present general arguments that helices of a specific pitch-to-radius ratio and hairpins and sheets emerge as the chosen structures when one considers a chain of effective atoms (say, the C_α atoms) with the prescribed three-body interactions. Strikingly, these structures are the building blocks of protein folds. These theoretical ideas are confirmed by the results of detailed simulations; on requiring that there be a certain minimum number of pairwise interactions between the effective atoms (or equivalently that there are at least that many pairs within a certain distance threshold), we determine the optimal conformation having the largest possible thickness. The conformations that emerge are space-filling in the interior and, therefore, are able to expel water from the core of the protein and include helices of the correct pitch-to-radius ratio and hairpins and sheets. Our results suggest a simple geometrical framework for thinking about proteins.

RESULTS

We have recently studied, in a general context, the self-interactions of a chain and have shown that a satisfactory continuum theory must entail discarding effective pair-wise potentials and employ, instead, a suitable three-body potential for chains.[1] (Such effective self-interactions are induced by the finer degrees of freedom, e.g., those of the solvent and the internal degrees of freedom not considered in a coarse-grained description.) A pair potential is unable to discriminate between whether the pair constitutes a local or a nonlocal interaction along the chain (see Methods). Note that the requirement that pairwise interaction be discarded is strictly needed only in a continuum description. For discrete chains, one might envision using different pair-wise interactions depending on the spacing along the chain. Nevertheless, such a scheme does not permit one to define the notion of the thickness of a chain conformation.

Armed with this insight obtained through a continuum analysis, let us return to a discrete system of particles forming a chain, and consider three particles that interact via the three-body potential. A characteristic length scale[1] entering such a potential is obtained by drawing a circle through the three particles and determining its radius (Fig. 1). The radius provides the contextual information

Grant sponsor: NASA; Grant sponsor: Cofin 99; Grant sponsor: NSF; Grant number: DMR-0080019.

*Correspondence to: J. R. Banavar, Dept. of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802. E-mail: banavar@psu.edu or A. Maritan, SISSA, Via Beirut 2–4, 34014 Trieste, Italy. Email: maritan@sissa.it

Received 21 August 2001; Accepted 11 December 2001

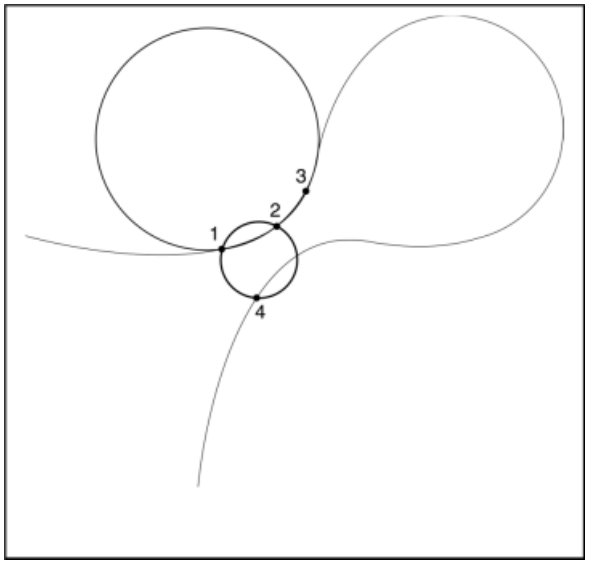


Fig. 1. Continuum description of a chain. The only length scale within the context of a two-body potential that one may define is proportional to the mutual distance between the particles, which vanishes when the two particles are in the immediate vicinity of each other, regardless of whether they are close along the chain [e.g., particles (1) and (2)] or not [e.g., particles (2) and (4)]. In order to distinguish between these two possibilities, one needs to consider at least three particles in a chain and the radius of the circle through them. When they are infinitesimally close to each other, particles (1), (2), and (3), which are contiguous on the chain, are characterized by a radius equal to the local radius of curvature of the chain. On the other hand, particles (4), (1), and (2) are characterized by a radius that provides a measure of the distance of approach of one part of the chain to another. A three-body potential that penalizes small values of the radii captures the physics of non-zero thickness, respects self-avoidance, and prohibits large curvature.

associated with the chain connectivity and can readily distinguish between local and non-local interactions along the chain. When the three particles are in contiguous locations in the chain, this is a measure of the local radius of curvature.[6] When one deals with non-local interactions along the chain, the radius provides a measure of the distance of approach of one part of the chain with the other.

The thickness of the chain may be defined by considering all triplets of particles and selecting the smallest among all the radii[7] and has proved to be useful in the theory of knots.[7,8] It has a simple interpretation of the maximum thickness of a uniform tube, whose axis is the chain, and that does not have sharp corners or self-intersect. The thickness provides a simple measure of the internal wiggle room or the free volume around the chain. A chain of non-zero thickness is perhaps the simplest way to capture the effects of steric constraints[2] in a generic manner and may be implemented by disallowing any conformations that lead to any three-body radius being less than the desired thickness. Indeed, in order to maximize the wiggle room, one may instead choose to maximize the smallest among the three-body radii to make the tube thickness as large as possible. Alternatively, in order to study compact conformations of a tube of non-zero thickness, one may sum over a Lennard Jones–like potential energy of all the triplets whose argument is the three-body radius.

Our computer simulations have been carried out with the main goal of determining the conformations of a tube subject to compaction. Compactness is promoted by requiring that there be at least a threshold number of pairwise contacts. Within this constraint, we seek tube conformations that have the largest possible tube thickness as measured by the three-body radii.

Before showing the results of our computer simulations, it is useful to think about what the consequences are on using such three-body potentials. A tube of unit radius corresponds to any of the three-body radii having a constraint of not being less than 1 unit and the forces promoting compaction in our thought experiment may be simply captured by requiring that as many triplets have a radius as close to unity as possible. (Note that the compactness criterion in our simulations, as described in the previous paragraph, is different from this and shows the generality of our result.)

Quite remarkably, for short segments of a protein, a helix with a special pitch-to-radius ratio[9,10] would be selected; both the radius of curvature of contiguous triplets and the radius associated with particles in successive turns of the helix would have the desired value of unity. While this special shape is determined by geometry alone and is in accord with the helix geometry found in proteins,[9,11] the helix radius (or pitch) is determined by the chemistry or more specifically, by the necessity to accommodate a hydrogen bond between backbone atoms of amino acids in successive turns. Such as helix, with a pitch about twice the preferred radius of 2.7 Å, is effective in squeezing out the solvent from within it. This is because a tube of this size is essentially space-filling in the interior.

The tight geometry of the helix poses a problem when there is steric incompatibility or equivalently an inability to form a tight-enough curve. The energy can then be lowered only by non-local triplets having the "right" radii. One way of accomplishing this might be to maintain the overall one-dimensional geometry and have multiple helices with a large local radius of curvature winding around each other.

A more versatile way to obtain non-local interactions is by means of a sheet. First, a strand in an extended conformation would form to locally accommodate the larger radius of curvature enforced by the steric incompatibility. In order to have as many triplets of the desired radius as possible (see Figs. 2 and 3), one would need interactions with a different part of the chain and the problem reduces to determining the optimal placement of two essentially independent parts of the chain. From symmetry considerations, one would expect the most favorable circumstances to occur when two such identical strands from different parts of the chain are in the vicinity of each other and when they both lie essentially in a plane. Note that two neighboring strands in a protein[12,13] are replicas of each other or mirror images of each other (corresponding to parallel and antiparallel sheets, respectively) in terms of the C_α atoms. They are located at a distance about twice the minimal radii, which allows the formation of a supporting framework for the assembly of the strands based on

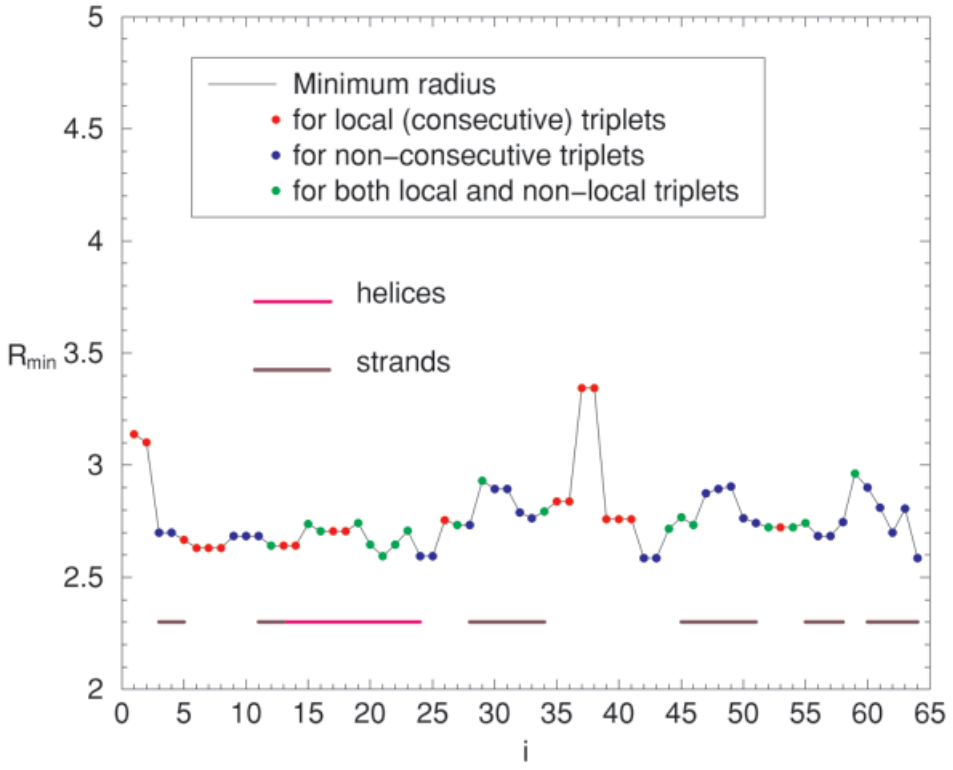


Fig. 2. Protein conformation viewed as a tube. We consider the backbone C^{α} atoms of protein (2CI2) in its native state. A given C^{α} atom is combined with all pairs of other C^{α} atoms to form triplets and the radii of circles passing through them are calculated. Shown is a plot of the minimum radius for each of the C^{α} atoms in the sequence. The red points of the curve signify that the smallest radius arises from three consecutive C^{α} atoms, the blue points denote the situation when non-consecutive triplets (for example, from strands forming a sheet) have the minimum radius, while the green points indicate the case when local as well as non-local triplets have similar radii within a 5% threshold (as is the case for the α-helix). **Bottom:** The secondary structure elements of the protein.[30]

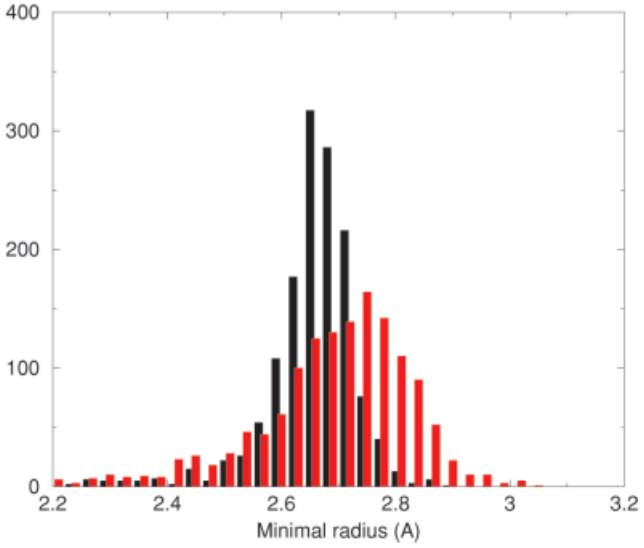


Fig. 3. Normalized histograms for the distribution of minimal radii for residues in α-helices (black) and β-sheets (red) in 50 unrelated globular proteins. Both distributions show small deviations from the average minimal radii (2% for helices and 5% for sheets). It is striking that for both classes of motifs (see also Fig. 2) the minimal radii cluster around half the typical distance (5.5 Å) between C_{α} atoms of residues involved in hydrogen bonds. This allows one to view a protein native state conformation as a harmoniously packed tube of approximately uniform thickness.

hydrogen bonds between atoms in neighboring strands. A sheet is formed by a repetition of the same process. Figure 3 shows that the minimal radii in β-sheets is larger than in α-helices, showing that the former may be thought of as a slightly larger tube than the latter.

When one considers longer segments of the proteins, it is not energetically favorable to have just one helix or one sheet because distant regions would not necessarily have triplets characterized by the preferred radius. Thus, there is a persistence length associated with a given secondary structure. In order to assemble the tertiary structure, which provides more energetically favored triplets from distinct secondary structure elements, one would need a mechanism for producing tight turns, which would entail having a small local radius of curvature. In proteins, this is facilitated by small amino acids like Glycine, which is often found in backward bends.

We have studied the energy landscape computationally and find that it is vastly different from what one obtains using conventional pair-wise interactions and a chain of infinitesmal thickness. The two commonly observed secondary structures, the sheet and the helix, arise from geometrical considerations dictated by the three-body description, which is needed to describe a tube of non-zero thickness, without any reference to specific sequences or hydrogen

bonding. Figure 4 shows the results of the optimal conformations of chains subjected to two competing constraints. In order to enforce compactness, we require that the chain made up of N particles adopt a conformation with at least N_c^* contacts. A contact is said to occur when two particles separated by at least three units along the chain are within a distance of 1.75 units from each other. (We have verified that our results are essentially the same, when this contact distance is varied between 1.2 and 1.9 units.) We also require that, within this constraint, the chain have as large a thickness, Δ, as possible. This is ensured by requiring that the smallest among the three-body radii be as large as possible.

The requirement of maximum thickness opposes the tendency to form compact conformations. Hence, the structures selected by our optimization procedure have just the required number, N_c^*, of contacts. We have explored the structures obtained for a range of N_c^* values for several values of the chain length, N. This allows one to capture, in the simplest manner, the effects of protein sequence heterogeneity which impacts, through steric effects on the maximum achievable local compactness.

The results are striking (see Figs. 4 and 5). Unlike the common experience with pair-wise potentials, in which essentially featureless compact conformations are selected as the ground states, we find two distinct regimes. For small values of N_c^*, one first gets a hairpin and then a sheet structure, whereas for larger values of N_c^*, a helix with a smaller thickness than the other motifs is obtained. For intermediate values of N_c^*, other structures are found (see Fig. 5). There is a balance between the benefits of having many contacts in the helix at the cost of a loss in conformational entropy. Earlier studies[9,10] have shown that a helix with a pitch-to-radius ratio, identical to that obtained here, is selected as the optimal shape of a string subject to a different compactness constraint of a maximum radius of gyration, underscoring the generality of the result. Furthermore, the geometry of the optimal helix is very similar to that observed in proteins.[9,11] The key point is that the number of candidate structures is extremely limited and all of them are approximately characterized by an equality of the local radius of curvature and the smallest non-local radius (this is, of course, exact for the helix). Furthermore, all the structures are space-filling and are, therefore, able to expel water from the interior of the protein.

There are several features of our results that are noteworthy. First, there are many fewer classes of ground state conformations with a three-body potential than for a pair-wise potential. The three-body interaction leads to a huge reduction in the large degeneracy of ground states and provides a selection mechanism in structure space. Two kinds of β-hairpins [Fig. 4(a) and (b)] emerge as degenerate ground states. More strikingly, the hairpin and the sheet are substantially planar, underscoring that the patterns employed as building blocks of a protein are low dimensional manifolds; the helix may be thought of as one-dimensional, whereas the sheet is two-dimensional. Such lower dimensional manifolds allow easy assembly in three-dimensional space. In all the conformations shown in Figure 4 (and to a slightly lesser extent for the conformations in Fig. 5), a curved tube of uniform cross-sectional area, whose axis coincides with the optimal conformation of a chain and with the maximum thickness possible without kinks and self-intersections, is found to be approximately space-filling; the available space is efficiently used and there are no empty nooks and crannies.

The ratio of the thickness of the helix to the sheet in the model calculation (Fig. 4) is approximately equal to 0.73, showing that larger amino acids can be more easily accommodated in a sheet. One, therefore, finds that the more compact helix with local interactions is the preferred secondary structure, unless it becomes sterically incompatible with the conformations of certain side chains.[14–17]

DISCUSSION

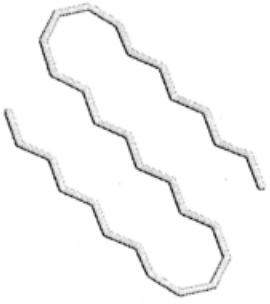
We now turn to a discussion of how many of the well-known properties of proteins may be understood within this framework. There is a huge number of sequences of amino acids that one may construct of even a modest length. For an overwhelming majority of these, one might expect that different parts of the sequence would attempt to take on conformations corresponding to pieces of secondary structure that simply do not fit together to form one of the folds. This inherent frustration[18] is absent for protein-like sequences and is responsible for selection in sequence space.

An energy landscape in structure space with relatively few energy minima has several consequences. Unlike the case with innumerable competing minima, the small number of folds is the premise for having a large basin of attraction for each of them. Second, the rich and varied repertoire of amino acids has undoubtedly been used by Nature in evolution to design sequences that are able to fold rapidly and reproducibly to just their native states. A given protein has only a few choices of putative folds that are predetermined based on geometrical considerations and not on chemical details. The selection from this limited menu, which is further reduced by the requirement that the native state be compatible with general chemical affinities and be free of steric clashes, is dictated by the appropriate choice of a small number of key residues that have distinctly better environments in the native state than in the competing folds. This scenario is consistent with the experimental evidence[19,20] showing that protein native states are usually robust against substitution of most residues with others of comparable volume and chemical attributes, except at a few key positions. Also, sequence design could be carried out in order to create a folding funnel[21,22] with the minimal amount of ruggedness in the energy landscape.

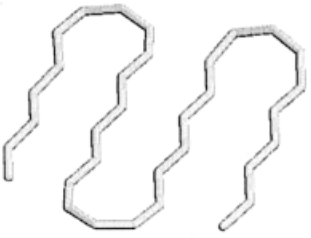
There are many sequences that map into a given structure because once a sequence has selected a fold as its native state, it is able to tolerate a significant degree of mutability except at certain key locations. The existence of such special folds would also explain the key role played by the native state topology[23–25] in determining many of the essential aspects of protein folding. Indeed, as stated

(a) $N = 31,\ N_c^* = 7,\ \Delta = 1.27$

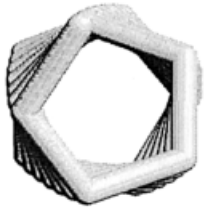
(b) $N = 31,\ N_c^* = 7,\ \Delta = 1.27$



(c) $N = 30,\ N_c^* = 8,\ \Delta = 1.27$



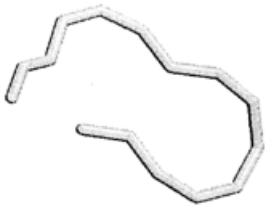
(d) $N = 33,\ N_c^* = 9,\ \Delta = 1.27$



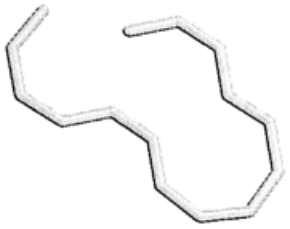
(e) $N = 33,\ N_c^* = 29,\ \Delta = 0.92$

(f) $N = 33,\ N_c^* = 29,\ \Delta = 0.92$

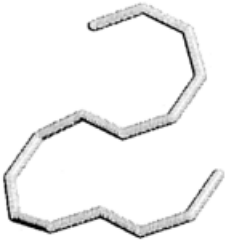
Fig. 4. Optimal conformations of a chain made up of N particles with a constraint that there be at least N_c^* contacts and as large a thickness, Δ, as possible. The values of N, N_c^*, and Δ are shown for each case. **e,f:** Two different views of the same helix. We have verified that similar structures are obtained for a range of N values, from 15 to 33. We have chosen to show only certain specific values of N and N_c here because they are representative of what one finds for other values. These types of structures (regular helices with a special pitch-to-radius ratio and the planar structures shown in the figure) are the only ones for which the interior is uniformly space-filling along the chain. The local radius of curvature and the smallest non-local radius are equal for all the particles. This allows the maximum efficiency for the expulsion of water from the interior of the conformation. Other types of conformations also emerge in our simulations (see Fig. 5).



(a) $N = 15$, $N_c^* = 2$, $\Delta = 1.27$

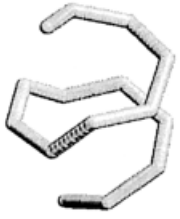


(b) $N = 15$, $N_c^* = 3$, $\Delta = 1.27$



(c) $N = 15$, $N_c^* = 3$, $\Delta = 1.27$

(d) $N = 15$, $N_c^* = 4$, $\Delta = 1.16$



(e) $N = 15$, $N_c^* = 6$, $\Delta = 1.06$

(f) $N = 15$, $N_c^* = 11$, $\Delta = 0.92$

Fig. 5. Optimal conformations for fixed length of the chain, $N = 15$, as a function of the number of contacts N_c^*. The values of N_c^* and Δ are shown for each case. Only the planar conformations with $\Delta = 1.27$ and the helix with $\Delta = 0.92$ are space-filling. Other optimal conformations are not and are representative of the structures chosen for other values of N as well. The maximum thickness conformation for $N = 15$ and $N_c^* = 2$ is a hairpin with a dangling end, which is not planar. Such trivially degenerate conformations arise when the numbers chosen for N and the constraints imposed by N_c^* do not fit each other. The transition between the planar and the helical regimes is smooth. For intermediate thickness values, optimal structures are somewhat less regular than helices or hairpins. On decreasing the thickness, one obtains first saddle-shaped conformations (a planar hairpin folded onto itself in a three-dimensional conformation) and then irregular though roughly helical-shaped conformations with an increasing number of contacts. The saddles and irregular helices are distinct from the conformations in Figure 4 in that the equality of the local and the smallest non-local radii hold for most but not all of the particles.

earlier, one would expect that the folding process is controlled by just a few key sites.[26] The powerful forces of evolution[27] operate within the fixed playground of these special folds, yielding "better" or more versatile sequences and multiple protein functionalities could arise within the context of a single fold.[28]

In summary, we have shown that the chain connectivity of a protein requires the use of a three-body description in order to capture the effect of steric constraints within the picture of a tube. We find the emergence of sheets and helices, which are the building blocks of protein tertiary structure. The energy landscape characterizing such a three-body interaction has relatively few minima and may explain the small number of protein folds. Indeed, if these folds result from geometrical considerations, one might expect that this universe of immutable folds forms the backdrop for protein evolution. Proteins can have their native states in just these folds and novel folds are not created by evolution.[27] Sequences that fold rapidly and reproducibly into the desired structure are promoted by evolution. The functionality of a protein is often dictated by a relatively small number of residues. Evolution can lead to proteins having the same fold but quite different functionality.[28]

METHODS

Three-Body Interaction in a Chain

Consider a chain of particles whose nearest neighbor separation is approximately 1 unit. The commonly used pair-wise potential considers a pair of particles but is unable to discern the context of the particles within the chain.[1] For example, if you were told that, in a given conformation, two particles are at a distance of 1 unit, you would not be able to know whether that was because they were neighboring particles with a distance constraint or whether two distant particles along the chain had strayed close to each other. A simple way to capture contextual information pertaining to the chain connectivity is by means of a three-body potential,[1] whose argument is the radius of the circle that passes through them (see Fig. 1). A small radius would imply either that the particles were from different parts of the chain that were in the vicinity of each other or the particles were neighbors along the chain in a region of high curvature. For a potential energy function that penalized small values of the radius and was constant for large radii, neighboring particles would be effectively non-interacting, if the local radius of curvature was high enough. This possibility of non-interacting neighboring particles is not afforded by the pair-wise potential.

The dynamics associated with a three-body potential are very likely to be quite different from that with a pair potential. In particular, frustrating influences that arise in pair-wise interactions, because they do not take into account the chain connectivity, are likely to be absent in many-body interactions.

Description of Computation

The conformations in Figures 4 and 5 were obtained starting from a random conformation of a chain made up of N equally spaced points (the spacing between neighboring points is defined to be 1 unit) and successively distorting the chain with pivot, crankshaft, and slithering moves commonly used in stochastic chain dynamics.[29] A Metropolis Monte Carlo procedure is employed with a thermal weight, $e^{+\Delta/T}$, where Δ is the thickness and T is a fictitious temperature set initially to a high value such that the acceptance rate is close to 1 and then decreased gradually to zero in several thousand steps. Self-avoidance of the optimal string is a natural consequence of the maximization of the thickness. The introduction of a hard-core repulsion between beads was found to significantly speed up convergence to the optimal solution and avoid trapping in self-intersecting structures. We have verified that the same values (within 1%) of the final thickness of the optimal strings are obtained starting from unrelated initial conformations.

ACKNOWLEDGMENTS

We are indebted to George Rose for advice and encouragement and to Oscar Gonzalez and John Maddocks for stimulating discussions. This work was supported in part by the Penn State MRSEC (NSF grant DMR-0080019).

REFERENCES

1. Banavar JR, Gonzalez O, Maddocks JH, Maritan A. Self-interactions of strands and sheets (unpublished).
2. Ramachandran GN, Sasisekharan V. Conformations of polypeptides and proteins. Adv Prot Chem 1968;23:283–438.
3. Hunt NG, Gregoret LM, Cohen FE. The origins of protein secondary structure. J Mol Biol 1994;241:214–225.
4. Yee DP, Chan HS, Havel TF, Dill KA. Does compactness induce secondary structure in proteins? J Mol Biol 1994;241:557–573.
5. Przytycka T, Srinivasan R, Rose GD. Recursive domains in proteins. Prot Sci 2002;11:409–417.
6. Rose and Seltzer [Rose GD, Seltzer JP. New algorithm for finding the peptide chain turns in a globular proteins. J Mol Biol 1977;113:153–164] have used the local radii of curvature of the backbone as an input in an algorithm for finding the peptide chain turns in a globular protein.
7. Gonzalez O, Maddocks JH. Global curvature, thickness and the ideal shapes of knots. Proc Natl Acad Sci USA 1999;96:4769–4773.
8. Katrich V, Olson WK, Pieranski P, Dubochet J, Stasiak A. Properties of ideal composite knots. Nature 1997;388:148–151.
9. Maritan A, Micheletti C, Trovato A, Banavar JR. Optimal shapes of compact strings. Nature 2000;406:287–290.
10. Stasiak A, Maddocks JH. Mathematics: best packing in proteins and DNA. Nature 2000;406:251–253.
11. Pauling L, Corey RB, Branson HR. The structure of proteins: two hydrogen-bonded helical conformations of the polypeptide chain. Proc Natl Acad Sci USA 1951;37:205–211.
12. Pauling L, Corey RB. Conformations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. Proc Natl Acad Sci USA 1951;37:729–740.
13. Richardson JS. β-sheet topology and the relatedness of proteins. Nature 1997;268:495–500.
14. Ramachandran and Sasisekharan [2] showed that steric considerations based on a hard sphere model leads to clustering of the backbone dihedral angles in two distinct α and β regions for non-glycyl and non-prolyl residues. The two backbone geometries that allow for systematic and extensive hydrogen bonding [11, 12] are the α-helix and the β-sheet obtained by a repetition of the backbone dihedral angles from the two regions respectively [15]. Poly-*L*-alanine, which is a good approximation to the backbone, readily forms a helix in water [16], but for heterogeneous side-chains the helix backbone sterically clashes with some side-chain conformers resulting in a loss of conformational entropy [17]. When the price in side-chain entropy is too large, an extended

backbone conformation results pushing the segment towards a β-strand structure [15].
15. Baldwin RL, Rose GD. Is protein folding hierarchic? I. Local structure and peptide folding. Trends Biochem Sci 1999;24:26–33.
16. Marqusee S, Robbins VH, Baldwin RL. Unusually stable helix formation in short alanine-based peptides. Proc Natl Acad Sci USA 1989;86:5286–5290.
17. Creamer TP, Rose GD. Side chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities. Proc Natl Acad Sci USA 1992;89:5937–5941.
18. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 1987;84: 7524–7528.
19. Kamtekar S, Schiffer JM, Xiong HY, Babik JM, Hecht MH. Protein design by binary patterning of polar and non-polar amino acids. Science 1993;262:1680–1685.
20. West MW, Wang WX, Patterson J, et al. De novo amyloid proteins from designed combinatorial libraries. Proc Natl Acad Sci USA 1999;96:11211–11216.
21. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways and the energy landscape of protein folding: A synthesis. Proteins 1995;21:167–195.
22. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nature Struct Biol 1997;4:10–19.
23. Micheletti C, Banavar JR, Maritan A, Seno F. Protein structures and optimal folding from a geometrical variational principle. Phys Rev Lett 1999;82:3372–3375.
24. Maritan A, Micheletti C, Banavar JR. Role of secondary motifs in fast folding polymers: a dynamical variational principle. Phys Rev Lett 2000;84:3009–3012.
25. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
26. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 1991;9:56–68.
27. Lesk AM, Chothia C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of globins. J Mol Biol 1980;136:225–270.
28. Holm L, Sander C. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. Proteins 1997;28:72–82.
29. Sokal AD. Monte Carlo methods for the self-avoiding walk. Nuclear Phys 1996;B47:172–179.
30. Daggett V, Li AJ, Itzhaki LS, Otzen DE, Fersht AR. Structure of the transition state for folding of a protein derived from experiment and simulation. J Mol Biol 1996;257:430–440.