WILEY **PROTEINS**

RESEARCH ARTICLE

The elixir phase of chain molecules

Tatjana Škrbić[1,2] | Trinh X. Hoang[3] | Amos Maritan[4] | Jayanth R. Banavar[2] | Achille Giacometti[1] (iD)

[1]Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia Campus Scientifico, Edificio Alfa, Venezia Mestre, Italy

[2]Department of Physics and Institute for Theoretical Science, University of Oregon, Eugene, Oregon

[3]Center for Computational Physics, Institute of Physics, Vietnam Academy of Science and Technology, Hanoi, Vietnam

[4]Dipartimento di Fisica e Astronomia, Università di Padova, and INFN, Padova, Italy

Correspondence
Jayanth Banavar, 1258 University of Oregon, Eugene OR 97403-1205 USA.
Email: banavar@uoregon.edu
and
Achille Giacometti, Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia, Campus Scientifico, Edificio Alfa, via Torino 155, 30170 Venezia Mestre, Italy.
Email: achille.giacometti@unive.it

Abstract

A phase of matter is a familiar notion for inanimate physical matter. The nature of a phase of matter transcends the microscopic material properties. For example, materials in the liquid phase have certain common properties independent of the chemistry of the constituents: liquids take the shape of the container; they flow; and they can be poured—alcohol, oil, and water as well as a Lennard-Jones computer model exhibit similar behavior when poised in the liquid phase. Here, we identify a hitherto unstudied "phase" of matter, the elixir phase, in a simple model of a polymeric chain whose backbone has the correct local cylindrical symmetry induced by the tangent to the chain. The elixir phase appears on breaking the cylindrical symmetry by adding side spheres along the negative normal direction, as in proteins. This phase, nestled between other phases, has multiple ground states made up of building blocks of helices and almost planar sheets akin to protein native folds. We discuss the similarities of this "phase" of a finite size system to the liquid crystal and spin glass phases. Our findings are relevant for understanding proteins; the creation of novel bioinspired nanomachines; and also may have implications for life elsewhere in the cosmos.

KEYWORDS

Monte Carlo simulations, polymer theory, protein folding

1 | INTRODUCTION

Polymer science,[1–4] the study of chain molecules (linear polymers), is a flourishing subject that has led to path-breaking advances in plastics, textiles, and a variety of other technologies. There are simple, yet powerful, paradigms for understanding chain molecules. A polymer chain is composed of many interacting monomers that are tethered together to form a linear array. If the only interaction is excluded volume or self-avoidance, a single chain is in a self-avoiding coil phase whose large scale behavior is distinct from that of a fully non-interacting chain in both two and three spatial dimensions. Upon adding self-attraction between pairs of non-adjacent monomers promoting chain compaction, there is a phase transition from a random coil phase at high temperatures to a highly degenerate compact phase at low temperatures.[1–3] Note that the notion of phases and phase transitions for a polymeric chain at non-zero temperatures strictly refer to a chain with an infinite number of monomers. Indeed, it is only in that limit that singularities in the thermodynamic potentials arise. However, real systems, and the polymer chain is no exception, are never infinite. Thus, we expect that, for long enough chains, we can identify the *memory* of the phases and the associated phase transitions in the infinite size limit. For finite size systems including an interacting many-body linear chain of finite length, phases as well as transitions between them can and do exist at zero temperature as some parameters of the system are varied. The existence of a first-order phase transition at zero temperature is typically signaled by the existence of metastable states at non-zero temperature that can persist for long time scales depending on the temperature and on the chain length. This is the context within which we carry out our analysis of a model of a protein, a finite length biopolymer.

A globular protein is a heteropolymer—a linear chain whose monomers can be of 20 types of naturally occurring amino acids with distinct side chains.[5] Thus, the number of possible sequences of a protein just 40 amino acids long is 20^{40}, an astronomically large number. Not all these sequences exhibit archetypical protein like behavior—folding rapidly and reproducibly under physiological conditions to their native state structures.[6] This is typically described in terms of two phases, the unfolded and folded, separated by a first-order transition, with the implicit meaning described above. Interestingly, protein structures are frequently tolerant to changes in amino acid sequence with mutations often causing only modest local changes in structure.[7] Remarkably, the total number of distinct protein native state folds is limited and only of the order of a few thousand.[8] This follows from the fact that native state folds are made of building blocks of tightly curled helices and strands assembled into almost planar sheet-like motifs. The number of distinct topologies in which these motifs can be assembled is limited.[9,10] The existence of these secondary motifs follows from two independent considerations: scaffolding provided by hydrogen bonds[11,12]; and the steric requirement to avoid atomic overlaps.[13]

Flory[1] wrote in 1969, *Synthetic analogs of globular proteins are unknown. The capability of adopting a dense globular configuration stabilized by self-interactions and of transforming reversibly to the random coil is characteristics peculiar to the chain molecules of globular proteins alone.* We note however that modern chemistry has made the engineering of synthetic analogs of globular proteins possible. The model proposed here provides a route for creating synthetic analogs of globular proteins and provides a bridge between polymer science and biomolecular science.

We begin with a standard classic homopolymer model. The linear chain comprises N main chain spheres of diameter σ in a railway train topology with, for simplicity, the distances between the centers of successive monomers set to be exactly b. To avoid spurious symmetries, we allow for the overlap of adjoining spheres ($\sigma \geq b$). Symmetry plays a key role in determining the phases of matter. For example, sensitive liquid crystal phases[14] form when a collection of spheres is replaced by a collection of anisotropic objects such as pencils or ellipsoids. The liquid crystal phase occurs at nonzero temperatures close to the liquid phase.

A sphere is isotropic and looks the same when viewed from any direction. In contrast, there is 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. Allowing adjoining spheres to overlap breaks the spherical symmetry of an individual main chain sphere. The overlap between adjoining main-chain spheres results in an entropic stiffness because of the reduction in the ability to bend the chain.[15] A hint that such symmetry breaking could be important is provided by the extreme case of a chain of coins, which, in the continuum limit, has a tube-like geometry.[16,17] And, as is well-known, a garden hose is often wound into a helical geometry in a hardware store. Quite remarkably, a tightly wound tube has roughly the same pitch to radius ratio[18] as an α-helix in a protein and two tightly wound tubes in a double helix have the same geometry[19] as a DNA double-helix. We further break the resulting cylindrical

symmetry by having side spheres of diameter σ_{sc} sticking out on each main sphere in the negative normal direction. Our model is achiral, for simplicity. A chiral version of the model would have the side spheres in the plane perpendicular to the tangent and sticking out at a nonzero angle to the negative normal of the chain. The side spheres not only break the cylindrical symmetry of the overlapping spheres but also provide additional steric hindrance. The need to have the correct symmetry of a chain molecule has often been overlooked in the vast polymer science literature.

The main chain monomers are subject to a generic short-range all/nothing attractive interaction. Any pair of monomers lying within a threshold distance, R_c, of each other is assigned a favorable unit of energy. To determine the ground state, one seeks to find the conformation with the minimum energy or, equivalently, the largest number of contacts within the threshold distance. There is one energy scale and four length scales in the model. The magnitude of the energy scale plays no role in determining the ground state phase diagram. Without loss of generality, we set the diameter of the main chain sphere to be 1 and measure the three other length scales in units of the main chain sphere diameter. The model that we study is entirely standard except for the somewhat unusual self-avoidance geometrical conditions, dictated by symmetry considerations. These alone lead to a rich and surprising phase diagram.

2 | MATERIALS AND METHODS

In this section, we shall provide the details of the formalism and the simulation protocol that have been used in the simulations.

2.1 | The model

Figure 1A gives a sketch of our model. It consists of N tethered spheres located at positions {$\mathbf{r}_1$, ..., $\mathbf{r}_N$}, each having diameter σ and a consecutive sphere-sphere distance along the chain of $b \leq \sigma$. The case $b = \sigma$ corresponds to neighboring spheres along the chain just touching each other. Here, consecutive main-chain spheres are in general allowed to interpenetrate to respect the correct (cylindrical) symmetry imposed by the local tangent to the curve and in this case $b/\sigma < 1$. Non-consecutive spheres have hard core interactions, as well as an attractive square-well interaction of range R_c capturing non-directional generic hydrophobic interactions. Side spheres of diameter σ_{sc} are added to each of the $N - 2$ internal spheres along the negative normal direction, $-\hat{\mathbf{N}}_i$, in the Frenet frame (see Figure 1B) to capture the steric hindrance of the side chains and to break the cylindrical symmetry in an achiral manner.

2.2 | The Frenet formalism

The Frenet frame is defined by a tangent vector[20]

$$\hat{\mathbf{T}}_i = \frac{\Delta_i + \Delta_{i+1}}{|\Delta_i + \Delta_{i+1}|}, \tag{1}$$

where $\Delta_i = \mathbf{r}_i - \mathbf{r}_{i-1}$, so that $|\Delta| = b$. For each of the non-terminal backbone spheres, one defines a normal vector
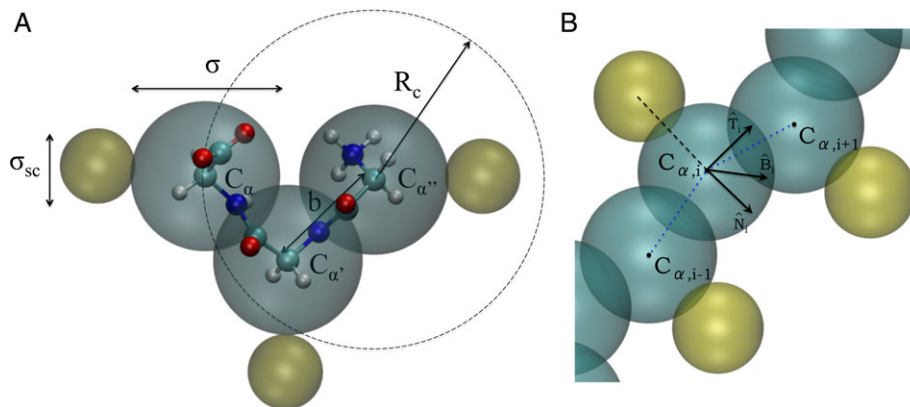
FIGURE 1 (A) Chain model. Each main chain sphere has diameter σ. The side sphere is in the negative normal direction and has a diameter σ_{sc}. The distance between successive main chain spheres is $b \leq \sigma$—consecutive spheres can, in general, partially overlap. Nonconsecutive main chain spheres experience a short range attractive constant potential if their separation is within the range of the attraction R_c. The atoms of glycine amino acids are shown as a guide. (B) Side sphere positions in Frenet frame [Color figure can be viewed at wileyonlinelibrary.com]

$$\hat{\mathbf{N}}_i = \frac{\mathbf{\Delta}_{i+1} - \mathbf{\Delta}_i}{|\mathbf{\Delta}_{i+1} - \mathbf{\Delta}_i|}, \tag{2}$$

where $i = 2, \ldots, N - 1$. The corresponding binormal vector is then given by

$$\hat{\mathbf{B}}_i = \hat{\mathbf{T}}_i \times \hat{\mathbf{N}}_i. \tag{3}$$

These equations are discrete versions of the continuum Frenet frame frequently used in polymer theory.[20]

Apart from their excluded volume, the side spheres do not interact either with the backbone spheres or with each other. The side spheres merely have steric interactions with main chain spheres and with each other.

The position of the side sphere is defined by

$$\mathbf{r}_i^{(sc)} = \mathbf{r}_i - \hat{\mathbf{N}}_i \frac{(\sigma + \sigma_{sc})}{2}; \tag{4}$$

the side chain sphere and the backbone sphere are tangent to each other. Amino acids in real proteins have an average tilt of roughly 41° with respect to the negative normal direction and have different sizes. Our model does not have any chirality built into it.

2.3 | Simulations protocol

We have studied the zero-temperature phase diagram using microcanonical Wang-Landau[21] and parallel tempering (also known as replica exchange) canonical Monte Carlo simulations,[22] always with consistent results. In the Wang-Landau simulations, we sample polymer conformations according to the microcanonical distribution by generating a sequence of chain conformations $A \rightarrow B$, and accepting the new configuration B with the microcanonical acceptance probability

$$P_{acc}(A \rightarrow B) = \min\left(1, \frac{w_B g(E_A)}{w_A g(E_B)}\right), \tag{5}$$

where w_A and w_B are weight factors ensuring the microscopic reversibility of the moves. The set of MC moves, that are accepted or rejected according to the probability (5) includes both local-type moves, such as single-sphere crankshaft, reptation, and end-point, as well as nonlocal-type moves, such as pivot, bond-bridging, and back-

bite moves, randomly sampled so that on average N spheres (or a multiple of it) are moved to complete a MC step.

The density of states $g(E)$ is then constructed iteratively by filling suitable energy histograms and controlling their flatness. However, to compute the ground state energy, the lowest energy was consecutively selected using the acceptance probability (5) with a bias toward less populated energy states. This corresponds to the usual preliminary calculation carried out without a low-energy cut-off in the usual Wang-Landau scheme. In the full Wang-Landau calculation, we typically assume convergence after 30 levels of iterations, corresponding to a multiplicative factor value of $f = 10^{-9}$. For the ground state calculation, each run is composed of at least 10^9 Monte Carlo moves per sphere.

In the parallel tempering simulations, we carried out individual canonical Monte Carlo simulations within a Metropolis scheme at fixed temperature, with periodic swapping between neighboring temperatures within a prescribed annealing schedule. The acceptance probability for the exchange between two replicas Γ_i and Γ_j at temperatures T_i and T_j, respectively, is given by the acceptance probability

$$P_{swap} = \min\left(1, \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)(E_i - E_j)\right]\right). \tag{6}$$

Twelve replicas were used with the replica temperatures expanding from those of the swollen phase to those of the compact phase. The total number of steps ranged from 10^8 to 10^9 sweeps per replica, depending on the system size, one step corresponding to N attempted MC moves, where N is the number of main chain spheres. A replica exchange is attempted every 10 Monte Carlo steps. The move sets include pivot, crankshaft, and reptation moves with probabilities 0.2, 0.7, and 0.1, respectively.

The results obtained are robust and independent of the technique used.

3 | RESULTS

The self-avoidance is unusual (and motivated by protein chemistry) compared to previous studies because of the absence of spurious

symmetries, which is a feature that has been often overlooked. The phase diagram is a result of the competition between the large number of self-avoiding conformations (entropy) and the few self-avoiding compact conformations, which avail of the attractive interactions and is a transition driven by temperature. The elixir phase persists over a range of temperatures and is a low temperature phase. Our model is for a single chain. Unlike a system of hard spheres where packing fraction is a key parameter, the density does not play a role here. The microcanonical Wang-Landau calculations allow one to measure the density of states and hence the free energy from which all thermodynamics can be obtained. The same method also provides a direct way to access the ground state of the system. In parallel tempering, we gradually lower the temperature and monitor the energy. Below the folding transition temperature, this is taken as the ground state energy and matches the value obtained via the Wang-Landau calculation. Taken together, these methods confirm that the elixir phase is stable over a range of low temperatures. When the system is cooled below the folding transition temperature, it achieves the analog of the native state in real proteins. We denote this as a ground state and Figure 2 displays the corresponding phase diagram. In addition to conventional polymer phases, the phase diagram has regions with a unique ground state including a single β-sheet, a single helix, and two helices wrapped around each other. Nestled between these phases lies a

particularly interesting phase, that we have denoted as the *elixir phase*, whose ground states are assembled structures of helices and strands. The elixir phase is degenerate with the assembled structures having nearly the same energy (see below). Intriguingly, the geometries of the building block strands and helices in the elixir phase, as well as those in a larger surrounding region identified by dotted lines in Figure 2D, are statistically the same as those of strands and helices in protein native state structures. Each ground state in the phase diagram was assigned to a specific phase on the basis of suitable order parameters[15] that includes the twist (for the helices) and the average triple scalar product of normal Frenet unit vectors (for the sheets). An additional fingerprint of the secondary structures stems from their characteristic representations in the contact maps. The figures label the phases and also depict the types of structures supported in each of the phases. The entire region between the coil and the globule phases shrinks to a single point for a conventional chain model with no overlap between consecutive main chain spheres and with no side spheres. This underscores the key roles of breaking, first, the spherical symmetry (capturing the correct local cylindrical symmetry associated with any chain molecule through the overlap) and then the cylindrical symmetry (through the side spheres). The phases labeled as helix and beta has ground states of a single helix and a single beta sheet, respectively.
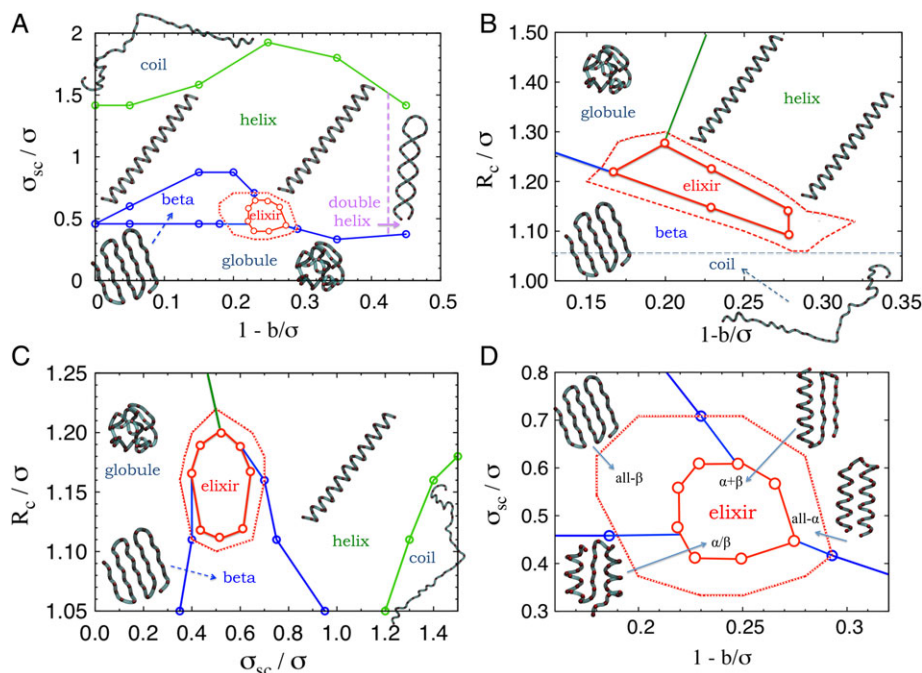

FIGURE 2   Phase diagram of chain model. The main chain sphere diameter σ sets the characteristic length scale of the model and is taken to be 5Å. The three dimensionless variables studied are: (1) the ratio of the diameters of the side-chain (σ_{sc}) and main chain (σ) spheres; (2) the ratio of the range of the attractive interaction (R_c) to the diameter of the main chain sphere (σ); and (3) the degree of overlap of consecutive main chain spheres measured by the difference between the diameter of the main chain sphere (σ) and the distance between the centers of consecutive main chain spheres along the chain (b) normalized by the diameter of the main chain spheres (σ). Plots A-C display the ground state phase diagrams in the three corresponding planes for a chain of length 40. Plot 2A: side-chain size σ_{sc}/σ-overlap $1 - b/\sigma$ plane. The range of interaction has been fixed at $R_c/\sigma = 1.16$, corresponding to $R_c \approx 6$Å. Plot 2B: range of attraction R_c/σ- overlap $1 - b/\sigma$ plane. The side chain sphere size has been fixed to $\sigma_{sc}/\sigma = 0.5$ corresponding to $\sigma_{sc} \approx 2.5$Å. Plot 2C: range of attraction R_c/σ-side-chain sphere size σ_{sc}/σ plane. The overlap value has been fixed at $1 - b/\sigma = 0.25$ corresponding to $b \approx 3.8$Å. In all cases, the central enclosed phase, denoted as the *elixir phase*, include nearly *degenerate* (in energy) conformations comprised of combinations of α helices and β strands. The larger dotted enclosed region, contain, in addition, conformations having either α helices or β strands with geometries matching those appearing in real proteins. These also include all-β and all-α conformations that are not part of the elixir phase (plot D) [Color figure can be viewed at wileyonlinelibrary.com]
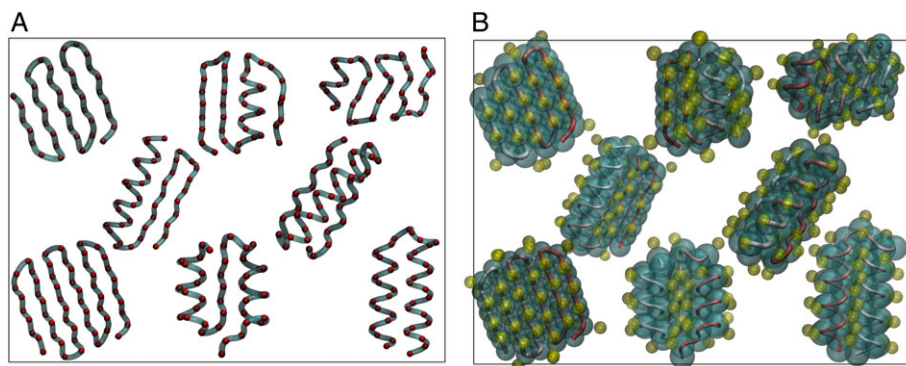
FIGURE 3 Plot A is a gallery of ground state structures found in the elixir phase and in the extended region surrounding it. All helices have radii (2.4 ± 0.1)Å and pitches (5.5 ± 0.5)Å matching those of protein helices within error bars. Likewise, all strands have distances between C_α^i and $C_\alpha^{i+2} \approx 6$Å, as in proteins. A tube-like representation of the chain has been used for better visualization. Plot 3B is the same as plot 3A, but with full display of the spheres. The ground state structures within the elixir phase and those shown in plots 3A and 3B are low energy structures with approximately the same number of contacts [Color figure can be viewed at wileyonlinelibrary.com]

The elixir phase is remarkable in several ways: it is stable for moderate length chains—we find no significant change in the nature of the phase on varying the chain length between 20 to around 200; the nearly degenerate ground state structures are geometrical sculptures assembled from building blocks of helices and sheets allowing for geometry based lock-key interactions so ubiquitous in living systems; the assembled structures in the elixir phase have topologies similar to the observed folds in proteins—$\alpha + \beta$ and α/β[23,24] (see Figure 3A,B for a gallery of conformations); and the elixir phase is surrounded by other phases and thus the structures within it are poised in the vicinity of phase transitions albeit for small systems.

Figure 4 shows the total number of ground state contacts, N_c, as a function of the three characteristic length scales of the model. The boundaries of the elixir phase in Figure 2A-D correspond to the vertical dashed lines reported in Figure 4. For positive increasing R_c/σ, the number of contacts N_c monotonically increases from very small values characteristic of the coil phase until a threshold value sufficiently large to form a β phase as shown in Figure 4C. The geometry of the strands,
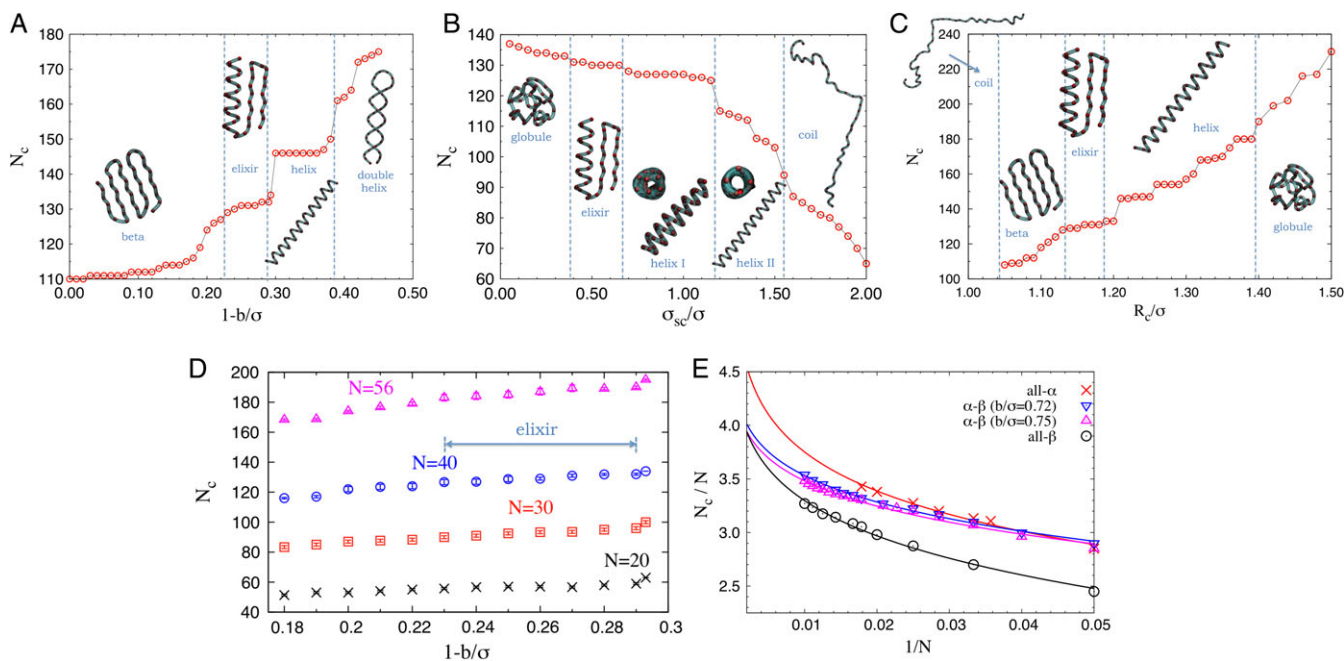

FIGURE 4 Number of ground state contacts N_c as a function of $1 - b/\sigma$ (A), σ_{sc}/σ (B), and R_c/σ (C) for a chain of length 40. The vertical lines indicate transitions to the phases housing conformations shown in the snapshots, and are consistent with the transitions of the phase diagrams in Figure 2. The transition helix I → helix II appearing in the central Plot is a structural transition between two different type of helices, and, for simplicity, is not explicitly displayed in the phase diagram of Figure 2. In all cases, the remaining two parameters have been set to the center of the elixir phase, as in Figure 2: $1 - b/\sigma = 0.25$, $\sigma_{sc}/\sigma = 0.5$, and $R_c/\sigma = 1.16$. Figure 4D shows the number of ground state contacts N_c as a function of $1 - b/\sigma$ for different chain lengths ($N = 20, 30, 40, 56$). Each point is the average over $5 - 10$ independent runs, with errors bars of the order of the size of each point. For $N = 40$ the range of the elixir phase is highlighted. Figure 4E shows the number of ground state contacts per sphere N_c/N for large N, showing that it tends to a value of ≈ 4. Note that the all β conformations tend to extrapolate to the same value, whereas the all-α conformations approach a slightly higher value of ≈ 4.5 [Color figure can be viewed at wileyonlinelibrary.com]
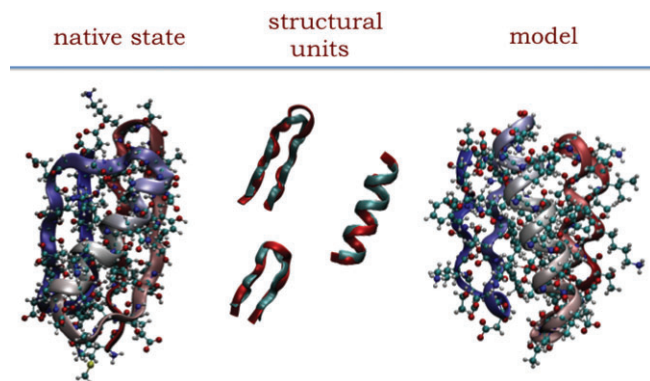
FIGURE 5 An example of comparison between ground state structures in the elixir phase and the native fold of a protein with $N = 56$. Protein G (PDBid 3GB1), whose native state is formed by two β antiparallel hairpins connected by a single α helix (left) and the topologically similar ground state in the elixir phase (right) with parameters: $\sigma = 5$Å, $1 - b/\sigma = 0.25$, $\sigma_{sc}/\sigma = 0.416$, and $R_c/\sigma = 1.16$. The root-mean-square deviation (RMSD) between the native state of the protein and the ground state of the model is ≈ 7Å. The central plot shows the fidelity of the overlap of the building blocks of the protein structure (α helices and β hairpins) to those in the elixir phase (RMSD) ≤ 2.0Å. Structural units of the real protein is shown in red; those from the model are in cyan

as measured by the angle subtended at each main chain sphere with its adjacent spheres along the chain, changes smoothly until it becomes similar to that of a protein strand at $R_c/\sigma \approx 1.1$. Upon increasing R_c/σ further, one enters the degenerate elixir phase with combined α–β structures, where the number of contacts is essentially independent of R_c/σ as indicated by the flatness of N_c in this region. At $R_c/\sigma \approx 1.20$, an abrupt upswing of N_c signals the end of the elixir phase and entry into a single helix phase.

A similar phenomenology is observed along the σ_{sc}/σ axis, as shown in Figure 4B. Here, N_c monotonically decreases from the globular phase until it becomes flat upon entering into the elixir phase. The discontinuities seen within the helix phase correspond to a structural transition between two different geometries of helix (that we denote as helix I and helix II). The behavior of N_c as a function of $1 - b/\sigma$, shown in Figure 4A, has a single beta phase for all $1 - b/\sigma > 0$ (with $\sigma_{sc}/\sigma = 0.5$ and $R_c/\sigma = 1.16$). Again, the strand geometry changes smoothly and matches that of protein strands around $1 - b/\sigma \approx 0.18$,

as one approaches the elixir phase. The initial assembled topology is an all-β conformation (see Figure 2D), and N_c progressively increases until a combined α–β ground state conformational topology is achieved, where N_c is again approximately flat. At $1 - b/\sigma \approx 0.29$, a drastic conformational change to a single helix is observed, where N_c becomes flat again in the range $0.29 \leq 1 - b/\sigma \leq 0.39$ characterized by maximally packed helices with consecutive turns lying one on top of the other.

Movies 1-3 included in the Supporting Information show how the final ground state is eventually achieved upon decreasing the energy. Movies 1 and 2 refer to state points corresponding to single β and single α conformation, respectively, none of which have the correct parameters matching those in real proteins. Movie 3 shows how one of the possible ground states in the elixir phase is reached, usually through multiple conformational changes.

A characteristic feature of the elixir phase is its approximate degeneracy, with many different ground state conformations with $N_c \approx 130$ for chain length $N = 40$ (see Figure 4D for different chain lengths N going beyond the boundaries of the elixir phase). Figure 4E shows that the number of ground state contacts per main chain sphere, N_c/N, tends to a value of ≈ 4 in the elixir phase, in the limit of large N. Having building blocks (helices and beta strands) with the correct interplay between the geometries and range of interactions,[25] distinct self-assembled ground state topologies have comparable energies. Thus, one can construct a library of many different topological folds in the elixir phase, all having energies within a small interval of $\approx 10\%$ spread (see Figure 3A,B for a representative sample).

Figure 5 illustrates the remarkable similarity between the native-fold of a protein and the corresponding ground state found in the elixir phase. Here, we have taken Protein G as representative example, but this result is valid for essentially *any* topology. The left plot of Figure 5 represents its real native state. The right plot of Figure 5 shows the ground state conformation in the elixir phase, where the specific sequence pertaining to protein G has been inserted by using the PULCHRA tool.[26] The central plot shows the overlap of the single units, both the α helix and the β sheets, illustrating how they have the correct geometries matching those of real proteins.

We stress the importance of selecting the correct value of $R_c/\sigma \approx 1.16$ (corresponding to $R_c \approx 6$Å for $\sigma \approx 5$Å) in order to observe the elixir phase. For either larger or smaller values of R_c/σ, the elixir phase
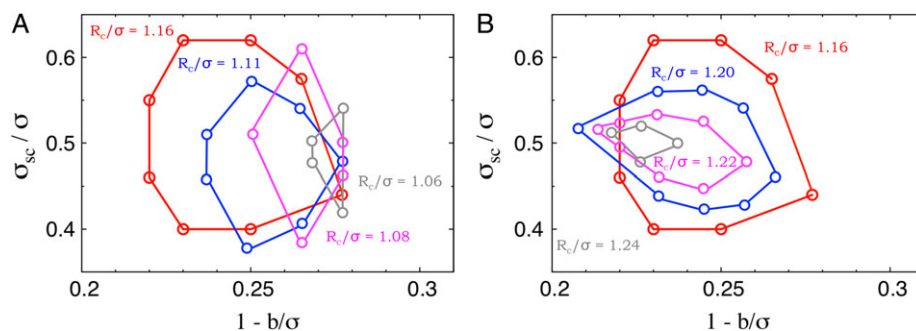

FIGURE 6 (A) Schematic view of the gradual shrinking of the elixir phase for a chain of length 40 upon decrease of the attraction range, R_c. The largest loop corresponds to the boundary of the elixir phase of Figure 2D. Note, the tendency to form a vertical line rather than a point in the limit $R_c/\sigma \to 1$. (B) Same upon increasing R_c. Here, unlike in the opposite limit, the elixir phase tends to shrink to a point and eventually disappear [Color figure can be viewed at wileyonlinelibrary.com]

gradually shrinks until it eventually disappears, as displayed in Figure 6. In the three-dimensional parameter space, $\{1 - b/\sigma, \sigma_{sc}/\sigma, R_c/\sigma\}$, the elixir phase is centered at the values of parameters found in proteins, and has a lemon shape with its two ends forming a meeting point for the helix, the beta, and the globular phases for large R_c/σ, and a somewhat extended line (the origin and the explanation of this line will be discussed elsewhere) for the helix, the beta, and the coil phases for small R_c/σ. In essence, the elixir phase is an extended coexistence region with a degenerate ground state.

We note that, while helices and beta strands have the correct geometries within the elixir phase (and not outside it), our simplified model has shortcomings in capturing some of the details of real protein structures. One is the out-of-phase arrangement between parallel strands of a β-sheet in contrast with the in-phase arrangement in the β-sheet of real proteins. Another is the difference in the number of residues per turn: 3.6 in real helices and 4 in elixir helices. Both discrepancies can be cured by introducing a binormal-binormal interaction in the model between the Frenet reference frames of main chain spheres in spatial proximity with each other. We stress that our simple model is adequate for identifying the elixir phase and our goal here is not to mimic the glorious complexities of protein structures by adding more details.

Form determines function for proteins. Furthermore, many globular proteins serve as extraordinarily powerful machines and catalysts. Bernal[27] wrote, *any effective picture of protein structure must provide at the same time for the common character of all proteins as exemplified by their many chemical and physical similarities, and for the highly specific nature of each protein type.* The amazing common characteristics of proteins along with our observation of the elixir phase in a simple chain model emboldens us to make a constructive hypothesis that protein native state structures may lie in a phase of matter, which confers these properties and the attendant advantages on them. Our model does not rely on quantum chemistry except in an emergent sense. It does not incorporate hydrogen bonds, the approximate planarity of the peptide bond, the heterogeneity of the side chains, the molecular nature of the rich variety of amino acids and the important role played by the solvent water molecules. A consequence of our hypothesis is that the menu of protein native state structures is determined not by chemistry but rather by general considerations of geometry and the absence of spurious symmetries. The role of the sequence then would be to choose its native-fold from this menu in a harmonious manner accounting for the highly specific nature of each protein type. Unlike an earlier study of Zhang et al.,[28] here we do not consider either an all-atom description of a protein or incorporate hydrogen bonds. It would be interesting to study the coil phase to elucidate the effective scaling exponent of the algebraic dependence of the radius of gyration on chain length[29] to assess whether the unfolded state of foldable sequences is expanded to suppress misfolding and aggregation.

4 | DISCUSSION

We find that the library of observed protein folds is a consequence of basic physical law rather than deriving from the considerations of sequence diversity and chemistry. The finiteness of the number of folds allows for conservation of the geometrical lock-key mechanism central to ligand binding, protein-protein interactions and functionality even as sequence evolution takes place. Our work suggests that Darwinian evolution shapes sequences and functionalities within the constraints of the immutable Platonic folds in the elixir phase. While many aspects of living systems here on earth can be understood within the framework of the theory of evolution, there are important constraints, enormous simplicities, as well as huge advantages deriving from the fact that basic physical law governs living matter.

Physical systems in the vicinity of a phase transition exhibit exquisite sensitivity to small perturbations of the right kind. Being poised in the vicinity of transitions to other phases affords significant advantages to proteins accounting for their power as molecular machines. We wish to emphasize that while our model is motivated by symmetry considerations and is generic and simple, it is directly inspired by protein chemistry. In the protein context, the absence of spurious symmetries arise from the presence of side chains, the self-attraction is mediated by water and the effect of hydrophobicity and the overlap of neighboring main chain spheres is a direct consequence of quantum chemistry and covalent bonding. What is truly remarkable is the stunning convergence (or fit) of the chemistry into the simple model we have studied. A corollary of our study is that protein-like behavior ought to be realizable in multiple unrelated ways opening a potential frontier in the creation of nifty machines and even artificial life. In summary, we find a new phase of matter, which has many attributes similar to those found in proteins. The elixir phase is observed over a range of temperatures. The ground state structures are zero temperature structures. The structures are not the same as protein native state structures. Rather, the elixir phase has certain features that make it a truly novel phase: it is a phase that is observed for a moderate length chain; the structures within it are comprised of helices and sheets; the phase is nestled between other phases and this confers exquisite sensitivity to this phase, which, in turn, makes it suitable for housing machines; the phase exhibits multiple nontrivially connected ground states allowing for switching between them; the ground states are geometrical structures allowing for a lock-key relationship that underlies life; the phase is observed with just self-avoidance and a simple self-attraction regardless of how they arise; the self-avoidance does not have spurious symmetries and the self attraction can be mediated by water molecules and hydrophobicity; hydrogen bonds and all-important chemical details including amino acid specificity are not considered but it is quite remarkable how well they fit into the generic scheme; the notion of a phase means that there may be a plethora of ways in which a system can be generated that is housed within the phase; our work therefore transcends the glories of proteins but includes it as a special case.

While, to the best of our knowledge, the elixir phase has not been identified in any previous studies, the model studied here, but with parameters poising the system outside the elixir phase, has been found to exhibit the characteristics of a machine by spontaneously switching between two distinct geometries, a single helix, and a double helix, merely because of thermal fluctuations.[17] A consequence of the existence of the elixir phase is that it can be exploited for the creation of nifty artificially made interacting nanomachines.

The elixir phase is distinct from conventional phases of matter in that the variety of ground state structures is geometry-based and

occurs for finite size systems. Our model may be thought of as a generalization of the liquid crystal phase in a chain topology. Just like liquid crystals, the elixir phase structures are stable yet sensitive. Unlike liquid crystals, which occur at nonzero temperatures, the sculptures in the elixir phase are ground states. The elixir phase structures are neither relatively open structures (as in the coil phase) nor are they maximally compact (as in the globular phase). They lie within a marginally compact phase. The ground states in the elixir phase are neither non-degenerate (as in the helix phase) nor do they have a huge degeneracy (as in the globular and especially the coil phases). The elixir phase has an intermediate degeneracy. These "Goldilocks"-like just-right characteristics make the elixir phase, an attractive candidate for facilitating functionality.

The elixir phase is distinct from what is known as the *molten globule* that is believed to be an intermediate state of the folding process.[30] Rather, the elixir phase ground states are the counterparts of the native states of globular proteins.

The elixir phase has characteristics vaguely similar to spin glasses.[31] Both a spin glass and the elixir phase have energy landscapes with multiple stable minima with barriers between them. This property of spin glasses arises from frustration in conflicting interactions and has been invoked to model content addressable memories[32] and prebiotic evolution.[33,34] The protein free energy landscape has been described to be minimally frustrated leading to a folding funnel geometry.[35–37] Of course, in the extreme limit of an unfrustrated system, one obtains, in standard spin models, a unique ground state along with its symmetric partners (e.g., all up spins or all down spins for an Ising ferromagnet). The elixir phase is novel in that there is no conventional frustration—rather, there are geometrical constraints that allow for distinct chain conformations (multiple ground states) to avail of roughly the same attractive energy; the system is not infinitely large; and the ground states are modular, geometry-based structures related by the distinct topologies of the assembled secondary structures. In fact, the presence of these ground state folds in a homopolymer model suggests that the top of the folding funnel is engineered with no sequence information into many broad basins and the latter part of dynamical folding entails the harmonious fitting of a minimally frustrated sequence into its native state basin. Our picture leads to additional simplicities for understanding and exploring the relative ease of the folding dynamics of globular proteins.[38,39] It would be interesting to look for experimental verification of the existence of the elixir phase in colloidal systems where exquisite control can be achieved.[40,41]

The elixir phase is the ground state of our highly simplified model. Thus, it is the analog of the native state in real proteins. However, it can also be thought of as an *approximate* description of the molten globule in real proteins,[30] that is a conformation within which secondary structures have still not been fully formed. Of course, this lack of refinement of secondary structures is the result of an incomplete model, for example, due to the absence of hydrogen bonds. Nevertheless, the approximate structures can be exploited by using the ground states of the elixir phase may be used as computationally efficient starting points for more refined calculations[42,43] that introduce additional chemical details. It would be extremely interesting to combine the present strategy with complementary coarse-grained approaches that have been proposed recently in the literature[44–47] to shed new light on the nature of protein folding pathway.[38,39]

A cell is not just a container of ordinary molecules—rather, it consists of incredibly powerful interacting molecular machines that orchestrate life. Proteins are amazing catalysts that speed up reactions by many orders of magnitude and carry out many of the functions of a living cell. They are essential ingredients of life. Darwin wrote about the origins of life: *But if we could conceive in some warm little pond with all sorts of ammonia and phosphoric salts, light, heat, electricity... present, that a protein compound was chemically formed, ready to undergo still more complex changes...* Our work suggests that Nature may have stumbled upon the elixir phase here on earth and eventually this resulted in life as we know it. This opens up the intriguing possibility that life elsewhere in our cosmos could well have a very different chemical basis with some of its vital molecules populating the same elixir phase of matter.

ORCID

Achille Giacometti https://orcid.org/0000-0002-1245-9842

REFERENCES

1. Flory P. *Statistical Mechanics of Chain Molecules*. New York: Interscience Publishers; 1969.
2. de Gennes P. *Scaling Concepts in Polymer Physics*. Ithaca: Cornell University Press; 1979.
3. Khokhlov AR, Grosberg AY, Pande VS. *Statistical Physics of Macromolecules (Polymers and Complex Materials)*. 1994th ed. New York: American Institute of Physics; 2002:3.
4. Rubinstein M, Colby RH. *Polymer Physics (Chemistry)*. 1st ed. London: Oxford University Press; 2003:6.
5. Creighton TE. *Proteins: Structures and Molecular Properties*. New York, 2nd ed. W. H. Freeman; 1992:8.
6. Anfinsen CB. Principles that govern the folding of protein chains. *Science*. 1973;181:223-230.
7. Baase WA, Liu L, Tronrud DE, Matthews BW. Lessons from the lysozyme of phage T4. *Protein Sci*. 2010;19:631-641.
8. Chothia C. Proteins 1000 families for the molecular biologist. *Nature*. 1992;357:543-544.
9. Chothia C, Levitt M, Richardson D. Structure of proteins: packing of alpha-helices and pleated sheets. *Proc Natl Acad Sci USA*. 1977;74(10): 4130-4134.
10. Przytycka T, Aurora R, Rose G. A protein taxonomy based on secondary structure. *Nat Struct Biol*. 1999;6:672-682.
11. Pauling L, Corey RB. The structure of synthetic polypeptides. *Proc Natl Acad Sci USA*. 1951;37:205-211.
12. Pauling L, Corey RB. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc Natl Acad Sci USA*. 1951;37:729-740.

13. Ramachandran G, Sasisekharan V. Conformation of polypeptide chains. *Adv Protein Chem*. 1968;23:283-437.

14. Chaikin P, Lubensky T. *Principles of Condensed Matter Physics*. London: Cambridge University Press; 2000.

15. Škrbić T, Hoang TX, Giacometti A. Effective stiffness and formation of secondary structures in a protein-like model. *J Chem Phys*. 2016;145: 084904.

16. Banavar JR, Maritan A. Physics of proteins. *Annu Rev Biophys Biomol Struct*. 2007;36:261-280.

17. Banavar JR, Cieplak M, Hoang TX, Maritan A. First-principles design of nanomachines. *Proc Natl Acad Sci USA*. 2009;106:6900-6903.

18. Maritan A, Micheletti C, Trovato A, Banavar J. Optimal shapes of compact strings. *Nature*. 2000;406:287-290.

19. Stasiak A, Maddocks J. Mathematics - best packing in proteins and DNA. *Nature*. 2000;406:251-253.

20. Kamien R. The geometry of soft materials: a primer. *Rev Mod Phys*. 2002;74:953-971.

21. Wang F, Landau D. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys Rev Lett*. 2001;86:2050-2053.

22. Frenkel D, Smit B. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science Series, Vol 1)*. 2nd ed. London: Academic Press; 2001:11.

23. Levitt M, Chothia C. Structural patterns in globular proteins. *Nature*. 1976;261(5561):552-558.

24. Richardson J, Richardson D, Tweedy N, et al. Looking at proteins - representations, folding, packing, and design - biophysical-Sociaty National Lecture, 1992. *Biophys J*. 1992;63:1186-1209.

25. Wang L, Qiao H, Cao C, Xu S, Zou S. An accurate model for biomolecular helices and its application to helix visualization. *PLoS One*. 2015; 10:1-13.

26. Rotkiewicz P, Skolnick J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J Comput Chem*. 2008;29:1460-1465.

27. Bernal J. Structure of proteins. *Nature*. 1939;143:663-667.

28. Zhang Y, Hubner IA, Arakaki AK, Shakhnovich E, Skolnick J. On the origin and highly likely completeness of single-domain protein structures. *Proc Natl Acad Sci USA*. 2006;103(8):2605-2610.

29. Riback JA, Bowman MA, Zmyslowski AM, et al. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science*. 2017;358:238-241.

30. Baldwin RL, Rose GD. Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol*. 2013;23:4-10.

31. Edwards SF, Anderson PW. Theory of spin glasses. *J Phys F: Metal Phys Ther*. 1975;5(5):965.

32. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA*. 1982;79(8): 2554-2558.

33. Anderson PW. Suggested model for prebiotic evolution: the use of chaos. *Proc Natl Acad Sci USA*. 1983;80(11):3386-3390.

34. Stein DL, Anderson PW. A model for the origin of biological catalysis. *Proc Natl Acad Sci USA*. 1984;81(6):1751-1753.

35. Onuchic J, LutheySchulten Z, Wolynes P. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem*. 1997;48: 545-600.

36. Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. *Annu Rev Biophys*. 2008;37:289-316.

37. Wales DJ. Energy landscapes: calculating pathways and rates. *Int Rev Phys Chem*. 2006;25:237-282.

38. Englander SW, Mayne L. The case for defined protein folding pathways. *Proc Natl Acad Sci USA*. 2017;114:8253-8258.

39. Baldwin RL. Clash between energy landscape theory and foldon-dependent protein folding. *Proc Natl Acad Sci USA*. 2017;114:8442-8443.

40. Zerrouki D, Baudry J, Pine D, Chaikin P, Bibette J. Chiral colloidal clusters. *Nature*. 2008;455:380-382.

41. Wang Y, Wang Y, Breed DR, et al. Colloids with valence and specific directional bonding. *Nature*. 2012;491:51-U61.

42. Shaw D, Maragakis P, Lindorff-Larsen K, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*. 2010;330: 341-346.

43. Ovchinnikov S, Park H, Vaghese N, et al. Protein structure determination using metagenime sequence data. *Science*. 2017;355:294-298.

44. Hoang T, Trovato A, Seno F, Banavar J, Maritan A. Geometry and symmetry presculpt the free-energy landscape of proteins. *Proc Natl Acad Sci USA*. 2004;101:7960-7964.

45. Craig A, Terentjev E. Folding of polymer chains with short-range binormal interactions. *J Phys A Math Gen*. 2006;39:4811-4828.

46. Coluzza I. A coarse-grained approach to protein design: learning from design to understand folding. *PLoS One*. 2011;6:1-8.

47. Bore SL, Milano G, Cascella M. Hybrid particle-field model for conformational dynamics of peptide chains. *J Chem Theory Comput*. 2018;14: 1120-1130.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.