

## Role of Water in the Selection of Stable Proteins at Ambient and Extreme Thermodynamic Conditions

Valentino Bianco,<sup>1</sup> Giancarlo Franzese,<sup>2</sup> Christoph Dellago,<sup>1</sup> and Ivan Coluzza<sup>1</sup>

<sup>1</sup>*Faculty of Physics, University of Vienna, Vienna 1090, Austria*

<sup>2</sup>*Secció de Física Estadística i Interdisciplinària—Departament de Física de la Matèria Condensada, Facultat de Física & Institute of Nanoscience and Nanotechnology (IN2UB), Universitat de Barcelona, Barcelona 08028, Spain*

(Received 21 October 2016; revised manuscript received 1 February 2017; published 26 June 2017)

Proteins that are functional at ambient conditions do not necessarily work at extreme conditions of temperature  $T$  and pressure  $P$ . Furthermore, there are limits of  $T$  and  $P$  above which no protein has a stable functional state. Here, we show that these limits and the selection mechanisms for working proteins depend on how the properties of the surrounding water change with  $T$  and  $P$ . We find that proteins selected at high  $T$  are superstable and are characterized by a nonextreme segregation of a hydrophilic surface and a hydrophobic core. Surprisingly, a larger segregation reduces the stability range in  $T$  and  $P$ . Our computer simulations, based on a new protein design protocol, explain the hydropathy profile of proteins as a consequence of a selection process influenced by water. Our results, potentially useful for engineering proteins and drugs working far from ambient conditions, offer an alternative rationale to the evolutionary action exerted by the environment in extreme conditions.

DOI: [10.1103/PhysRevX.7.021047](https://doi.org/10.1103/PhysRevX.7.021047)

Subject Areas: Biological Physics, Soft Matter, Statistical Physics

### I. INTRODUCTION

Proteins are molecules made of a sequence of amino acid residues that fold to target structures (native states) for a range of temperatures  $T$  and pressures  $P$ . Different proteins, with different sequences, have different  $T$ – $P$  ranges of stability for their native state. As a matter of fact, many life forms survive under extreme  $T$ – $P$  conditions [1,2], implying that their proteins fit that environment. Recent works [3–10] have evidenced that changes in  $T$  and  $P$  significantly alter the water-mediated hydrophilic and hydrophobic interactions of the residues along the chain, with an effect that depends on the sequence. However, a direct observation of how the protein selection responds to extreme changes in the aqueous conditions is still lacking. One of the reasons for this is that, for the few notable studies accounting for explicit water [11–16], an exhaustive sampling of the  $T$ – $P$  stability region of the selected proteins is beyond the reach of current computers. Here, we fill this gap. We present a computational study based on a novel strategy that mimics the adaptation process of solvated proteins in a wide range of thermodynamic conditions. Our results reveal the border beyond which artificial protein design and natural evolution fail. Moreover, we show that

high- $T$  adaptation alone selects for superstable sequences that are highly resistant to both cold and pressure denaturation. Here, we define the proteins as superstable if their average stability region encompasses the average stability region of proteins designed at ambient conditions. Our results confirm, for the first time, the hypothesis that a strong stability of the folded state at high  $T$  corresponds to strong stability also at low  $T$  and high  $P$  [17–20].

We develop a protein design strategy that focuses on the relation between sequence and folded structure, allowing us to calculate the stability of many proteins in a wide range of  $T$  and  $P$  in explicit water, a formidable task that is currently infeasible for atomistic models. Following the standard approach introduced by Shakhnovich and Gutin [21,22], we adopt a coarse-grained lattice representation of proteins that is computationally effective and yields results that can also be extended to off-lattice proteins, as recently demonstrated for the implicit solvent case [23–25]. The main difference between our strategy and the standard approach is that here, instead of considering the phenomenology of water implicitly, we explicitly include water in our coarse-graining. As an explicit solvent, we adopt a many-body water model that has recently been shown to provide relevant information on the role of water in the cold- and pressure-denaturation mechanisms of proteins [9].

### II. DESIGN APPROACH

Protein and solvent interactions are described by the Hamiltonian (see Appendix A for details)  $\mathcal{H} \equiv \mathcal{H}_{R,R} + \mathcal{H}_{R,w} + \mathcal{H}_{w,w}^{(b)} + \mathcal{H}_{w,w}^{(h)}$ , where the first term accounts for the

---

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

residue-residue interactions and the second term for the residue-hydrophobic or hydrophilic interaction with the solvent, respectively. The third term,  $\mathcal{H}_{w,w}^{(b)}$ , accounts for the water-water interaction in the bulk, including the isotropic van der Waals interaction, as well as the directional and cooperative components of the hydrogen bonds (HBs) [8,9,32–36]. The last term,  $\mathcal{H}_{w,w}^{(h)}$ , describes the water-water interaction in the hydration layer—the first layer of water molecules in contact with the protein—and accounts for the effect of the protein’s hydrophobic and hydrophilic residues on the interaction and structure of the hydration water [9].

By assuming that the protein interface affects the water properties in the hydration shell, we use a design strategy based on the enthalpy associated with the hydrated protein, regardless of the bulk contribution,  $H_R^{(h)} \equiv \mathcal{H}_{R,R} + \mathcal{H}_{R,w} + \mathcal{H}_{w,w}^{(h)} + P(V - V^{(b)})$ , with a bias that enhances the sequence heterogeneity to avoid homopolymer design solutions [23,24,29–31]. In the above expression for  $H_R^{(h)}$ ,  $V$  is the total volume of the system and  $V^{(b)}$  is the volume associate with the bulk water, i.e., the water not in direct contact with the protein. Since our optimization accounts for changes in the aqueous conditions, the enthalpies of each tested sequence must be averaged over the water configurations. We then envisage two possible design protocols based on measuring the average sequence enthalpies  $\langle H_R^{(h)} \rangle$ , where the angular brackets  $\langle \dots \rangle$  refer to the thermodynamic average calculated over equilibrium water configurations, including the bulk. The first strategy consists in minimizing the average enthalpy  $\langle H_{R,f}^{(h)} \rangle$  of the hydrated protein in its folded ( $f$ ) conformation. The second consists in maximizing the enthalpic gap  $\Delta H_R^{(h)}$  between the  $f$  state and the unfolded ( $u$ ) protein conformation, both hydrated. As a  $u$  state, we use a completely stretched structure that is a representative of all equivalent unbounded configurations. For both protocols, the enthalpy of the hydration water is averaged over equilibrium configurations, accounting for the energy, density, and entropy of the solvent at the given thermodynamic conditions and protein interface. We refer to the two protocols as MIN ENTHALPY and MAX GAP, respectively. In order to quantify the solvent contribution, we compare the results obtained with these two protocols with those of a simpler method (IMPLICIT SOLVENT) based on the minimization of the residue interactions without any water contribution.

### III. RESULTS AND DISCUSSION

#### A. Thermal and pressure stability of proteins

We focus our attention on ten different target proteins with lengths between 30 and 90 amino acids that are compact enough that the exposed surface to the solvent is the smallest possible. For sake of clarity, here we show only

the calculations for the 30 amino acid proteins. The same results hold for longer proteins. In each protocol, we design the protein at a given temperature  $T_d$  and pressure  $P_d$  and test its stability in a wide range of  $T$  and  $P$  [37]. We choose  $T_d$  and  $P_d$  uniformly over values ranging from the liquid-gas transition to the glass-transition temperature, and from negative pressure to  $\approx 1$  GPa. For each  $(T_d, P_d)$  and each protocol, we identify 5–15 optimized sequences, for a total of more than  $1.5 \times 10^3$  optimizations, a number far beyond the capability of any fully atomistic protein model. For each designed sequence, we test the stability in  $T$ – $P$  by checking, with Monte Carlo simulations, if the average equilibrium conformation coincides with the protein native-target structure. Then, we average the stability regions of all sequences optimized for a given target-native state. For all designed proteins, we find stability regions that resemble those predicted by theory [38], previous numerical works [5,7,9,39], and those observed in experiments [18,40–52] (Fig. 1). Throughout the paper, we use IU for  $T$  and  $P$  defined in Methods. We find that the boundaries of the stability regions extend from low  $T$ , where proteins undergo cold denaturation, to high  $T$ , where thermal denaturation occurs. The  $T$  range of stability depends on the protein sequence and size. Small proteins undergo cold denaturation for  $T < 0.1$  IU [53], consistent with previous results [9]. Longer proteins, with higher content of hydrophobic residues, cold-denature at higher  $T$ . All the sequences we consider are unstable above the threshold pressure  $P \approx 0.7$  IU (Figs. 5 and 6 in Appendix D), qualitatively consistent with what is observed in Refs. [40,54,55]. We find that denaturation occurs also at low or negative  $P$ , consistent with recent findings [9,39,56].

Experimental data for different proteins [Fig. 1(b)] show that the higher the thermal stability of the proteins, the higher the pressure-stability region [17,18]. Our simulations have the same trend. In particular, we find that the stability region is more extended for sequences designed at higher  $T_d$  and intermediate  $P_d$ , corroborating the hypothesis that stability ranges in  $T$  and  $P$  are positively correlated: Proteins with a pronounced thermal stability are also more stable with respect to pressure. Furthermore, we observe that sequences designed at high  $T_d$  are stable both at high and low  $T$ , while those designed at lower  $T_d$  can fold only at low  $T$ . For example, sequences designed at  $P = 0$  and  $T_d = 0.1$  IU are stable up to  $T_1 \approx 0.3$  IU, while to get a sequence that also folds at higher  $T_2 = 0.4$  IU, we need to design it at  $T_d = T_2$  [Fig. 1(a)]. These results are consistent with experiments comparing the stability range of mesophilic and thermophilic proteins [57] and with data revealing the higher resistance of thermophilic proteins to cold denaturation [19,20]. More generally, we find that proteins that are designed at high  $T_d$  are super-stable; i.e., they also show a remarkable stability at low  $T$  and at high  $P$ .

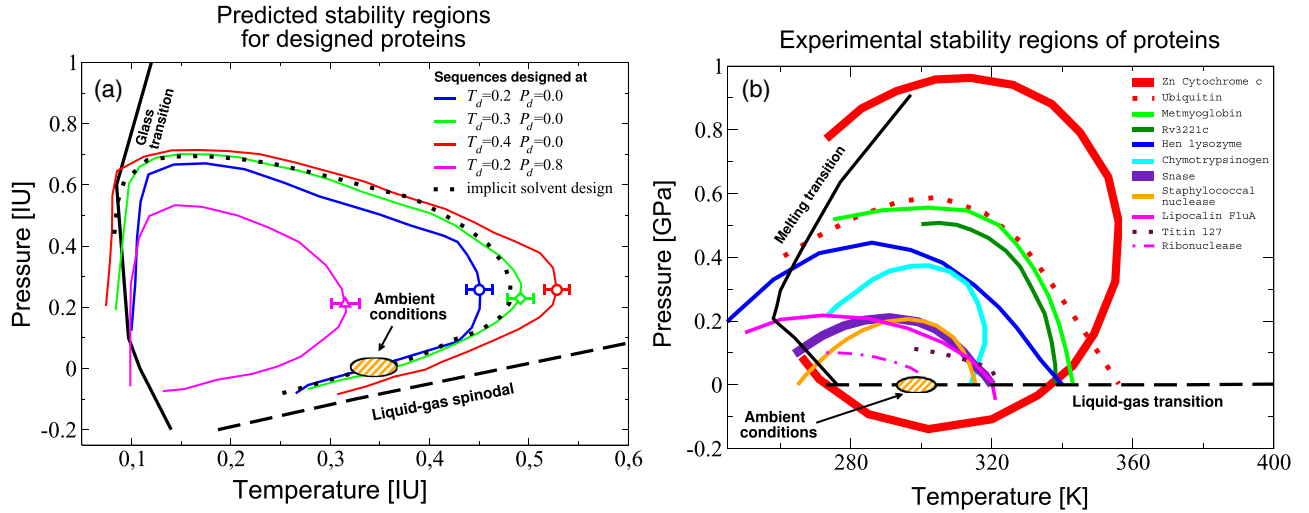


FIG. 1.  $T$ - $P$  stability regions of designed and experimental proteins. (a) Average stability regions (closed curves) within which the protein sequences designed with the MIN ENTHALPY protocol at different  $T_d$  and  $P_d$  fold into the native state. The regions for proteins designed at  $P_d = 0$  (red, green, and blue curves), close to ambient pressure, are enclosed one into another with an extension that is proportional to  $T_d$ . All of these regions enclose the stability region of proteins designed at higher  $P_d$  (pink curve). The dotted line shows the stability region for a sequence calculated with the IMPLICIT SOLVENT protocol. The “glass-transition” (solid black) line defines the temperatures below which the system does not equilibrate. The long-dashed line represents the limit of stability (spinodal) of the liquid with respect to the gas phase. Pressure and temperature are expressed in internal units (IU). (b) Experimental stability region for different proteins indicated in the legend (adapted from Refs. [40–49]). The long-dashed line is the liquid-gas transition line. The data show a clear positive correlation between the  $T$  range and  $P$  range of stability. Pressure denaturation is observed in a range of  $-0.1 \lesssim P/\text{GPa} \lesssim 0.6$ , while stability at higher  $P$  is reached by introducing artificial covalent bridges between the amino acids, as in the case of Zn cytochrome  $c$  [43]. In both panels, the shaded region corresponds to ambient conditions.

## B. Water effect on the protein surface and core hydrophathy profiles

We rationalize this fundamental property of protein evolution in terms of the relative fraction of hydrophobic and hydrophilic residues in the sequence and how it changes depending on the thermodynamic state of the solvent. We calculate the fraction  $\mathcal{P}_{\text{surf}}^{\text{PHI}}(T_d, P_d)$  of hydrophilic (PHI) amino acids on the protein surface—number of hydrophilic amino acids on the surface divided by the total number of surface amino acids—and the fraction  $\mathcal{Q}_{\text{core}}^{\text{PHO}}(T_d, P_d)$  of hydrophobic (PHO) amino acids into the protein core—number of hydrophobic amino acids divided by the total number of core amino acids (Figs. 2 and 3). The comparison between  $\mathcal{P}_{\text{surf}}^{\text{PHI}}(T_d, P_d)$  and  $\mathcal{Q}_{\text{core}}^{\text{PHO}}(T_d, P_d)$  characterizes the level of segregation of the sequence in the folded state. By segregation, we mean the tendency of the optimized sequence to have a surface exposed to the solvent that is rich in hydrophilic amino acids and a core that is abundant in hydrophobic amino acids. For all designed proteins and for both design protocols with solvent, we find that the optimized sequences at ambient conditions ( $P_d \approx 0.7 \times 10^{-4}$  IU,  $T_d \approx 0.35$  IU) are mostly hydrophilic on the surface and hydrophobic into the core, as expected. However, upon lowering  $T_d$ , both protocols with solvent generate protein sequences that, in their folded state, have a lower segregation between a hydrophilic

surface and a hydrophobic core (Fig. 3). These proteins with less segregation in their sequence are less stable at high  $T$  [Fig. 1(a)].

This finding is consistent with previous studies on thermophilic proteins, i.e., proteins stable at high  $T$ , showing that the higher thermostability is correlated with a stronger segregation between the polar surface [60,61] and the hydrophobic core [62,63]. In particular, a systematic analysis of the hydrophathy profile of thermophilic [58], mesophilic [58] (i.e., stable at intermediate  $T$ ), and ice-binding proteins (i.e., proteins that interfere with ice growth, Fig. 7) [59] reveals that these categories of proteins have, on average, compositions that are close to those that we calculate with our design strategies at high  $T$  ( $T \gtrsim 0.45$  IU and  $P \approx 0$ ), ambient  $T$  ( $T = 0.35$  IU and  $P \approx 0$ ), and low  $T$  ( $T \lesssim 0.15$  IU and  $P \approx 0$ ). Indeed, the data [58] (Fig. 3) show that  $\approx 78.8\%$  of the surface of thermophilic proteins is hydrophilic and  $\approx 54.5\%$  of their core is hydrophobic, while for mesophilic proteins, the percentages are  $\approx 78.4\%$  and  $\approx 54.7\%$ , respectively. Moreover, our direct analysis of ice-binding proteins [59] (Fig. 7) shows that  $\approx 67.0\%$  of their surface is hydrophilic and  $\approx 48.5\%$  of their core is hydrophobic. Hypothesizing that the protein’s composition is optimized at the environmental conditions at which the protein works, we observe that the real proteins have a segregation

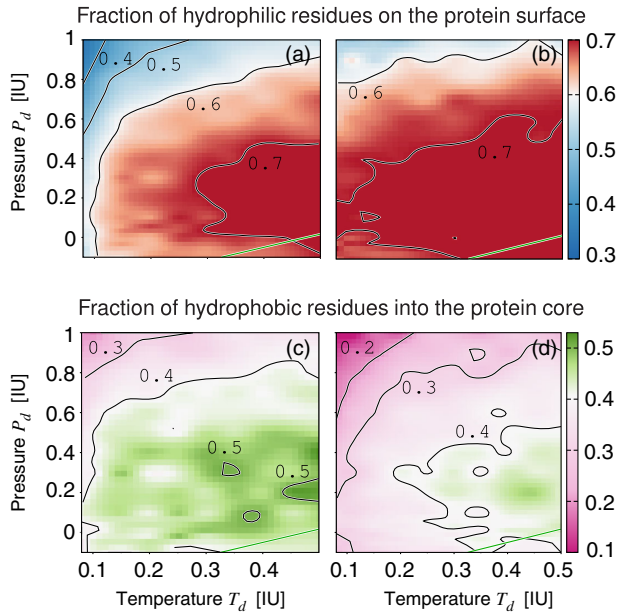


FIG. 2. Hydropathy profile of the designed protein surface and core. The color-coded contour plots of the fraction  $\mathcal{P}_{\text{surf}}^{\text{PHI}}(T_d, P_d)$  of surface hydrophilic amino acid [panels (a) and (b)] and the fraction  $\mathcal{Q}_{\text{core}}^{\text{PHO}}(T_d, P_d)$  of core hydrophobic amino acid [panels (c) and (d)] as a function of the design temperature  $T_d$  and pressure  $P_d$ , calculated with both design protocols, MIN ENTHALPY [panels (a) and (c)] and MAX GAP [panels (b) and (d)], show a stronger segregation between the hydrophilic surface and the hydrophobic core for proteins designed at high  $T_d$  and low or intermediate  $P_d$ . The straight (green) line in the lower-right corner of each panel marks the liquid-gas spinodal line of the solvent.  $T_d$  and  $P_d$  are expressed in internal units.

between hydrophilic surface and hydrophobic core that follows the same qualitative trend as our prediction (Figs. 3 and 8). Hence, both our theory and the real data show that the higher the temperature of stability of the protein, the stronger the segregation of the amino acid sequence into the hydrophilic surface and hydrophobic core, although an extreme segregation does not imply the stability of the folded state at higher  $T$ , as shown by the sequences generated with the implicit solvent design [Figs. 1(a) and 3].

To get insight into the role of water in the evolutionary process of sequence segregation, we perform a detailed comparison of the results of the two design protocols MIN ENTHALPY and MAX GAP. We find that they select similar sequences at high  $T_d$  but not at low  $T_d$ .

In particular, at high  $T_d$  and intermediate  $P_d$ , e.g.,  $T_d = 0.45$  IU and  $P_d = 0.2$  IU, both protocols select sequences with strong segregation between the hydrophilic surface and the hydrophobic core (Fig. 2). Instead, at low  $T_d$ , e.g.,  $T_d = 0.1$  IU and  $P_d = 0.2$  IU, the MIN ENTHALPY protocol selects sequences with a large number of surface hydrophobic residues with minor changes in the core with respect to the high- $T_d$  case [Figs. 2(a) and 2(c)], while the

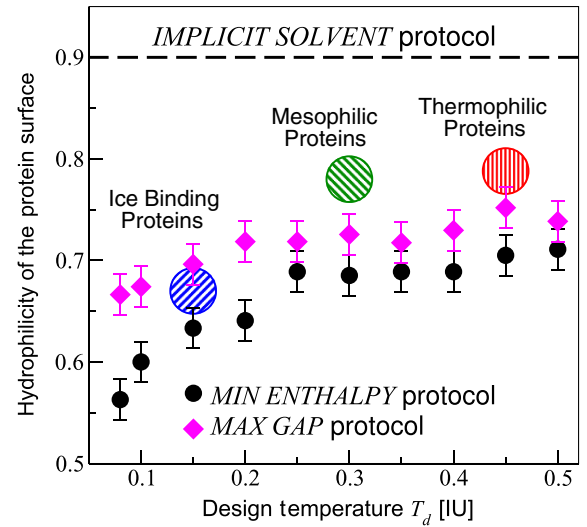


FIG. 3. Hydrophilic profile of a protein surface designed at ambient pressure. The fraction  $\mathcal{P}_{\text{surf}}^{\text{PHI}}(T_d, P_d)$  of surface hydrophilic amino acid decreases with the design temperature  $T_d$ , for both the design protocols MIN ENTHALPY (black circles) and MAX GAP (diamonds). Data are from Figs. 2(a) and 2(b) at  $P = 0$ . Our results qualitatively follow the trend observed for the average hydropathy of real thermophilic [58] (large red circle), mesophilic [58] (large green circle), and ice-binding proteins [59] (large blue circle), plotted at high  $T$  ( $T = 0.45$  IU), intermediate  $T$  ( $T = 0.3$  IU), and low  $T$  ( $T = 0.15$  IU), respectively, to ease the comparison. Such values of  $T$  resemble the average working temperature of real thermophilic, mesophilic, and ice-binding proteins. The hydropathy profile (independent of  $T_d$ ) of sequences designed with implicit solvent (dashed line) is unable to reproduce the observed trend.

MAX GAP protocol leads to proteins with a large number of hydrophilic residues in the core without major changes on the surface with respect to the high- $T_d$  case [Figs. 2(b) and 2(d)].

These differences in the selection are a consequence of the fact that at low  $T$  the hydrophobic amino acids are more soluble [3,9,64,65] because the energetic gain of forming water-water HBs at the hydrophobic interface compensates for the loss of water-hydrophilic residue HBs [9]. Hence, on the one hand, the first protocol minimizes the enthalpy of the  $f$  state by increasing the number of exposed hydrophobic residues.

On the other hand, the increased solubility at low  $T$  of a largely hydrophobic core would decrease the enthalpy of the  $u$  state and, as a consequence, the relative enthalpic gain of the  $f$  state. Hence, to maximize this gain, the second protocol selects sequences with a less hydrophobic core with respect to the high- $T$  case and without changes of the surface composition, i.e., of the enthalpy of the  $f$  state.

Remarkably, both categories of sequences—those with less hydrophilic surface and those with less hydrophobic core—have similar stability at low  $T$  [Figs. 1(a) and 9(a)]. Hence, at low  $T$ , there is a variety of folding—less

segregated—sequences that are larger than at high  $T$ . Furthermore, we observe that proteins designed at low  $T_d$ —e.g., at ( $T_d = 0.2$  IU,  $P_d = 0$ )—have a stability range that includes ambient conditions [Fig. 1(a)]. Hence, the lack of sequence segregation found in naturally evolved proteins [66,67] is a way to maximize the number of folding sequences at ambient conditions, taking advantage of the  $T$ - $P$ -dependent water properties.

This observation allows us to reconsider the usual understanding of the high number of hydrophobic residues on the surface of natural proteins as a necessary compromise between solubility and functionality [66,67]. To clarify this point, we apply the IMPLICIT SOLVENT protocol, which uses implicit water without accounting for the water contribution to the enthalpy; we find that it generates only sequences that are highly segregated—with  $\approx 90\%$  hydrophilic residues on the surface and  $\approx 55\%$  hydrophobic residues in the core—and that are unstable at extreme conditions, as well as those designed at ambient  $T$  and  $P$  [Fig. 1(a)] [68]. Therefore, our results suggest that at extreme  $T$  and  $P$ , proteins are more stable if not completely segregated in their sequence, and this is the result of an adaptation process in water since water is a solvent that changes properties with temperature and pressure.

### C. High-pressure design and stability

We now focus on the role of water in selecting sequences that fold at high  $P$ . While proteins designed at high  $T_d$  are stable in their  $f$  state in a range of  $T > T_d$ , we find that those generated by the protocols at high  $P_d$  fold only at  $P < P_d$  [Fig. 1(a)]. Furthermore, for  $P \gtrsim 0.7$  IU, no protein folds.

These findings are surprising because, although the design protocols at high  $P_d$  select for proteins that are more homogeneous than those generated at lower  $P_d$ —with surfaces less hydrophilic and cores less hydrophobic (Fig. 2)—the generated sequences are such that the hydrated protein enthalpy of the  $f$  state is much less than that of the  $u$  state (Fig. 4, inset). Therefore, the larger instability of the folded protein is due to all the terms in the Gibbs free energy that are not explicitly included in our calculation of the hydrated protein enthalpy, i.e., to the bulk-water contributions.

To clearly show this result, we calculate, as a function of  $P$  and  $T$ , the denaturation free-energy gain  $\Delta_G \equiv G^f - G^u$ , where  $G^f$  and  $G^u$  are the free energy for the entire system—protein, hydration water, and bulk water—calculated for the  $f$  state and for a completely stretched conformation representative of the  $u$  state, respectively. This quantity is given by the difference  $\Delta_G = \Delta_H - \Delta_{TS}$  between the denaturation enthalpy gain  $\Delta_H \equiv H^f - H^u$  and the folding entropy surplus  $\Delta_{TS} \equiv T(S^f - S^u)$ , where  $H^\alpha$  ( $S^\alpha$ ) are the enthalpy (entropy) of the  $\alpha$  state with  $\alpha = f, u$ . A positive value of  $\Delta_H$  favors the denaturation, while upon protein folding, there is an entropy surplus,  $\Delta_{TS} > 0$ , due to

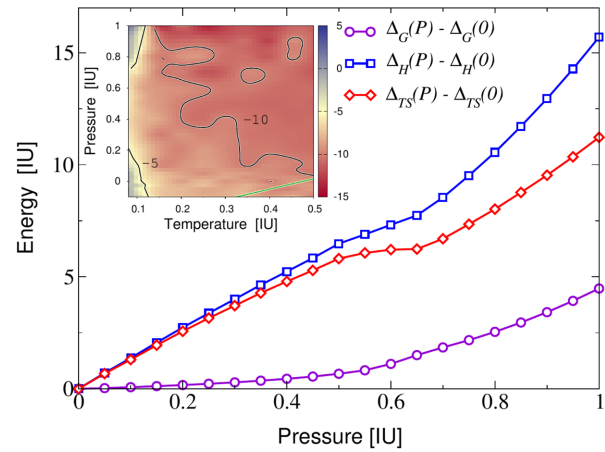


FIG. 4. Isothermal free energy, enthalpy, and entropy variation. Inset: Color-coded (in IU) maximum value of  $\langle \Delta H_R^{(h)} \rangle$  found for the sequences designed with the MAX GAP protocol. Main panel: The variations of the denaturation free-energy gain  $\langle \Delta_G(P) - \Delta_G(0) \rangle$ , of the denaturation enthalpy gain  $\langle \Delta_H(P) - \Delta_H(0) \rangle$ , and of the folding entropy surplus  $\langle \Delta_{TS}(P) - \Delta_{TS}(0) \rangle$  increase with  $P$  along the isotherm  $T = 0.15$  IU for a protein designed at  $P_d = 0.8$  IU and  $T_d = 0.2$  IU. We calculate  $\langle \Delta_H \rangle$  as thermodynamic average over equilibrium configurations,  $\langle \Delta_G(P) - \Delta_G(0) \rangle = \int_0^P \langle V^f(P) - V^u(0) \rangle dP$  and  $\langle \Delta_{TS}(P) - \Delta_{TS}(0) \rangle = \langle \Delta_H(P) - \Delta_H(0) \rangle - \langle \Delta_G(P) - \Delta_G(0) \rangle$ .

the larger bulk-water contribution. Hence, the balance  $\Delta_H - \Delta_{TS}$  regulates the protein stability, favoring the  $u$  state when  $\Delta_G > 0$  and the  $f$  state when  $\Delta_G < 0$ .

We find that for proteins designed at high  $P_d$  and low  $T_d$ , e.g.,  $P_d = 0.8$  IU and  $T_d = 0.2$  IU (Fig. 4), both  $\Delta_H$  and  $\Delta_{TS}$  become larger for increasing  $P$  at constant  $T$ , with the denaturation enthalpy gain increasing faster than the folding entropy surplus inducing the unfolding of the protein. The increase of  $\Delta_{TS}$  is expected because, for increasing  $P$ , the number of HBs decreases, making the entropy difference between the  $f$  and the  $u$  state larger when the water-protein interface decreases upon folding. On the other hand, our results (Fig. 4, inset) suggest that the protein and hydration-water contributions to  $\Delta_H$  are negative for the considered range of  $P$  and  $T$  and are slowly changing with  $P$  at  $T = 0.15$  IU. Therefore, the increase of  $\Delta_H$  implies that the contribution coming from bulk water is increasingly larger for increasing  $P$ . We conclude that the proteins designed at high  $P_d$  are unstable in their native state under pressurization because of the bulk-water contribution to the total free energy  $G$  of the system. This contribution completely dominates  $G$  for those pressures for which the number of HBs is vanishing—in our case, at  $P > 0.6$  IU [69,70].

Hence, according to our simulations, both the cold- and the pressure-denaturation boundaries are due to the free-energy contribution of the hydration and bulk water [71]. This limit is a consequence of the specific features of the HB, representing a natural barrier above which folding is

not expected. More importantly, such boundaries are also barriers for evolution and can be crossed only by introducing artificial interactions between the amino acids to stabilize the protein native state (Fig. 1) [72].

#### IV. CONCLUSIONS

We present a computationally efficient model capable of describing the artificial evolution of protein stability regions as a function of the thermodynamic properties of water. With our method, we study a large number of scenarios and demonstrate that the resulting stability regions are qualitatively similar to those of natural proteins.

Our results elucidate the role that water has in the selection process of protein sequences. In particular, we show that the maximum denaturation pressure  $\approx 1$  GPa above which all proteins denature in experiments is a consequence of the specific features of the hydration and bulk-water hydrogen bonds.

We adopt design protocols that mimic the natural selection of proteins and find that the proteins selected at high  $T_d$  are superstable (Fig. 1), i.e., are stable in a wide range of  $T$  and  $P$ , while those selected at lower  $T_d$  are stable only in a reduced range of  $T$  and  $P$ . We observe that proteins designed to be stable at different conditions of  $T$  and  $P$  are characterized by sequences with different degrees of segregation between the hydrophilic surface and the hydrophobic core (Figs. 2 and 3). The optimal degree of segregation is selected spontaneously as a consequence of the free-energy balance of the protein aqueous solution without the necessity of an external active process of selection. In particular, we find that the segregation decreases by decreasing  $T_d$ , and at ambient conditions, it is moderate ( $\approx 70\%$  of the surface is hydrophilic and  $\approx 50\%$  of the core is hydrophobic). This result is consistent with a trend observed in the composition of thermophilic, mesophilic, and ice-binding proteins (Fig. 3). The broader stability of high- $T$  designed proteins implies that these segregated sequences are a subset of those designed at lower  $T_d$ .

Furthermore, the general observation that many proteins expose the solvent to a high percentage of hydrophobic residues [66,67], as predicted by our model, suggests that such an exposure is not a compromise between stability and biological functionality of the proteins but rather a natural consequence of the water properties. As a matter of fact, selecting artificial sequences with an extreme segregation does not increase the stability of proteins but rather reduces it. We believe that our findings could potentially improve the engineering of artificial biopolymers since the aggregation can be prevented or enhanced in different thermodynamic conditions, according to the hydrophilic-hydrophobic ratio of the protein surface, although an experimental validation of the proposed design strategies remains crucial.

Putting the hydrophobic effect in an evolutionary perspective, our results substantiate an intriguing hypothesis: Many features observed in natural proteins generally arise

when the selection process explicitly takes into account the thermodynamic properties of the solvent.

#### ACKNOWLEDGMENTS

We are thankful to L. Rovigatti, E. Locatelli, and M. Goethe for helpful discussions and suggestions. V.B. and I.C. acknowledge support from the Austrian Science Fund (FWF), Grant No. P 26253-N27. V.B. acknowledges support from the FWF, Grant No. M 2150-N36. G.F. acknowledges support from Spanish MINECO Grants No. FIS2012-31025 and No. FIS2015-66879-C2-2-P. The computational results presented here have been achieved using the Vienna Scientific Cluster (VSC).

#### APPENDIX A: PROTEIN-WATER MODEL

At any  $P$  and  $T$ , we partition the total volume  $V(P, T)$  of our system into  $N$  regular, nonoverlapping cells of volume  $v \equiv V/N \geq v_0$ , where  $v_0$  is the water-excluded volume, each occupied by one of the  $N_C$  protein's residues or by water molecules (Fig. 10). We adopt a coarse-grain representation of the protein that follows previous works [29–31] and has been extensively used in the literature to get a qualitative understanding of protein properties (e.g., Refs. [4,5,9,73,74]) and protein design [21,22,75–78]. For the solvent, we consider a coarse-grain model capable of reproducing in a qualitative way, at least, the properties of water [8,32–34,36,79] and its contribution to the protein folding [9,35].

The Hamiltonian for the entire system is given by

$$\mathcal{H} \equiv \mathcal{H}_{R,R} + \mathcal{H}_{R,w} + \mathcal{H}_{w,w}, \quad (\text{A1})$$

where each term is defined in the following. As in Refs. [29–31], we assume that the protein's residues have only nearest-neighbor interactions,

$$\mathcal{H}_p \equiv \mathcal{H}_{R,R} + \mathcal{H}_{R,w} \equiv \sum_i^{N_C} \left[ \sum_{j \neq i}^{N_C} C_{ij} S_{ij} + \sum_{j'}^{N_W} C_{ij'} S_i^W \right], \quad (\text{A2})$$

where the indices  $i$  and  $j$  run over the residues and  $j'$  runs over the  $N_W \leq N - N_C$  water molecules in the hydration shell [80];  $C$  is a contact matrix, with  $C_{l,k} = 1$ , if  $l$  and  $k$  are first neighbors, and 0 otherwise;  $S_{ij}$  are elements of the Miyazawa-Jernigan residue-residue interaction matrix  $S$  [29,81,82] (solvent independent—see Table 1 in Appendix C) accounting for the correlations between real amino acids, and for the protein-water interaction  $S_i^W = \epsilon^{\text{PHO}}$  if the residue  $i$  is hydrophobic and  $S_i^W = \epsilon^{\text{PHI}}$  if it is hydrophilic [9].

Each cell accommodates, at most, one molecule, with the average O—O distance between next-neighbor water molecules given by  $r = v^{1/3}$ . We associate a variable  $n_i = 1$

with each cell, if the cell  $i$  is occupied by a water molecule and has  $v_0/v > 0.5$ , and  $n_i = 0$  otherwise. Hence,  $n_i$  is a discretized density field replacing the water translational degrees of freedom.

The Hamiltonian of bulk water is

$$\mathcal{H}_{\text{w,w}}^{(b)} \equiv \sum_{ij} U(r_{ij}) - JN_{\text{HB}}^{(b)} - J_{\sigma}N_{\text{coop}}^{(b)}. \quad (\text{A3})$$

The first term, with  $U(r) \equiv \infty$  for  $r < r_0 \equiv v_0^{1/3} = 2.9 \text{ \AA}$  (water van der Waals diameter),  $U(r) \equiv 4\epsilon[(r_0/r)^{12} - (r_0/r)^6]$  for  $r \geq r_0$ , with  $\epsilon \equiv 2.9 \text{ kJ/mol}$ , and  $U(r) \equiv 0$  for  $r > r_c \equiv 3r_0$  (cutoff), accounts for the O—O van der Waals interaction between molecules  $i$  and  $j$ . The sum runs over all possible water-molecule couples.

The second term represents the directional component of the HB interaction, where  $N_{\text{HB}}^{(b)} \equiv \sum_{(ij)} n_i n_j \delta_{\sigma_{ij}, \sigma_{ji}}$  is the total number of bulk HBs. The sum is over each nearest-neighbor pair, and the argument is nonzero if the following conditions are satisfied: (i)  $n_i n_j = 1$ , i.e.,  $r_{ij} \leq 2^{1/3} r_0 = 3.6 \text{ \AA}$ , the maximum distance that is conventionally associated with a HB (while  $n_i n_j = 0$  otherwise), and (ii)  $\delta_{\sigma_{ij}, \sigma_{ji}} = 1$ , with  $\delta_{ab} = 1$  if  $a = b$ , and 0 otherwise. Here,  $\sigma_{ij} = 1, \dots, q$  is a bonding index. Conventionally, a HB is broken if  $\widehat{\text{OOH}} > 30^\circ$ , implying that only 1/6 of the entire range of values [0,360°] for the  $\widehat{\text{OOH}}$  angle is associated with a bonded state. Thus, choosing  $q = 6$ , we correctly account for the entropy variation due to HB formation and breaking. With this definition, each molecule can form up to four HBs with its neighbors. Bifurcated HBs are excluded.

The third term accounts for the HB cooperativity due to the quantum many-body interaction [83,84] and leads to the low- $P$  tetrahedral structure [85]. The cooperativity is defined as the sum  $N_{\text{coop}}^{(b)} \equiv \sum_i n_i \sum_{lk} \delta_{\sigma_{ik}, \sigma_{il}}$ , over all the water molecules  $i$  and over all the  $lk$  pairs of the bonding indices  $\sigma_{il}$  and  $\sigma_{ik}$  of the molecule  $i$ . We choose  $J_{\sigma} \ll J$  to guarantee an asymmetry between the two HB terms.

The formation of a HB in the bulk leads to the local volume increase  $v_{\text{HB}}^{(b)}/v_0$ , with an enthalpic variation  $-J + P v_{\text{HB}}^{(b)}$ , which accounts for the  $P$ -disrupting effect on the HB network. Here,  $v_{\text{HB}}^{(b)}/v_0$  represents the average volume increase between high-density ices VI and VIII and low-density (tetrahedral) ice Ih, and it is chosen as an approximation of the average volume variation per HB when a tetrahedral HB network is formed. Hence, the volume of bulk molecules is  $V^{(b)} = Nv + N_{\text{HB}}^{(b)} v_{\text{HB}}^{(b)}$ .

The presence of the hydrophobic or hydrophilic protein interface affects the water-water hydrogen bonding in the hydration shell [7,86–94] (water molecules that are first neighbors of the protein amino acids). Hence, the Hamiltonian for water, including hydration molecules and the many-body effect on HB formation close to the protein interface, reads

$$\mathcal{H}_{\text{w,w}} \equiv \mathcal{H}_{\text{w,w}}^{(b)} + \mathcal{H}_{\text{w,w}}^{(h)}, \quad (\text{A4})$$

where

$$\mathcal{H}_{\text{w,w}}^{(h)} \equiv -[J^{\text{PHO}} N_{\text{HB}}^{\text{PHO}} + J^{\text{PHI}} N_{\text{HB}}^{\text{PHI}}] - [J_{\sigma}^{\text{PHO}} N_{\text{coop}}^{\text{PHO}} + J_{\sigma}^{\text{PHI}} N_{\text{coop}}^{\text{PHI}}]. \quad (\text{A5})$$

Within the square brackets of Eq. (A5), we have explicitly indicated the contribution from the water molecules at the protein-interface ( $h$ ) neighboring hydrophobic or hydrophilic residues.

The hydrophobic interface strengthens the water-water hydrogen bonding in the first hydration shell [7,86–90] and increases the local water density upon pressurization [90,95–97]. Therefore, we assume  $J^{\text{PHO}} > J$  for HBs between water molecules at the hydrophobic interface and  $J_{\sigma}^{\text{PHO}} > J_{\sigma}$  for the cooperative component. This condition guarantees that the solvation free energy of a hydrophobic amino acid decreases at low  $T$  [65]. Following Ref. [9], we express the average volume change per water-water HB at the hydrophobic interface as a series expansion in  $P$  up to the linear term  $v_{\text{HB}}^{\text{PHO}}/v_{\text{HB},0}^{\text{PHO}} \equiv 1 - k_1 P$  [98], where  $v_{\text{HB},0}^{\text{PHO}}$  is the volume change associated with the HB formation in the hydrophobic hydration shell at  $P = 0$  and  $k_1 > 0$ . Therefore, the volume contribution  $V^{\text{PHO}}$  to total volume  $V$  due to the HBs in the hydrophobic shell is  $V^{\text{PHO}} \equiv N_{\text{HB}}^{\text{PHO}} v_{\text{HB}}^{\text{PHO}}$ , where  $V^{\text{PHO}}$  and  $N_{\text{HB}}^{\text{PHO}}$  are the hydrophobic hydration shell volume and number of HBs, respectively. We assume that the water-water hydrogen bonding and the water density at the hydrophilic interface are not affected by the protein since they are equal to the bulk-water values. Therefore,  $J^{\text{PHI}} = J$ ,  $J_{\sigma}^{\text{PHI}} = J_{\sigma}$ , and  $v_{\text{HB}}^{\text{PHI}} = v_{\text{HB}}^{(b)}$ . Moreover, the polarization effect of hydrophilic residues on the HB network in the hydrophilic shell is not included here because, in our coarse-grain description, it does not show a qualitative change in protein denaturation mechanisms [9], consistent with previous observations [99]. A HB between two water molecules, one hydrating a PHO amino acid and the other a PHI amino acid, is formed with a coupling constant  $(J^{\text{PHO}} + J^{\text{PHI}})/2$ , and it leads to a local increase of volume  $(v_{\text{HB}}^{\text{PHO}} + v_{\text{HB}}^{\text{PHI}})/2$ .

Hence, the total volume is

$$V \equiv V^{(b)} + V^{(h)}, \quad (\text{A6})$$

where  $V^{(h)} \equiv V^{\text{PHO}} + V^{\text{PHI}}$  is the volume due to the HBs in the hydration shell.

In order to favor the visualization and the understanding of our results, we adopt a model representation in two dimensions [4,5,7,75–78,100]. We choose the parameters in Eq. (A1) in such a way as to get proteins that are stable for  $250 \lesssim T/\text{K} \lesssim 350$  and  $P < 1 \text{ GPa}$ , consistent with experimental observations [40–51,54,55,101–103]. We express all the quantities in IU: adopting  $8\epsilon$  as the energy unit,  $v_0$  as the volume unit,  $8\epsilon/k_B$  as the temperature unit, and  $8\epsilon/v_0$  as the

pressure unit. Accordingly, we fix the interaction constants (all expressed in IU)  $J = 0.3$  and  $J_\sigma = 0.05$  for bulk water,  $J^{\text{PHI}} = J$  and  $J_\sigma^{\text{PHI}} = J_\sigma$  for water at hydrophilic interfaces, and  $J^{\text{PHO}} = 1.2$  and  $J_\sigma^{\text{PHO}} = 0.2$  for water at hydrophobic interfaces. Moreover, we fix  $k_1 = 4$  and  $v_{\text{HB},0}^{\text{PHO}} = 2$  (in IU). These choices have two effects: (i) They balance the residue-residue, residue-water, and water-water interactions and make the proteins stable for thermodynamic conditions comprised in the (stable and metastable) liquid phase, including ambient conditions; (ii) they account for the lower surface volume ratio in the two-dimensional system with respect to a three-dimensional one, enhancing the interface interactions. Changes in parameters combine in a nontrivial way, resulting in a shift, broadening, or reduction of the stability regions of proteins, but leaving the qualitative scenario unaffected [104]. Lastly, since our preliminary results on the three-dimensional many-body water model [105] show a phase diagram qualitatively similar to the one in two dimensions [36], we expect that our findings will remain substantially unaltered in a more realistic three-dimensional version of the model.

## APPENDIX B: DESIGN AND FOLDING SIMULATIONS

The concept of protein design refers to an optimization scheme for the sequence of amino acids, aiming to maximize the probability of folding into a specific target conformation. The design simulation consists in a broad sampling of the space of sequences, on top of a fixed protein structure that defines the native-target conformations (Fig. 10). We perform Monte Carlo simulations, keeping  $P$ ,  $T$ ,  $N_C$ , and  $N$  constant, with point mutation of the sequence and residue swapping moves with the following acceptance probability:

$$\mathcal{P}_{\text{acc}}^{\text{min}} \equiv \min\{1, e^{-\Delta/T_o}\} \min\{1, (\mathcal{N}_p^n/\mathcal{N}_p^o)^{\epsilon_p}\}, \quad (\text{B1})$$

where  $\Delta$  is the difference between the new and the old configurations in  $\langle H_R^{(h)} \rangle$  or  $\langle \Delta H_R^{(h)} \rangle$  or  $\mathcal{H}_{R,R}$ , depending on the design protocol;  $T_o = 0.05$  IU is the optimization temperature,  $\mathcal{N}_p^n$  and  $\mathcal{N}_p^o$  are the number of permutations for the new ( $n$ ) and old ( $o$ ) amino acid sequences, respectively; and  $\epsilon_p = 14$  is a weighting parameter. The term  $(\mathcal{N}_p^n/\mathcal{N}_p^o)^{\epsilon_p}$  is added to bias towards highly heterogeneous sequences, which are better folders [24,30]. Therefore, we minimize the enthalpy of the folded structure via a Monte Carlo scheme with separate acceptance criteria for the water moves and the sequence moves. While the water is simulated at  $T_d$  and  $P_d$ , the sequences are sampled at low optimization temperature  $T_o$ . In Fig. 8, we show that tuning  $\epsilon_p$  does not significantly affect the hydrophathy profile of proteins, although a low value of  $\epsilon_p$  generates

homopolymer solutions. We set  $T_o \ll T$ , the design temperature, because we look for sequences with either minimum or maximum values of the enthalpy. A Monte Carlo step consists in an attempt to modify the protein sequence followed by a number of water moves that is large enough to equilibrate the solvent around the fixed protein. For the MAX GAP design protocol, we also perform a simulation with a completely stretched protein, whose sequence is identical to the folded protein, in order to calculate the enthalpy difference between the two conformations (native and unfolded). For both the water-dependent design protocols, the protein enthalpy that we consider is  $\langle H_R^{(h)} \rangle \equiv \langle \mathcal{H}_p \rangle + \langle \mathcal{H}_{w,w}^{(h)} \rangle + P \langle V^{(h)} \rangle$ .

We perform the design for a large number of thermodynamic state points, sampling  $[3; 5] \times 10^8$  independent sequences for each  $T$ ,  $P$  and protein target-native structure. Each water-dependent design is performed on a grid of  $\approx 90$  different  $T$ - $P$  points. Implicit water design is performed once per structure. We sample the averages  $\langle H_R^{(h)} \rangle$  and  $\langle \Delta H_R^{(h)} \rangle$  for each sequence over  $\approx 10^2$  water configurations. We sort, in ascending order, the sequences according to their values of  $\langle H_R^{(h)} \rangle$ ,  $-\langle \Delta H_R^{(h)} \rangle$ , and  $\mathcal{H}_{R,R}$ , and consider, for characterization and stability analysis, only the top 5; 15 sequences from each list. Overall, we perform more than 1500 independent designs.

We test the validity of the design by folding the protein with Monte Carlo simulations at constant values of  $P$ ,  $T$ ,  $N_C$ , and  $N$ . We start from a stretched protein conformation and, keeping the amino acid sequence, allow the protein to move using local corner flip, pivot, and crankshaft algorithms [106]. Water is equilibrated using cluster moves [36]. For each sequence and each state point, we calculate the free energy as a function of the number of native contacts. A protein is defined to be stable if, at the free-energy minimum, 90% of its native contacts are formed. This definition guarantees that a stable protein folds in its native state. Computing the free energy on a grid of values of  $T$  and  $P$  yields the stability region in the  $T$ - $P$  plane. By averaging the stability region over all sequences designed at the same  $T_d$  and  $P_d$ , we calculate the average stability curve.

## APPENDIX C: RESIDUE-RESIDUE INTERACTION MATRIX S

In Table I we report the interaction matrix between the amino acids. This matrix does not include solvent contributions, since these are explicitly stated in the Hamiltonians  $\mathcal{H}_{R,w}$  and  $\mathcal{H}_{w,w}^{(h)}$ , and has been scaled by a factor 2 to balance with the water-water HB interaction. Amino acids are indicated with letters following the FASTA code. The amino acids I, V, L, F, C, M, and A are assumed to be hydrophobic, according to the Kyte-Doolittle hydrophathy scale [82].



TABLE I. Residue-residue interaction energy expressed in internal units.

|   | A     | C     | D    | E     | F     | G     | H     | I     | K     | L     | M     | N     | P     | Q     | R     | S     | T     | V     | W     | Y     |
|---|-------|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| A | -0.26 | 0     | 0.24 | 0.52  | 0.06  | -0.14 | 0.68  | -0.44 | 0.28  | -0.02 | 0.5   | 0.56  | 0.2   | 0.16  | 0.86  | -0.12 | -0.18 | -0.2  | -0.18 | 0.18  |
| C |       | -2.12 | 0.06 | 1.38  | -0.46 | -0.16 | -0.38 | 0.32  | 1.42  | -0.16 | 0.38  | 0.26  | 0     | 0.1   | 0.48  | -0.04 | 0.38  | 0.12  | 0.16  | 0.08  |
| D |       |       | 0.08 | -0.3  | 0.78  | -0.44 | -0.78 | 1.18  | -1.52 | 1.34  | 1.3   | -0.6  | 0.08  | -0.34 | -1.44 | -0.62 | -0.58 | 1.16  | 0.48  | 0     |
| E |       |       |      | -0.06 | 0.54  | 0.5   | -0.9  | 0.7   | -1.94 | 0.86  | 0.88  | -0.64 | -0.2  | -0.34 | -1.48 | -0.52 | 0     | 0.68  | 0.58  | -0.2  |
| F |       |       |      |       | -0.88 | -0.76 | -0.32 | -0.38 | 0.88  | -0.6  | -0.84 | 0.36  | 0.4   | -0.58 | 0.82  | 0.58  | 0.62  | -0.44 | -0.32 | 0     |
| G |       |       |      |       |       | -0.76 | 0.4   | 0.5   | 0.22  | 0.46  | 0.38  | -0.28 | -0.22 | -0.12 | -0.08 | -0.32 | -0.52 | 0.32  | 0.36  | 0.28  |
| H |       |       |      |       |       |       | -0.58 | 0.98  | 0.44  | 0.32  | 1.98  | -0.48 | -0.42 | -0.04 | -0.24 | -0.1  | -0.38 | 0.38  | -0.24 | -0.68 |
| I |       |       |      |       |       |       |       | -0.44 | 0.72  | -0.82 | -0.56 | 1.06  | 0.5   | 0.72  | 0.84  | 0.42  | 0.28  | -0.5  | 0.04  | 0.22  |
| K |       |       |      |       |       |       |       |       | 0.5   | 0.38  | 0     | -0.66 | 0.22  | -0.76 | 1.5   | -0.26 | -0.18 | 0.88  | 0.44  | -0.42 |
| L |       |       |      |       |       |       |       |       |       | -0.54 | -0.4  | 0.6   | 0.84  | 0.52  | 0.7   | 0.5   | 0.4   | -0.58 | -0.18 | 0.48  |
| M |       |       |      |       |       |       |       |       |       |       | 0.08  | 0.16  | -0.68 | 0.92  | 0.62  | 0.28  | 0.38  | -0.28 | -1.34 | -0.26 |
| N |       |       |      |       |       |       |       |       |       |       |       | -1.06 | -0.36 | -0.5  | -0.28 | -0.28 | -0.22 | 1     | 0.12  | -0.4  |
| P |       |       |      |       |       |       |       |       |       |       |       |       | 0.52  | -0.84 | -0.76 | 0.02  | -0.14 | 0.18  | -0.56 | -0.66 |
| Q |       |       |      |       |       |       |       |       |       |       |       |       |       | 0.58  | -1.04 | -0.28 | -0.28 | 0.48  | 0.16  | -0.4  |
| R |       |       |      |       |       |       |       |       |       |       |       |       |       |       | 0.22  | 0.34  | -0.7  | 0.6   | -0.32 | -0.5  |
| S |       |       |      |       |       |       |       |       |       |       |       |       |       |       |       | -0.4  | -0.16 | 0.36  | 0.68  | 0.18  |
| T |       |       |      |       |       |       |       |       |       |       |       |       |       |       |       |       | 0.06  | 0.5   | -0.44 | 0.26  |
| V |       |       |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       | -0.58 | -0.14 | 0.04  |
| W |       |       |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | -0.24 | -0.08 |
| Y |       |       |      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       | -0.12 |

### APPENDIX D: ADDITIONAL FIGURES

In this Appendix we report supplementary figures.

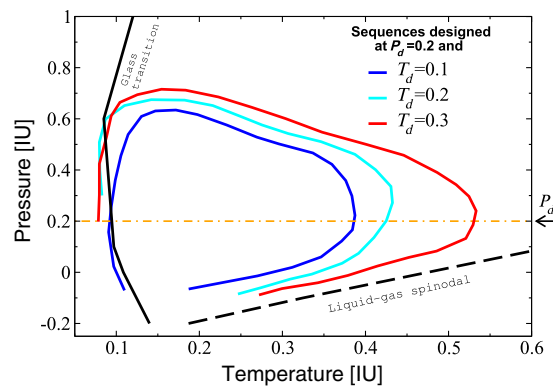


FIG. 5. Average stability regions for proteins designed at constant pressure according to the MIN ENTHALPY protocol. The proteins are designed along the isobar  $P_d = 0.2$  IU. We observe how designed sequences at high  $T_d$  are more resistant to thermal fluctuations.

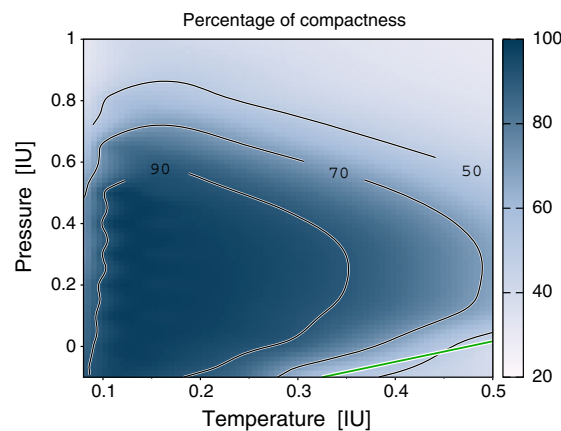


FIG. 6. Average compactness of the designed proteins. The compactness is calculated as the average number of residue-residue contact points, regardless of if they correspond to the native contacts, as a function of  $T$  and  $P$ . It is expressed as a (color-coded) percentage over the maximum possible number of contacts and averaged over proteins designed at different values of  $P_d$  and  $T_d$ . The straight (green) line on the bottom represents the liquid-gas spinodal line.

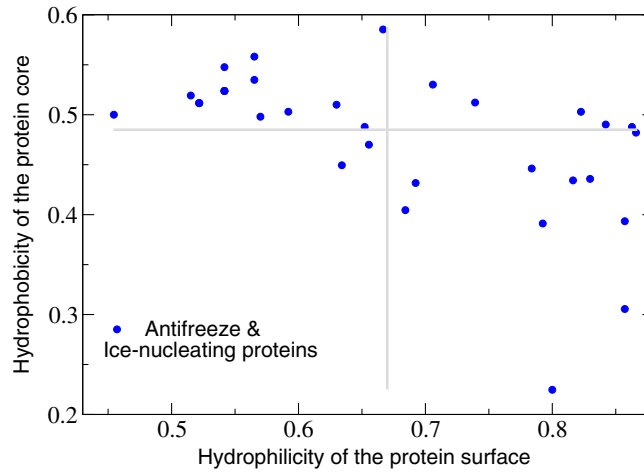


FIG. 7. Hydropathy of the protein surface and the protein core of the ice-binding proteins.

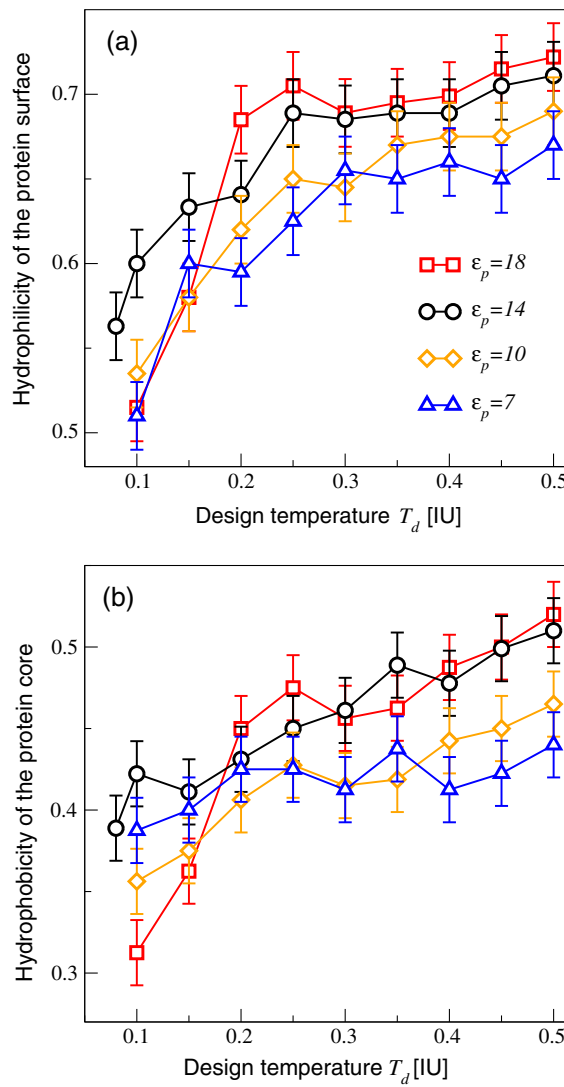


FIG. 8. Hydropathy profile of proteins designed at ambient pressure for different values of  $\epsilon_p$ . Here, we report how the hydrophilic profile of the protein surface and the hydrophobic profile of the protein core, respectively shown in panels (a) and (b), changes as a function of  $T_d$ , at ambient pressure, according to the parameter  $\epsilon_p$ , appearing in Eq. (B1). The data are calculated following the protocol MIN ENTHALPY. In all cases, the segregation between the hydrophilic surface and the hydrophobic core decreases with  $T$ .

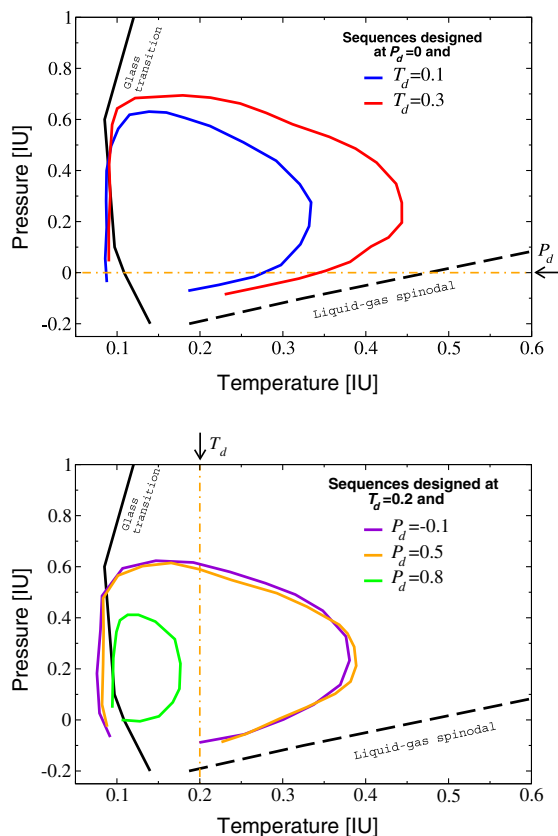


FIG. 9. Average stability regions for proteins designed according to the MAX GAP protocol. Average stability regions for proteins designed along the isobar  $P_d = 0$  and the isotherm  $T_d = 0.2$  IU, respectively, shown in panels (a) and (b). Arrows point at the design pressure  $P_d = 0$  IU and to the design temperature  $T_d = 0.2$  IU.

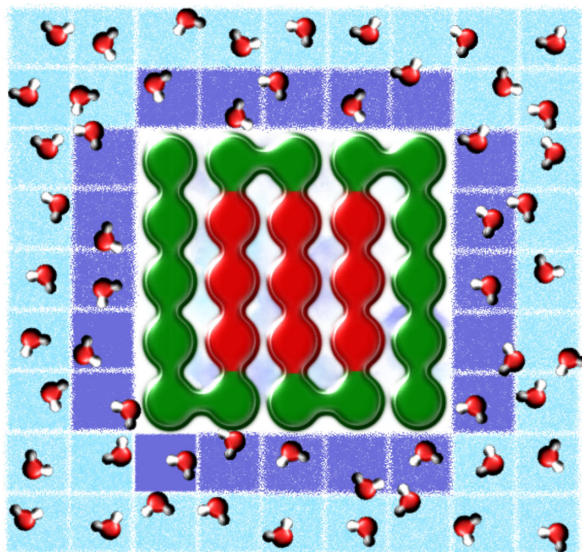


FIG. 10. Schematic representation of the water-protein model. The position of the water molecules is coarse-grained, assigning a cell to each water molecule. The cell size coincides with the average distance between first-neighbor molecules and fluctuates according to the Boltzmann weight of isobaric-isothermal ensemble. The conformational state of water molecule  $i$  is described via four bonding indices  $\sigma_{ij}$ , each one accounting for the bonding conformation of molecule  $i$  with respect to the neighbor in the direction  $j$ . Water molecules around the protein form the hydration shell (dark blue cells). Protein is modeled as a self-avoiding lattice chain, composed of 20 different amino acids. Each amino acid can be hydrophobic or hydrophilic, according to the Kyte-Doolittle hydrophathy scale [82]. Red and green amino acids in the figure refer to core and surface elements, respectively.

- [1] K. E. Zachariassen and E. Kristiansen, *Ice Nucleation and Antinucleation in Nature*, *Cryobiology* **41**, 257 (2000).
- [2] L. J. Rothschild and R. L. Mancinelli, *Life in Extreme Environments*, *Nature (London)* **409**, 1092 (2001).
- [3] G. Salvi, S. Mölbert, and P. De Los Rios, *Design of Lattice Proteins with Explicit Solvent*, *Phys. Rev. E* **66**, 061911 (2002).
- [4] M. I. Marqués, J. M. Borreguero, H. E. Stanley, and N. V. Dokholyan, *Possible Mechanism for Cold Denaturation of Proteins at High Pressure*, *Phys. Rev. Lett.* **91**, 138103 (2003).
- [5] B. A. Patel, P. G. Debenedetti, F. H. Stillinger, and P. J. Rossky, *A Water-Explicit Lattice Model of Heat-, Cold-, and Pressure-Induced Protein Unfolding*, *Biophys. J.* **93**, 4116 (2007).
- [6] I. N. Berezovsky, K. B. Zeldovich, and E. I. Shakhnovich, *Positive and Negative Design in Stability and Thermal Adaptation of Natural Proteins*, *PLoS Biol.* **3**, e52 (2007).
- [7] C. L. Dias, T. Ala-Nissila, M. Karttunen, I. Vattulainen, and M. Grant, *Microscopic Mechanism for Cold Denaturation*, *Phys. Rev. Lett.* **100**, 118101 (2008).
- [8] G. Franzese and V. Bianco, *Water at Biological and Inorganic Interfaces*, *Food Biophys.* **8**, 153 (2013).
- [9] V. Bianco and G. Franzese, *Contribution of Water to Pressure and Cold Denaturation of Proteins*, *Phys. Rev. Lett.* **115**, 108101 (2015).
- [10] S. B. Kim, J. C. Palmer, and P. G. Debenedetti, *Computational Investigation of Cold Denaturation in the Trp-Cage Miniprotein*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8991 (2016).
- [11] J. R. Desjarlais and T. M. Handel, *De Novo Design of the Hydrophobic Cores of Proteins*, *Protein Sci.* **4**, 2006 (1995).
- [12] B. I. Dahiyat, *De Novo Protein Design: Fully Automated Sequence Selection*, *Science* **278**, 82 (1997).
- [13] B. I. Dahiyat and S. L. Mayo, *Probing the Role of Packing Specificity in Protein Design*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10172 (1997).
- [14] B. Kuhlman *et al.*, *Design of a Novel Globular Protein Fold with Atomic-Level Accuracy*, *Science* **302**, 1364 (2003).
- [15] D. Rothlisberger *et al.*, *Kemp Elimination Catalysts by Computational Enzyme Design*, *Nature (London)* **453**, 190 (2008).
- [16] L. Jiang *et al.*, *De Novo Computational Design of Retro-Aldol Enzymes*, *Science* **319**, 1387 (2008).
- [17] D. J. Hei and D. S. Clark, *Pressure Stabilization of Proteins from Extreme Thermophiles*, *Appl. Environ. Microbiol.* **60**, 932 (1994).
- [18] J. Torrent, P. Rubens, M. Ribó, K. Heremans, and M. Vilanova, *Pressure Versus Temperature Unfolding of Ribonuclease A: An FTIR Spectroscopic Characterization of 10 Variants at the Carboxy-Terminal Site*, *Protein Sci.* **10**, 725 (2001).
- [19] S. Kumar, C. J. Tsai, and R. Nussinov, *Temperature Range of Thermodynamic Stability for the Native State of Reversible Two-State Proteins*, *Biochemistry* **42**, 4864 (2003).
- [20] B. N. Dominy, H. Minoux, and C. L. Brooks, *An Electrostatic Basis for the Stability of Thermophilic Proteins*, *Proteins* **57**, 128 (2004).
- [21] E. I. Shakhnovich and A. M. Gutin, *A New Approach to the Design of Stable Proteins*, *Protein Eng.* **6**, 793 (1993).
- [22] E. I. Shakhnovich and A. M. Gutin, *Engineering of Stable and Fast-Folding Sequences of Model Proteins*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7195 (1993).
- [23] I. Coluzza, *Transferable Coarse-Grained Potential for De Novo Protein Folding and Design*, *PLoS One* **9**, e112852 (2014).
- [24] I. Coluzza, *A Coarse-Grained Approach to Protein Design: Learning from Design to Understand Folding*, *PLoS One* **6**, e20853 (2011).
- [25] In order to optimize a sequence for a given target structure, a common strategy, known as negative design, consists in minimizing the free energy of the target conformation with respect to the free energy of all the possible compact conformations [26–28]. Negative design is normally necessary when, due to the limitations of the model used (e.g., reduced alphabet size or coarse-graining), the standard procedure cannot find a sequence that folds to the target structure. In all the models we considered so far, we could always directly design on- and off-lattice structures using an alphabet of 20 letters [29–31].
- [26] C. Micheletti, F. Seno, A. Maritan, and J. R. Banavar, *Design of Proteins with Hydrophobic and Polar Amino Acids*, *Proteins* **32**, 80 (1998).
- [27] I. Samish, C. M. Macdermaid, J. M. Perez-Aguilar, and J. G. Saven, *Theoretical and Computational Protein Design*, *Annu. Rev. Phys. Chem.* **62**, 129 (2011).
- [28] N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, and D. Baker, *Principles for Designing Ideal Protein Structures*, *Nature (London)* **491**, 222 (2012).
- [29] I. Coluzza and D. Frenkel, *Designing Specificity of Protein-Substrate Interactions*, *Phys. Rev. E* **70**, 051917 (2004).
- [30] I. Coluzza and D. Frenkel, *Monte Carlo Study of Substrate-Induced Folding and Refolding of Lattice Proteins*, *Biophys. J.* **92**, 1150 (2007).
- [31] I. Coluzza and J.-P. Hansen, *Transition from Highly to Fully Stretched Polymer Brushes in Good Solvent*, *Phys. Rev. Lett.* **100**, 016104 (2008).
- [32] K. Stokely, M. G. Mazza, H. E. Stanley, and G. Franzese, *Effect of Hydrogen Bond Cooperativity on the Behavior of Water*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1301 (2010).
- [33] M. G. Mazza, K. Stokely, S. E. Pagnotta, F. Bruni, H. E. Stanley, and G. Franzese, *More than One Dynamic Crossover in Protein Hydration Water*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19873 (2011).
- [34] G. Franzese, V. Bianco, and S. Iskrov, *Water at Interface with Proteins*, *Food Biophys.* **6**, 186 (2011).
- [35] V. Bianco, S. Iskrov, and G. Franzese, *Understanding the Role of Hydrogen Bonds in Water Dynamics and Protein Stability*, *J. Biol. Phys.* **38**, 27 (2012).
- [36] V. Bianco and G. Franzese, *Critical Behavior of a Water Monolayer under Hydrophobic Confinement*, *Sci. Rep.* **4**, 4440 (2014).
- [37] Here,  $T_d$  defines the fluctuations of the water's energy, in units of the Boltzmann constant  $k_B$ , over which the enthalpy is averaged. This parameter is different from

- what is called “design temperature” in V. S. Pande, A. Y. Grosberg, and T. Tanaka, *Biophys. J.* **73**, 3192 (1997), which in our case has a constant and very low value.
- [38] S. A. Hawley, *Reversible Pressure-Temperature Denaturation of Chymotrypsinogen*, *Biochemistry* **10**, 2436 (1971).
- [39] H. W. Hatch, F. H. Stillinger, and P. G. Debenedetti, *Computational Study of the Stability of the Miniprotein Trp-Cage, the GB 1 $\beta$ -Hairpin, and the AK16 Peptide, under Negative Pressure*, *J. Phys. Chem. B* **118**, 7761 (2014).
- [40] L. Smeller, *Pressure-Temperature Phase Diagrams of Biomolecules*, *Biochim. Biophys. Acta* **1595**, 11 (2002).
- [41] R. Ravindra and R. Winter, *On the Temperature-Pressure Free-Energy Landscape of Proteins*, *Chem. Phys. Chem.* **4**, 359 (2003).
- [42] H. Herberhold and R. Winter, *Temperature- and Pressure-Induced Unfolding and Refolding of Ubiquitin: A Static and Kinetic Fourier Transform Infrared Spectroscopy Study*, *Biochemistry* **41**, 2396 (2002).
- [43] H. Lesch, H. Stadlbauer, J. Friedrich, and J. M. Vanderkooi, *Stability Diagram and Unfolding of a Modified Cytochrome c: What Happens in the Transformation Regime?*, *Biophys. J.* **82**, 1644 (2002).
- [44] J. Wiedersich, S. Köhler, A. Skerra, and J. Friedrich, *Temperature and Pressure Dependence of Protein Stability: The Engineered Fluorescein-Binding Lipocalin FluA Shows an Elliptic Phase Diagram*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5756 (2008).
- [45] A. Zipp and W. Kauzmann, *Pressure Denaturation of Metmyoglobin*, *Biochemistry* **12**, 4217 (1973).
- [46] M. W. Lassalle, H. Yamada, and K. Akasaka, *The Pressure-Temperature Free Energy-Landscape of Staphylococcal Nuclease Monitored by (1)H NMR*, *J. Mol. Biol.* **298**, 293 (2000).
- [47] A. Maeno, H. Matsuo, and K. Akasaka, *The Pressure-Temperature Phase Diagram of Hen Lysozyme at Low pH*, *Biophysik* **5**, 1 (2009).
- [48] J. Somkuti, Z. Mártonfalvi, M. S. Z. Kellermayer, and L. Smeller, *Different Pressure-Temperature Behavior of the Structured and Unstructured Regions of Titin*, *Biochim. Biophys. Acta* **1834**, 112 (2013).
- [49] J. Somkuti, S. Jain, S. Ramachandran, and L. Smeller, *Folding-Unfolding Transitions of Rv3221c on the Pressure-Temperature Plane*, *High Press. Res.* **33**, 250 (2013).
- [50] F. Meersman, L. Smeller, and K. Heremans, *Pressure-Assisted Cold Unfolding of Proteins and Its Effects on the Conformational Stability Compared to Pressure and Heat Unfolding*, *High Press. Res.* **19**, 263 (2000).
- [51] A. Pastore, S. R. Martin, A. Politou, K. C. Kondapalli, T. Stemmler, and P. A. Temussi, *Unbiased Cold Denaturation: Low- and High-Temperature Unfolding of Yeast Frataxin under Physiological Conditions*, *J. Am. Chem. Soc.* **129**, 5374 (2007).
- [52] J. Roche, J. A. Caro, D. R. Norberto, P. Barthe, C. Roumestand, J. L. Schlessman, A. E. Garcia, B. Garcia-Moreno E., and C. A. Royer, *Cavities Determine the Pressure Unfolding of Proteins*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6945 (2012).
- [53] With the adopted units, 1 atm correspond to  $P \approx 0.7 \times 10^{-4}$  IU. For such values of  $P$ , the liquid-gas transition occurs at  $T_{LG} \approx 0.44$  IU. Therefore, the ambient temperature is  $T = 0.8 \times T_{LG} \approx 0.35$  IU.
- [54] W. B. Floriano, M. A. Nascimento, G. B. Domont, and W. A. Goddard, *Effects of Pressure on the Structure of Metmyoglobin: Molecular Dynamics Predictions for Pressure Unfolding through a Molten Globule Intermediate*, *Protein Sci.* **7**, 2301 (1998).
- [55] M. Gross and R. Jaenicke, *Proteins under Pressure. The Influence of High Hydrostatic Pressure on Structure, Function and Assembly of Proteins and Protein Complexes*, *Eur. J. Biochem.* **221**, 617 (1994).
- [56] E. Larios and M. Gruebele, *Protein Stability at Negative Pressure*, *Methods* **52**, 51 (2010).
- [57] J. Hollien and S. Marqusee, *A Thermodynamic Comparison of Mesophilic and Thermophilic Ribonucleases H*, *Biochemistry* **38**, 3831 (1999).
- [58] S. Fukuchi and K. Nishikawa, *Protein Surface Amino Acid Compositions Distinctively Differ Between Thermophilic and Mesophilic Bacteria*, *J. Mol. Biol.* **309**, 835 (2001).
- [59] We analysed the PDB files of the following ice-binding proteins: 1MSI, 1AME, 1B71, 1B7J, 1B7K, 1C3Z, 1C8A, 1EKL, 1EZG, 1HG7, 1JAB, 1KDF, 1OPS, 1WFB, 1XUZ, 2AME, 2BRD, 2MSI, 2PY2, 2ZDR, 2ZIB, 3P4G, 3ULT, 3UYV, 3WP9, 4DT5, 4KDV, 4NU3, 4UR4, and 4UR6. We assume that an amino acid is exposed to the solvent if the average accessible surface area is  $\geq 0.3$  of the total accessible area of the amino acid (calculations are performed with the MDTRAJ program at <http://mdtraj.org/1.7.2>). Protein structures are taken from the PDB database at <http://www.rcsb.org/pdb/home/home.do>.
- [60] G. Vogt, S. Woell, and P. Argos, *Protein Thermal Stability, Hydrogen Bonds, and Ion Pairs*, *J. Mol. Biol.* **269**, 631 (1997).
- [61] G. Vogt and P. Argos, *Protein Thermal Stability: Hydrogen Bonds or Internal Packing?*, *Folding Des.* **2**, S40 (1997).
- [62] V. Z. Spassov, A. D. Karshikof, and R. Ladenstein, *The Optimization of Protein-Solvent Interactions: Thermostability and the Role of Hydrophobic and Electrostatic Interactions*, *Protein Sci.* **4**, 1516 (1995).
- [63] P. Haney, J. Konisky, K. K. Koretke, Z. Luthey-Schulten, and P. G. Wolynes, *Structural Basis for Thermostability and Identification of Potential Active Site Residues for Adenylate Kinases from the Archaeal Genus Methanococcus*, *Proteins* **28**, 117 (1997).
- [64] E. van Dijk, P. Varilly, T. P. J. Knowles, D. Frenkel, and S. Abeln, *Consistent Treatment of Hydrophobicity in Protein Lattice Models Accounts for Cold Denaturation*, *Phys. Rev. Lett.* **116**, 078101 (2016).
- [65] M. S. Moghaddam and H. S. Chan, *Pressure and Temperature Dependence of Hydrophobic Hydration: Volumetric, Compressibility, and Thermodynamic Signatures*, *J. Chem. Phys.* **126**, 114507 (2007).
- [66] L. Lins, A. Thomas, and R. Brasseur, *Analysis of Accessible Surface of Residues in Proteins*, *Protein Sci.* **12**, 1406 (2003).
- [67] S. Moelbert, E. Emberly, and C. Tang, *Correlation Between Sequence Hydrophobicity and Surface-Exposure Pattern of Database Proteins*, *Protein Sci.* **13**, 752 (2004).

- [68] The sequences designed with  $\mathcal{H}_{RR}^{\min}$  are independent of  $T_d$  and  $P_d$ .
- [69] The condition for a vanishing number of HBs is  $P \approx J/v_{\text{HB}}$  (where  $J$  and  $v_{\text{HB}}$  are parameters of the model described in the Methods section), which in Ref. [70] gives  $P \geq 1$  IU, while here it gives  $P \geq 0.6$  IU  $\approx 1$  GPa when converted to real units.
- [70] G. Franzese and H. E. Stanley, *The Widom Line of Supercooled Water*, *J. Phys. Condens. Matter* **19**, 205126 (2007).
- [71] Details of our results reveal that the protein interface affects up to the second hydration shell and that this contributes significantly to the enthalpy difference between the  $u$  and the  $f$  states.
- [72] Depending on the specific set of residue-residue interactions, the exact boundary can move, but the fact that both in experiments and in our simulations the high pressure boundary is  $\approx 1$  GPa implies that the values we are using to represent the interactions between the amino acids have the correct order of magnitude.
- [73] G. Caldarelli and P. De Los Rios, *Cold and Warm Denaturation of Proteins*, *J. Biol. Phys.* **27**, 229 (2001).
- [74] S. Matysiak, P. G. Debenedetti, and P. J. Rossky *Role of Hydrophobic Hydration in Protein Stability: A 3D Water-Explicit Protein Model Exhibiting Cold and Heat Denaturation*, *J. Phys. Chem. B* **116**, 8095 (2012).
- [75] K. F. Lau and K. A. Dill, *A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins*, *Macromolecules* **22**, 3986 (1989).
- [76] K. Yue and K. A. Dill, *Inverse Protein Folding Problem: Designing Polymer Sequences*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4163 (1992).
- [77] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, *Computational Characterization of the Sequence Landscape in Simple Protein Alphabets*, *Proteins* **62**, 232 (2006).
- [78] B. A. Patel, P. G. Debenedetti, F. H. Stillinger, and P. J. Rossky, *The Effect of Sequence on the Conformational Stability of a Model Heteropolymer in Explicit Water*, *J. Chem. Phys.* **128**, 175102 (2008).
- [79] F. de los Santos and G. Franzese, *Understanding Diffusion and Density Anomaly in a Coarse-Grained Model for Water Confined Between Hydrophobic Walls*, *J. Phys. Chem. B* **115**, 14311 (2011).
- [80] Here, for the sake of simplicity, following Ref. [9], we assume that, in each volume  $v$ , there is at most one water molecule. This approximation could be removed by generalizing the model in such a way that the water variables associated with each cell could account for the state of more than one water molecule at a time.
- [81] S. Miyazawa and R. L. Jernigan, *Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-chemical Approximation*, *Macromolecules* **18**, 534 (1985).
- [82] J. Kyte and R. F. Doolittle, *A Simple Method for Displaying the Hydrophobic Character of a Protein*, *J. Mol. Biol.* **157**, 105 (1982).
- [83] L. H. de la Peña and P. G. Kusalik, *Temperature Dependence of Quantum Effects in Liquid Water*, *J. Am. Chem. Soc.* **127**, 5246 (2005).
- [84] U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, *Interaction Energies of Large Clusters from Many-Body Expansion*, *J. Chem. Phys.* **135**, 224102 (2011).
- [85] A. K. Soper and M. A. Ricci, *Structures of High-Density and Low-Density Water*, *Phys. Rev. Lett.* **84**, 2881 (2000).
- [86] N. Giovambattista, P. J. Rossky, and P. G. Debenedetti, *Effect of Pressure on the Phase Behavior and Structure of Water Confined Between Nanoscale Hydrophobic and Hydrophilic Plates*, *Phys. Rev. E* **73**, 041604 (2006).
- [87] C. Petersen, K.-J. Tielrooij, and H. J. Bakker, *Strong Temperature Dependence of Water Reorientation in Hydrophobic Hydration Shells*, *J. Chem. Phys.* **130**, 214511 (2009).
- [88] Y. I. Tarasevich, *State and Structure of Water in Vicinity of Hydrophobic Surfaces*, *Colloid J.* **73**, 257 (2011).
- [89] J. G. Davis, K. P. Gierszal, P. Wang, and D. Ben-Amotz, *Water Structural Transformation at Molecular Hydrophobic Interfaces*, *Nature (London)* **491**, 582 (2012).
- [90] S. Sarupria and S. Garde, *Quantifying Water Density Fluctuations and Compressibility of Hydration Shells of Hydrophobic Solutes and Proteins*, *Phys. Rev. Lett.* **103**, 037803 (2009).
- [91] K. A. Dill, T. M. Truskett, V. Vlachy, and B. Hribar-Lee B, *Modeling Water, the Hydrophobic Effect, and the Ion Solvation*, *Annu. Rev. Biophys. Biomol. Struct.* **34**, 173 (2005).
- [92] A. Badasyan, Y. S. Mamasakhlisov, R. Podgornik, and V. A. Parsegian, *Solvent Effects in the Helix-Coil Transition Model Can Explain the Unusual Biophysics of Intrinsically Disordered Proteins*, *J. Chem. Phys.* **143**, 014102 (2015).
- [93] C. J. Fennell, C. W. Kehoe, and K. A. Dill, *Modeling Aqueous Solvation with Semi-Explicit Assembly*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3234 (2011).
- [94] C. Camilloni, D. Bonetti, A. Morrone, R. Giri, C. M. Dobson, M. Brunori, S. Gianni, and M. Vendruscolo, *Towards a Structural Biology of the Hydrophobic Effect in Protein Folding*, *Sci. Rep.* **6**, 28285 (2016).
- [95] P. Das and S. Matysiak, *Direct Characterization of Hydrophobic Hydration During Cold and Pressure Denaturation*, *J. Phys. Chem. B* **116**, 5342 (2012).
- [96] T. Ghosh, A. E. García, and S. Garde, *Molecular Dynamics Simulations of Pressure Effects on Hydrophobic Interactions*, *J. Am. Chem. Soc.* **123**, 10997 (2001).
- [97] C. L. Dias and H. S. Chan, *Pressure-Dependent Properties of Elementary Hydrophobic Interactions: Ramifications for Activation Properties of Protein Folding*, *J. Phys. Chem. B* **118**, 7488 (2014).
- [98] This approximation implies that our calculations are valid only for  $P < 1/k_1$ .
- [99] M. Kurnik, L. Hedberg, J. Danielsson, and M. Oliveberg, *Folding without Charges*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5705 (2012).
- [100] P. De Los Rios and G. Caldarelli, *Putting Proteins Back into Water*, *Phys. Rev. E* **62**, 8449 (2000).
- [101] G. Hummer, S. Garde, A. E. García, M. E. Paulaitis, and L. R. Pratt, *The Pressure Dependence of Hydrophobic Interactions Is Consistent with the Observed Pressure*

- Denaturation of Proteins*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1552 (1998).
- [102] F. Meersman, C. M. Dobson, and K. Heremans, *Protein Unfolding, Amyloid Fibril Formation and Configurational Energy Landscapes under High Pressure Conditions*, *Chem. Soc. Rev.* **35**, 908 (2006).
- [103] N. V. Nucci, B. Fuglestad, E. A. Athanasoula, and A. J. Wand, *Role of Cavities and Hydration in the Pressure Unfolding of T4 Lysozyme*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13846 (2014).
- [104] V. Bianco, N. P. Gelabert, I. Coluzza, and G. Franzese, *How the Stability of a Folded Protein Depends on Interfacial Water Properties and Residue-Residue Interactions*, [arXiv:1704.03370](https://arxiv.org/abs/1704.03370).
- [105] L. E. Coronas, V. Bianco, A. Zantop, and G. Franzese, *Liquid-Liquid Critical Point in 3D Many-Body Water Model*, [arXiv:1610.00419](https://arxiv.org/abs/1610.00419).
- [106] D. Frenkel and B. Smit, *Understand Molecular Simulations* (Academic Press, San Diego, London, 2002).