

Exemplars of proteins

Jayanth R. Banavar^{1,*} and Amos Maritan^{2,†}

¹Department of Physics, 104 Davey Lab, The Pennsylvania State University, University Park PA 16802, USA

²INFN and Dipartimento di Fisica 'G. Galilei', Università di Padova, Via Marzolo 8, 35131 Padova, Italy

A unified framework for understanding proteins is presented, which provides links between the fields of protein science, polymer physics and the physics of liquid crystals.

SIMPLE paradigms often play an invaluable role in helping one understand complex systems. Living cells¹ are complex systems whose components interact selectively, nonlinearly and in a temporally orchestrated manner to yield coherent and robust behaviour. Proteins^{2–5}, a fundamental component of all living cells, fold into thousands of native-state structures under physiological conditions. For proteins, form determines functionality and the rich variety of observed forms underscores the versatility of proteins. Proteins act as enzymes; sometimes speeding up reactions by more than a factor of 10 billion, they interact with each other and other substrates and finally feedback into the gene to regulate the synthesis of other proteins. Francis Crick⁶ wrote about the challenge of the protein folding problem: ‘Nature performs these folding “calculations” effortlessly, accurately, and in parallel, a combination we cannot hope to imitate exactly. Moreover, evolution will have found good strategies for exploring many of the possible structures in such a way that shortcuts can be taken on the path to the correct fold. The final structure is a delicate balance between two numbers, the energy of attraction between the atoms, and the energy of repulsion. Each of these is very difficult to calculate accurately, yet to estimate the free energy of any structure we have to estimate their difference. The fact that it usually happens in aqueous solution, so that we have to allow for many water molecules bordering the protein, makes the problem even more difficult’. Our focus in this article, is on understanding protein behaviour by means of some paradigms.

Spin glasses

The idea of a spin glass⁷, originating in condensed matter physics, has spawned new ways of thinking in many distinct areas of science⁸. Spin glasses are characterized by random interactions between spins, resulting in the absence of any spin arrangement in which all the interactions are satisfied. This frustration is the hallmark of spin

glasses and leads to an energy landscape with multiple local minima and no long-range order in the spin arrangement. Challenging optimization problems such as the travelling salesman problem and the optimal placement of the components of a microchip can be mapped into the problem of finding the low-lying energy minima of a spin glass. Neural networks⁹, related to spin glasses, have been demonstrated to function as content addressable memories or to mimic pattern recognition. The stability and diversity inherent in the spin-glass energy landscape¹⁰ with multiple minima have been exploited to study prebiotic evolution.

Spin glasses and proteins

Under physiological conditions, proteins curl up rapidly and reproducibly¹¹ into a somewhat compact form called their native-state to shield the hydrophobic side chains of some of the amino acids from the surrounding water. Experiments¹² probing the dynamics of proteins revealed that the motions of proteins occur on a rugged landscape analogous to that found in spin glasses¹³. A random sequence of amino acids would have random and possibly conflicting interactions during an attempt at folding, leading to a rough energy landscape. The principle of minimal frustration¹⁴ provides an explanation for the rapid folding of proteins by suggesting a selection mechanism – proteins are well-designed sequences of amino acids for which the conflicts are minimized. Models in statistical mechanics have been used to show that proteins are characterized by a funnel-like energy landscape¹⁵ and spin-glass techniques have proved fruitful¹⁶ for the evaluation of approximate interaction energies for protein structure recognition and for the design of optimal sequences characterized by a pronounced funnel. Indeed, the common belief is that the amino acid sequence of a protein plays an all-important role in shaping its funnel-like landscape.

Common character of proteins

Globular proteins share many common characteristics^{17,18}. These include their ability to fold rapidly and reproducibly into their native-state structures¹¹; the total number of topologically distinct folds is only of the order of a few thousand^{19,20}; these structures are modular forms made up of simple building blocks, helices of a specific pitch-to-

*For correspondence. (e-mail: jayanth@phys.psu.edu)

†For correspondence. (e-mail: maritan@pd.infn.it)

radius ratio and almost planar sheets assembled from zigzag strands; these structures are flexible²¹ accounting for the ability of these proteins to carry out a wide variety of tasks; proteins are able to interact with each other and with ligands in a versatile yet robust manner; proteins are able to act as molecular targets of evolution; and proteins have a tendency to aggregate and form amyloids^{22,23}, which is implicated in human diseases such as Alzheimer's. Pauling, Corey and Branson^{24,25} and Ramachandran, Ramakrishnan and Sasisekharan²⁶ considered the common backbone of all proteins and showed that helices and sheets were indeed the preferred building blocks of proteins based on two independent mechanisms: the stabilization of the native-state structures by hydrogen bonds and the avoidance of steric clashes in the compact folded state. Proteins seem to occupy a phase of matter characterized not just by stability (as exemplified by their ability to fold reproducibly) and diversity (as exemplified by the many distinct folds) as in spin glasses, but also by sensitivity to the right types of perturbations (as needed for their functionality). There are several issues that arise from these remarkable facts. Are these stunning similarities an accident or is there a deeper underlying reason as to why proteins share these amazing attributes? Living matter is after all governed by physical laws. What is the phase of matter employed by nature to house these protein native-state structures? What are the essential ingredients that one must incorporate in order to develop a unified framework for understanding proteins?

Liquid crystals

Perhaps the simplest model of matter is a collection of hard spheres. Hard spheres exhibit both the fluid phase and the crystalline phase with a first-order phase transition between them on varying the packing fraction²⁷. Liquid crystals^{28,29} are well known to form several distinct structures which are stable, yet sensitive to changes in temperature, electric and magnetic fields. They exist in a phase that opens up in the vicinity of the phase transition between a liquid with no translational order and a crystal with translational order in all three directions. This is accomplished by breaking the isotropy of the constituent particles – the molecules of liquid crystals are rod-like or disc-like and are not spherical and can have translational order in fewer than three dimensions and/or orientational order. The fact that the liquid crystal phase is sandwiched between the liquid and the crystal phases confers on it its exquisite sensitivity. Thus, unlike spin glasses, liquid crystals are characterized by stability, diversity and sensitivity.

The phase of matter we seek is the analogue of the liquid crystal phase but for a chain molecule, which has the added benefit of providing contextual information. A tethered chain is the simplest flexible manifold that proteins can use in order to maintain the proximity of their constituents in the crowded cell environment.

Chain molecule viewed as a tube

The liquid crystal molecules are anisotropic objects. Such an anisotropy is inherent in any generic chain molecule because each of its constituents has a special axis (direction) associated with it, defined by the local tangent vector of the chain. A simple caricature of such a molecule is a chain of coins, which in a continuum description resembles a tube of non-zero thickness. The self-avoidance of a tube^{30,31} can be captured with a suitable three-body interaction: one takes all triplets of points along the axis of the tube and draws circles through them and ensures that none of the radii is smaller than the tube radius. This is a generalization of the self-avoidance criterion for hard spheres, for which one considers all pairs of spheres and ensures that their centres are no closer than the sphere diameter.

Emergent building blocks of protein structures

A tube of non-zero radius is a coarse-grained description of the protein backbone. It captures the avoidance of steric clashes – the room within the tube can be used to place the side chains of the amino acids. The inherent anisotropy of a tube leads to a preferential placement of nearby tube segments parallel to each other, a feature observed in sheets and helices and respecting the geometrical constraints of hydrogen-bond placement. This effect is enhanced when the tube size is comparable to the range of the self-attraction (due to hydrophobicity) that serves to promote compaction of the tube. These length scales are comparable to each other for a protein because both are effectively determined by the side chains of the constituent amino acids, which interact with each other through short-ranged interactions screened by the water. Indeed, the folding of a protein requires that tube segments snap into place nearly simultaneously to avail of the attraction accounting for the cooperative nature of the folding transition^{18,32}.

On squeezing a short chain of coins or tube (Figure 1), one obtains two kinds of structures: a helix with a pitch-to-radius ratio equal to that observed in protein structures (Figure 2) and zigzag strands assembled into an almost planar sheet. Helices are characterized by uniaxial anisotropy and a β -strand by biaxial anisotropy.

Proteins and liquid crystals

How does nature exploit the advantages of a liquid crystal phase, starting with the chain molecules of proteins? As we have seen, liquid crystal molecules are anisotropic as are the emergent building blocks of protein structures. Native-state protein structures are finite-size assemblies of these self-generated, anisotropic building blocks and may be thought of as nanodrops of liquid crystals. The

variety of protein native-state structures arises simply from the number of ways of arranging these building blocks together in the folded state, yet maintaining the desired sensitivity.

Symmetry, geometry and protein native-state structures

In accord with our observations, it has been recently demonstrated³³ that the key ingredients of symmetry (the choice of a coin as the basic constituent of the chain molecule) and geometry (arising from the constraints imposed by sterics and hydrogen-bond formation) lead to novel phase behaviour of a homopolymer chain (Figure 3). Holding several parameters of the coarse-grained Hamiltonian, such as hydrogen-bond energies and an energy reward for cooperativity of hydrogen bonds, fixed and on varying just two parameters, an overall pairwise attraction meant to mimic the hydrophobicity and an energy penalty for tight local turns (a flexible tube cannot be locally distorted into a conformation whose local radius of curvature is less than the tube radius), a rich phase diagram for a short homopolymer is obtained. Adjoining the swollen phase, one finds the marginally compact phase characterized by a variety of putative native-state struc-

tures, including a single helix, helix bundles, sheets and assembled structures of helices and sheets with a variety of topologies.

This result demonstrates that the free energy landscape of proteins is pre-sculpted by considerations of geometry and symmetry – the role of the sequence of a protein is the selection, from this predetermined menu, of its native state. Conversely, the principle of minimal frustration¹⁴ can be thought of as the mechanism for identifying protein-like sequences which fit well within this menu.

A consequence of the fixed menu of possible folds accounts for the fact that many sequences can fold into the same structure³⁴. From this point of view, the distinction between sequences is somewhat blurred in their native-state conformations. This makes the task of identifying the native-state of a given sequence a somewhat difficult one. In contrast, the denatured state^{35,36} ought to be sensitive to the sequence at the local (along the sequence) level, thereby revealing the individuality of the amino acids along the sequence. The native-state conformation can then be thought of as an unfrustrated global arrangement compatible with the local propensities³⁷. A dramatic simplification³² of the folding problem would arise, should the denatured state of a sequence contain key information regarding its native-state.

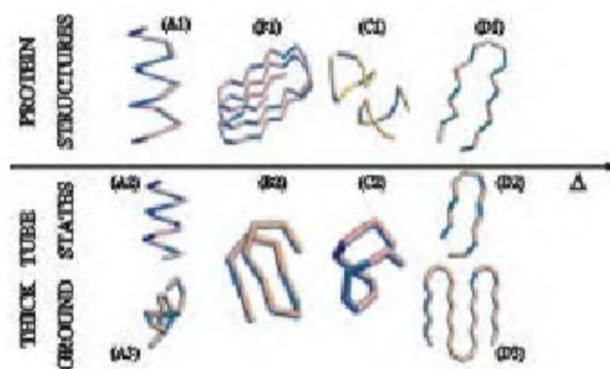


Figure 1. Building blocks of biomolecules and ground-state structures associated with the marginally compact phase of a short tube corresponding to a discrete chain of tethered coins of radius Δ . The axis in the middle indicates the direction along which the tube thickness Δ increases. (Top) Building blocks of biomolecules. (Bottom) Corresponding structures obtained as ground-state conformations of a short tube. (A1) is an α -helix of a naturally occurring protein, while (A2) and (A3) are helices obtained in our calculations – (A2) is a regular helix, whereas (A3) is a distorted helix in which the distance between successive atoms along the helical axis is not constant but has period 2. (B1) is a helix of strands in the alkaline protease of *Pseudomonas aeruginosa*, whereas (B2) shows the corresponding structure obtained in our computer simulations. (C1) shows the ‘kissing’ hairpins of RNA and (C2) the corresponding conformation obtained in our simulations. (D1) and (D2) are two instances of quasi-planar hairpins. The first structure is from the same protein as before (the alkaline protease of *P. aeruginosa*), while the second is a typical conformation found in our simulations. The sheet-like structure (D3) is obtained for a longer tube (see Banavar *et al.*⁵¹ for more details). The biomolecular structures are shown in the C α representation for proteins, and in the P representation for RNA kissing-hairpins.

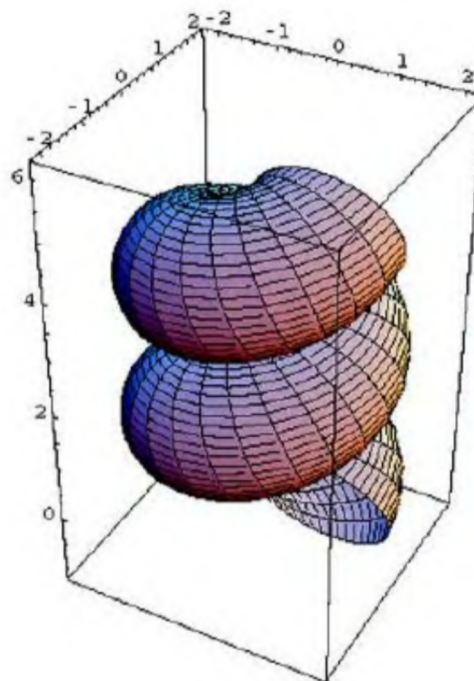


Figure 2. Space-filling optimal helix with a pitch-to-radius ratio $c^* \approx 2.512$ (drawn using Mathematica). This optimal value is determined by requiring that the radius of curvature of the helical curve is equal to half the minimum distance of closest approach between different turns of the helix. It can be shown⁵¹ that the planarity of hairpins and sheets is a consequence of this optimal space-filling criterion. The same geometrical feature is strikingly found to hold, within 3%, for α -helices occurring in the native-states of natural proteins⁵².

Amyloid phase

A range of human diseases such as Alzheimer's, spongiform encephalopathies and light-chain amyloidosis lead to degenerative conditions and involve the deposition of plaque-like material in tissue arising from the aggregation of proteins^{22,23,38,39}. We have found in computer simulations³² that multiple chains readily aggregate in extensive sheets with biaxial symmetry corresponding to an analogue of the biaxial nematic phase^{40,41} in liquid crystals (Figure 4). Our ideas suggest a unifying theme in the behaviour of proteins. Just as the class of cross-linked β -structures are determined from geometrical considerations⁴², the menu of protein native-state structures is also determined by the common attributes of globular proteins: the inherent anisotropy associated with a chain molecule and the geometrical constraints imposed by hydrogen bonds and steric considerations.

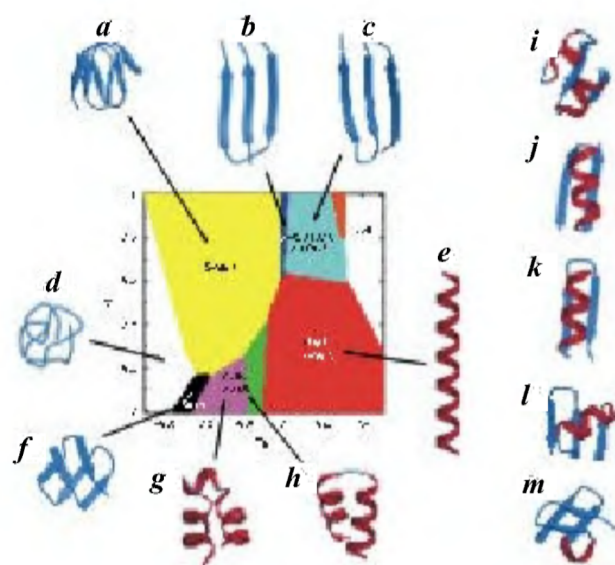


Figure 3. Phase diagram of ground-state conformations of a homopolymer. These conformations were obtained by means of Monte-Carlo simulations of chains of 24 C^α atoms. e_R and e_W denote the local radius of curvature energy penalty and the solvent-mediated interaction energy respectively. Over 600 distinct local minima were obtained in our simulations in different parts of the parameter space, starting from a randomly generated initial conformation. The temperature is set initially at a high value and then decreased gradually to zero. *a–c*, *e–h* are Molscript representations of conformations which are the ground states in different parts of the parameter space, as indicated by arrows. Helices and strands are assigned depending on the type, local or non-local, and number of hydrogen bonds. *i–m* are competitive local minima. In the orange phase, the ground state is a two-stranded β -hairpin (not shown). Two distinct topologies of a three-stranded β -sheet (dark and light blue phases) are found corresponding to conformations shown in (*b*) and (*c*) respectively. The white region in the left of the phase diagram has large attractive values of e_W and ground state conformations are compact globular structures with a crystalline order induced by hard sphere packing considerations⁵³ and not by hydrogen bonding (conformation *d*).

Evolution and natural selection

Crick⁶ wrote, 'Biologists must constantly keep in mind that what they see was not designed, but rather evolved'. Traditionally, the framework of evolution in life works through two aspects of organization called the genotype and the phenotype. The genotype is the heritable information encoded in the DNA, which is translated through the RNA molecules into proteins. The phenotype is valuable for adaptation and, at the molecular level, plays a key role in natural selection. One conventionally assumes that there is a selection of phenotypes which leads to an enhancement in the numbers of the genotype. Furthermore, mutations of the genotype lead to the possibility of new phenotypes.

Let us consider the situation at two levels: the sequence level (which is the genotype because it is a direct translation from the evolving DNA molecules) and the structure level, which we can think of as the phenotype. As pointed out by Maynard-Smith⁴³, as the sequence undergoes mutation, there must be a continuous network that the mutated sequences can traverse without passing through any intermediaries that are non-functioning. Thus, one seeks a connected network in sequence space for evolution by natural selection to occur. There is considerable evidence, accumulated since the pioneering suggestion of Kimura⁴⁴, and King and Jukes⁴⁵, that much of evolution is neutral. The experimental data strongly support the view that the

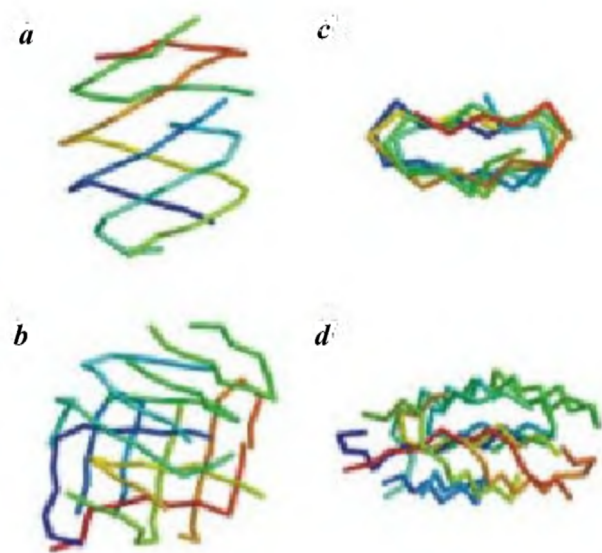


Figure 4. Aggregated structures³² formed by five and ten chains of length 12 with hydrophobic parameter $e_W = -0.08$ and local bending penalty $e_R = 0.2$. We show the lowest energy conformations obtained in long simulations for five chains at $T = 0.18$ (*a*), and for ten chains at $T = 0.2$ (*b*). The five-chain system is confined within a cubic box of side $L = 80$ Å, whereas the ten-chain system is confined within a cubic box of side $L = 100$ Å. The conformations shown in (*c*) and (*d*) are the same as those in (*a*) and (*b*), but viewed from a different angle.

'random fixation of selectively neutral or very slightly deleterious mutants occurs far more frequently in evolution than selective substitution of definitely advantageous mutants'⁴⁶. Also 'those mutant substitutions that disrupt less the existing structure and function of a molecule (conservative substitutions) occur more frequently in evolution than more disruptive ones'⁴⁶. Thus, while one has a 'random walk' in sequence space that forms a connected network, there is no similar continuous variation in structure space.

These facts are in accord with our result of a pre-sculpted free-energy landscape that is shared by all proteins and has thousands of local minima corresponding to putative native-state structures – not too few because that would not lead to sufficient diversity and not too many because that would lead to too rugged a landscape, with little hope that a protein could fold reproducibly and rapidly into its native-state structure. Indeed, many proteins share the same native-state fold³⁴ and often the mutation of one amino acid into another does not lead to radical changes in the native-state structure, underscoring the fact that it is not the details of the amino acid side chains that sculpt the free-energy landscape, but aspects of symmetry and geometry that are common to all proteins. In this respect, the phase of matter that comprises the native-state structures is one that is determined by physical law, rather than by the plethora of microscopic details, in analogy with the limited menu of possible crystal structures.

Summary

We conclude by revisiting the classic theoretical work of Pauling^{24,25} and Ramachandran²⁶. Both of them considered the protein backbone, which is the common part of all proteins. Pauling and coworkers explored the types of structures that are consistent with both the backbone geometry and the formation of hydrogen bonds. They predicted that helices and sheets are the structures of choice in this regard (Figure 5 *a* and *b*). Ramachandran and coworkers carried out their pioneering work more than a decade after Pauling. They considered the role of excluded volume or steric interactions between the adjacent amino acids in reducing the available conformational phase space (Figure 5 *c*). Astonishingly, the two significantly populated regions of the Ramachandran plot correspond to the α -helix and the β -strand. Even though hydrogen bonds and sterics are not related to each other, they are both promoters of helices and sheets. Is this concurrence of events a mere accident? The marginally compact phase of short tubes has helices and sheets as its preferred structures. In order for nature to take advantage of this phase of matter, proteins, which obey physical laws, may have been selected to conform to the tube geometry. Hydrogen bonds serve to enforce the parallelism of nearby tube segments, a feature of both helices and sheets, while ster-

ics emphasizes the non-zero thickness of the tube and serves to position it in the marginally compact phase. Because the marginally compact phase is a finite size effect, proteins tend to be relatively short compared to conventional macromolecules, including DNA. What is remarkable, however, is that the lengths of the covalent and hydrogen bonds and the rules of quantum chemistry conspire to provide a perfect fit to the basic structures in this novel phase. Indeed, proteins seem to be a vivid example of the adaptation of nature to her own laws.

Biochemist Arthur Kornberg⁴⁷ wrote: 'What chemical feature most clearly enables the living cell and organism to function, grow and reproduce? Not the carbohydrate stored as starch in plants or glycogen in animals, nor the depots of fat. It is not the structural proteins that form muscle, elastic tissue, and the skeletal fabric. Nor is it DNA, the genetic material. Despite its glamour, DNA is simply the construction manual that directs the assembly of the cell's proteins. The DNA is itself lifeless, its language cold and austere. What gives the cell its life and personality are enzymes. They govern all body processes; malfunction of even one enzyme can be fatal. Nothing in

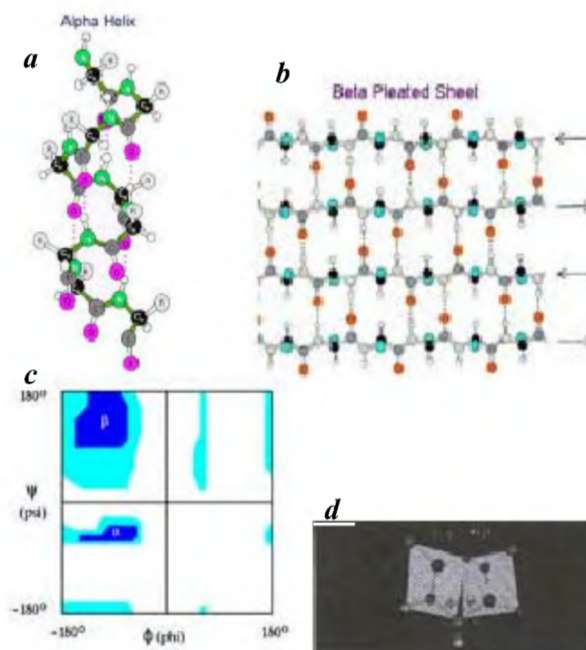


Figure 5. Pauling and Ramachandran revisited: (Top) 'Classic' structures of α -helix (*a*) and pleated β -sheet (*b*). Main-chain backbone atoms and C^β atoms of the side-chain groups are shown (colour codes are different for *a* and *b*). Hydrogen bonds, which stabilize the structures are shown as dashed lines. (Bottom) Ramachandran plot (*c*) describing how the torsional degrees of freedom (ψ , ϕ), the backbone dihedral angles within an all-atom representation, are constrained by steric effects. Coloured areas in the plot correspond to allowed regions in conformational space. The structures (*a*) and (*b*) stabilized by hydrogen bonding indeed lie squarely within the sterically accessible regions. An example of a dipeptide conformation disallowed because of steric hindrance is shown in (*d*).

nature is so tangible and vital to our lives as proteins, and yet so poorly understood and appreciated by all but a few scientists'.

Earlier, Bernal¹⁷ had stated the challenge of the protein problem: 'Any effective picture of protein structure must provide at the same time for the common character of all proteins as exemplified by their many chemical and physical similarities, and for the highly specific nature of each protein type'.

Attempts at creating a framework for understanding proteins using ideas from polymer physics were unsuccessful, as stated by Flory⁴⁸: 'Synthetic analogs of globular proteins are unknown. The capability of adopting a dense globular configuration stabilized by self-interactions and of transforming reversibly to the random coil are peculiar to the chain molecules of globular proteins alone'.

The ideas presented here bridge the gap between polymer physics and protein science through the recognition that chain molecules ought to be viewed as being made up of anisotropic objects. The self-generated building blocks of protein structures, the α -helix and the, β -strand with uniaxial and biaxial symmetries respectively, allow the creation of liquid-crystal-like phases of chain molecules, the proteins. Our findings are in complete accord with the words of Anfinsen¹¹: 'Biological function appears to be more a correlate of macromolecular geometry than of chemical detail'.

Ziman⁴⁹ wrote: 'In science, to echo Beethoven's dictum about music, "Everything should be both surprising and expected"'. The ideas presented here represent a different way of thinking about the protein problem, but the results of the analysis are entirely in accord with the wealth of information from protein experiments.

We conclude with an admonition to theoretical physicists from Francis Crick, 'Physicists are all too apt to look for the wrong sorts of generalizations, to concoct theoretical models that are too neat, too powerful, and too clean. Not surprisingly, these seldom fit well with data. To produce a really good biological theory, one must try to see through the clutter produced by evolution to the basic mechanisms. What seems to physicists to be a hopelessly complicated process may have been what nature found simplest, because nature could build on what was already there'. Indeed, protein native-state structures are determined by the overarching features of geometry and symmetry and provide a fixed backdrop⁵⁰ for evolution to act in shaping sequences and functionalities.

1. Alberts, B. *et al.*, *Molecular Biology of the Cell*, Garland Pub., New York, 2002; Voet, D. and Voet, J. G., *Biochemistry*, John Wiley, New York, 2003.
2. Creighton, T. E., *Proteins: Structures and Molecular Properties*, W. H. Freeman and Co., New York, 1993.
3. Fersht, A., *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman and Co., New York, 1999.
4. Lesk, A. M., *Introduction to Protein Architecture: The Structural Biology of Proteins*, Oxford University Press, Oxford, 2000.
5. Finkelstein, A. V. and Ptitsyn, O., *Protein Physics: A Course of Lectures*, Academic Press, New York, 2002.
6. Crick, F., *What Mad Pursuit*, Basic Books, New York, 1988.
7. Binder, K. and Young, A. P., Spin glasses: Experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.*, 1986, **58**, 801; Mezard, M., Parisi, G. and Virasoro, M., *Spin Glass Theory and Beyond*, World Scientific, Singapore, 1987.
8. Anderson, P. W., Spin glass VII: Spin glass as paradigm. *Phys. Today*, 1990, **43**, 9.
9. Hopfield, J. J., Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 1982, **79**, 2554.
10. Anderson, P. W., Suggested model for prebiotic evolution: The use of chaos. *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 3386.
11. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science*, 1973, **181**, 223.
12. Frauenfelder, H., Sligar, S. G. and Wolynes, P. G., The energy landscape and motions of proteins. *Science*, 1991, **254**, 1598.
13. Stein, D. L., A model of protein conformational substrates. *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 3670.
14. Bryngelson, J. D. and Wolynes, P. G., Spin glasses and the statistical-mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 7524.
15. Leopold, P. E., Montal, M. and Onuchic, J. N., Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 8271; Wolynes, P. G., Onuchic, J. N. and Thirumalai, D., Navigating the folding routes. *Science*, 1995, **267**, 1619; Dill, K. A. and Chan, H. S., From Levinthal to pathways to funnels. *Nature Struct. Biol.*, 1997, **4**, 10.
16. Goldstein, R. A., Luthey-Schulten, Z. A. and Wolynes, P. G., Protein tertiary structure recognition using optimized Hamiltonians with local interactions. *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 9029; Bryngelson, J. D., Onuchic, J. N., Succi, N. D. and Wolynes, P. G., Funnels, pathways and the energy landscape of protein folding: A synthesis. *Proteins*, 1995, **21**, 167; Klimov, D. K. and Thirumalai, D., A criterion that determines the foldability of proteins. *Phys. Rev. Lett.*, 1996, **76**, 4070; Chan, H. S. and Dill, K. A., Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins*, 1998, **30**, 2.
17. Bernal, J. D., Structure of proteins. *Nature*, 1939, **143**, 663.
18. Banavar, J. R. and Maritan, A., Colloquium: Geometrical approach to protein folding: A tube picture. *Rev. Mod. Phys.*, 2003, **75**, 23.
19. Chothia, C. and Finkelstein, A. V., The classification and origins of protein folding patterns. *Annu. Rev. Biochem.*, 1990, **59**, 1007.
20. Chothia, C., One thousand families for the molecular biologist. *Nature*, 1992, **357**, 543.
21. Jacobs, D. J., Rader, A. J., Kuhn, L. A. and Thorpe, M. F., Protein flexibility predictions using graph theory. *Proteins*, 2001, **44**, 150; Rader, A. J., Hespeneheide, B. M., Kuhn, L. A. and Thorpe, M. F., Protein unfolding: Rigidity lost. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 3540.
22. Kelly, J. W., The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.*, 1998, **8**, 101.
23. Dobson, C. M., Protein folding and disease: A view from the first horizon symposium. *Nature Rev. Drug Discovery*, 2003, **2**, 154.
24. Pauling, L., Corey, R. B. and Branson, H. R., The structure of proteins: two hydrogen-bonded helical conformations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA*, 1951, **37**, 205.
25. Pauling, L. and Corey, R. B., Conformations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. USA*, 1951, **37**, 729.
26. Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V., Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*,

- 1963, **7**, 95; Ramachandran, G. N. and Sasisekharan, V., Conformations of polypeptides and proteins. *Adv. Protein Chem.*, 1968, **23**, 283.
27. Chaikin, P. M. and Lubensky, T. C., *Principles of Condensed Matter Physics*, Cambridge University Press, Cambridge, 2000.
28. Chandrasekhar, S., *Liquid Crystals*, Cambridge University Press, Cambridge, 1977.
29. de Gennes, P. G. and Prost, J., *The Physics of Liquid Crystals*, Oxford University Press, Oxford, 1995.
30. Banavar, J. R., Gonzalez, O., Maddocks, J. H. and Maritan, A., Self-interactions of strands and sheets. *J. Stat. Phys.*, 2003, **110**, 35.
31. Gonzalez, O. and Maddocks, J., Global curvature, thickness and the ideal shapes of knots. *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 4769. The theorem proved in the context of knot topologies shows the remarkable connection between the three-point recipe and the tube thickness.
32. Banavar, J. R., Hoang, T. X., Maritan, A., Seno, F. and Trovato, A., A unified perspective on proteins – a physics approach. *Phys. Rev. E* (submitted).
33. Hoang, T. X., Trovato, A., Seno, F., Banavar, J. R. and Maritan, A., Geometry and symmetry prescript the free-energy landscape of proteins. *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 7960.
34. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. and Sauer, R. T., Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science*, 1990, **247**, 1306; Lim, W. A. and Sauer, R. T., The role of internal packing interactions in determining the structure and stability of a protein. *J. Mol. Biol.*, 1991, **219**, 359; Heinz, D. W., Baase, W. A. and Matthews, B. W., Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 3751; Matthews, B. W., Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.*, 1993, **62**, 139.
35. Rose, G. D., Getting to know u. *Adv. Protein Chem.*, 2002, **62**, xv; Baldwin, R. L., A new perspective on unfolded proteins. *Adv. Protein Chem.*, 2002, **62**, 361.
36. Shortle, D., The expanded denatured state: An ensemble of conformations trapped in a locally encoded topological space. *Adv. Protein Chem.*, 2002, **62**, 1.
37. The importance of local information is underscored by the fact that many schemes for secondary-structure prediction do moderately well with just local sequence information. See also Srinivasan, R. and Rose, G. D., LINUS: A hierarchic procedure to predict the fold of a protein. *Proteins*, 1995, **22**, 81, on the importance of local interactions in protein folding.
38. Radford, S. E. and Dobson, C. M., From computer simulations to human disease: Emerging themes in protein folding. *Cell*, 1999, **97**, 291; Bucciantini, M. *et al.*, Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature*, 2002, **416**, 507; Dumoulin, M. *et al.*, A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. *Nature*, 2003, **424**, 783; Chiti, F., *et al.*, Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, 2003, **424**, 805.
39. Prusiner, S. B., Prions. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 13363.
40. Freiser, M. J., Ordered states of a nematic liquid. *Phys. Rev. Lett.*, 1970, **24**, 1041; Madsen, L. A., Dingemans, T. J., Nakata, M. and Samulski, E. T., Thermotropic biaxial nematic liquid crystals. *Phys. Rev. Lett.*, 2004, **92**, 145505; Acharya, B. R., Primak, A. and Kumar, S., Biaxial nematic phase in bent-core thermotropic mesogens. *Phys. Rev. Lett.*, 2004, **92**, 145506.
41. Pitard, E., Garel, T. and Orland, H., Protein folding, anisotropic collapse and blue phases. *J. Phys. I Fr.*, 1997, **7**, 1201.
42. Fandrich, M. and Dobson, C. M., The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.*, 2002, **21**, 5682.
43. Maynard Smith, J., Natural selection and the concept of protein space. *Nature*, 1970, **225**, 563.
44. Kimura, M., Evolutionary rate at the molecular level. *Nature*, 1968, **217**, 624.
45. King, J. L. and Jukes, T. H., Non-Darwinian evolution. *Science*, 1969, **164**, 788.
46. Kimura, M., How genes evolve: A population geneticist's view. *Ann. Genet.-Paris*, 1976, **19**, 153.
47. Komberg, A., *For the Love of Enzymes*, Harvard University Press, Cambridge, 1989.
48. Flory, P. J., *Statistical Mechanics of Chain Molecules*, Wiley, New York, 1969, p. 301.
49. Ziman, J., *Reliable Knowledge*, Cambridge University Press, Cambridge, 1978, p. 71.
50. Denton, M. and Marshall, C., Laws of form revisited. *Nature*, 2001, **410**, 417.
51. Banavar, J. R., Flammini, A., Marenduzzo, D., Maritan, A. and Trovato, A., Geometry of compact tubes and protein structures. *ComplexUs*, 2003, **1**, 4.
52. Maritan, A., Micheletti, C., Trovato, A. and Banavar, J. R., Optimal shapes of compact strings. *Nature*, 2000, **406**, 287.
53. Zhou, Y., Hall, C. K. and Karplus, M., A first-order disorder-to-order transition in an isolated homopolymer model. *Phys. Rev. Lett.*, 1996, **77**, 2822.

ACKNOWLEDGEMENTS. We are indebted to Trinh Hoang, Flavio Seno and Antonio Trovato for their invaluable contributions to the work presented here. We are grateful to Marek Cieplak, Alessandro Flammini, Oscar Gonzalez, Tim Lezon, John Maddocks, Davide Marenduzzo, Cristian Micheletti and Saraswathi Vishveshwara and Igor Volkov for useful discussions and collaboration. This work was supported by PRIN 2003, FISIR 2001, INFM, NASA, NSF IGERT grant DGE-9987589 and NSF MRSEC at Penn State.

Received 31 May 2004; accepted 15 June 2004