

A Tale of Two Chains: Geometries of a Chain Model and Protein Native State Structures

Tatjana Škrbić^{1,2}, Achille Giacometti^{3,4}, Trinh X. Hoang⁵, Amos Maritan⁶ and Jayanth R. Banavar⁷

¹ Ca' Foscari University of Venice, Department of Molecular Sciences and Nanosystems, Venice, Italy; tatjana.skrbic@unive.it

² University of Oregon, Department of Physics and Institute for Fundamental Science, Eugene, Oregon, USA; tskrbic@uoregon.edu

³ Ca' Foscari University of Venice, Department of Molecular Sciences and Nanosystems, Venice, Italy; achille.giacometti@unive.it

⁴ European Centre for Living Technology (ECLT), Ca' Bottacin, Dorsoduro 3911, Calle Crosera, 30123 Venice, Italy; achille.giacometti@unive.it

⁵ Vietnam Academy of Science and Technology, Institute of Physics, Hanoi, Vietnam; hoang@iop.vast.ac.vn

⁶ University of Padua, Department of Physics and Astronomy, Padua, Italy; amos.maritan@unipd.it

⁷ University of Oregon, Department of Physics and Institute for Fundamental Science, Eugene, Oregon, USA; banavar@uoregon.edu

* Correspondence: tatjana.skrbic@unive.it

Abstract

Linear chain molecules play a central role in polymer physics with innumerable industrial applications. They are also ubiquitous constituents of living cells. Here we highlight the similarities and differences between two distinct ways of viewing a linear chain. We do this, on the one hand, through the lens of simulations for a standard polymer chain of tethered spheres at low and high temperatures and, on the other hand, through published experimental data on an important class of biopolymers, proteins. We present detailed analyses of their local and non-local structures as well as the maps of their closest contacts. We seek to reconcile the startlingly different behaviors of the two types of chains based on symmetry considerations.

Keywords: symmetry; geometry; local and nonlocal structure; contact maps

1. Introduction

Polymer science [1-4], the study of chain molecules including linear polymers, is a flourishing subject that has led to life-changing progress in several technologies including plastics, textiles, and the design of novel materials. At the same time, linear chain molecules form the very basis of life including both the DNA molecule, whose information is translated into the sequence of amino acids, and proteins, which serve as amazing molecular machines in living cells. While conventional polymer models with stiffness have proved to be adequate for describing the relevant physics of the DNA molecule [5-7], the physical behavior of canonical polymers and proteins are strikingly different. In contrast to DNA, the structural changes during protein folding occur on multiple length scales at once, making it difficult to separate the relative contributions of the myriad interactions [8-13].

A linear chain is composed of many interacting monomers that are tethered together in a railway train topology. If the only interaction is self-avoidance, a single chain is in a coil phase whose large-scale behavior is in the same class as a self-avoiding walk. Upon adding an attractive interaction between pairs of non-adjacent monomers, the chain undergoes compaction into a highly degenerate compact phase at low temperatures [14-44]. While the notion of phases and phase transitions for a polymeric chain strictly refer to a chain with an infinite number of monomers, proteins are modest length chains, which yet exhibit several common characteristics. Notably, the compact state of proteins is modular and made up of two kinds of secondary building blocks, topologically one-dimensional helices [45], and two-dimensional sheets made up of zig-zag strands [46]. The helices and the strands are connected by turns or loops [47-49]. The nature of the ground states of compact polymers is qualitatively distinct from that of proteins and ordinarily do not exhibit any secondary motifs. The common characteristics of proteins are believed to derive from the shared backbone of distinct amino acid sequences.

Here we present an analysis of these two distinct classes of behaviors to understand their similarities and distinctions. We do this in two complementary ways. For conventional polymers, we study the simplest model of tethered hard spheres of diameter σ and a bond length equal to the sphere diameter. Following the standard nomenclature, we call this the tangent sphere model. We impose an attractive interaction between all pairs of non-adjacent monomers through a square well of range 1.6σ and depth -1, which sets the energy scale without loss of generality.

A sphere is isotropic and looks the same when viewed from any direction. There is nevertheless a preferred axis at the location of each main chain sphere corresponding to the tangent along the chain or the direction along which the chain is oriented at that location. Replacing the spheres with objects such as unidirectional coins, uniaxial discs, allowing neighboring spheres to overlap, or adding side chains to the

spheres along the main chain are all steps that break the spherical symmetry and yield ground state structures, which resemble protein structures to varying degrees [50-61]. To avoid clouding the issue by studying an approximate model for proteins, we resort instead to a careful analysis of experimental data of over 4,000 protein native state structures (see Materials and Methods). A side-by-side comparison of the local structures and non-local contacts of proteins to those of the tangent sphere model at both low and high temperatures provides a vivid picture of the different views of a chain molecule. Our goal here is to assess how well the tangent sphere model describes the protein backbone.

2. Materials and Methods

2.1. Our protein dataset

Our protein data set consists of 4,391 globular protein structures from the Protein Data Bank (PDB), a subset of Richardsons' Top 8000 set [62] of high-resolution, quality-filtered protein chains (resolution $< 2\text{\AA}$, 70% PDB homology level), that we further distilled out to exclude structures with missing backbone atoms, as well as amyloid-like structures (for full list of the PDB identifiers of protein structures in our database see Table S1 in the Supplementary Information of Refs. [58,63]). The program DSSP (CMBI version 2.0) [64] has been used to determine the backbone hydrogen bonding pattern and thus place each protein residue in context within a protein chain: within an α -helix, it is labeled an ' α -residue'; within a β -strand, it is labeled a ' β -residue', or elsewhere, it is tagged as a 'loop-residue'.

2.2. Numerical simulations of a chain of tethered tangent spheres

To obtain a set of independent equilibrium configurations of a chain of tethered tangent spheres comprised of $N=80$ spherical beads, subject to an attractive potential for a wide range of temperatures, we have employed standard replica exchange (RE) (or parallel tempering) canonical simulations [65,66]. Most of our simulations were carried out with a chain of 80 hard spheres of diameter σ . The model is a tangent sphere model because the bond length is also constrained to be equal to σ . At any temperature, the spheres do not self-intersect and are hard. We introduce a generic attractive square-well attraction of range $R_{\text{att}} = 1.6\sigma$, between all pairs of spheres, and magnitude ϵ , which sets the characteristic energy scale. The attractive interaction causes the chain to become compact at low temperatures.

The RE calculation [65,66] relies on a set of canonical simulations run in parallel at a set of M carefully chosen different temperatures, T_i , $i = 1, 2, \dots, M$. Each simulation represents a replica, or a system copy in thermal equilibrium. The key advantage is the possibility of swapping replicas at different temperatures without affecting the equilibrium condition at each temperature. This permits rapid equilibration even when there is a rugged free energy landscape. In each Monte Carlo (MC) simulation of

one replica, new moves are accepted with the standard Metropolis acceptance probabilities [67]. We ensure that the number of swaps that entail exchange of replicas is large enough to ensure the fidelity of the statistics. The efficiency of the RE scheme depends on the number of replicas, the selected set of temperatures, as well as of the swap moves frequency. For best performance, the acceptance rate of swaps is tuned to be around 20% [68]. The RE simulations results are conveniently analyzed using the weighted histogram analysis method [69]. We employed 30 replicas, with a finer temperature mesh at lower values of the reduced temperatures ($k_B T/\epsilon$) in the range $k_B T/\epsilon = 0.3\text{-}0.5$ with a separation of neighboring temperatures of 0.02. In the $k_B T/\epsilon = 0.5\text{-}1$ interval, the separation of neighboring temperatures was 0.05, and for $k_B T/\epsilon = 1\text{-}4$, the separation interval was 0.2. We allowed for the RE swaps only between neighboring temperatures. The exchange moves were attempted every 100 MC steps per monomer. The length of the simulations was 10^9 MC steps per monomer and per replica.

For sampling chain configurations at infinite temperature, we have used the standard Metropolis MC algorithm [67] in which all proposed updates of the chain configuration that respect its self-avoidance are accepted. In both simulation protocols, standard local moves, including crankshaft, reptation (or slithering-snake) moves, endpoint moves, and the nonlocal pivot move [70] are employed with equal probabilities.

The, $N=80$ beads, tangent sphere polymer exhibits two continuous (second order) ‘transitions’ (see inset of Figure 5(b)). At a temperature $k_B T/\epsilon \sim 3$ (ϵ denotes the energy scale of attraction), there is a coil-to-globule transition (signaled by a kink in the specific heat per bead C_v/Nk_B). This is a finite-size counterpart of the θ -point for this system. At low temperatures $k_B T/\epsilon \sim 0.4$ there is a second ‘transition’ into a compact globule phase, signaled by the maxima of the specific heat per bead C_v/Nk_B . The results we will present for the tangent sphere model are at infinite temperature and $k_B T/\epsilon = 0.3$.

In the next section, we will present some definitions and general observations pertaining to the local structure of a chain, especially in the continuum limit. We then go on to the Results Section, which is divided into two parts. First, we depict some similarities and some critical differences between the polymer model and protein structures in terms of power law behavior of certain geometrical measures of local structure. In the second part, we will highlight some striking differences between the geometries of the model polymer and protein native state structures. We then conclude with a brief discussion.

3. General considerations

Three-body interactions and the self-avoidance of a continuum tube

We will begin our analysis with some definitions and general observations. For fixed bond length, a local chain conformation is specified by two angles θ and μ . θ is a measure of bond bending, a straight conformation has $\theta=\pi$. The angle between successive binormals, μ , is the second angular coordinate and is the dihedral or torsional angle [63].

The standard model of a linear chain in polymer science, represented by tethered spheres, does not lend itself, in a natural manner, to housing helices, which are recurrent motifs in biomolecules [71-74]. In contrast, a tube, such as a garden hose, which can be thought of as a chain of discs, can be wound readily into a helix. The protein α -helix has a geometry akin to that of a tube wound tightly into a space-filling helix [75].

Consider a collection of uniform untethered hard spheres. The standard prescription for ensuring that spheres do not overlap is to ensure that the distance between the centers of every pair of spheres is no smaller than the sphere diameter. Pairwise interactions often capture the essence of interacting systems. The simplest generalization to a chain topology is tethered hard spheres (with the same self-avoidance constraint), which, as we have noted, is the simplest conventional model of polymer physics. In contrast, the alternative description of a chain as a tube leads to an unusual condition of self-avoidance. It has been shown that the correct way to determine whether a tube *in the continuum limit* is self-avoiding or not entails discarding pairwise interactions and invoking appropriate many-body interactions [75,76]. This is illustrated by considering the self-avoidance of a tube of non-zero thickness (see Figure 1). Knowledge of the distance between a pair of points on the tube axis (say A and B or B and D) does not discriminate between the two contexts of nearby points along the axis or in different parts along a chain. In the continuum limit, points A, B, and C, locally positioned along the axis, can become infinitesimally close to each other. This cannot be the case, however, for non-local points, where self-avoidance is a prime consideration, and one must ensure that the two points do not approach each other too closely. Knowledge of the coordinates of a pair of points does not inform you about the context that the two points are in – are they points locally along the axis (which can of course be arbitrarily close to each other) or are they non-local points (that may have come too close to each other possibly signaling an intersection)? This inability to discriminate between local and non-local pairs of points is at the root of the problem.

The standard method in polymer physics of ensuring self-avoidance in the continuum limit is to first make the tube infinitesimally thin (a natural limiting case for a continuum chain of tethered spheres) and then use a singular δ -function potential [77] interaction: there is no energy cost as long as two points on

the axis do not overlap exactly but there is an infinite energy cost when there is in fact an overlap. There are at least three problems with this [78]. First, the δ -function potential is singular (unlike say a familiar Lennard-Jones potential) and one needs to use renormalization group theory to introduce an artificial cut-off length scale to carry out the calculations and then demonstrate that this length scale can safely approach zero and yet retain the validity of the results. Second, the δ -function pairwise potential does not preserve the topology of a closed string – the number of knots is not necessarily conserved. Finally, in standard polymer physics, there is no description of the self-avoidance of a continuum self-avoiding tube (or surface) of non-zero thickness.

These problems are deftly averted by discarding pairwise interactions and working with a suitable many-body potential. In the continuum limit, there is a simple geometrical condition [75] to ascertain both whether a tube (and by extension a surface) is self-avoiding non-locally and is not too tightly wound locally. One can draw a circle through any triplet of points along the tube axis and measure the radius. The prescription for self-avoidance of a continuum tube is to consider all possible triplets, local or otherwise, and ensure that every one of the three body radii is greater than or equal to the tube radius. The radius of the circle passing through the triplet of points (A,B,C) in Figure 1 is the local radius of curvature as the points approach each other and results in a kink in the tube when the radius of curvature becomes smaller than the tube radius. On the other hand, the radius of the circle drawn through (A,B,D) is a measure of the distance of approach of two parts of the tube and must not be smaller than the tube thickness in order to respect self-avoidance. Likewise, the self-avoidance of a surface or layer (or a sheet of paper) of non-zero thickness necessarily entails discarding both pairwise and three-body interactions and working with a suitable four-body potential. One considers all quartets of points on the symmetry plane of a surface and draws spheres through each quartet. A surface is self-avoiding if the sphere radius of each quartet (local or non-local) is larger than the thickness of the surface. We note that this many-body prescription is strictly needed only in the continuum limit [76].

For a discrete chain, such as the ones we focus on in this paper, the pair-wise distance between a local pair of points is the bond length and there is no singular behavior at the local level because of a natural cut-off length scale. Also, a minimum threshold of the non-local three body radius can readily be measured by directly accessing the non-local pair-wise distance. We will make use of these simplifications for a discrete chain in our analysis below. It is important to note that there are myriad local interactions, including hydrogen bonds, as well as entropic effects that will play a key role in governing the local behavior of a real chain. Despite the absence of an imperative need to invoke many body interactions for a discrete chain, the three-body and four-body radii do yield interesting information. An infinitely large

three-body radius signals co-linearity of the three points (bond bending angle θ equal to 180° ; $\theta=0^\circ$ is excluded because of steric overlap), whereas an infinite four-body radius is associated with planarity of the quartet of points (μ equal to 0° , 180° , or -180°).

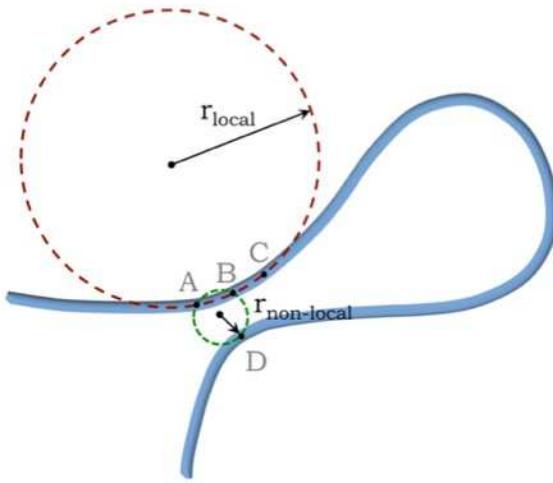


Figure 1. Sketch of the axis of a self-avoiding continuum tube (depicted in blue). The points A, B, and C lie alongside each other on the tube axis whereas point D is a nearby point from another part of the tube. The three-body prescription is to draw circles through all triplets of points on the tube axis and ensure that none of the radii is smaller than the tube radius. For a local triplet of points, one obtains the local radius of curvature whereas the non-local radius is a measure of the distance of closest approach of two parts of the tube.

4. Results

4.1. Power law scaling

Power laws often signify scale invariance and are a signature of an absence of a characteristic scale [79]. A liquid-vapor system at its critical point exhibits critical opalescence. The system appears milky white because light of all wavelengths scatter from the droplets and bubbles of liquid and vapor of all sizes thoroughly interspersed among each other. Another example of a non-trivial power law is the fractal dimension of a self-avoiding walk in three dimensions of around $5/3$ [2]. Here we discuss a somewhat trivial but surprising realization of ‘universal’ power law behavior arising in the statistics of the local conformation of a discrete chain molecule. We alert the reader that, unlike in critical phenomena, here there is neither any many-body emergent behavior nor the need to invoke a system in the thermodynamic limit.

The procedure that we follow is simple. For a set of three (four) points, one can readily draw a circle (sphere) passing through them. The center of a circle (sphere) is the point equally distant from all three (four) points and can be determined as a solution of a suitable system of linear equations. In the case of three points, the solution is simple, and the radius of the circle R is related to the area A of the triangle passing through the three points and its sides a , b , and c : $R=abc/4A$. For four points, we merely solve the equations on a computer to obtain the radius R of the sphere.

Our goal here is to measure the radii R associated with many realizations of these points and obtain cumulative probability distribution functions of the inverse radius $X=1/R$. Figure 2 shows plots of the cumulative distribution function (CDF) of the inverse radii ($X=1/R$) (the probability P ($1/r < 1/R$) as a function of R that a given radius r is larger than R). The circles (spheres) in question are drawn through three (four) points (chosen consecutively along a chain and randomly in some instances) in three dimensions.

We have studied: (a) consecutive triplets (quartets) of C_α atoms along the backbones of globular proteins when the Ramachandran ω angles characterizing a consecutive triplet have canonical values of $|\omega| \sim 180^\circ$. Case (a) occurs in around 99.7% of the cases in globular proteins yielding the *trans* isomeric conformation of a peptide backbone, where the two neighboring C_α atoms are on opposite sides of the peptide bond, with a bond length approximately equal to 3.81Å [63]; (b) consecutive triplets (quartets) of C_α atoms in globular proteins in which at least one of the two Ramachandran ω angles has a rare non-canonical value of $|\omega| \approx 0^\circ$ that occurs in $\sim 0.3\%$ of cases. This happens when two neighboring C_α atoms are on the same side of the peptide bond, resulting in a shorter bond length of around $\sim 2.95\text{\AA}$. [63]. This corresponds to the so-called *cis*-conformation of a protein backbone [80]. Cases (a) and (b) are combined for the quartets because they show very similar behavior; (c) three (four) points selected from a two-step (three-step) self-avoiding random walk of a tangent sphere model (hard spheres of diameter 3.81Å and bond length 3.81Å with no other interaction besides the non-overlapping of hard spheres (polymer at infinite temperature); (d) three (four) points selected from a two-step (three-step) self-avoiding random walk of a tangent sphere model (spheres of diameter 3.81Å and bond length 3.81Å) subject to an attractive square-well interaction of range $R_{\text{att}} = 1.6\sigma \approx 6\text{\AA}$ (polymer at low temperature); (e) three (four) points chosen randomly within a three-dimensional sphere of unit radius; and (f) three (four) points selected as points on a two-step (three-step) random walk (no self-avoidance or steric constraints) in three dimensions with a fixed bond length of 3.81Å (corresponding to the distance between consecutive C_α atoms in proteins [63]).

In the case of triplets ('3-body' case) shown in Figure 2(a), we see that the last five systems (b)-(f) exhibit power law behavior with approximately the same exponent of 2, but the canonical protein triplet does *not* (case (a) in Figure 2(a)).

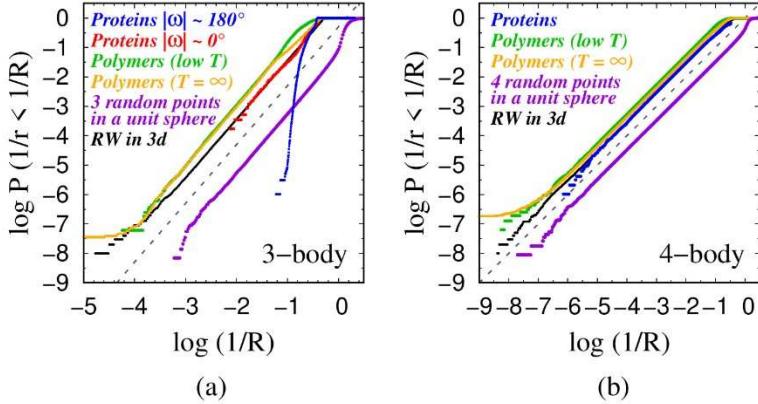


Figure 2. (a) Cumulative probability distributions of the inverse radii $X=1/R$ of circles drawn through three consecutive points along different classes chains: (blue) 965,122 triplets of the backbones of globular proteins in our data set (defined by C_α atoms) when both the Ramachandran ω angles characterizing a consecutive triplet have canonical values of $|\omega| \approx 180^\circ$; (red) 5,774 consecutive triplets of C_α atoms in globular proteins in which at least one of the two Ramachandran ω angles is $|\omega| \approx 0^\circ$; (green) 16,391,622 triplets taken from $\approx 200,000$ low temperature ($k_B T/\epsilon = 0.3$) configurations, obtained using replica-exchange (RE) simulations, of a chain of 80 tangent spheres of diameter σ with an attractive square well potential of range $R_{\text{att}} = 1.6\sigma \approx 6\text{\AA}$; (orange) 25,598,976 triplets obtained from MC simulations at $T = \infty$; (purple) 143,557,206 triplets of points chosen uniformly from within a unit sphere in three dimensions; and (black) 100,000,000 two-step random walks in three dimensions with fixed bond length of 3.81\AA . The gray dashed line has a slope of 2 and is a guide to the eye. (b) Cumulative probability distributions of the inverse radii $X=1/R$ of spheres, whose surface passes through four consecutive points in different classes of chains: (blue) 957,723 local quartets along the backbones of globular proteins (employing C_α atoms); (green) 16,181,473 local quartets selected from $\approx 200,000$ chain configurations, obtained from replica-exchange (RE) simulations, of a chain of 80 tangent spheres of diameter σ subject to an attractive square-well potential of range $R_{\text{att}} = 1.6\sigma \approx 6\text{\AA}$ in the low temperature phase ($k_B T/\epsilon = 0.3$); (orange) 25,270,784 local quartets obtained from MC simulations at $T = \infty$; (purple) 112,754,340 quartets of points chosen

uniformly within a unit sphere in three dimensions; and (black) 100,000,000 three-step random walks in three dimensions with a fixed bond length of 3.81Å. The gray dashed line is a guide to the eye and has a slope of 1. In all simulations, the bond lengths have been chosen to be 3.81Å equal to the mean value of the distance between the two consecutive C_α atoms along the protein chain. The distinctive behaviors of the purple curves (corresponding to the random points cases) occur because one can obtain circles of arbitrarily small radii, a situation precluded in the other cases due to steric considerations. The behaviors of the tangent polymer model at high and low temperatures are essentially the same. The local behavior is governed by the same steric constraints in both cases and the CDF does not change. In contrast, for real polymers, recent experimental studies [81,82] have shown the importance of the mechanical properties in determining the local curvature in the context of super-lubricity at the single molecule level.

For a canonical protein backbone in its *trans* conformation, quantum chemistry does not allow the bond bending angle θ to be greater than $\approx 150^\circ$, thereby preventing too large a value of R . In contrast, for a non-canonical protein backbone in its *cis* conformation [80], two consecutive C_α atoms along the protein chain are much closer to one another and the backbone in many of these cases has PRO residues. This stiffens up the protein backbone with respect to the canonical case permitting large bond bending angles, θ , that can almost reach $\approx 180^\circ$.

The second panel in Figure 2, Figure 2(b), shows similar ‘universal’ power law behavior for the CDF in the case of quartets (‘4-body’ case) of the inverse radius ($X=1/R$) of a sphere for various cases, detailed in the caption, this time with an exponent 1. We provide a simple rationalization of these findings in the next section for a triplet of points.

Rationalization of the power-law exponent

To illustrate the origin of the power law behavior, we present here a simple derivation of the probability distribution $P(X=1/R)$ of the inverse radii of circles drawn through three points of a two-step d -dimensional random walk with a fixed bond length b , which provides a natural length scale. We do not present a similar derivation for the radius of sphere passing through four points because it is more complex and is best handled numerically (which is what we have done). The quantity Xb is a dimensionless quantity, which enters in the derivation below. The input is $p(\theta)$, the probability distribution

of the bond bending angle θ . For a d-dimensional random walk, the probability distribution $p(\theta)$ scales as [83],

$$p(\theta) \sim (\sin(\theta))^{d-2} \quad (1)$$

R and θ are related by

$$R = b/(2 \cos(\theta/2)) \quad (2)$$

or equivalently:

$$X = 1/R = (2 \cos(\theta/2))/b \quad (3)$$

Noting that:

$$P(X)dX = p(\theta)d\theta \quad (4)$$

one obtains:

$$P(X) \sim (Xb)^{d-2} [1 - (Xb/2)^2]^{(d-3)/2} \quad (5)$$

One thus obtains asymptotically (when $Xb \ll 1$ or in the large radius R limit) a power law behavior of the probability distribution $P(1/R)$ with an exponent $(d-2)$ for $d > 2$ with a power law correction. This means that the cumulative probability distribution $P(1/r < 1/R)$, being an integral of the probability $P(1/R)$, displays a power law behavior with an exponent $(d-1)$ ($=2$ in three dimensions) with a power law correction. This correction is relatively small when $(Xb)^2$ is much smaller than 1 and yields good power law behavior as observed in Figure 2(a) and Figure 3(a). The pivotal quantity that determines the asymptotic exponent is the behavior of $p(\theta)$ when θ approaches 180° and the three points become collinear. The difference in behavior in 2 and 3 dimensions is shown in Figure 3(b). The numerical simulations are in good accord with the prediction of Eq. (1).

The '4-body' case works in a similar manner to the '3-body' case, except that the radius now depends on two independent variables. An unexpected sensitivity of the power law exponent to the choice of the 4 points is demonstrated in Figure 3(c). Quartets derived from a plain 3-step random walk in three dimensions (3d) exhibit power law behavior with an exponent of 1 (in accord with the results shown in Figure 2(b)). We have also considered a simple variant of the plain random walk that we call a constrained random walk. Here we define the first two points to lie along the x-axis. We then place the third point

randomly on a pre-determined x-y plane. Superficially, this may not seem to be an onerous constraint, because any three points will necessarily lie in a plane. Nevertheless, a constrained random walk still shows power law behavior but with a distinct exponent, behaving as though it is in a fractal dimension regime. This is because the sampling of phase space is now altered in a relevant manner. We note that this regime may occur when a polymer system happens to be in the vicinity of a surface or a solid wall.

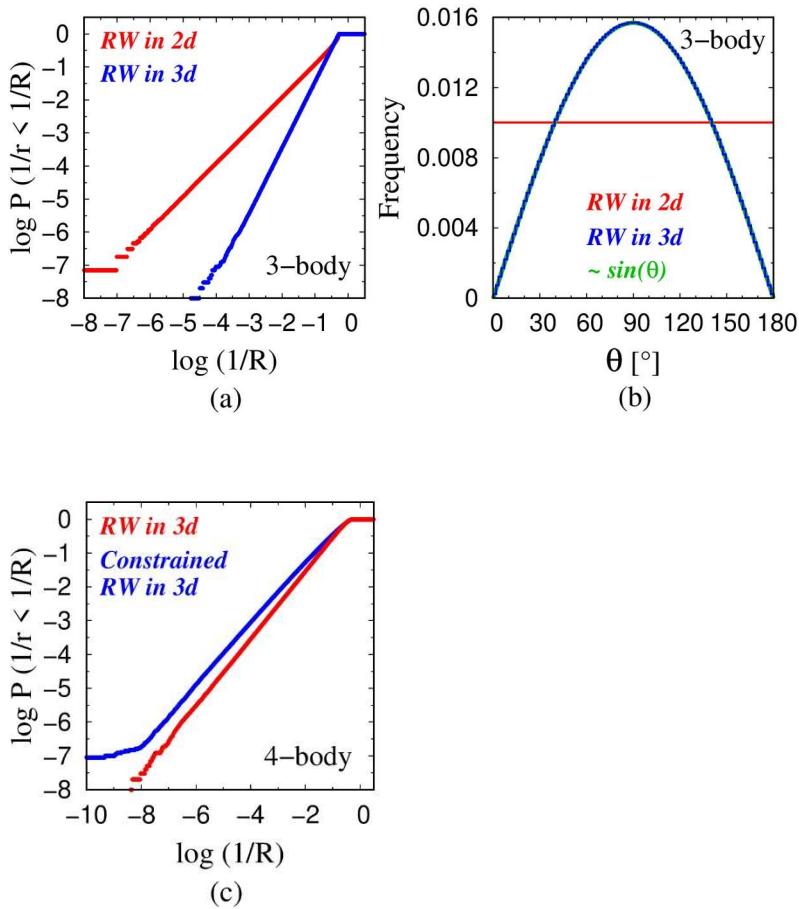


Figure 3. (a) Cumulative probability distributions of the inverse radii $X=1/R$ of the circles drawn through three points of a two-step random walk with fixed bond length of 3.81\AA , when the walk is performed in two dimensions (red) and in three dimensions (blue). (b) The distribution of the bond bending angle θ in the two-step random walk in two dimensions is uniform $p(\theta) = \text{const.}$ (red histogram), while, in three dimensions, $p(\theta) = \sin \theta$ (blue histogram and the green line). (c) Cumulative probability distributions of the inverse radii $X= 1/R$ of the spheres drawn through four points of a three-step random walk with fixed bond length of 3.81\AA , when the walk is performed in three dimensions (red) and for a constrained random walk. The constraint arises because the

first three points (first two steps of the random walk) are sampled from a plane (sampling in two dimensions – after all, any three points do lie in a plane), while the final step (fourth point) is sampled in full three-dimensional space (blue). The constrained random walk corresponds to a fractal dimension $2 < d < 3$ for the effective sampling of the variables controlling $1/R$ and reduces the steepness of the power law.

4.2 Chain geometries

We begin with an analysis of distinct local chain geometries. We compare the behaviors of the tangent sphere model at low and high temperatures, on one hand, and that of the native states of globular proteins, on the other. The local structures of these cases are shown in Figure 4 through their characteristic (θ, μ) plots [63]. These plots are drawn by measuring local pairs of bond-bending and dihedral angles for nearly a million monomers (for the polymer models) and residues (for protein native state structures). In contrast to the plots of model polymers (Figures 4(a) and (b)), the (θ, μ) plot for globular protein native state structures (Figure 4(c)) exhibits significant structure (distinct from the features present in the low temperature tangent sphere model) and signal the presence of the building blocks of helices and sheets.

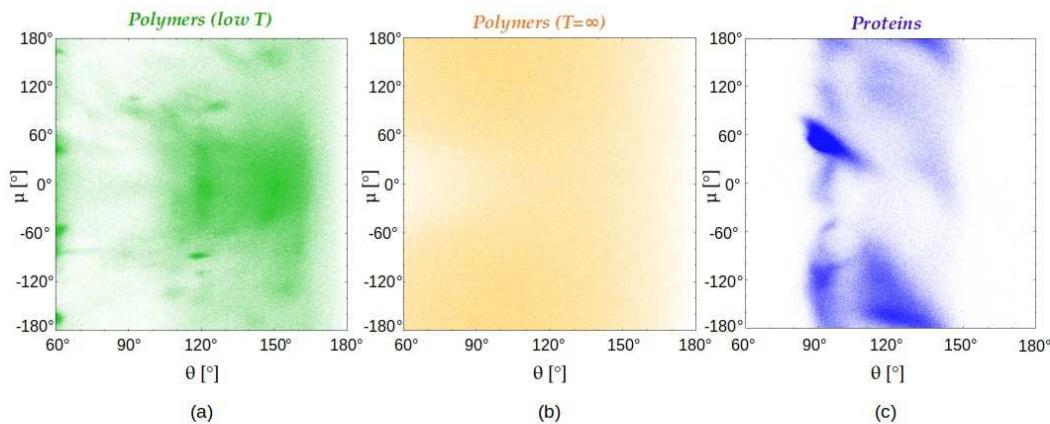


Figure 4. (θ, μ) cross plots of the bond bending angle θ versus dihedral angle μ for 966,505 randomly chosen monomers belonging to two different polymer classes and for the same number of residues in protein native state structures. (a) (light green) Polymer chain consisting of 80 tangent spheres of diameter σ subject to an attractive square well potential of range $R_{\text{att}} = 1.6\sigma \approx 6\text{\AA}$ at low temperature ($k_B T/\epsilon = 0.3$) studied using RE simulations. (b) (orange) Polymer chain consisting of 80 tangent spheres of diameter σ at infinite temperature accessed by means

of MC simulations with the only interaction being steric avoidance of all pairs of spheres. (c) (blue) For 966,505 residues of the 4,391 globular proteins in our data set.

The trends in Figure 4 are analyzed in Figure 5, which depicts a histogram of how far the nearest non-local monomer or residue is along the chain sequence. A non-local contact is defined to be one that is separated by at least three positions along the chain with the two beads nearest to each other (i, j) satisfying $|i-j| \geq 3$. For both helices and turns, there are sharp maxima of close-by neighbors at a sequence separation of 3.

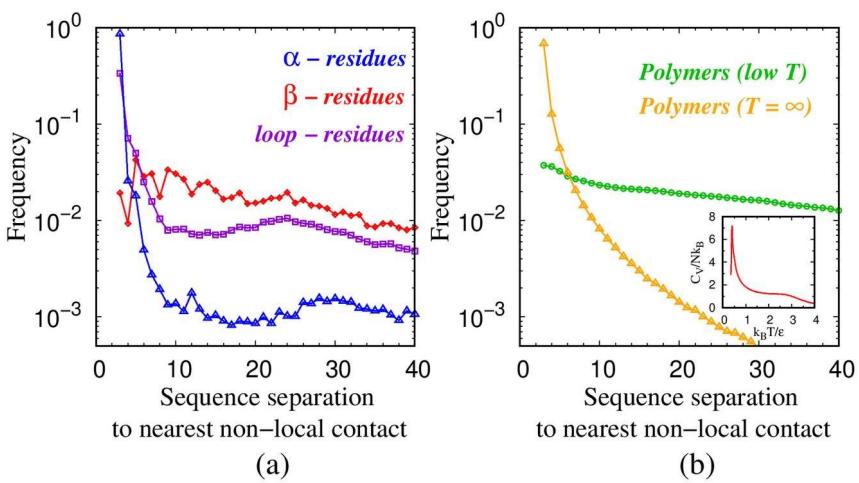


Figure 5. Frequency distribution of the sequence separation $|i-j|$ along the chain, of the nearest non-local contact of bead i found at location j . In Panel (a), the blue points indicate 313,574 α -residues out of a total of 975,287 residues in our data set from 4,391 globular proteins; the red points are 214,501 β -beads; and the purple points denote an analysis of 442,821 loop-residues. (b) the green and yellow points show the distinct smooth behaviors of a tangent polymer model at low ($k_B T/\epsilon = 0.3$) and infinite temperatures. The behavior is monotonic with the closest non-local contacts always being close along the sequence. The inset shows the specific heat per bead C_V/Nk_B for a 80 beads long tangent sphere polymer as a function of the reduced temperature $k_B T/\epsilon$. There are two continuous (second order) 'transitions': at a high temperature ~ 3 , there is a coil-to-globule transition and, at ~ 0.4 , there is a transition into a compact globule.

There are at least two characteristic length scales associated with the conformation of a chain. One is the local behavior as measured by the local radius of curvature of a triplet of contiguous points. The

second is the distance to the nearest non-local contact. For the space-filling conformation of a continuum tube of non-zero thickness, these two length scales become equal [75]. Figure 6 shows a histogram of the two scales, local and non-local, for several situations. The local radius is $R = b/(2 \cos(\theta/2))$, where θ is the bond bending angle associated with a local triplet. Structures in the histogram of local radii denote a preference for certain angles of θ . The α -residues exhibit a sharp peak corresponding to $\theta_\alpha \approx 92^\circ$, the β residues for $\theta_\beta \approx 120^\circ$, and the loop residues have a pronounced maximum close to the helical value and a less prominent maximum around 111° , see Figure 6(a). These peaks are also reflected in the high-density regions in the (θ, μ) cross plot of protein native state structures, shown in Figure 4c. Figure 6b shows the histograms of the relevant non-local length scales for all five cases. All five curves exhibit a single peak denoting a relevant non-local length scale. Table 1 is a compilation of these characteristic local and non-local length scales.

Table 1. Characteristic local and non-local length scales for three residue types in proteins and for monomers of a tangent polymer chain in the low and high temperature phases.

Monomers	Most frequent value of local radius [Å]	Relevant non-local length scale [Å]
α – residues in proteins	2.73	5.06
β – residues in proteins	3.6	4.67
loop – residues in proteins	2.75	5.26
Spheres in tangent polymers at low T	2.21	3.81
Spheres in tangent polymers at $T = \infty$	2.47	7.72

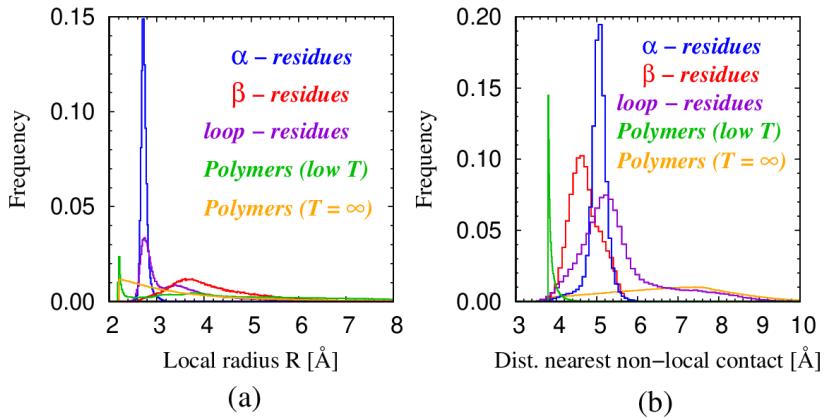


Figure 6. (a) Frequency distribution of the local radius of curvature R for five different classes of consecutive triplets. We consider only pure protein triplets in which all three residues are in the *same* structural class. There are 256,154 α triplets (blue) comprising $\sim 26\%$ of all 970,896 triplets. There are 134,643 β triplets (red) and 313,923 loop triplets (purple). There are 16,391,622 low temperature polymer triplets (green) and 25,270,784 infinite temperature triplets (orange). (b) Frequency distributions of the distances to the closest non-local contact, defined as $|j-i| \geq 3$, of monomers belonging to different classes, with the same color code as in panel (a).

Figure 7 shows five representative chain conformations (denoted A-E), each having 80 monomers. We will present the key characteristics of these conformations to highlight similarities and especially the differences. Figure 7 also presents the contact maps for each of these five conformations to show for each monomer (labeled 1-80) the nearest, in distance, monomer (separated at least 3 positions along the chain) indicating in red, when the pairwise distance is greater than 6Å. Such distant contacts are rare in the polymer models unlike for the three protein chains. There is little structure in the globular polymer structure A. The infinite temperature polymer B is characterized by the nearest contacts being nearby in sequence, as evidenced by the points in its contact map being close to the diagonal.

This feature of locality of the closest contacts is also seen in protein α -helices C where coordinated closest contacts are of the $(i,i+3)$ type. Furthermore, a distinctive pattern is seen for β sheets D, that display coordinated (i,j) contacts in which bead index $j \geq i+4$ and takes on a coordinated pattern that is consistent with the situation of two strands coming together and forming parallel β sheets (when points in the contact map are parallel to the principal diagonal but shifted away from it) or anti-parallel β -sheets (when points in the contact map are placed along the directions that are perpendicular to the diagonal). The mixed α/β protein E has the features present in both α -helices and two types of β -sheets (parallel and anti-parallel). The similarity of the α helix contact map and that of the infinite temperature polymer is in accord with expectations that helices are more prone to nucleate from the coil phase than β sheets because of the prevalence of short-range contacts.

Figure 8 shows the relative importance of the local radius of curvature and the relevant non-local distance in determining the nature of compact conformations in the five cases. In all the five panels, we have scaled the quantities with the appropriate characteristic length scale defined in Table 1. Even though the scaling factors are clearly different, the behavior of the polymer chain is superficially similar at high and low temperatures. The scaled value of the closest non-local distance is substantially flat around a value of 1 with the fluctuations being a bit larger at infinite temperatures. In contrast, the local radius exhibits bigger swings in the low temperature conformation compared to that at infinite temperature. The α helix is special in that both the local and closest non-local scaled distances are close to each other and to the value 1. In the continuum limit, this equality would result in a space-filling helix [75]. The β regions and the loops do not exhibit this kind of behavior with the non-local scaled distance often being smaller than the scaled local radius of curvature.

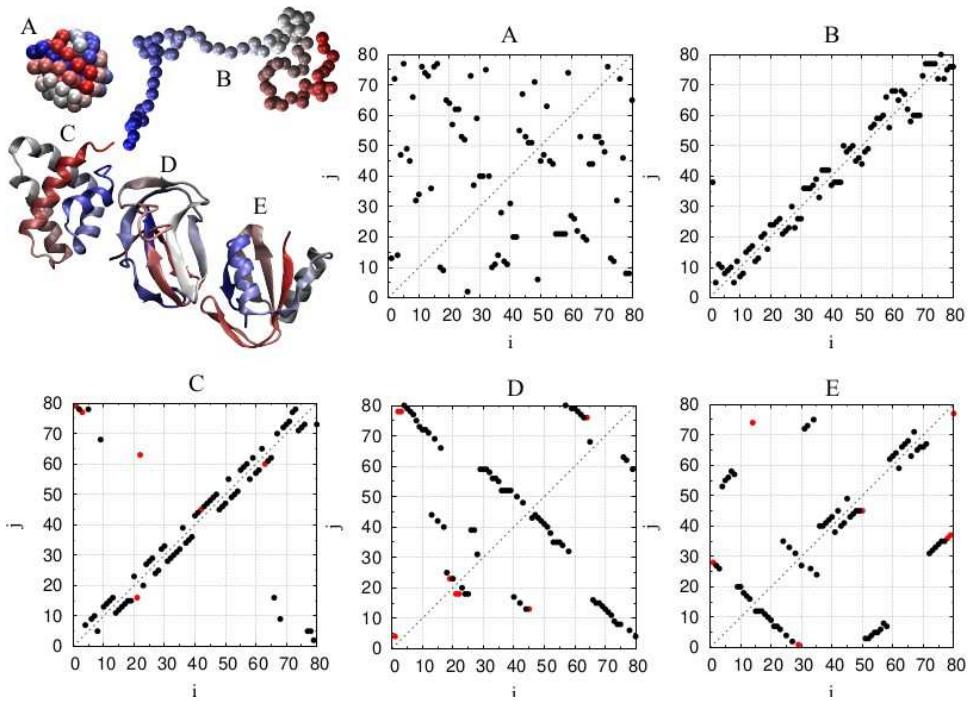


Figure 7. Representative conformations of chains of length 80: A) a low temperature globule and B) an infinite-temperature coil of a generic tangent polymer chain; C) all- α protein [PDB code: 3bqp, chain B]; D) all- β protein [PDB code: 1bdo, chain A]; and E) α/β protein [PDB code: 3l9, chain X]. The color coding of the conformation does not represent secondary motifs but rather depicts how far a monomer is along the sequence. The chain beginning is depicted in red color that morphs towards blue at the chain end. The contact maps shown in the other panels are those indicating the closest non-local contact j of a given monomer. The black points indicate those that are found within 6 Å and the red points those that are found further away than 6 Å, of these configurations. The choice of 6 Å is dictated by the fact that the radial distribution function of proteins exhibits a pronounced minimum at this value [60].

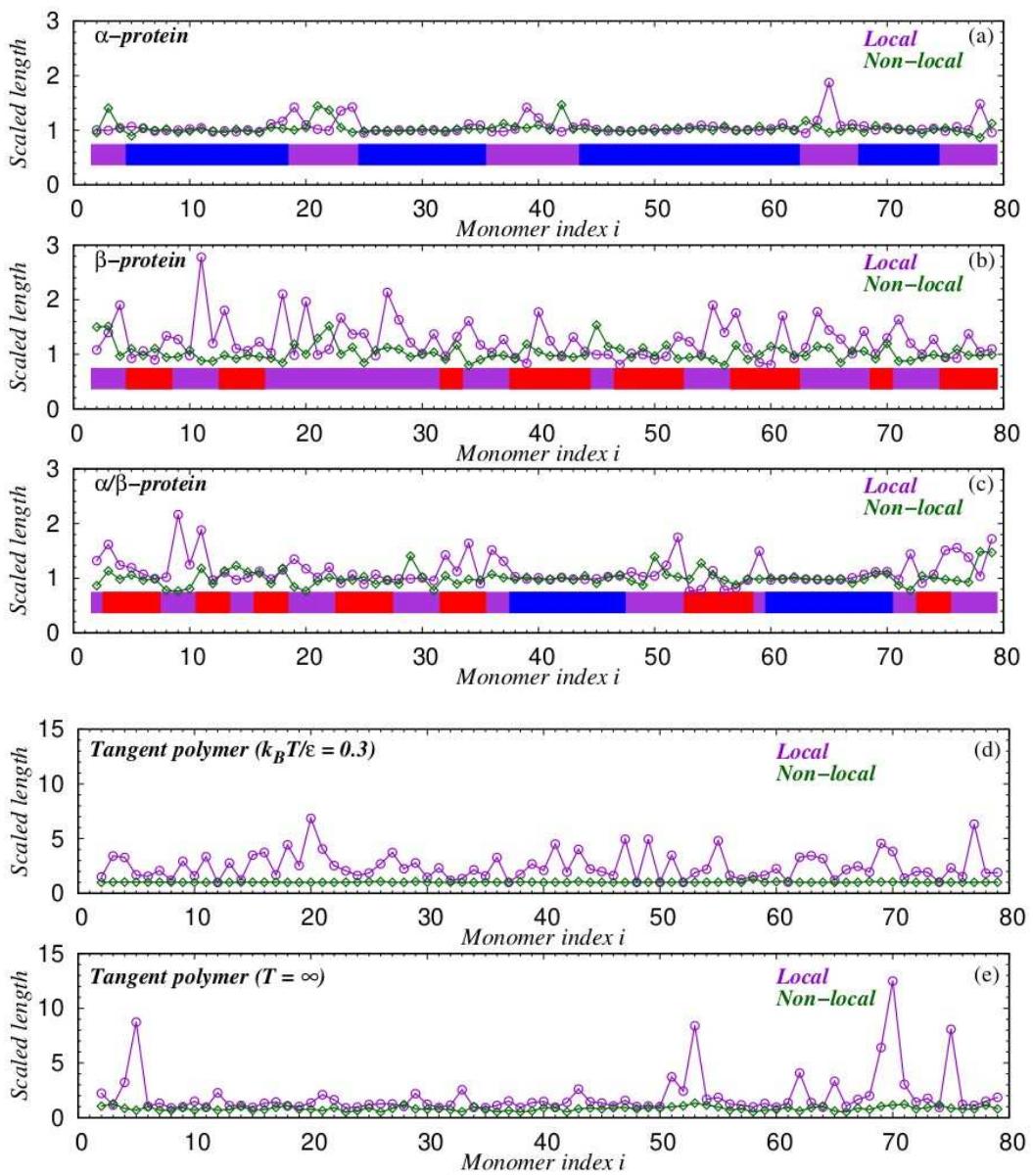


Figure 8. Scaled values of the local radius and minimal non-local distance for the five conformations shown in Figure 7. For each residue of the three proteins, depending on its type (' α ', ' β ', or 'loop') these quantities are appropriately scaled with the value of the corresponding characteristic length scale presented in Table 1. (a) all- α protein 80 residues long [PDB code: 3bqp, chain B]. (b) all- β protein 80 residues long [PDB code: 1bdo, chain A]. (c) α/β protein 80 residues long [PDB code: 3l9, chain X]. In the top three panels, the blue ribbons indicate α -helical parts of the sequence, the red ribbons indicate β -sheets, and the purple ribbons indicate

loop regions; (d) a tangent polymer conformation at $k_B T/\epsilon = 0.3$; (e) a tangent polymer conformation at $T = \infty$.

5. Discussion

Our principal focus in this paper was to study the similarities and differences between chains viewed in two separate ways. The first, a baby model in polymer science, is a tangent sphere model subject to a square-well attraction. We have carried out extensive simulations of the model in the high and low temperature phases. We compare the behaviors with those of experimentally determined (and presented in the Protein Data Bank) structures of more than 4,000 proteins. The relevant local behavior can be measured by determining the radius of a sphere passing through a local triplet or the radius of a sphere, whose surface passes through a set of four consecutive monomers. Surprisingly, we found power law behavior of the probability distribution functions of the radii (with the notable exception of amino acid triplets in proteins). We presented a simple rationalization of this behavior.

We then went on to underscore the numerous distinctions between the model results and protein data. Proteins are complex molecules, which follow the rules of quantum chemistry [45,46,84,85] and are influenced by the interactions with the surrounding solvent molecules [86-97]. A protein is a distinct sequence of amino acids. Yet, proteins show remarkable common characteristics. They generally fold reproducibly and rapidly into their native state structures. These structures are directly implicated in protein function. The native state conformations are modular and made up of building blocks, notably helices and sheets comprised of zig-zag strands. These common characteristics arise because proteins share the same backbone despite having distinct sequences. An important open challenge is the determination of the simplest chain model that would suitably describe the striking features of the common backbones of protein native state structures.

Author Contributions: Conceptualization, J.R.B, A.M., and T.Š.; methodology, J.R.B, A.M., and T.Š.; software, T.X.H. and T.Š.; validation, J.R.B, A.M., and T.Š.; formal analysis, T.Š.; investigation, J.R.B, A.M., T.Š.; resources, J.R.B, A.G., T.X.H., A.M., and T.Š.; data curation, T.Š.; writing—original draft preparation, J.R.B.; writing—review and editing, J.R.B, A.G., T.X.H., A.M., and T.Š.; visualization, J.R.B, A.M., and T.Š.; supervision, J.R.B; project administration, J.R.B.; funding acquisition, J.R.B, A.G., T.X.H., A.M., and T.Š. All authors have read and agreed to the published version of the manuscript.

Funding: This project received funding from AdiR (Assegnazioni Dipartimentali per la Ricerca, at Ca' Foscari University of Venice), and European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant Agreement No. 894784 (T.Š.). J.R.B. was supported by a Knight Chair at the University of Oregon. AG acknowledges support from the Grant PRIN-COFIN 2022JWAF7Y. TXH was supported by the Vietnam Academy of Science and Technology under grant No. NVCC05.02/24-25.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Flory P. *Statistical Mechanics of Chain Molecules*. Inter-science Publishers: New York, USA, 1969.
2. de Gennes P.G. *Scaling Concepts in Polymer Physics*. Cornell University Press: Ithaca, USA, 1979.
3. Khokhlov A.R.; Grosberg A.Y.; Pande V.S. *Statistical Physics of Macromolecules (Polymers and Complex Materials)*, 4th ed.; American Institute of Physics: New York, USA, 2002.
4. Rubinstein M.; Colby R.H. *Polymer Physics (Chemistry)*, 1st ed; Oxford University Press: London, UK, 2003.
5. Kratky, O.; Porod, G. *Röntgenuntersuchung gelöster Fagenmoleküle*. Rec. Trav. Chim. Pays-Bas **1949**, 68, 1106–1123. DOI: [10.1002/recl.19490681203](https://doi.org/10.1002/recl.19490681203)
6. Schellman, J. A. *Flexibility of DNA*. Biopolymers **1974**, 13, 217–226. DOI: [10.1002/bip.1974.360130115](https://doi.org/10.1002/bip.1974.360130115)
7. Bustamante C.; Bryant Z.; Smith S.B. *Ten years of tension: single-molecule DNA mechanics*. Nature **2003**, 421, 423-427. DOI: [10.1038/nature014051](https://doi.org/10.1038/nature014051)
8. Kauzmann W. *Some factors in interpretation of protein denaturation*. Adv. Protein Chem. **1959**, 14, 1–63. DOI: [10.1016/S0065-3233\(08\)60608-7](https://doi.org/10.1016/S0065-3233(08)60608-7)
9. Privalov P. L.; Gill S.J. *Stability of protein structure and hydrophobic interaction*. Adv. Protein Chem. **1988**, 39, 191–234. DOI: [10.1016/S0065-3233\(08\)60377-0](https://doi.org/10.1016/S0065-3233(08)60377-0)
10. Dill K.A. *Dominant forces in protein folding*. Biochemistry **1990**, 29, 7133–7155. DOI: [10.1021/bi00483a001](https://doi.org/10.1021/bi00483a001)
11. Stickle D.F.; Presta L.G.; Dill K.A.; Rose G.D. *Hydrogen bonding in globular proteins*. J. Mol. Biol. **1992**, 226, 1143–1159. DOI: [10.1016/0022-2836\(92\)91058-w](https://doi.org/10.1016/0022-2836(92)91058-w)
12. Horinek D.; Serr A.; Geisler M.; Pirzer T.; Slotta U.; Lud S.Q.; Garrido J.A.; Scheibel T.; Hugel T. *Peptide absorption on a hydrophobic surface results from interplay of solvation, surface, and*

intrapeptide forces. Proc. Natl. Acad. Sci. USA **2008**, *105*, 2842-2847. DOI: [10.1073/pnas.0707879105](https://doi.org/10.1073/pnas.0707879105)

13. Newberry R.W.; Raines R.T. *Secondary Forces in Protein Folding*. ACS Chem. Biol. **2019**, *14*, 1677–1686. DOI: [10.1021/acschembio.9b00339](https://doi.org/10.1021/acschembio.9b00339)

14. Khokhlov A.R. *On the θ -behaviour of a polymer chain*. Journal de Physique **1977**, *38*, 845-849. DOI: [10.1051/jphys:01977003807084500](https://doi.org/10.1051/jphys:01977003807084500)

15. de Gennes P.G. *Collapse of a flexible polymer chain II*. J. Physique Lett. **1978**, *39*, 299-301. DOI: [10.1051/jphyslet:019780039017029900](https://doi.org/10.1051/jphyslet:019780039017029900)

16. Ceperley D.; Kalos M.H.; Lebowitz J.L. *Computer Simulation of the Dynamics of a Single Polymer Chain*. Phys. Rev. Lett. **1978**, *41*, 313. DOI: [10.1103/PhysRevLett.41.313](https://doi.org/10.1103/PhysRevLett.41.313)

17. Fixman M. Simulation of polymer dynamics. I. General theory. J. Chem. Phys. **1978**, *69*, 1527–1537. DOI: [10.1063/1.436725](https://doi.org/10.1063/1.436725)

18. Williams C.; Brochard F.; Frisch H.L. *Polymer Collapse*. Annu. Rev. Phys. Chem. **1981**, *32*, 433-451. DOI: [10.1146/annurev.pc.32.100181.002245](https://doi.org/10.1146/annurev.pc.32.100181.002245)

19. Webman I.; Lebowitz J.L.; Kalos H. *A Monte Carlo Study of the Collapse of a Polymer Chain*. Macromol. **1981**, *14*, 1495-1501. DOI: [10.1021/ma50006a069](https://doi.org/10.1021/ma50006a069)

20. Bruns W.; Bansal R. Molecular dynamics study of a single polymer chain in solution. J. Chem. Phys. **1981**, *74*, 2064–2072. DOI: [10.1063/1.441253](https://doi.org/10.1063/1.441253)

21. Dünweg B.; Kremer K. *Molecular dynamics simulation of a polymer chain in solution*. J. Chem. Phys. **1993**, *99*, 6983–6997. DOI: [10.1063/1.465445](https://doi.org/10.1063/1.465445)

22. Tanaka G.; Wayne L. *Chain Collapse by Atomistic Simulation*. Macromol. **1995**, *28*, 1049-1059. DOI: [10.1021/ma00108a036](https://doi.org/10.1021/ma00108a036)

23. Wittkop M.; Kreitmeier S.; Göritz D. The collapse transition of a single polymer chain in two and three dimensions: A Monte Carlo study. J. Chem. Phys. **1996**, *104*, 3373–3385. DOI: [10.1063/1.471041](https://doi.org/10.1063/1.471041)

24. Binder K.; Paul W. *Monte Carlo Simulations of Polymer Dynamics: Recent Advances*. J. Polym. Sci.: Part B: Polymer Physics **1997**, *35*, 1 – 31. DOI: [10.1002/\(SICI\)1099-0488\(19970115\)35:1%3C1::AID-POLB1%3E3.0.CO;2-%23](https://doi.org/10.1002/(SICI)1099-0488(19970115)35:1%3C1::AID-POLB1%3E3.0.CO;2-%23)

25. Michel A.; Kreitmeier S. *Molecular dynamics simulation of the collapse of a single polymer chain*. Comput. Theor. Polym. Sci. **1997**, *7*, 113-120. DOI: [10.1016/S1089-3156\(97\)00018-4](https://doi.org/10.1016/S1089-3156(97)00018-4)

26. J. D. Honeycutt J.D. *A general simulation method for computing conformational properties of single polymer chains*. Comput. Theor. Polym. Sci. **1998**, *8*, 1-8. DOI: [10.1016/S1089-3156\(97\)00025-1](https://doi.org/10.1016/S1089-3156(97)00025-1)

27. Aksimentiev A.; Hołyst R. *Single-chain statistics in polymer systems*. *Prog. Polym. Sci.* **1999**, *24*, 1045–1093. DOI: [10.1016/S0079-6700\(99\)00023-4](https://doi.org/10.1016/S0079-6700(99)00023-4)
28. Mavrntzas V.G.; Boone T.D.; Zervopoulou E.; Theodorou D.N. *End-Bridging Monte Carlo: A Fast Algorithm for Atomistic Simulation of Condensed Phases of Long Polymer Chains*. *Macromol.* **1999**, *32*, 5072-5096. DOI: <https://doi.org/10.1021/ma981745g>
29. Sikorski A. *Computer Simulation of Adsorbed Polymer Chains with a Different Molecular Architecture*. *Macromol. Theory. Simul.* **2001**, *10*, 38–45. DOI: [10.1002/1521-3919\(20010101\)10:1%3C38::AID-MATS38%3E3.0.CO;2-114](https://doi.org/10.1002/1521-3919(20010101)10:1%3C38::AID-MATS38%3E3.0.CO;2-114)
30. Kreer T.; Baschnagel J.; Müller M.; Binder K. *Monte Carlo Simulation of Long Chain Polymer Melts: Crossover from Rouse to Reptation Dynamics*. *Macromol.* **2001**, *34*, *4*, 1105–1117. DOI: [10.1021/ma001500f](https://doi.org/10.1021/ma001500f)
31. Fujiwara S.; Sato T. *Structure formation of a single polymer chain. I. Growth of trans domains*. *J. Chem. Phys.* **2001**, *114*, 6455–6463. DOI: [10.1063/1.1356440](https://doi.org/10.1063/1.1356440)
32. Müller-Plathe F. *Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back*. *Chem. Phys. Phys. Chem.* **2002**, *3*, 754-769. DOI: [10.1002/1439-7641\(20020916\)3:9%3C754::AID-CPHC754%3E3.0.CO;2-U](https://doi.org/10.1002/1439-7641(20020916)3:9%3C754::AID-CPHC754%3E3.0.CO;2-U)
33. Abrams C.F.; Lee N.-K.; Obukhov S.P. *Collapse dynamics of a polymer chain: Theory and simulation*. *Europhys. Lett.* **2002**, *59*, 391. DOI: [10.1209/epl/i2002-00207-5](https://doi.org/10.1209/epl/i2002-00207-5)
34. Auhl R.; Everaers R.; Grest G.S.; Kremer K; Plimpton S.J. *Equilibration of long chain polymer melts in computer simulations*. *J. Chem. Phys.* **2003**, *119*, 12718–12728. DOI: [10.1063/1.1628670](https://doi.org/10.1063/1.1628670)
35. Rampf F.; Binder K.; Paul W. *The Phase Diagram of a Single Polymer Chain: New Insights From a New Simulation Method*. *J. Polym. Sci. Part B: Polymer Physics* **2006**, *44*, 2542–2555. DOI: [10.1002/POLB.20908](https://doi.org/10.1002/POLB.20908)
36. Binder K.; Baschnagel J.; Müller M.; Paul W.; Rampf F. *Simulation of Phase Transitions of Single Polymer Chains: Recent Advances*. *Macromol. Symp.* **2006**, *237*, 128–138. DOI: [10.1002/masy.200650514](https://doi.org/10.1002/masy.200650514)
37. Binder K.; Paul W.; Strauch T.; Rampf F.; Ivanov V.; Luettmer-Strathmann J. *Phase transitions of single polymer chains and of polymer solutions: insights from Monte Carlo simulations*. *J. Phys.: Condens. Matter* **2008**, *20*, 494215. DOI: [10.1088/0953-8984/20/49/494215](https://doi.org/10.1088/0953-8984/20/49/494215)
38. Taylor M.P.; Paul W.; Binder K. All-or-none proteinlike folding transition of a flexible homopolymer chain. *Phys. Rev. E* **2009**, *79*, 050801. DOI: [10.1103/PhysRevE.79.050801](https://doi.org/10.1103/PhysRevE.79.050801)

39. Taylor M.P.; Paul W.; Binder K. Phase transitions of a single polymer chain: A Wang–Landau simulation study. *J. Chem. Phys.* **2009**, *131*, 114907. DOI: [10.1063/1.3227751](https://doi.org/10.1063/1.3227751)
40. Rossi G.; Monticelli L.; Puisto S.R.; Vattulainen I.; Alla-Nissilla T. *Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case*. *Soft Matter* **2011**, *7*, 698-708. DOI: [10.1039/C0SM00481B](https://doi.org/10.1039/C0SM00481B)
41. Taylor M.P.; Paul W.; Binder K. Applications of the Wang–Landau Algorithm to Phase Transitions of a Single Polymer Chain. *Polymer Sci. Series C* **2013**, *55*, 23-28. DOI: [10.1134/S1811238213060040](https://doi.org/10.1134/S1811238213060040)
42. Jiang Z.; Dou W.; Sun T.; Shen Y.; Cao D. *Effects of chain flexibility on the conformational behavior of a single polymer chain*. *J. Polym. Res.* **2015**, *22*, 236. DOI: [10.1007/s10965-015-0875-3](https://doi.org/10.1007/s10965-015-0875-3)
43. Tzounis P.-N.; Anogiannakis S.D.; Theodorou D.N. *General Methodology for Estimating the Stiffness of Polymer Chains from Their Chemical Constitution: A Single Unperturbed Chain Monte Carlo Algorithm*. *Macromol.* **2017**, *50*, 4575–4587. DOI: [10.1021/acs.macromol.7b00645](https://doi.org/10.1021/acs.macromol.7b00645)
44. Gartner T.E. III; Jayaraman A. *Modeling and Simulations of Polymers: A Roadmap*. *Macromol.* **2019**, *52*, 3, 755–786. DOI: [10.1021/acs.macromol.8b01836](https://doi.org/10.1021/acs.macromol.8b01836)
45. Pauling L.; Corey R.B.; Branson H.R. *The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain*. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 205. DOI: [10.1073/pnas.37.4.205](https://doi.org/10.1073/pnas.37.4.205)
46. Pauling L.; Corey R.B. *The pleated sheet, a new layer configuration of polypeptide chains*. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 251. DOI: [10.1073/pnas.37.5.251](https://doi.org/10.1073/pnas.37.5.251)
47. Richardson J.S. The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* 1981; *34*: 167. DOI: [10.1016/S0065-3233\(08\)60520-3](https://doi.org/10.1016/S0065-3233(08)60520-3)
48. Rose G.D.; Young W.B.; Giersch L.M. Interior turns in globular proteins. *Nature* 1983; *304*: 654. DOI: [10.1038/304654a0](https://doi.org/10.1038/304654a0)
49. Rose G.D.; Giersch L.M.; Smith J.A. Turns in peptides and proteins. *Adv. Protein Chem.* 1985; *37*: 1-109. DOI: [10.1016/S0065-3233\(08\)60063-7](https://doi.org/10.1016/S0065-3233(08)60063-7)
50. Škrbić T.; Badasyan A.; Hoang T.X.; Podgornik R.; Giacometti A. *From polymers to proteins: the effect of side chains and broken symmetry on the formation of secondary structures within a Wang–Landau approach*. *Soft Matter* **2016**, *12*, 4783-4793. DOI: [10.1039/C6SM00542J](https://doi.org/10.1039/C6SM00542J)
51. Škrbić T.; Hoang T.X.; Giacometti A. *Effective stiffness and formation of secondary structures in a protein-like model*. *J. Chem. Phys.* **2016**, *145*, 0849041-08490410. DOI: [10.1063/1.4961387](https://doi.org/10.1063/1.4961387)
52. Škrbić T.; Hoang T.X.; Maritan A.; Banavar J.R.; Giacometti A. *The elixir phase of chain molecules*. *Proteins* **2019**, *87*, 176-187. DOI: [10.1002/prot.25619](https://doi.org/10.1002/prot.25619)

53. Škrbić T.; Hoang T.X.; Maritan A.; Banavar J.R.; Giacometti A. *Local symmetry determines the phases of linear chains: a simple model for the self-assembly of peptides*. *Soft Matter* **2019**, *15*, 5596-5613. DOI: [10.1039/C9SM00851A](https://doi.org/10.1039/C9SM00851A)
54. Škrbić T.; Banavar J.R.; Giacometti A. *Chain stiffness bridges conventional polymer and biomolecular phases*. *J. Chem. Phys.* **2019**, *151*, 174901. DOI: [10.1063/1.5123720](https://doi.org/10.1063/1.5123720)
55. Škrbić T.; Hoang T.X.; Giacometti A.; Maritan A.; Banavar J.R. *Spontaneous dimensional reduction and ground state degeneracy in a simple chain model*. *Phys. Rev. E* **2021**, *104*, L0121011-L0121017. DOI: [10.1103/PhysRevE.104.L012101](https://doi.org/10.1103/PhysRevE.104.L012101)
56. Škrbić T.; Hoang T.X.; Giacometti A.; Maritan A.; Banavar J.R. *Marginally compact phase and ordered ground states in a model polymer with side spheres*. *Phys. Rev. E* **2021**, *104*, L0125011-L0125017. DOI: [10.1103/PhysRevE.104.L012501](https://doi.org/10.1103/PhysRevE.104.L012501)
57. Škrbić T.; Maritan A.; Giacometti A.; Rose G.D.; Banavar J.R. *Building blocks of protein structures: Physics meets biology*. *Phys. Rev. E* **2021**, *104*, 0144021-0144027. DOI: [10.1103/PhysRevE.104.014402](https://doi.org/10.1103/PhysRevE.104.014402)
58. Škrbić T.; Hoang T.X.; Giacometti A.; Maritan A.; Banavar J.R. *Proteins — a celebration of consilience*. *Intl. J. Mod. Phys. B* **2022**, *36*, 2140051. DOI: [10.1142/S0217979221400518](https://doi.org/10.1142/S0217979221400518)
59. Banavar J.R.; Giacometti A.; Hoang T.X.; Maritan A.; Škrbić, T. *A geometrical framework for thinking about proteins*. *Proteins* **2023**. DOI: [10.1002/prot.26567](https://doi.org/10.1002/prot.26567)
60. Škrbić T.; Giacometti A.; Hoang T.X.; Maritan A.; Banavar J.B. *II. Geometrical framework for thinking about globular proteins: The power of poking*. *Proteins* **2023**. DOI: [10.1002/prot.26566](https://doi.org/10.1002/prot.26566)
61. Škrbić T.; Giacometti A.; Hoang T.X.; Maritan A.; Banavar J.B. *III. Geometrical framework for thinking about globular proteins: Turns in proteins*. *Proteins* **2024**. DOI: [10.1002/prot.26671](https://doi.org/10.1002/prot.26671)
62. 3D Macromolecule Analysis & Kinemage Home Page at Richardson Laboratory. <http://kinemage.biochem.duke.edu/databases/top8000/>
63. Škrbić T.; Maritan A.; Giacometti A.; Banavar J.R. *Local sequence-structure relationships in proteins*. *Protein Sci.* **2021**, *30*, 818-829. DOI: [10.1002/pro.4032](https://doi.org/10.1002/pro.4032)
64. Kabsch W.; Sander C. *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features*. *Biopolymers* **1983**, *22*, 2577-2637. DOI: [10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211)
65. Swendsen R.H.; Wang J.S. *Replica Monte Carlo Simulation of Spin Glasses*. *Phys. Rev. Lett.* **1986**, *57*, 2607. DOI: [10.1103/PhysRevLett.57.2607](https://doi.org/10.1103/PhysRevLett.57.2607)
66. Geyer C.J., *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, p. 156, New York, 1991. American Statistical Association.

67. Metropolis N.; Rosenbluth A.W.; Rosenbluth M.N.; Teller A.H. *Equation of State Calculations by Fast Computing Machines*. *J. Chem. Phys.* **1953**, *21*, 1087-1092. DOI: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114)
68. Rathore N.; Chopra M.; de Pablo J.J. *Optimal allocation of replicas in parallel tempering simulations*. *J. Chem. Phys.* **2005**, *122*, 024111. DOI: [10.1063/1.1831273](https://doi.org/10.1063/1.1831273)
69. Ferrenberg A.M.; Swendsen R.H. *Optimized Monte Carlo data analysis*. *Phys. Rev. Lett.* **1989**, *63*, 1195-1198. DOI: [10.1103/PhysRevLett.63.1195](https://doi.org/10.1103/PhysRevLett.63.1195)
70. Madras N.; Sokal A.D. *The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk*. *J. Stat. Phys.* **1988**, *50*, 109-186. DOI: [10.1007/BF01022990](https://doi.org/10.1007/BF01022990)
71. Creighton T.E. *Proteins: Structures and Molecular Properties*. W. H. Freeman: New York, USA, 1983.
72. Lesk A.M. *Introduction to Protein Science: Architecture, Function and Genomics*. Oxford University Press: Oxford, UK, 2004.
73. Bahar I.; Jernigan R.L.; Dill K.A. *Protein Actions*. Garland Science: New York, USA, 2017.
74. Berg J.M.; Tymoczko J.L.; Gatto G.J. Jr; Stryer L. *Biochemistry*. Macmillan Learning: New York, USA 2019.
75. Maritan A.; Micheletti C.; Trovato A.; Banavar J.R. *Optimal shapes of compact strings*. *Nature* **2000**, *406*, 287-290. DOI: [10.1038/35018538](https://doi.org/10.1038/35018538)
76. Banavar J.R.; Gonzalez O.; Maddocks J.H.; Maritan A. *Self-interactions of strands and sheets*. *J. Stat. Phys.* **2003**, *110*, 35-50. DOI: [10.1023/A:1021010526495](https://doi.org/10.1023/A:1021010526495)
77. Doi M.; Edwards S.F. *The theory of polymer dynamics*. Clarendon Press: New York, USA, 1993.
78. Marenduzzo D.; Flammini A.; Trovato A.; Banavar J.R.; Maritan A. *Physics of thick polymers*. *J. Polym. Sci. Part B: Polym. Phys.* **2005**, *43*, 650-679. DOI: [10.1002/polb.20363](https://doi.org/10.1002/polb.20363)
79. Stanley H.E. *Scaling, universality, and renormalization: Three pillars of modern critical phenomena*. *Rev. Mod. Phys.* **1999**, *71*, S358. DOI: [10.1103/RevModPhys.71.S358](https://doi.org/10.1103/RevModPhys.71.S358)
80. Ramachandran G.N.; Mitra A.K. *An explanation for the rare occurrence of cis peptide units in proteins and polypeptides*. *J. Mol. Biol.* **1976**, *107*, 85. DOI: [10.1016/s0022-2836\(76\)80019-8](https://doi.org/10.1016/s0022-2836(76)80019-8)
81. Vilhena J.G.; Pawlak R.; D'Astolfo P.; Liu X.; Gnecco E.; Kisiel M.; Glatzel T.; Pérez R.; Häner R., Decurtins S.; Baratoff A.; Prampolini G.; Liu S.-X.; Meyer E. *Flexible Superlubricity Unveiled in Sidewinding Motion of Individual Polymeric Chains*. *Phys. Rev. Lett.* **2022**, *128*, 216102. DOI: [10.1103/PhysRevLett.128.216102](https://doi.org/10.1103/PhysRevLett.128.216102)
82. Cai W.; Trefs J.L.; Hugel T.; Balzer B.N. *Anisotropy of π - π Stacking as Basis for Superlubricity*. *ACS Materials Lett.* **2023**, *5*, 172-179. DOI: [10.1021/acsmaterialslett.2c00974](https://doi.org/10.1021/acsmaterialslett.2c00974)

83. Blumenson, L. E. *A Derivation of n-Dimensional Spherical Coordinates*. Am. Math. Mon. **1960**, 67, 63–66. DOI: [10.2307/2308932](https://doi.org/10.2307/2308932)
84. Rose G.D. *Protein folding-seeing is deceiving*. Protein Sci. **2021**, 30, 1606-1616. DOI: [10.1002/pro.4096](https://doi.org/10.1002/pro.4096)
85. Rose G.D. *From propensities to patterns to principles in protein folding*. Proteins **2023**. DOI: [10.1002/prot.26540](https://doi.org/10.1002/prot.26540)
86. Lesk A.M.; Chothia C. *Solvent Accessibility, Protein Surfaces, and Protein Folding*. Biophys. J. **1980**, 32, 35-57. DOI: [10.1016/S0006-3495\(80\)84914-9](https://doi.org/10.1016/S0006-3495(80)84914-9)
87. Kyte J.; Doolittle R.F. *A simple method for displaying the hydropathic character of a protein*. J. Mol. Biol. **1982**, 157, 105-132. DOI: [10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
88. Rose G.D.; Geselowitz A.R.; Lesser G.J.; Lee R.H.; Zehfus M.H. *Hydrophobicity of Amino Acid Residues in Globular Proteins*. Science **1985**, 229, 834-838. DOI: [10.1126/science.4023714](https://doi.org/10.1126/science.4023714)
89. Ben-Naim A. *Solvent effects on protein association and protein folding*. Biopolymers **1990**, 29, 567-596. DOI: [10.1002/bip.360290312](https://doi.org/10.1002/bip.360290312)
90. Alonso D.O.V.; Dill K.A. *Solvent Denaturation and Stabilization of Globular Proteins*. Biochemistry **1991**, 30, 5974-5985. DOI: [10.1021/bi00238a023](https://doi.org/10.1021/bi00238a023)
91. Southall N.T.; Dill K.A.; Haymet A.D.J. *A View of the Hydrophobic Effect*. J. Phys. Chem. B **2002**, 106, 521–533. DOI: [10.1021/jp015514e](https://doi.org/10.1021/jp015514e)
92. Hayashi T.; Yasuda S.; Škrbić T.; Giacometti A.; Kinoshita M. *Unraveling protein folding mechanism by analyzing the hierarchy of models with increasing level of detail*. J. Chem. Phys. **2017**, 147, 125102. DOI: [10.1063/1.4999376](https://doi.org/10.1063/1.4999376)
93. Davis C.M.; Gruebele M.; Sukenik S. *How does solvation in the cell affect protein folding and binding?* Curr. Opin. Struct. Biol. **2018**, 48, 23–29. DOI: [10.1016/j.sbi.2017.09.003](https://doi.org/10.1016/j.sbi.2017.09.003)
94. Hayashi T.; Inoue M.; Yasuda S.; Petretto E.; Škrbić T.; Giacometti A.; Kinoshita M. *Universal effects of solvent species on the stabilized structure of a protein*. J. Chem. Phys. **2018**, 149, 0451051-04510517. DOI: [10.1063/1.5042111](https://doi.org/10.1063/1.5042111)
95. Škrbić T.; Zamuner S.; Hong R.; Seno F.; Laio A.; Trovato A. *Vibrational entropy estimation can improve binding affinity prediction for non-obligatory protein complexes*. Proteins **2018**, 86, 393-404. DOI: <https://doi.org/10.1002/prot.25454>
96. Dongmo Foumthuim, C.J.; Carrer M.; Houvet M.; Škrbić T.; Graziano G.; Giacometti A. *Can the roles of polar and non-polar moieties be reversed in non-polar solvents?* Phys. Chem. Chem. Phys. **2020**, 22, 25848-25858. DOI: [10.1039/D0CP02948C](https://doi.org/10.1039/D0CP02948C)

97. Carrer M.; Škrbić T.; Bore S.L.; Milano G.; Cascella M.; Giacometti A. *Can Polarity-Inverted Surfactants Self-Assemble in Nonpolar Solvents?* J. Phys. Chem. B **2020**, *124*, 6448-6458. DOI: [10.1021/acs.jpcb.0c04842](https://doi.org/10.1021/acs.jpcb.0c04842)