Tubes near the edge of compactness and folded protein structures

2003 J. Phys.: Condens. Matter 15 S1787

(http://iopscience.iop.org/0953-8984/15/18/312)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **15** (2003) S1787–S1796 PII: S0953-8984(03)54816-0

Tubes near the edge of compactness and folded protein structures*

J R Banavar[1], A Flammini[2], D Marenduzzo[2], A Maritan[2,3] and A Trovato[4,5]

[1] Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA
[2] International School for Advanced Studies (SISSA) and INFM, V. Beirut 2-4, 34014 Trieste, Italy
[3] The Abdus Salam Centre for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy
[4] Niels Bohr Institutet, Blegdamsvej 17, 2100 København Ø, Denmark
[5] INFM—Dipartimento di Fisica 'G Galilei', Università di Padova, Padova, Italy

Received 15 October 2002
Published 28 April 2003
Online at stacks.iop.org/JPhysCM/15/S1787

Abstract

The protein folding problem has remained unsolved despite many decades of effort. Motivated by many of the puzzles in this field, we have carried out studies of the zero-temperature geometry of a short tube of non-zero thickness subjected to attractive interactions promoting its compaction. Here we present a review of our recent research on this subject. On varying a parameter, the ratio of the tube thickness to the range of attractive interactions, we find a 'transition' between two highly degenerate phases. The transition region in between is characterized by a marked decrease in the degeneracy—the ground state structures are 'marginally compact' and there is efficient space filling of the core region by the tube. Several classes of structures, especially helices and sheets, emerge as the chosen ground state conformations, many of which are building blocks of biomolecules. We also present results for the temperature–thickness phase diagram of longer tubes. We benchmark our results against experimental observations of proteins.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Since the pioneering experiment of Anfinsen [1], and the work of Ramachandran and Sasisekharan [2], the formidably complex protein folding problem has resisted solution in spite of concerted efforts over many decades. Previous investigations have ranged from coarse grained models (see e.g. [3, 4] for some examples) to atomistic detail simulations [5]

* Talk presented by A Maritan.

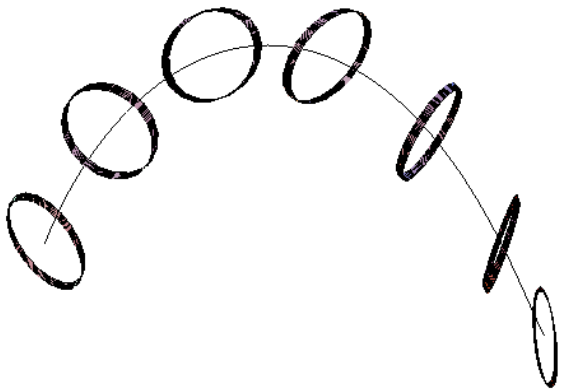

Figure 1. A portion of the protein is replaced by a tube with a non-zero thickness, which is discretized, for example, by means of a chain of coins.

(see [6] for a recent review). The best hope for a solution to this long-standing problem is an identification of the essential features underlying it. Our aim here is to present a brief review of our recent research in the development of a unified framework for understanding the common characteristics of small globular proteins.

Polymer chains and strings have been the subject of much study in recent years. In contrast, there has been little, if any, work on tubes of non-zero thickness. A major motivation for our study is a simplified picture of a segment of a protein as a short tube of non-zero thickness (figure 1) arising from the presence of side chains and the need to avoid steric overlaps [2]. Physically, the additional steric constraints given by the presence of side chains is better captured by a self-avoiding tube than by a conventional chain of spheres. This heuristic idea might be given a more quantitative basis by studying the local thickness of a protein residue for protein in their native structures. Plotting the probability distribution of the local thicknesses, one finds a sharply peaked function around its mean value, which is $\sim$0.27 nm [7]. In an ideal tube such a distribution would be a delta function: this lends credibility to a model of a uniform flexible tube as a first order approximation to a protein backbone. Inhomogeneities in the side chain do play a role, e.g. glycine is a very small aminoacid which does not impose severe steric constraints and indeed in real structures it is often positioned in loop regions [8]. Recent work has shown that, in order to provide a continuum description of a tube of non-zero thickness, one must discard pair-wise interactions and work with appropriately chosen three-body interactions as the basic interacting unit [9]. The key difficulty with pair-wise interactions is that the knowledge of the mutual distance between a pair of interacting particles does not provide any information on how far apart the particles are along the chain.

How might one define the thickness of a tube associated with a chain conformation? A simple procedure would be to construct a tube whose axis coincides with the chain and inflate the tube uniformly until it intersects with itself or has sharp corners. A natural definition of the thickness is then the radius of this largest tube. A tube with a large thickness has more space for internal rearrangements of the side chains than a thinner tube. This thickness can also be modelled using three-body interactions. For any three constituent particles of a chain, one defines, as the characteristic length scale, the radius of the circle passing through them. The thickness is obtained by computing the radius associated with all triplets (contiguous or otherwise) and selecting the smallest among these radii [10]. Thus one may simply choose a potential energy which is a sum of three-body terms whose argument is the radius and whose form has a hard core at short distances—any radius (local when one is dealing with contiguous

triplets, or non-local otherwise) is forbidden to take on a value less than the thickness of the tube [9].

Consider a discrete representation of a uniform tube of non-zero thickness undergoing compaction promoted by pair-wise attractive interactions (which are compatible with a discrete description) of a given range. We mimic the tube by considering discrete equally spaced particles lying along the axis of the tube (these could be e.g. the α-carbon atoms of the protein backbone or the backbone atoms of an RNA molecule). The three-body constraints, as described above, may be used to capture the notion of a non-zero thickness (induced by other degrees of freedom) and the attractive interactions are assumed to operate between the discrete particles.

When X, the ratio of the thickness of the tube to the range of the attractive interactions, is very large compared to unity, the tube is so fat that it is unable to benefit from the attractive interactions. One then obtains a swollen phase—all self-avoiding conformations of the tube are equally likely and a vast majority is not effective in filling the space in the core of the structure (recall that globular proteins fold in order to squeeze out water from the hydrophobic core). On the other hand, for a tube with a very small X compared to unity, one obtains many conformations, leading to a glassy energy landscape [11][6]. In this phase, the compacting interactions dominate and the three-body radii constraints do not play any role.

On varying X, we thus expect two regimes, the phase in which the tube thickness is not relevant and the swollen phase. Both phases are highly degenerate. One expects a 'transition' (the quotation marks are meant to alert the reader that our focus is on short tubes and true phase transitions occur only in the infinite size limit) region for intermediate values of X. In this region, there is a rich interplay of the pair-wise attractive interactions and the constraints imposed by the three-body interaction. In order for the tube to avail itself of the benefits of the attractive interaction, it must position itself selectively leading to a dramatic decrease in the degeneracy. In this region of parameter space, the forces promoting compaction just set in. As a result, one obtains 'marginally compact structures' which are space filling in the interior.

In order to determine the nature of these marginally compact structures, we model the axis of the tube of non-zero thickness R_0 as a one-dimensional chain, whose bonds are of fixed length (set equal to unity without loss of generality—all other lengths will be measured in these units from now on) and that connect N balls. The thickness of the tube is captured by disallowing conformations for which $R_0 > \min_{i\neq j\neq k} R_{i,j,k}$, where $R_{i,j,k}$ is the radius of the circle going through the centres of the balls i, j and k. Note that $R_{i,j,k} = \frac{r_{i,j}r_{j,k}r_{i,k}}{4A_{i,j,k}}$, where $A_{i,j,k}$ is the area of the triangle through i, j and k and $r_{i,j}$ is the distance between the centres of the ith and the jth balls. The interaction between non-consecutive balls is modelled via a two-body potential with a hard core, enforcing self-avoidance, and a square well allowing the formation of a contact only up to a threshold distance, R_1:

$$V(r_{i,j}) = \begin{cases} \infty & \text{if } r_{i,j} < 2R_{hc} \\ -1 & \text{if } 2R_{hc} < r_{i,j} < R_1 \\ 0 & \text{if } R_1 < r_{i,j}. \end{cases} \tag{1}$$

For the results that we show here, R_1 will be set to 1.6 and R_{hc} to 0.55. These values have been selected in order to mimic the protein backbone formed by the α-carbon atoms, in which the distance between successive Cα is on average 0.38 nm and the hydrogen bond mean length is 0.6 nm. At the same time, we have verified that our results are robust to reasonable variations in these values. A Lennard-Jones potential for the two-body interaction

[6] In the extreme limit of truly long range interactions, all the conformations of the swollen phase are again ground states but this time availing of the attractive interactions.

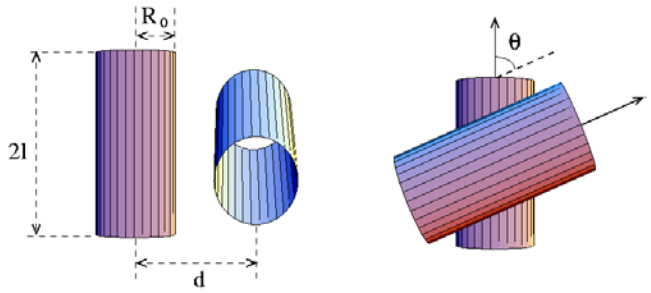

Figure 2. Interacting segments of a tube. The cylinders have a thickness R_0 and their axes lie in two planes parallel to each other at a mutual distance d.

energy, with an equilibrium distance slightly larger than the neighbouring site distance, is expected to give similar results to those presented here, provided the potential is truncated in order to remove the unphysical effects of a long range tail. Also note that in our formalism the two-body potential form has been chosen primarily for its simplicity and no conceptual changes are needed if e.g. the beads are charged and the tube describes a polyelectrolyte chain (of, course, the potential energy of interaction will have to be modified accordingly). The three-body potential instead must retain the hard constraint form discussed above in order to describe a hard tube. A Lennard-Jones form for the three-body constraint ought to be useful for molecular dynamics studies of soft tubes.

Before reporting the results of our numerical calculations, it is useful to discuss the key physical consequences of considering a tube instead of a chain of spheres. Let us consider the situation depicted in figure 2 where two segments of a cylindrical polymer lie in parallel planes (which are at a distance d apart), with their axes making a mutual angle θ between one another. The effects of the compacting potential may be measured by estimating the lengths of the tube axes whose points lie within a distance of R_1, the attraction range. The mutual interaction energy of the tubes is thus given by the double integral (extending over the length of the two portions):

$$\epsilon(\theta) = \int \mathrm{d}x \int \mathrm{d}y \, H\left(R_1 - \sqrt{d^2 + x^2 + y^2 - 2xy\cos\theta}\right), \tag{2}$$

where H is the Heaviside function, defined to be equal to unity if its argument is positive, and zero otherwise. The interaction energy $\epsilon(\theta)$ is plotted for two limiting cases in figure 3 and it should be compared with the case of a chain of spheres for which it is constant. When d is much smaller than R_1, the situation does not really differ from that of a conventional polymer with only hard sphere constraints, whereas when d is of the order of R_1, the attractive interaction energy has a marked minimum when the tubes are parallel. Though this oversimplified reasoning neglects the fact that the two portions are not independent of each other (they are, after all parts of the same tube) and thus their mutual configuration must satisfy additional constraints, we expect this reasoning to describe rather realistically the situation of dense tubes or of a tube which is in the collapsed low energy phase. Recent analysis of the native state structures of proteins [12] lends support to this picture with selective relative alignment of different portions of the protein backbone which are in contact with each other.

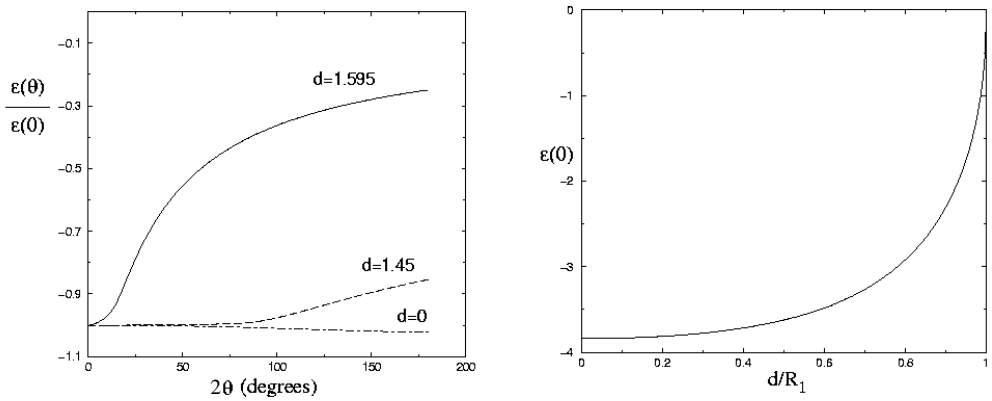

Figure 3. The left-hand plot shows the function $\epsilon(\theta)/\epsilon(0)$ for three values of d/R_1. For $d/R_1 \ll 1$ the interaction is practically isotropic; whereas the anisotropy becomes important as d/R_1 increases. The right-hand plot shows the interaction energy of two parallel tube segments as a function of d/R_1. The scale of the interaction energy becomes close to zero as one approaches the marginally compact limit and thus one would expect to have a transition at relatively low temperatures. This, in turn, suggests that the entropic effects near the transition are not significant, leading to the possibility of an all or nothing transition with few intermediates or competing conformations, as is commonly observed for small globular proteins.

2. Ground states of short tubes

In this section, we report on the ground state conformations for a short tube on varying the tube thickness. Let us consider N beads in the backbone of a short (discretized) tube. We denote by $N_c(N, R_0)$ the maximum number of contacts (each gaining an energy of -1 according to equation (2)) that can be made respecting the hard-core and the three-body constraints. On general grounds, we expect that, for a fixed N, N_c is a decreasing function of R_0 and that the decrease occurs in discrete steps corresponding to an inability to form the same number of contacts as before. Physically, for a given contact energy, one would choose the largest possible thickness in order to provide as much internal wiggle room (for the sidechains) within the tube as possible. This suggests that, for every window of thickness in which the number of contacts is fixed, the optimal conformation for a physical polymer stays at the maximum possible thickness.

In order to find the ground states, we have performed Monte Carlo simulated annealing for several values of N. The polymer is moved according to a fictitious dynamics consisting of pivot, reptation and crankshaft moves as is also common in the studies of conventional polymers [13]. The configurations in figure 4 result from numerical simulations with $N = 14$. Similar results are obtained for values of N between six and 20–25.

In order to improve the optimization efficiency, we have also used more complex algorithms based on parallel tempering techniques. The results shown here have proved to be algorithm independent. We have also compared the resulting structures with the optimal configurations to be found in sets of curves that can be treated 'exactly'. In particular we considered regular and generalized helices. In a regular discrete helix, whose axis, with no loss of generality, coincides with the z axis, two successive beads are at a unit distance from each other. The helix is therefore characterized uniquely by the rise per bead, v, and by the azimuthal (around z) rotation angle per bead, π/m. Here m, not necessarily an integer, corresponds to the number of beads per turn of the helix. The position of the ith bead, $\boldsymbol{r}(i)$,

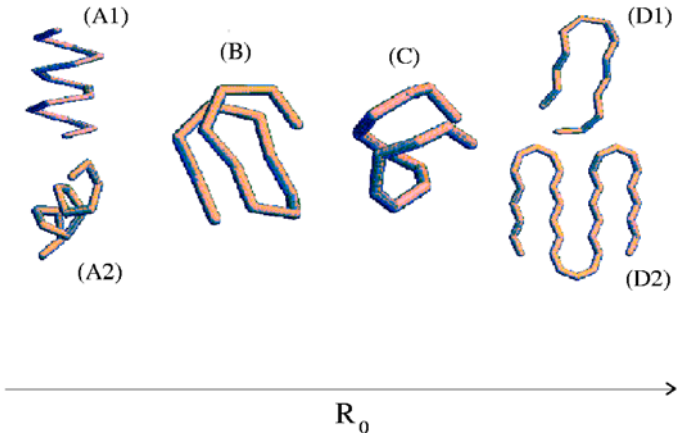

Figure 4. Five 'marginally compact' conformations for different tube thicknesses. (A1) and (A2) are the helices obtained in our calculations—(A1) has a regular contact map whereas (A2) is a distorted helix in which the distance between successive balls along the helical axis is not constant but has period two, i.e. the distance has the same value every other ball. (B) shows a helix of strands as obtained in our computer simulations. (C) Displays one typical saddle conformation we obtained in our simulations with $R_0 = 0.95$. Finally (D1) is an instance of quasi-planar hairpins, while (D2) gives an example of a longer chain assembled into a sheet.

$(i = 1, \ldots, N)$, is thus

$$r(i) = \begin{pmatrix} R\cos(2\pi \mathrm{i}/m) \\ R\sin(2\pi \mathrm{i}/m) \\ \mathrm{i}v \end{pmatrix}, \tag{3}$$

where the radius R is fixed by the condition that two successive beads have a relative distance equal to unity, i.e.

$$2R^2(1 - \cos(2\pi/m)) + v^2 = 1. \tag{4}$$

Generalized helices are obtained by allowing a variable rise per bead. We will consider only those helices with a rise per bead that depends periodically on the bead index.

The picture coming out from the numerical optimizations is the following.

For small R_0 one gets a highly degenerate phase with $N_c = 45$ for $N = 14$—the conformations that the tube adopts depend rather strongly on the details of the pair-wise potential. For R_0 between ∼0.75 and 0.8, there is an energy plateau in which the degeneracy is greatly reduced and helices are the ground states. Furthermore, for the tube with the largest thickness in this plateau, one obtains a specific helix as the unique ground state (see figure 4(A1)). For R_0 between 0.8 and 0.98, several classes of conformations compete. As R_0 increases, they are generalized helices (figure 4(A2)), helices made of strands (figure 4(B)) and saddles (which are planar hairpin conformations distorted into three-dimensional structures) (see figure 4(C)). Other more disordered conformations also compete but, for any given number of contacts, the thickest ground state structure is regular and almost non-degenerate.

Helices, whether regular or generalized, are excluded from being the ground states when the tube thickness exceeds $R_0 \sim 0.943$ which is obtained when two parallel straight lines (successive turns of the helix treated as circles with infinite radius) are parallel and face each other at the bond length R_1.

For $R_0 > 0.98$, the ground state structures become more and more planar, first locally then globally. For large N, the winning planar structures entail the combinations of strands

into a sheet structure. (We find that sheet structures persist for N as large as 24, whereas the persistence length for helices should be smaller.) For two zig-zag antiparallel strands facing each other, the maximum thickness is obtained (leaving aside the edge effect of how the strands are connected together in a hairpin) when one has a space-filling conformation, where the local and the smallest non-local radii of curvature have the same value. This happens at $R_0 \sim 1.2124$, when $R_1 = 1.6$.

The swollen phase occurs for even larger values of tube thickness and is characterized by two sets of configurations. The first set consists of all configurations with their ends in contact—the thickest tube which is able to make one contact is a closed polygon with N edges of unit length and one of length R_1. The second set consists of all configurations with no contacts: the maximum thickness is obtained with a straight, infinitely fat tube. Indeed, starting from the zig-zag conformation, all the (non-degenerate) conformations corresponding to the largest possible thickness at a fixed energy share the intriguing property that the local and the smallest non-local radius are exactly equal. It is worth noting that the optimal helix of figure 4(A1) has a ratio of the smallest non-local to the local radius of around 0.97 which is very close to the corresponding value 1.01 which characterizes α helices occurring in proteins [14].

Let us now discuss to what extent our ground states resemble secondary motifs in biomolecules. Our calculations predict the occurrence of helices and sheets, the building blocks of protein structures [15, 16]. In addition to these, we found other marginally compact conformations that bear a qualitative resemblance to secondary folds in biopolymers. Helices analogous to figure 4(A2) with an irregular contact map occur, e.g., in the HMG protein NHP6a [17] with pdb code 1CG7. Moreover, the 'kissing hairpins' [18] of RNA (pdb code 1KIS) are made of two distorted and twisted hairpin structures, very similar to the saddle conformation in figure 4(C), which is a hairpin distorted into a three-dimensional structure. Finally, helices of strands also known as β helices are found experimentally in many protein native states, see e.g. zinc metalloprotease [19] (pdb code: 1KAP).

3. Phase diagram and properties of longer tubes

We turn now to the study of thermodynamic properties of a longer thick polymer. We simulated tubes of up to 60 beads with a parallel tempering algorithm, adopting the same moves as used in the simulated annealing. In figure 5 we show the phase diagram for a 40-bead tube, as inferred by the specific heat peaks. We can distinguish two regimes. The first occurs for thin tubes ($R_0 < 0.8$ roughly): in this case we observe two transitions. The first critical temperature separates the high T swollen phase from a disordered globular phase, and the second the disordered globular phase from a compact one in which the different portions of the tube that are in contact tend to orient parallel to each other, similarly to thermotropic liquid crystal molecules in the nematic phase. A second regime is found for thicker tubes. In this case there is only one transition, from a swollen to a globular ordered phase. Our calculations indicate that the latter transition and the aforementioned second transition for a thin tube are first order, while the first transition (for a thin tube) is second order.

A similar situation was observed in [20], where the authors show (by means of Monte Carlo calculations) that a stiff polymer undergoes one or two transitions according to whether the polymer rigidity exceeds a critical threshold. Even though it is usually assumed that the folding transition for short polypeptides is first-order-like, experiments and recent theoretical results on the thermodynamics of proteins [21, 22] indicate that real proteins may also collapse slightly before folding occurs as happens in the present model for R_0 slightly smaller than 0.8. The effective thickness of real protein native structures viewed as tubes is $R_0 \sim 0.71$ in our units [7].

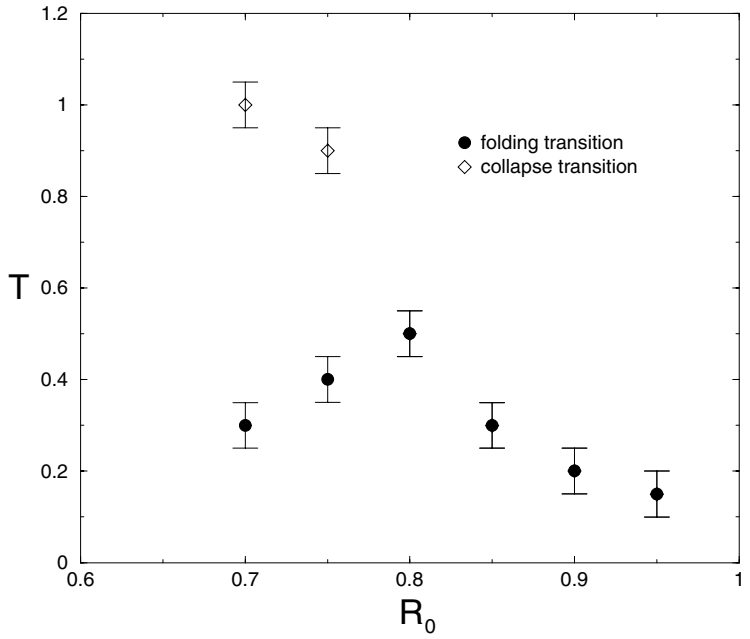

Figure 5. Phase diagram for a thick polymer. Points represent peaks in the specific heat (obtained for chains with 40 beads). Full circles represent first order transitions to a phase with broken rotational symmetry similar to the one of the nematic phase in a liquid crystal. Empty circles represent second order transitions from a swollen to a disordered globular phase.

4. Discussion of the results

We conclude with a discussion of the relevance of our results to the study of proteins [23]. Proteins are long, linear chains of amino acids that, under physiological conditions, fold rapidly and reversibly [1] into a somewhat compact state, expelling water from the core of the structure. The shape of this compact conformation (also called the native state structure) determines the protein functionality [23]. The rich repertoire of amino acids allows many sequences to share the same native state structure. As an example, it has been estimated that humans have more 100 000 proteins, but the number of distinct folds that these proteins adopt ranges in the few thousands [24]. Furthermore, these folds are special [25, 26]—they are not just any compact form but, rather, are made up of building blocks such as helices [15] and sheetlike planar structures [16] with tight loops connecting these secondary motifs.

In a coarse grained picture, segments of the protein backbone may be thought of as short tubes of non-zero thickness (the non-zero thickness of the tube approximately captures the effects of steric constraints [2]) subject to pair-wise attractive interactions which attempt to squeeze out the water from the hydrophobic core in the folded state of the protein. Because space-filling structures occur in the transition region, one may infer that the interactions in a protein are self-tuned to a situation in which there is a rich interplay between the pair-wise attractive interactions and the constraints imposed by the non-zero thickness of the tube. This is indeed the case because the atomistic scale interactions are short range and, at the microscopic scale, the squeezing out of water is facilitated by the outer atoms of nearby side chains coming together.

Our results for the ground state structures suggest that proteins, which are smart and versatile machines that are able to perform an amazing array of functions, employ native

state structures poised near a phase transition to exploit the great sensitivity that such systems possess. Indeed, it is well known that proteins use conformational flexibility to achieve optimal catalytic properties [23].

The relatively low degeneracy associated with the marginally compact phase leads to an energy landscape in structure space with few energy minima associated with the protein folds and has important consequences. First, because there are relatively few folds, each would be expected to have a correspondingly large basin of attraction. Furthermore, because the structures are marginally compact, one might expect the folding process to be relatively fast.

We have also shown that a thick homopolymer can be used to understand the thermodynamic properties of proteins. The phase diagram we obtain is in agreement with both experimental and theoretical views of the folding transition. Strikingly, the notion of a tube of nonzero thickness, combined with a novel three-body description of the self-interaction, provides a unified framework for understanding the common features of proteins.

Acknowledgments

This work was supported by INFM, MURST cofin99, the Penn State MRSEC under NSF grant DMR-0080019 and NASA. We are indebted to George Rose for many stimulating discussions. We thank P Carloni and S Pantano for helpful advice.

References

[1] Anfinsen C 1973 Principles that govern the folding of protein chains *Science* **181** 223–30
[2] Ramachandran G N and Sasisekharan V 1968 Conformations of polypeptides and proteins *Adv. Protein Chem.* **23** 283–438
[3] Wolynes P G, Onuchic J N and Thirumalai D 1995 *Science* **267** 1619
[4] Garstecki P, Hoang T X and Cieplak M 1999 *Phys. Rev.* E **60** 3219
[5] Carloni P, Rothlisberger U and Parrinello M 2002 *Accounts Chem. Res.* **35** 455
[6] Mirny L and Shakhnovich E 2001 *Annu. Rev. Biophys. Biomol. Struct.* **30** 361
[7] Banavar J R, Maritan A, Micheletti C and Trovato A 2002 Geometry and physics of proteins *Proteins* **47** 315
[8] Branden C and Tooze J 1999 *Introduction to Protein Structure* 2nd edn (New York, NY: Garland)
[9] Banavar J R, Gonzalez O, Maddocks J H and Maritan A 2003 Self-interactions of strands and sheets *J. Stat. Phys.* **110** 35
[10] Gonzalez O and Maddocks J H 1999 Global curvature, thickness and the ideal shapes of knots *Proc. Natl Acad. Sci. USA* **96** 4769–73
[11] Dokholyan N V, Pitard E, Buldyrev S V and Stanley H E 2002 Glassy behavior of a homopolymer from molecular dynamics simulations *Phys. Rev.* E **65** 030801
[12] Banavar J R, Maritan A and Seno F 2002 Anisotropic effective interactions in a coarse-grained tube picture of proteins *Proteins* **49** 246
[13] Sokal A D 1996 Monte Carlo methods for the self-avoiding walk *Nucl. Phys.* B *(Suppl)* **47** 172
[14] Maritan A, Micheletti C, Trovato A and Banavar J R 2000 Optimal shapes of compact strings *Nature* **406** 287–90
[15] Pauling L, Corey R B and Branson H R 1951 The structure of proteins: two hydrogen-bonded helical conformations of the polypeptide chain *Proc. Natl Acad. Sci. USA* **37** 205–11
[16] Pauling L and Corey R B 1951 Conformations of polypeptide chains with favored orientations around single bonds: two new pleated sheets *Proc. Natl Acad. Sci. USA* **37** 729–40
[17] Allain F H T, Yen M, Masse J E, Schultze P, Dieckmann T, Johnson R C and Feigon J 1999 Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non sequence specific binding *EMBO J.* **18** 2563
[18] Chang K Y and Tinoco I 1997 The Structure of an RNA 'kissing' hairpin complex of the HIV tar hairpin loop and its complement *J. Mol. Biol.* **269** 52
[19] Baumann U, Wu S, Flaherty K M and Mckay D B 1993 Three-dimensional structure of the alkalyne protease of pseudomonas aeruginosa: a two-domani protein with a calcium binding parallel beta roll motif *EMBO J.* **12** 3357
[20] Pitard E, Garel T and Orland H 1997 Protein folding, anisotropic collapse and blue phases *J. Physique* I **7** 1201

[21] Chahine J, Nymeyer H, Leite V B P, Socci N D and Onuchic J N 2002 Specific and nonspecific collapse in protein folding funnels *Phys. Rev. Lett.* **88** 169101
[22] Plaxco K W *et al* 1999 Chain collapse can occur concomitantly with the rate-limiting step in protein folding *Nat. Struct. Biol.* **6** 554
Ptitsyn O B 1995 Molten globule and protein folding *Adv. Prot. Chem.* **47** 83
[23] Fersht A R 1998 *Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding* (New York: Freeman)
[24] Chothia C 1992 One thousand families for the molecular biologist *Nature* **357** 543–4
[25] Chothia C 1984 Principles that determine the structure of proteins *Annu. Rev. Biochem.* **53** 537–72
[26] Levitt M and Chothia C 1976 Structural patterns in globular proteins *Nature* **261** 552–8