



# A Coarse-Grained Protein Model in a Water-like Solvent

SUBJECT AREAS:

COMPUTATIONAL  
BIOPHYSICS

COMPUTATIONAL MODELS

BIOLOGICAL PHYSICS

BIOPHYSICAL CHEMISTRY

Sumit Sharma<sup>1\*</sup>, Sanat K. Kumar<sup>1</sup>, Sergey V. Buldyrev<sup>2</sup>, Pablo G. Debenedetti<sup>3</sup>, Peter J. Rossky<sup>4</sup>  
& H. Eugene Stanley<sup>5</sup>

<sup>1</sup>Department of Chemical Engineering, Columbia University, New York, NY 10027, <sup>2</sup>Department of Physics, Yeshiva University, New York, NY 10033, <sup>3</sup>Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, <sup>4</sup>Department of Chemistry and Biochemistry, University of Texas, Austin, TX 78712, <sup>5</sup>Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215.

Received  
8 April 2013

Accepted  
29 April 2013

Published  
14 May 2013

Correspondence and  
requests for materials  
should be addressed to  
S.K.K. (sk2794@  
columbia.edu)

\* Current address:  
2501 NW 229<sup>th</sup> Ave,  
Ronles Acres 3, Intel  
Corporation,  
Hillsboro, OR 97124

Simulations employing an explicit atom description of proteins in solvent can be computationally expensive. On the other hand, coarse-grained protein models in implicit solvent miss essential features of the hydrophobic effect, especially its temperature dependence, and have limited ability to capture the kinetics of protein folding. We propose a free space two-letter protein (“H-P”) model in a simple, but qualitatively accurate description for water, the Jagla model, which coarse-grains water into an isotropically interacting sphere. Using Monte Carlo simulations, we design protein-like sequences that can undergo a collapse, exposing the “Jagla-philic” monomers to the solvent, while maintaining a “hydrophobic” core. This protein-like model manifests heat and cold denaturation in a manner that is reminiscent of proteins. While this protein-like model lacks the details that would introduce secondary structure formation, we believe that these ideas represent a first step in developing a useful, but computationally expedient, means of modeling proteins.

The biological function of proteins is closely linked to their folded, three-dimensional structure<sup>1,2</sup>. Thus, the ability to predict the folded structure of a protein from knowledge of the sequence of amino acids opens many possibilities – such as predicting de-stabilizing mutations<sup>3</sup> and design of synthetic proteins<sup>4</sup>.

The thermodynamic stability of proteins is primarily attributed to the hydrophobic effect<sup>5,6</sup>. Mutations swapping one amino acid with another of a similar hydrophobic character have been shown to frequently result in a similar folded state<sup>7–9</sup>. Guided by these observations, models of proteins have been proposed in which amino acids are simply classified as hydrophobic, polar or neutral<sup>5,10</sup>. To further reduce complexity, the solvent is usually treated implicitly, including the hydrophobic effect as an effective attraction between hydrophobic amino acid residues. These models are computationally fast and have played an important role in developing an understanding of the thermodynamics of protein folding<sup>11</sup>, mis-folding<sup>12,13</sup>, aggregation<sup>14–16</sup> and adsorption<sup>17,18</sup>. However, because of the absence of explicit solvent, they can only use empirical parameterizations to capture the temperature and pressure-dependent changes in the hydrophobic effect, which is thought to be responsible for phenomena like cold and pressure denaturation<sup>19</sup>. Simulation studies also show that the solvent plays an important role in the kinetics of the hydrophobic collapse<sup>20</sup>. All-atom simulations (e.g. Ref. 21–25) of proteins in water can circumvent these issues, but at substantial computational cost. There is clearly a need for intermediate-level coarse-grained models that are computationally tractable, while explicitly accounting for water, and thus capturing the behavior of real proteins with fidelity.

Protein molecules are special heteropolymers<sup>26,27</sup>. Although the outer surface of a protein is patchy, with both hydrophobic and polar amino acids<sup>28</sup>, and there are stable charged groups in the protein’s interior<sup>29</sup>, the molecules fold so as to predominantly shield their hydrophobic amino acids while exposing the polar residues to the solvent<sup>30</sup>. This kind of spatial arrangement of amino acids would be expected to prevent protein aggregation, unlike the behavior of oil molecules in water<sup>31</sup>.

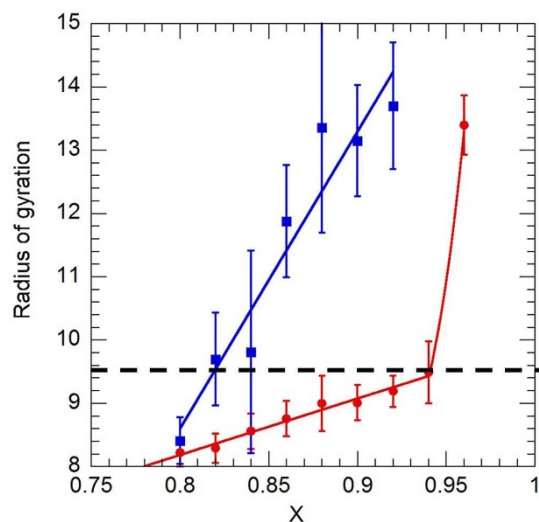
The folding of a protein molecule involves a global hydrophobic collapse of the entire protein chain and the formation of local secondary structures<sup>32–34</sup>. The kind of secondary structure that a protein segment forms is a function of the sequence of hydrophobic and polar amino acids in that segment<sup>35,36</sup>. The sequence-dependence of kinetics and thermodynamics of the hydrophobic collapse of a heteropolymer chain has not been systematically studied in experiments and computationally has been restricted to implicit solvent models<sup>37–39</sup> with a few exceptions<sup>40</sup>.



The goal of this work is to design a protein-like heteropolymer (composed of hydrophobic and polar-like groups) in an explicit water-like medium that has the potential to capture the essential features of hydrophobic-driven collapse, and the predominant exposure of polar groups to the solvent. Many lattice-based algorithms have been proposed for designing sequences that are capable of folding into a unique or a low degeneracy conformation in an implicit solvent<sup>41–45</sup>. These ideas are clearly not applicable in our case, and instead we use a combination of real space Monte Carlo (MC) and sequence space simulated annealing MC simulations (see Methods). This methodology, which is inspired by early ideas of Khokhlov and Khlatur<sup>45</sup>, allows us to design sequences that have many protein-like features, such as a collapsed state with a hydrophobic core and a polar exterior. It will be seen that this protein model exhibits cold and heat denaturation, as found in real proteins. Further, we show that the hydrophobic collapse is a function of the sequence of amino acids in the protein chain. As a caveat we emphasize that no attempt has been made to capture the well-known fact that proteins develop secondary structures, and we defer this important development to future work.

## Results

Fig. 1 shows how the root mean squared radius of gyration,  $\langle R_g^2 \rangle^{1/2}$  of the designed and the random 1000-mer heteropolymer, varies as a function of  $x$  ( $=V_{polar}/V_{att}$ ), determined by averaging the  $\langle R_g^2 \rangle^{1/2}$  from MC simulations of three different sequences for both the designed and the random heteropolymer. (For the definition of parameters of the model, see Methods section) We find that simulations of the random heteropolymer starting with a collapsed conformation equilibrates to an open conformation for  $x > 0.82$ . The designed heteropolymer, on the other hand, can attain a collapsed conformation for  $x \leq 0.94$ . Further, we find that the collapsed state of the designed heteropolymer always has a hydrophobic core and a polar exterior (Figs. 2(b) and 2(d)), while the random sequence does not show such segregation (Figs. 2(a) and 2(c)). To understand the



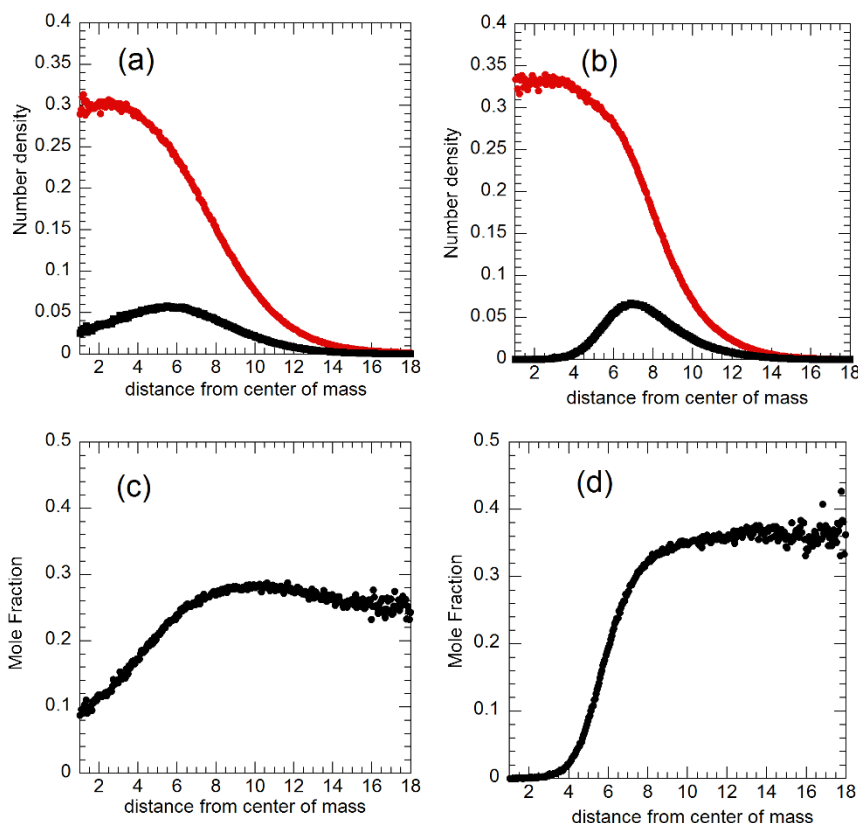
**Figure 1** | Root-mean-squared radius of gyration,  $\langle R_g^2 \rangle^{1/2}$  of designed (red) and random (blue) heteropolymers as a function of  $x$  at  $T = 0.7$  and  $\rho = 0.257$  determined using canonical ensemble MC simulations.  $x$  is the ratio of the potential well depth of the polar monomers with respect to the solvent monomers ( $V_{polar}/V_{att}$ ). Three different simulations with different starting conformations were averaged for the calculation of  $\langle R_g^2 \rangle^{1/2}$ . For both the designed and random heteropolymer, three different sequences were chosen for the simulations. The dotted line is at  $\langle R_g^2 \rangle^{1/2} = 9.53$ . The configurations for which the  $\langle R_g^2 \rangle^{1/2}$  lies below the dotted line are considered collapsed (see text). The error bars represent the standard deviation of the  $\langle R_g^2 \rangle^{1/2}$  from the three simulations. The lines are guide to the eye.

random sequence data we note that a chain comprised purely of hydrophobic monomers is collapsed under these state conditions<sup>46</sup>. Since the collapsed state of the designed sequence is stable for a broader range of  $x$  than the random sequence, we speculate that the tendency of the designed heteropolymer to collapse for  $x \leq 0.94$  probably reflects a micro-phase separation of the chain because of the partial segregation of the polar and hydrophobic monomers in the sequence. These interaction-dependent results are consistent with the expectation that the hydrophobic collapse of a protein is a function of its sequence.

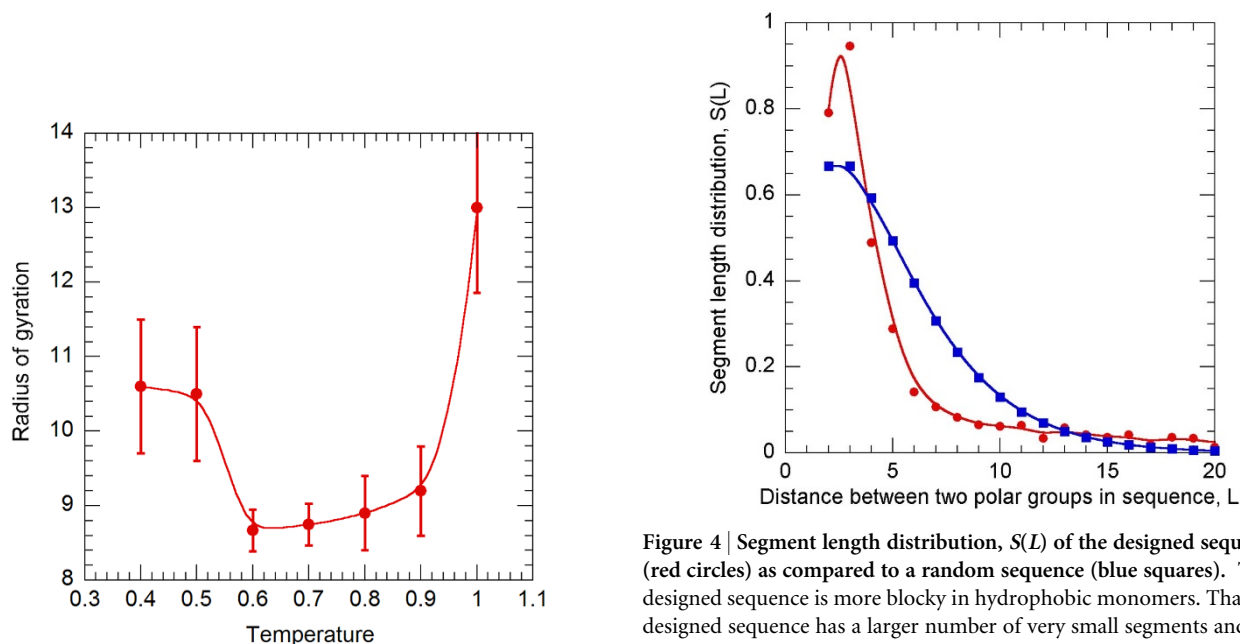
Fig. 2 shows the spatial distribution of the hard sphere (“hydrophobic”) and Jagla (“polar”) monomers of a random sequence and a designed heteropolymer, with  $x = 0.8$ , in their collapsed conformations after long canonical ensemble MC simulations. Let us first examine the random sequence (Figs. 2(a) and 2(c)). Polar monomers are preferred at the surface, but it is apparent that they are not excluded from the core. This supports the picture that the chain with a random sequence collapses non-specifically, as may be expected from a homopolymer chain in poor solvent conditions. This behavior contrasts strongly with that of the designed sequence, in which the polar monomers are completely excluded from the core (Figs. 2(b) and 2(d)). While both polar and hydrophobic monomers are present on the surface, there is a strong segregation of polar monomers to the surface of the heteropolymer. Further, this sequence can be collapsed from a random, open conformation, to a spatial distribution similar to that represented in Figs. 2b and 2d. This supports the assertion that the collapsed state of this sequence is the thermodynamic equilibrium state under these conditions. We note that this collapsed conformation is not unique, and, as noted, it does not acquire secondary structure, because our model only employs flexible bond angles, has no torsional potentials, and does not have any side groups which can direct this process. However, the ensemble of conformations in the collapsed state always has a hydrophobic core and a polar exterior and therefore has the basic structural feature of protein molecules in their native state.

The designed heteropolymer is seen to display both cold and heat denaturation (Fig. 3). For  $T \leq 0.5$  and  $T > 0.9$ , the designed sequence displays an open conformation. Upon heat denaturation, the  $\langle R_g^2 \rangle^{1/2}$  of the designed heteropolymer becomes greater than the  $\langle R_g^2 \rangle^{1/2}$  of a 1000-mer Gaussian chain ( $\sim 12.9$ ). Upon cold denaturation, the  $\langle R_g^2 \rangle^{1/2}$  of the heteropolymer increases but does not attain a random walk state. This behavior is analogous to the behavior of homopolymer<sup>47</sup>. It has been shown that at pressure  $P$  close to 0.0, the solubility of hard spheres in the Jagla solvent increases and the solvent-separated configuration of two hard spheres becomes the most stable configuration as  $T$  is decreased below 0.75<sup>46,48</sup>. It is also consistent with peaked distribution of the angle between the two subsequent bonds of a swollen homopolymer, with a maximum near 60 degrees<sup>47</sup>. The increase in  $\langle R_g^2 \rangle^{1/2}$  of the heteropolymer for  $T \leq 0.5$  is a consequence of the increase in solubility of hard spheres in Jagla solvent. We have defined the formation of the ground state of the designed heteropolymer simply by the value of  $\langle R_g^2 \rangle^{1/2}$  and not necessarily by the loss of secondary structure or native contacts, as is more typical for experiment and more detailed models. It is now well accepted that proteins can lose secondary structure, but still maintain a molten globule shape over a broader temperature range. Based on these facts, the stability range of the native state of a protein as defined by its size is expected to be broader than would be the case when using a metric that only incorporates secondary structure changes. This might explain the relatively large temperature range over which the collapsed state of the heteropolymer is stabilized.

We now explore the molecular origins of these results. We define a hydrophobic segment length as the number of hydrophobic monomers that separate two polar monomers along the chain backbone. The segment length distribution,  $S(L)$ , (Fig. 4) was calculated



**Figure 2** | Comparison of the spatial arrangement of monomers in the collapsed state of a designed sequence and of a random sequence at  $x = 0.8$ ,  $T = 0.7$  and  $\rho = 0.257$  obtained from a canonical ensemble MC simulation. Similar spatial profiles were obtained from simulations with different starting conformations and different designed and random sequences. The radial number density of the hydrophobic monomers (red) and the polar monomers (black) for (a) the random heteropolymer and (b) the designed heteropolymer as a function of the distance from their center of mass. The mole fraction of polar monomers in (c) the random heteropolymer and (d) the designed heteropolymer as a function of the distance from their center of mass.



**Figure 3** | Mean squared radius of gyration of the designed heteropolymer as a function of temperature for a number density  $\rho = 0.257$  and  $x = 0.86$ . The data is averaged over four simulations with different designed sequences and started from different random conformations. The designed heteropolymer displays cold and heat denaturation. The error bars represent the standard deviation of  $\langle R_g^2 \rangle^{1/2}$  from the four simulations. The line is guide to the eye.

**Figure 4** | Segment length distribution,  $S(L)$  of the designed sequence (red circles) as compared to a random sequence (blue squares). The designed sequence is more blocky in hydrophobic monomers. That is, the designed sequence has a larger number of very small segments and large segments (see text for the definition of a segment). The  $S(L)$  of random sequences of polar and hydrophobic monomers was determined theoretically by assuming a chain of infinite length. The theoretical distribution deviates only slightly from  $S(L)$  of random sequences computed by randomly generating  $10^4$  sequences of length 1000. The  $S(L)$  of the designed sequence was computed from configurations that were generated during the sequence-space MC simulation when the  $\theta = 0$  condition was reached. The lines are guides to the eye.



by multiplying the length of a segment,  $L$ , by its probability of occurring in a sequence. The  $S(L)$  of the random sequence of polar and hydrophobic monomers was determined by assuming that we have  $xN$  beads of P-H type and  $(1 - 2x)N$  beads of H type, which are mixed randomly in the sequence. Random sequences of length  $N$  can be generated by randomly selecting beads of type P-H with probability  $x/(1 - 2x)$  and beads of type H with probability  $(1 - 2x)/(1 - x)$ . The probability to find a sequence of  $(L - 1)$  H in a row (which corresponds to distance  $L$  between two neighboring P) will be then a pure geometric distribution  $P(L) = [(1 - 2x)/(1 - x)]^{L-2}x/(1 - x)$ .  $S(L)$  for random sequences is simply  $L P(L)$ . The actual distribution of  $S(L)$  for random sequences determined by generating  $10^4$  random sequences slightly deviates from the above theoretical distribution for  $S(L)$  because of the finite size of the sequence. The  $S(L)$  of the designed sequence was computed from configurations that were generated during the sequence-space MC simulation when the  $\theta = 0$  condition was reached. As compared to the  $S(L)$  of a random heteropolymer, the  $S(L)$  of the designed sequences has a larger peak at small  $L$  and consequently a broader tail. This implies that the designed sequence contains statistically both smaller and larger blocks of hydrophobic units than a random sequence. This result is similar to that found by Khokhlov and Khalatur<sup>45</sup>. We rationalize it as follows. If a polar monomer is at the surface of the collapsed protein, then the next polar monomer should be relatively close to it in the sequence so that it too can be placed at the surface without much entropic loss. On the other hand, if a hydrophobic sequence goes into the core of the collapsed chain, then it must be a long sequence since polar units are preferentially excluded from the core. The formation of a protein-like structure with a hydrophobic core and the preferential placement of polar monomers at the surface, therefore appears to be at the heart of the behavior of  $S(L)$ .

We also explored the possibility of observing pressure denaturation in our model. We know from previous work<sup>46,47</sup> that a hydrophobic homopolymer can show pressure induced swelling in a Jagla solvent, thus implying that pressure induced denaturation should also be observed for these designed heteropolymers. For the pressure range  $0.0 \leq P \leq 1.4$ , we find that a swollen chain undergoes a collapse implying that the collapsed state is the thermodynamically stable state at these conditions. However, in the simulations with  $P > 1.4$ , within the length of our simulations, a swollen chain does not undergo any collapse, and at the same time, a collapsed chain does not attain a swollen state. The hysteresis in the results for  $P > 1.4$  implies that as the density of the solvent increases, the activation barriers between the “folded” and “unfolded” states of the heteropolymer become large, thus preventing us from clearly identifying the pressure denaturation transition without a more sophisticated or intensive sampling. More work needs to be performed to explore this interesting finding and quantify the free energy barrier that apparently becomes large at high pressure.

## Discussion

We have developed a simple model of a protein-like heteropolymer in an explicit solvent medium having water-like behavior. In contrast to “real” water, which has three atoms and explicit hydrogen bonding, our model represents “water” as a single sphere that interacts with all other atoms through a spatially isotropic potential called the Jagla potential. The Jagla potential has been shown previously to be quite successful in displaying water-like thermodynamic, dynamic and structural anomalies<sup>49,50</sup> as well as water-like solvation thermodynamics<sup>46,48</sup>. The Jagla potential has effectively two length scales – a repulsive ramp and a hard-sphere core. At high densities and temperatures, the hard-sphere core defines the effective radius of the Jagla particles. Upon reducing pressure or decreasing temperature, the interparticle separation of Jagla pairs changes from close to the hard-core diameter to a distance more comparable to the repulsive ramp diameter, thus allowing the system to achieve a more open

structure. This behavior is reminiscent of water at low temperature and density, wherein an open-structured liquid is formed because of the formation of H-bonds. These simplifications yield significant computational advantages, but without the loss of qualitative hydrophobic behavior of solutes. Because of the presence of explicit solvent, the model is able to capture *a priori* the phenomena of cold and heat denaturation reminiscent of real proteins. The model results show that a sequence that is able to attain a spatial arrangement in which, predominantly, hydrophobic monomers are in the interior and polar monomers are on the surface of the collapsed state has a much stronger tendency to undergo a hydrophobic collapse than does a random heteropolymer. This is consistent with the expectation that hydrophobic collapse in natural proteins is aided by their sequence. The results also show that the evolved sequences producing collapse have longer than statistical runs of hydrophobic groups, consistent with the needed topology of a quasi-spherical collapsed state with a predominantly solvophilic surface. In this study, we did not attempt to include features in the model that would lead to secondary structure, and we have only touched on the influence of high pressure. Both of these aspects will be topics of future studies.

## Methods

**Model.** We model water molecules through a spherically symmetric Jagla potential<sup>46,51</sup>. The Jagla potential has a hard sphere core of radius,  $\sigma/2$  (defining the characteristic length scale for the model), a repulsive, linear ramp from  $r = \sigma/2$  to  $r = a/2$  and an attractive tail from  $r = a/2$  to  $r = b/2$ . At  $r = \sigma/2$ , the potential of the repulsive ramp is a positive value,  $V_{rep}$  and at  $r = a/2$ , the potential is chosen to be  $V_{att}$ , defining the characteristic energy scale for the model. In this work, we measure length in units of  $\sigma$ , energy in units of  $|V_{att}|$ , pressure in units of  $|V_{att}|\sigma^{-3}$ , temperature in units of  $|V_{att}|k_B^{-1}$  and number density,  $\rho$  in units of  $\sigma^{-3}$ . From  $r = a/2$  to  $r = b/2$  the attractive tail increases linearly to zero. The potential is zero for  $r > b/2$ . The parameter values invoked here are those used by Buldyrev et al.<sup>46</sup> ( $a = 1.72$ ,  $b = 3$  and  $V_{rep} = 3.5$ ). The Jagla model has been shown to display water-like thermodynamic, dynamic and structural anomalies<sup>49</sup> and water-like solvation thermodynamics<sup>46,52</sup>. The water-like anomalies of the Jagla potential are due to its two length scales – the repulsive ramp and the hard-sphere core. At high densities and temperatures, the hard-sphere core defines the effective radius of the Jagla particles. Upon reducing pressure or decreasing temperature, the inter-particle separation of Jagla nearest-neighbor pairs changes from close to the hard-core diameter ( $r = 1$ ) to a distance more comparable to the repulsive ramp diameter ( $r = 1.72$ ). This allows the system to achieve a more open structure, a behavior that is reminiscent of water’s at low temperature and density, wherein an open-structured liquid is formed because of the formation of H-bonds. This provides one interpretation for the success of the Jagla model in describing water-like anomalies<sup>50,53</sup>.

We model a protein molecule as a linear flexible chain of polar and hydrophobic monomers. The polar monomers interact with each other through the Jagla potential. The parameters of the Jagla potential between a pair of polar monomers is taken to be the same as that of the solvent particles except for the potential well depth, which can vary with a scaling parameter  $x$ ;  $V_{polar} = x V_{att}$ . The interaction potential between a solvent molecule and a polar monomer is taken to be Jagla potential, with the well depth of the Jagla potential =  $x^{0.5} V_{att}$ , as determined using Lorentz-Berthelot mixing rule. The hydrophobic monomers interact with all the particles, including each other, through a hard sphere potential defined by a hydrophobic particle hard core diameter equal to 1.0. The bond length,  $\ell$  is fixed to 1.0. No torsional or bond-angle potentials are included in the present model. The polymer molecule is taken to be of length  $N = 1000$  monomers.

We pick this relatively long chain so that a clear distinction can be made between the surface and the core of the chain in the collapsed state. Chains of length 176, for example, are too short to make this distinction<sup>54</sup>. We choose to limit this study of sequence to a fixed fraction of polar monomers in the “protein” chain of 0.25.

Finally, in the present study, a design constraint was imposed in which polar monomers were not allowed to be adjacent to each other in the sequence. This constraint was imposed to ensure that a complete segregation of polar and hydrophobic monomers does not occur. This is a strong constraint as polar monomers do occur sequentially adjacent to each other in proteins. A less stringent constraint could be to have a finite but decreasing probability for having  $N_p$  polar monomers adjacent to each other in sequence. Such alternatives are appropriate for further investigations.

**Designing a protein-like sequence.** Simulations were performed in a box size of  $36 \times 36 \times 36$  (in reduced units; see *Model* section). Periodic boundary conditions were used in the  $x$ ,  $y$  and  $z$  directions. The temperature was taken to be  $T = 0.7$  and the total number density,  $\rho$  of the system was 0.257. The pressure  $P$  for this system was found to be  $0.01 \pm 0.01$ . Insensitivity to finite size effects was demonstrated by doing simulations in a box size of  $50 \times 50 \times 50$  and the same  $\rho$  of 0.257 (number of solvent particles = 31,125). The results from the larger box size were found to match the ones from the smaller box size. A freely jointed chain of  $N = 1000$  monomers has a root mean squared radius of gyration (RMSRG),  $\langle R_g^2 \rangle^{1/2} = (N/6)^{1/2} \ell = 12.9$ . At  $T = 0.7$ ,





the RMSRG of a 1000-mer hard-sphere homopolymer collapsed with an internal (or “core”) density  $\rho = 0.257$  and an  $\langle R_g^2 \rangle^{1/2}$  of  $7.4 \pm 0.1$ , which is consistent with a simple spherical geometric estimate of  $\langle R_g^2 \rangle^{1/2} = (3/5)^{1/2}(3N/4\pi\rho)^{1/3}\sigma = 7.55$ . We can make the result more relevant to our context by realizing that the effective size of a Jagla monomer ( $\approx a/2 = 0.86$ ) is larger than that of the hydrophobic monomers (0.5). After accounting for these size differences, and for the probability of the occurrence of a Jagla monomer in a peptide chain,  $p = 0.25$  (see *Model* section), we can then write a better estimate for the size of the “protein” chain if fully collapsed:

$$\langle R_g^2 \rangle^{1/2} = \left[ \frac{pr_j^3 + (1-p)r_j^3}{r_j^3} \right]^{1/3} \times 7.55 \approx 9.53 \quad (1)$$

We thus expect that the  $\langle R_g^2 \rangle^{1/2}$  of the heteropolymer will fall somewhere between 7.55 and 9.53. Thus, the chain size corresponding to the collapsed state is small enough, relative to the dimensions of the simulation cell, to avoid a collapsed chain interacting with its own image. The chain dimensions obtained in the collapsed state are thus expected to be reliable, as is the predicted unfolding temperature, while the dimensions of less ordered states may be quantitatively less reliable.

A protein-like sequence was designed using a two-step procedure. In the first step, random heteropolymers containing 25% of polar monomers were generated by randomly selecting monomers from the sequence and labeling them as polar. The remaining monomers were labeled as hydrophobic (or hard core monomers). The random heteropolymers with  $x$  varying from 0 to 1 (recall that  $x$  is defined as  $V_{\text{polar}}/V_{\text{att}}$ ), were investigated using canonical ensemble Monte Carlo (MC) simulation at  $T = 0.7$  and  $\rho = 0.257$ . It was found that for  $x < 0.82$ , random heteropolymers adopt collapsed equilibrium configurations, i.e., characterized by  $\langle R_g^2 \rangle^{1/2} \leq 9.53$ . In the second step, we chose  $x = 0.8$  and, starting from a random sequence and a random conformation, performed a canonical ensemble MC simulation ( $T = 0.7$  and  $\rho = 0.257$ ) combined with a simulated annealing MC simulation in sequence space to generate the optimal sequences of a collapsed “protein”-like sequence. In the sequence space MC, a polar and a hydrophobic monomer were chosen at random and an attempt was made to swap their positions in the sequence, consistent with the design constraint that polar monomers cannot be adjacent to each other in the sequence. The attempt to swap was accepted with the probability  $= \exp\left(-\frac{\Delta E}{\theta}\right)$ , where  $\theta$  is the sequence-space “temperature”, and  $\Delta E$  is the change in energy accompanying the swap. The sequence space MC was started with a large value of  $\theta (= 0.8)$ . After 4000 attempted swap moves,  $\theta$  was decreased by 0.1. The above steps were continued until  $\theta = 0$ . The sequence space MC moves were attempted with a probability of 0.02 after every real space canonical ensemble MC cycle. Overall,  $6 \times 10^6$  real space MC cycles were performed. Once this simulation was complete, we found that the heteropolymer always attained a collapsed state with hydrophobic groups predominantly in the core and the polar groups close to the surface. The surface of the heteropolymer in its collapsed state was identified by determining the Gibbs surface as follows: The spherically averaged spatial density profile of monomers around the center of mass of the heteropolymer in its collapsed state was determined. We assumed a “vapor” density of zero and a “liquid” density equal to the monomer density in the inner core of the protein. The “mid” plane where the net surface depletion of monomers on the liquid side exactly compensates the surface excess on the vapor side is the location of the Gibbs plane,  $r_{\text{Gibbs}}$ .  $r_{\text{Gibbs}}$  was found by solving the below equation,

$$\int_0^{r_{\text{Gibbs}}} 4\pi r^2 [\rho(0) - \rho(r)] dr = \int_{r_{\text{Gibbs}}}^{\infty} 4\pi r^2 \rho(r) dr \quad (2)$$

In the above equation, the center of mass of the protein is at  $r = 0$  and  $\rho(r)$  is the number density of the monomers as a function of  $r$ . The sequence of the heteropolymer achieved after this simulation was called the “designed heteropolymer”.

Several realizations of the above simulation were performed. The final sequence of the designed heteropolymer was not unique, but the ensemble of sequences thus obtained showed similar properties with respect to the distribution of monomers in the sequence and the collapse tendency. In the above sequence design procedure,  $x$  was chosen to be 0.8, as this ensures that the heteropolymer attains a collapsed state during the design procedure (Fig. 1). If too large a value of  $x$  is used during the sequence design procedure, the heteropolymer attains an open conformation and hence good sequence design is not achieved. However, as described below, this design procedure is capable of generating sequences that remain stable in the collapsed state even for the range of  $x$  where a random heteropolymer is not ( $x \geq 0.82$ ). The behavior of the designed heteropolymer was studied as a function of  $x$ ,  $T$  and  $P$  using real space canonical ensemble MC simulations.

- Anfinsen, C. B. Principles That Govern Folding of Protein Chains. *Science* **181**, 223–230 (1973).
- Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
- Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).
- Mutter, M. & Vuilleumier, S. A Chemical Approach to Protein Design - Template-Assembled Synthetic Proteins (Tasp). *Angew. Chem.* **28**, 535–554 (1989).
- Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **29**, 7133–7155 (1990).
- Honig, B. & Yang, A. S. Free-Energy Balance in Protein-Folding. *Adv. Protein Chem.* **46**, 27–58 (1995).
- Kamtekar, S., Schiffer, J. M., Xiong, H. Y., Babik, J. M. & Hecht, M. H. Protein Design by Binary Patterning of Polar and Nonpolar Amino-Acids. *Science* **262**, 1680–1685 (1993).
- West, M. W. & Hecht, M. H. Binary Patterning of Polar and Nonpolar Amino-Acids in the Sequences and Structures of Native Proteins. *Protein Sci.* **4**, 2032–2039 (1995).
- Hecht, M. H. *et al.* Design or libraries of de novo proteins by binary patterning of polar and nonpolar amino acids. *Protein Eng.* **10**, 20–20 (1997).
- Guo, Z. Y., Thirumalai, D. & Honeycutt, J. D. Folding Kinetics of Proteins - a Model Study. *J. Chem. Phys.* **97**, 525–535 (1992).
- Yue, K. & Dill, K. A. Forces of tertiary structural organization in globular-proteins. *Proc. Natl. Acad. Sci. USA* **92**, 146–150 (1995).
- Gupta, P., Hall, C. K. & Voegler, A. Computer simulation of competition between protein folding and aggregation. *Fluid Phase Equilib.* **158**, 87–93 (1999).
- Lattman, E. E., Fiebig, K. M. & Dill, K. A. Modeling Compact Denatured States of Proteins. *Biochemistry* **33**, 6158–6166 (1994).
- Cellmer, T., Bratko, D., Prausnitz, J. M. & Blanch, H. Protein-folding landscapes in multichain systems. *PNAS* **102**, 11692–11697 (2005).
- Anand, G., Sharma, S., Kumar, S. K. & Belfort, G. Stability of Tethered Proteins. *Langmuir* **25**, 4998–5005 (2009).
- Dima, R. I. & Thirumalai, D. Exploring protein aggregation and self-propagation using lattice models: Phase diagram and kinetics. *Protein Sci* **11**, 1036–1049 (2002).
- Anand, G., Sharma, S., Dutta, A. K., Kumar, S. K. & Belfort, G. Conformational Transitions of Adsorbed Proteins on Surfaces of Varying Polarity. *Langmuir* **26**, 10803–10811 (2010).
- Sharma, S., Berne, B. J. & Kumar, S. K. Thermal and Structural Stability of Adsorbed Proteins. *Biophys. J.* **99**, 1157–1165 (2010).
- Patel, B. A., Debenedetti, P. G., Stillinger, F. H. & Rossky, P. J. A water-explicit lattice model of heat-, cold-, and pressure-induced protein unfolding. *Biophys. J.* **93**, 4116–4127 (2007).
- Ferguson, A. L., Panagiotopoulos, A. Z., Kevrekidis, I. G. & Debenedetti, P. G. Systematic determination of order parameters for chain dynamics using diffusion maps. *Proc. Natl. Acad. Sci. U.S.A* **107**, 13597–13602 (2010).
- Nivon, L. G. & Shakhnovich, E. I. Thermodynamics and kinetics of the hairpin ribozyme from atomistic folding/unfolding simulations. *J. Mol. Biol.* **411**, 1128–1144 (2011).
- Duan, Y. & Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**, 740–744 (1998).
- Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
- Simmerling, C., Strockbine, B. & Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **124**, 11258–11259 (2002).
- Snow, C. D., Nguyen, N., Pande, V. S. & Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**, 102–106 (2002).
- Pande, V. S., Grosberg, A. Y. & Tanaka, T. Nonrandomness in Protein Sequences - Evidence for a Physically Driven Stage of Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 12972–12975 (1994).
- Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
- Haynes, C. A. & Norde, W. Structures and stabilities of adsorbed proteins. *J. Colloid Interface Sci.* **169**, 313–328 (1995).
- Kajander, T. *et al.* Buried charged surface in proteins. *Structure* **8**, 1203–1214 (2000).
- Meirovitch, H. & Scheraga, H. A. Empirical-Studies of Hydrophobicity. 2. Distribution of the Hydrophobic, Hydrophilic, Neutral, and Ambivalent Amino-Acids in the Interior and Exterior Layers of Native Proteins. *Macromolecules* **13**, 1406–1414 (1980).
- Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.* **20**, 1325–1336 (2003).
- Daggett, V. & Fersht, A. R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18–25 (2003).
- Tanford, C. Contribution of Hydrophobic Interactions to Stability of Globular Conformation of Proteins. *J. Am. Chem. Soc.* **84**, 4240 (1962).
- Ptitsyn, O. B. Protein Folding - Hypotheses and Experiments. *J. Protein Chem.* **6**, 273–293 (1987).
- Bellesia, G., Jewett, A. I. & Shea, J. E. Sequence periodicity and secondary structure propensity in model proteins. *Protein Sci.* **19**, 141–154 (2010).
- Degrado, W. F., Wasserman, Z. R. & Lear, J. D. Protein Design, a Minimalist Approach. *Science* **243**, 622–628 (1989).
- Ashbaugh, H. S. Tuning the Global Assembly of Hydrophobic/Hydrophilic Heteropolymer Sequences. *J. Phys. Chem. B* **113**, 14043–14046 (2009).
- Pham, T. T., Dunweg, B. & Prakash, J. R. Collapse Dynamics of Copolymers in a Poor Solvent: Influence of Hydrodynamic Interactions and Chain Sequence. *Macromolecules* **43**, 10084–10095 (2010).



39. van den Oever, J. M. P. *et al.* Coil-globule transition for regular, random, and specially designed copolymers: Monte Carlo simulation and self-consistent field theory. *Phys. Rev. E* **65** (2002).
40. Patel, B. A., Debenedetti, P. G., Stillinger, F. H. & Rosicky, P. J. The effect of sequence on the conformational stability of a model heteropolymer in explicit water. *J. Chem. Phys.* **128** (2008).
41. Yue, K. & Dill, K. A. Inverse Protein Folding Problem - Designing Polymer Sequences. *Proc. Nat. Acad. Sci. USA* **89**, 4163–4167 (1992).
42. Shakhnovich, E. I. & Gutin, A. M. Engineering of Stable and Fast-Folding Sequences of Model Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7195–7199 (1993).
43. Seno, F., Vendruscolo, M., Maritan, A. & Banavar, J. R. Optimal protein design procedure. *Phys. Rev. Lett.* **77**, 1901–1904 (1996).
44. Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. Improved design of stable and fast-folding model proteins. *Folding & Design* **1**, 221–230 (1996).
45. Khokhlov, A. R. & Khalatur, P. G. Conformation-dependent sequence design (engineering) of AB copolymers. *Phys. Rev. Lett.* **82**, 3456–3459 (1999).
46. Buldyrev, S. V., Kumar, P., Debenedetti, P. G., Rosicky, P. J. & Stanley, H. E. Water-like solvation thermodynamics in a spherically symmetric solvent model with two characteristic lengths. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20177–20182 (2007).
47. Buldyrev, S. V., Kumar, P., Sastry, S., Stanley, H. E. & Weiner, S. Hydrophobic collapse and cold denaturation in the Jagla model of water. *J. Phys.: Condens. Matter* **22**, 284109 (2010).
48. Maiti, M., Weiner, S., Buldyrev, S. V., Stanley, H. E. & Sastry, S. Potential of mean force between hydrophobic solutes in the Jagla model of water and implications for cold denaturation of proteins. *J. Chem. Phys.* **136**, 044512 (2012).
49. Yan, Z. Y., Buldyrev, S. V., Giovambattista, N., Debenedetti, P. G. & Stanley, H. E. Family of tunable spherically symmetric potentials that span the range from hard spheres to waterlike behavior. *Phys. Rev. E* **73**, 051204 (2006).
50. Yan, Z. Y., Buldyrev, S. V., Kumar, P., Giovambattista, N. & Stanley, H. E. Correspondence between phase diagrams of the TIP5P water model and a spherically symmetric repulsive ramp potential with two characteristic length scales. *Phys. Rev. E* **77**, 042201 (2008).
51. Jagla, E. A. Core-softened potentials and the anomalous properties of water. *J. Chem. Phys.* **111**, 8980–8986 (1999).
52. Su, Z. Q., Buldyrev, S. V., Debenedetti, P. G., Rosicky, P. J. & Stanley, H. E. Modeling simple amphiphilic solutes in a Jagla solvent. *J. Chem. Phys.* **136** (2012).
53. Yan, Z. Y. *et al.* Structure of the first- and second-neighbor shells of simulated water: Quantitative relation to translational and orientational order. *Phys. Rev. E* **76**, 051201 (2007).
54. Sharma, S. Conformational changes in proteins upon adsorption to surfaces. *Columbia University, Ph.D. thesis* (2010).

## Acknowledgments

We thank C.A. Angell, M. Marqués, S. Sastry, and Z. Yan for helpful discussions. S.S. and S.K.K. acknowledge the DOE – Basic Engineering Sciences for funding this research. P.G.D. gratefully acknowledges the support of the National Science Foundation (Grant CHE-1213343). P.J.R. gratefully acknowledges the support of the National Science Foundation (Collaborative Research Grants CHE-0908265 and CHE-0910615). Additional support from the R.A. Welch Foundation (F-0019) to P.J.R. is also gratefully acknowledged. HES thanks the NSF Chemistry Division for support through grants CHE 0911389, CHE 0908218 and CHE-1213217. S.V.B. acknowledges the partial support of this research through the Dr Bernard W. Gamson Computational Science Center at Yeshiva College.

## Author contributions

All authors designed research. S.S. performed the simulations. S.K.K. supervised the project. All authors wrote the paper.

## Additional information

**Competing financial interests:** The authors declare no competing financial interests.

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

**How to cite this article:** Sharma, S. *et al.* A Coarse-Grained Protein Model in a Water-like Solvent. *Sci. Rep.* **3**, 1841; DOI:10.1038/srep01841 (2013).