How the stability of a folded protein depends on interfacial water properties and residue-residue interactions

Valentino Bianco\textsuperscript{a,\*}, Neus Pagès-Gelabert\textsuperscript{b}, Ivan Coluzza\textsuperscript{a}, Giancarlo Franzese\textsuperscript{b,c,\*}

\textsuperscript{a} Universität Wien, Sensengasse 8/10, Vienna 1090, Austria
\textsuperscript{b} Secció de Física Estadística i Interdisciplinària–Departament de Física de la Matèria Condensada, Facultat de Física Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain
\textsuperscript{c} Institute of Nanoscience and Nanotechnology (IN2UB), Universitat de Barcelona, Av. Joan XXIII S/N, Barcelona 08028, Spain

A R T I C L E   I N F O

Article history:
Received 11 April 2017
Received in revised form 4 August 2017
Accepted 6 August 2017
Available online 10 August 2017

Keywords:
Protein folding
Cold denaturation
Pressure denaturation
Interfacial water
Explicit water
Biological physics

A B S T R A C T

Proteins tend to adopt a single or a reduced ensemble of configurations at natural conditions \cite{1}, but changes in temperature $T$ and pressure $P$ induce their unfolding. Therefore for each protein there is a stability region (SR) in the $T$–$P$ thermodynamic plane outside which the biomolecule is denatured. It is known that the extension and shape of the SR depend on i) the specific protein residue-residue interactions in the native state of the amino acids sequence and ii) the water properties at the hydration interface. Here we analyze by Monte Carlo simulations the different coarse-grained protein models in explicit water how changes in i) and ii) affect the SR. We show that the solvent properties ii) are essential to rationalize the SR shape at low $T$ and high $P$ and that our findings are robust with respect to parameter changes and with respect to different protein models, representative of the ordered and disordered proteins. These results can help in developing new strategies for the design of novel synthetic biopolymers.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The capability of the single components to independently organize in pattern and structures without an external action fulfills a crucial role in the supramolecular organization and assembling of the biological matter \cite{2,3}. To cite some examples, self-assembly is observed in bio–molecules \cite{4}, in DNA and chromosomes \cite{5–8}, in lipid membranes \cite{9,10}, in the cytoskeleton \cite{11}, in cells and tissues \cite{12,13}, in virus and bacteria \cite{14,15}, and in proteins \cite{16,17}. In particular, the protein folding represents one of the most challenging and elusive biochemical processes where a chain of amino acids organizes itself into a unique native and folded structure \cite{18,19}. The protein folding is a spontaneous process driven by intra-molecular (residue-residue) van der Walls interactions and hydrogen bonds which overcome the conformational entropy. It depends also on the presence of co-factors as the chaperones \cite{20} and, in particular, the properties of the solvent, i.e. water \cite{21}, and the co-solutes \cite{22} that regulate the pH level and the salt concentration, for example.

Although water has no influence on the primary structure (the protein sequence), it affects the protein in all the other level of organization \cite{23–25}. Indeed, i) water forms H-bonds with the polar/charged residues of the side chains, influencing the adoption of secondary structures like alpha helices or beta sheets which expose the most hydrophilic residues to water; ii) the hydrophobic effect drives the collapse of the protein core and stabilizes the tertiary protein structure; iii) water induces the aggregation of proteins since they usually present hydrophobic regions on their surface (quaternary structure).

The stability of a protein, i.e. its capability to keep the folded conformation, is usually reduced by factors which destabilize H-bonding and other forces that contribute to secondary and tertiary protein structure, as, for example, crowding effects and variations of pH or ionic strength. In particular, experiments have clearly documented that proteins maintain their native structure in a limited range of temperatures $T$ and pressures $P$ \cite{26–41} showing an elliptic-like stability region (SR) in the $T$–$P$ plane, as accounted by a Hawley’s theory \cite{42}. Outside its SR a protein unfolds, with a consequent loss of its tertiary structure and functionality.

At high $T$ the protein unfolding is due to the thermal fluctuations which disrupt the protein structure. Open protein conformations increases the entropy $S$ minimizing the global Gibbs free energy $G = H − TS$, where $H$ is the total enthalpy. Upon cooling, if the nucleation of water is avoided, some proteins cold–denaturate \cite{27,29,34,36,43–46}. Usually such a phenomenon is observed below the melting line of water, although in some cases cold denaturation occurs above the
0 C, as in the case of the yeast frataxin [36]. Protein denaturation is observed, or predicted, also upon pressurization [26,28,35,41,47]. A possible explanation of the high-P unfolding is the loss of internal cavities, sometimes presents in the folded states of proteins [48]. Denaturation at negative P has been experimentally observed [49] and simulated recently [21,49,50]. Pressure denaturation is usually observed for 100 MPa \( \leq P \leq 600 \) MPa, and rarely at higher P unless the tertiary structure is engineered with stronger covalent bonds [33]. Cold- and P-denaturation of proteins have been related to the equilibrium properties of the hydration water [21,51-60]. However, the interpretations of the mechanism is still largely debated [47,48,61-71].

Here we investigate by Monte Carlo simulations of different coarse-grained protein models in explicit water how the SR is affected by changes in i) the specific protein-residue-residue interactions in the native state of the amino acids sequence and ii) the solvent properties at the hydration interface, focusing on water energy and density fluctuations. In particular, after introducing the model and the numerical method in Section 2, we study in a broad range of P and T how the conformational space of proteins depends on the model’s parameters for the hydration water in Section 3.1 and how it depends on the residue-residue interactions in Section 3.2. Next, we discuss the possible relevance of our results in the framework of protein design in Section 4 and, finally, we present our concluding remarks in Section 5.

2. Models and methods

The extensive exploration with atomistic models of protein conformations in explicit solvent at different thermodynamic conditions, including extreme low T and high P, is a very demanding analysis. To overcome this limitation, we adopt a coarse-grain model for protein-water interaction based on A) the many-body water model [21,60,67,72-80], combined with B) a lattice representation of the protein.

The many-body water model has been proven to reproduce—in at least qualitative way—the thermodynamic [72,80] and dynamic [78] behavior of water, the properties of water in confinement [73,74,77,79] and at the inorganic interfaces [67]. Its recent combination with the lattice representation of the protein has given a novel insight into the water-protein interplay [21,60,75,76,81].

As we will describe later, for the protein we consider a model that, in its general formulation as polar protein, follows the so-called “Go-models”, a common approach in protein folding. In their seminal paper Go and Taketomi [82] employed non-transferable potentials tailored to the native structure. The interactions were designed to have a sharp minimum only at the native residue-residue distance, guaranteeing that the energy minimum is reached only by the native structure. The Go-proteins thus successfully fold, and have a smooth free-energy landscape with a single global minimum in the native structure [83]. Hence, Go-models are equivalent to having an infinite variety of pair interactions among the residues (alphabet \( \mathcal{A} \)), such that each amino acid interacts selectively with a subset of residues defined by the distances in the native configuration. If the size of the alphabet is reduced, the construction of folding proteins requires an optimization step of the amino acid sequence along the chain [63,84,85]; for this reason these methods are often referred to as “protein design”. Comparing designed proteins with Go-proteins, Coluzza recently showed that, close to the folded state, Go and designed proteins behave in a very similar manner [86]. Since we are interested in measuring the stability regions defined by the environmental condition at which the trial protein is at least 90% folded, Go-models are an appropriate protein representation, and, at this stage, we do not require to perform the laborious work of protein design to obtain general results. We will discuss later the possibility to extend our model to the case of a limited alphabet \( \mathcal{A} \) of residues (20 amino acids).

2.1. The bulk many-body water model

We consider the coarse-grain many-body bulk water at constant P, constant T and constant number \( N^{(b)} \) of water molecules, while the total volume \( V^{(b)} \) occupied by water is a function of P and T. Because in the following we will consider the model with water at the hydration protein interface and (bulk) water away from the interface, for the sake of clarity here we introduce the notation with a superscript \( (b) \) for quantities that refer to the bulk.

We replace the coordinates and orientations of the water molecules by a continuous density field and discrete bonding variables, respectively. The density field is defined based on a partition of the available volume \( V^{(b)} \) into a fixed number \( N_0 = N^{(b)} \) of cells, each with volume \( \nu^{(b)} \equiv \nu^{(b)}/N^{(b)} \geq \nu_0 \), where \( \nu_0 \equiv \nu_i^0 \) is the water excluded volume with \( \nu = 9 \AA \) (water van der Waals diameter). For the sake of simplicity we assume that, when the water molecules are not forming hydrogen bonds (HBs), the (dimensionless) density is homogeneous in each cell and equal to \( \rho^{(b)} \equiv \nu_0/\nu^{(b)} \). As we will discuss later, the density is, instead, locally inhomogeneous when water molecules form HBs. Specifically, the density depends on the number of HBs, therefore \( \rho^{(b)} \) only represents the average bulk density.

The Hamiltonian of the bulk water is

\[
\mathcal{H}^{(b)} = \sum_q U(r_{ij}) - J^N_{\text{HB}} - J_0 N_{\text{HB}}^{(b)}.
\]

The first term represents the isotropic part of the water-water interaction and accounts for the van der Waals interaction [87]. It is modeled with a Lennard-Jones potential

\[
\sum_q U(r_{ij}) \equiv 4\epsilon \sum_q \left[ \left( \frac{r_{ij}^6}{r_{ij}^6} \right) - \frac{1}{2} \left( \frac{r_{ij}^6}{r_{ij}^6} \right)^{12} \right]
\]

where \( \epsilon \equiv 5.8 \) kJ/mol and the sum runs over all the water molecules i and j at O–O distance \( r_{ij} \) calculated as the distance between the centers of the two cells i and j where the molecules belong. We assume a hard-core exclusion \( U(r) \equiv \infty \) for \( r < r_0 \) and a cutoff for \( r > r_c \equiv 6r_0 \).

The second term in Eq. (1) represents the directional (covalent) component of the HB, where

\[
N_{\text{HB}}^{(b)} \equiv \sum_{(ii)} n_i n_j \delta_{\sigma_i \sigma_j}
\]

is the number of bulk HBs and the sum runs over neighbor cells occupied by water molecules. Here we introduce the label \( n_i = 1 \) if the cell i has a water density \( \rho^{(b)} > 0.5 \) and \( n_i = 0 \) otherwise. In the homogeneous bulk this condition guarantees that two water molecules can form a HB only if their relative distance is \( r < 2^{1/3}r_0 \equiv 3.66 \AA \), corresponding to the range of a water’s first coordination shell as determined from the O-O radial distribution function from 220 to 673 K and at pressures up to 400 MPa [88].

The variable \( \sigma_q \equiv 1, \ldots, q \) in Eq. (3) is the bonding index of the water molecule in cell i with respect to the neighbor molecule in cell j and \( \delta_{\sigma_i \sigma_j} = 1 \) if \( a = b \), or 0 otherwise, is a Kronecker delta function. Each water molecule has as many bonding variables as neighbor cells, but can form only up to four HBs. Therefore, if the molecule has more than four neighbors, e.g., in a cubic lattice partition of \( V^{(b)} \), an additional condition must be applied to limit to four the HBs participated by each molecule.

The parameter \( q \) in the definition of \( \sigma_q \) is determined by the entropy decrease associated to the formation of each HB. Each HB is unbroken if the hydrogen atom H is in a range of \([-30, 30]\) with respect to the O–O axes [89]. Hence, only 1/6 of the entire range of values \([0, 360]\) for the O–H angle is associated to a bonded state. Therefore, in the zero-order approximation of considering each
HB independent, a molecule that has \(4 - n\) HBs, with \(n = 1, \ldots, 4\), has an orientational entropy that is \(S^o_{\text{HB}} \equiv n \ln 6\) above that of a fully bonded molecule with \(S^o_{\text{HB}} = 0\), where \(k_B\) is the Boltzmann constant. As a consequence, the choice \(q = 6\) accounts correctly for the entropy variation due to HB formation and breaking given the standard definition of HB.

The third term in Eq. (1) is associated to the cooperativity of the HBs due to the quantum many-body interactions [72,90]. Indeed, the formation of a new HB affects the electron probability distribution around the molecule favoring the formation of the following HB in a local tetrahedral structure [91]. We assume that the energy gain due to this effect is proportional to the number of cooperative HBs in the system

\[
N_{\text{coop}} \equiv \sum_i n_i \sum_{(l,k)} \delta_{\tau_l \tau_k},
\]

where \(n_i\) assures that we include this term only for liquid water. With this definition and with the choice \(J_0/4\epsilon \ll J\) the term mimics a many-body interactions among the HBs participated by the same molecule. Indeed, the condition \(J_0/4\epsilon \ll J\) guarantees that the interaction takes place only when the water molecule \(i\) is forming several HBs. The inner sum is over \((l,k)\), indicating each of the six different pairs of the four indices \(\tau_l\) of the molecule \(i\).

The formation of HBs leads to an open network of molecules, giving rise to a lower density state. We include this effect into the model assuming that for each HB the volume \(V^{(b)}\) increases of \(\delta V_{\text{HB}}/V_0 = 0.5\). This value is the average volume increase between high-density ices VI and VIII and low-density (tetrahedral) ice I. As a consequence, the average bulk density is

\[
\rho^{(b)} = \frac{N V_0}{V^{(b)} + N V_{\text{HB}}^{(b)}},
\]

We assume that the HBs do not affect the distance \(r\) between first neighbor molecules, consistent with experiments [91]. Hence, the water-water distances \(r\) is calculated only from \(V^{(b)}\).

As discussed in Ref. [21] a good choice for the parameters that accounts for the ions in a protein solution is \(\epsilon = 5.8\) kJ/mol, \(J/4\epsilon = 0.3\) and \(J_0/4\epsilon = 0.05\) that give an average HB energy \(\sim 20\) kJ/mol. In the following we consider two protein models, a simpler one used to understand the molecular mechanisms through which water contributes to the unfolding, and a more detailed model which includes the effect of polarization. For the sake of simplicity, we present here the result for a system in two dimension. Preliminary results for the model in three dimensions of both bulk water [80] and protein folding show results that are qualitatively similar to those presented here.

### 2.2. Hydrophobic protein model

The protein is modeled as a self-avoiding lattice polymer, embedded into the cell partition of the system. Despite its simplicity, lattice protein models are still widely used in the contest of protein folding [21,52,53,59,71,92-94] because of their versatility and the possibility to develop coarse-grained theories and simulations for them. Each protein residue (polymer bead) occupies one cell. In the present study, we do not consider the presence of cavities into the protein structure.

To simplify the discussion in this first part of the work, we assume that (i) there is no residue-residue interaction, (ii) the residue-water interaction vanishes, unless otherwise specified and (iii) all the residues are hydrophobic. This implies that the protein has multiple ground states, all with the same maximum number \(n_{\text{max}}\) of residue-residue contacts. As shown by Bianco and Franzese [21], the results hold also when the hypothesis (i), (ii) and (iii) are released, as we will discuss in the following.

Our stating hypothesis is that the protein interface affects the water-water properties in the hydration shell, here defined as the layer of first neighbor water molecules in contact with the protein (Fig. 1). There are many numerical and experimental evidences supporting this hypothesis. In particular, it has been shown that the water-water HBs in the protein hydration shell are more stable and more correlated with respect to the bulk HBs by using theoretical calculations [95], experiments [96,98,99] and atomistic simulations [97,99,100]. We account for this by replacing \(J_0\) of Eq. (1) with \(J_0 > J\) for the water-water HBs at the hydrophobic (\(\Phi\)) interface. Another possibility, discussed later, would be to consider that the cooperative interaction \(J_{\Phi,\Phi}\) at the \(\Phi\)-interface, directly related to the tetrahedral order of the water molecules, is stronger with respect to the bulk. This case would be consistent with the assumption that water forms ice-like cages around \(\Phi\)-residues [101]. Both choices, according to Muller discussion [102], would ensure the water enthalpy compensation during the cold-denaturation [60].

At the \(\Phi\)-interface, besides the stronger/stabler water-water HB, we consider also the larger density fluctuations with respect to the bulk. These larger densities fluctuations have been observed in extensive coarse-grained molecular dynamics simulations including explicit solvation [65,103] and extensive atomistic molecular simulations [97] of hydrated \(\Phi\)-solutes.

Although it is still a matter of debate if, at ambient conditions, the average density of water at the \(\Phi\)-interface is larger or smaller with respect to the average bulk water density [104–108], there are evidences showing that such density fluctuations reduce upon pressurization [65,97,109,110]. We include this effect in the model by assuming that the volume change \(V^{(\Phi)}_{\text{HB}}\) associated to the HB formation in the \(\Phi\) hydration shell can be expanded as a series function of \(P\)

\[
\frac{V^{(\Phi)}_{\text{HB}}}{V^{(\Phi)}_{\text{HB},0}} \equiv 1 - k_1 P - k_2 P^2 - k_3 P^3 + O(P^4)
\]

where \(V^{(\Phi)}_{\text{HB},0}\) is the value of the change when \(P = 0\). Here the coefficients \(k_1, k_2, k_3\) are such that \(\partial V^{(\Phi)}_{\text{HB}}/\partial P\) is always negative. As first approximation, we study the linear case, with \(k_i = 0\ \forall i > 1\). We discuss later how the protein stability is affected by considering...
the quadratic terms in Eq. (6). Our initial choice implies that we can study the system only when $P < 1/k_1$. As we will discuss in the next section, this condition does not limit the validity of our results. The total volume $V$ of the system is, therefore,

$$V = Nv_0 + N_{\text{HB}}^{(b)}v_{\text{HB}}^{(b)} + N_{\text{HB}}^{(o)}v_{\text{HB}}^{(o)},$$

(7)

where $N_{\text{HB}}^{(o)}$ is the number of HBs in the Φ shell.

2.3. Polar protein model

In order to account for the effect of the hydrophilic residues on the water-water hydrogen bonding in the hydration shell, we consider also the case in which the protein is modeled as a heteropolymer composed of hydrophobic (Φ) and hydrophilic (ζ) residues. In this case it is worth introducing residue-residue interactions that lead to a specific folded (native) state for the protein.

We fix the native state by defining the interaction matrix $A_{ij} = \epsilon_i \delta$ if residues $i$ and $j$ are n.n. in the native state, 0 otherwise. To simplify our model we set all the residues in contact with water in the native state as hydrophilic, and all those buried into the protein core as hydrophobic. The water interaction with Φ- and ζ-residues is given by the parameters $\epsilon_{w,\Phi}$ and $\epsilon_{w,\zeta}$ respectively, where we assume $\epsilon_{w,\Phi} < J$ and $\epsilon_{w,\zeta} > J$.

The polar ζ residues interfere with the formation of HB of the surrounding molecules, disrupting the tetrahedral order and distorting the HB network. Thus we assume that each ζ residue has a preasigned bonding state $q_\zeta = 1, \ldots, q$, different and random for each ζ residue. In this way, a water molecule $i$ can form a HB with a ζ residue, located in the direction $j$, only if $\epsilon_{ij} = q_\zeta$.

In the polar protein model, the formation of water-water HBs in the hydration shell is described by the parameters $J_{\Phi}$ and $J_{\Phi,\Phi}$ (directional and cooperative components of the HB) if both molecules hydrate two Φ-residues; $J_{\Phi}$ and $J_{\Phi,\zeta}$ if both molecules hydrate two ζ-residues; $J_{\Phi,\zeta} = (J_{\Phi} + J_{\zeta})/2$ if $J_{\Phi,\Phi}$ and $J_{\Phi,\zeta}$ if the two water molecules are in contact one with a Φ-residue and another with a ζ-residue, forming a Φ-ζ-interface. Accordingly, the volume associated to the formation of HB in the hydration shell is $v_{\text{HB}}^{(b)}$, $v_{\text{HB}}^{(o)}$ and $v_{\text{HB}}^{(c)}$. Then, we assume that $v_{\text{HB}}^{(b)}$ changes with $P$ following the Eq. (6). Due to the condition $\epsilon_{w,\zeta} > J$, we assume that the density fluctuations near a ζ-residue are comparable, or smaller, than those in bulk water, therefore we set $v_{\text{HB}}^{(o)} = v_{\text{HB}}^{(b)}$. Finally, we define $v_{\text{HB}}^{(c)} = (v_{\text{HB}}^{(b)} + v_{\text{HB}}^{(o)})/2$.

2.4. Simulations' details

We study proteins of 30 residues with Monte Carlo simulations in the isobaric-isothermal ensemble, i.e., constant $P$, constant $T$ and constant number of particles. Along the simulation we calculate the average number of residue-residue contacts to estimate the protein compactness, sampling $\sim 10^5$ independent protein conformations for each thermodynamic state point. For the hydrophobic protein model, we assume that the protein is folded if the average number of residue-residue contacts is $n_{rr} \geq 50\% n_{\text{max}}$. While for the polar protein model, having a unique folded state, we fix the threshold at $n_{rr} \geq 90\% n_{\text{max}}$.

For the sake of simplicity, we consider our model in two dimensions. Although this geometry could appear as not relevant for experimental cases, our preliminary results for the three dimensional system show no qualitative difference with the case presented here. We understand this finding as a consequence of the peculiar property of bulk water of having, on average, not more than four neighbors. This coordination number is preserved if we consider a square partition of a two dimensional system. Differences between the two dimensional and the three dimensional models could arise from the larger entropy in higher dimensions for the protein, however our preliminary results in 3D show that they can be accounted for by tuning the model parameters.

3. Results and discussion

3.1. Results for the hydrophobic protein model

Bianco and Franzese show [21] that the hydrophobic protein model, with parameters $k_1 = v_0/4\epsilon$ (and $k_2 = k_3 = 0$), $v_{\Phi}^{(b)}/v_0 = 0.5$ and $J_{\Phi}^{(b)}/\epsilon = 0.55$, $J_{\Phi,\Phi} = J_{\Phi}$, has a SR that is elliptic in the $T$–$P$ plane. This finding is consistent with the predictions of the Hawley theory [31,42] accounting for the thermal, cold and pressure denaturation [Fig. 2].

They find that at high $T$ the large entropy associated to open protein conformations keeps the protein unfolded. By isobaric decrease of $T$, the energy cost of an extended water-protein interface can no longer be balanced by the entropy gain of the unfolded protein, and the protein folds to minimize the number of hydrated Φ-residues, as expected.

By further decreasing of $T$ at constant $P$, the number of water-water HBs increases both in bulk and at the protein interface. At low enough $T$, the larger stability, i.e., larger energy gain, of the HBs at the Φ-interface drives the cold denaturation of the protein.

Upon isothermal increase of $P$, the enthalpy of the system increases for the increasing $PV$ term. Therefore, a mechanism that reduces $V$ would reduce the total enthalpy. Here the mechanism is provided by the water compressibility that is larger at the Φ-interface than in bulk. Therefore, the larger water density at the protein interface drives the unfolding, which leads to a larger Φ-interface and enthalpy gain.

Finally, when the system is under tension, i.e., at $P < 0$, the total enthalpy is minimized when $V$ in Eq. (7) is maximized. However, the increase of average separation between water molecules breaks the HBs. In particular, bulk HBs break more than those at the Φ-interface because the first are weaker than the latter. Hence, $N_{\Phi}^{(b)}$ vanishes when $N_{\Phi}^{(b)} > 0$. As a consequence, the maximization of $V$ is
achieved by maximizing $N_{\text{HB}}^{(0)}$, i.e., by exposing the maximum number of $\Phi$-residues, leading to the protein denaturation under tension.

Once it is clear that the model can reproduce the protein SR, allowing us to understand the driving mechanism for the denaturation at different thermodynamic conditions, it is insightful to study how the SR depends on the model parameters. Therefore, in the following of this work we show our new calculations about the effect of varying one by one the model parameters.

### 3.1.1. Varying the water-water HB directional component $J_b$ at the $\Phi$-interface

Changing the (covalent) strength $J_b$ of the interfacial HB has a drastic effect on the SR. As discussed above, having $J_b / J > 1$, as in the reference case, drives the cold unfolding as a consequence of the larger gain of HB energy near the $\Phi$-interface. Instead, by setting $J_b / J < 1$ (Fig. 3a) the folded protein becomes more stable at low $T$ than in the reference case, because there is a larger energy gain in forming as many bulk HB as possible, i.e., in reducing the number of those near $\Phi$-residues. Hence, there is a larger free-energy gain in reducing the exposed $\Phi$-interface with respect to the reference case.

As a matter of fact, with our choice $J_b / 4e = 0.20$, we find cold denaturation only for $P < 0$. This is a consequence of the fact that the free energy has a term with $N_{\text{HB}}^{(0)}$ multiplying $-J_b + P V_{\text{HB}}$, hence for $P < 0$ the free energy decreases if $N_{\text{HB}}^{(0)}$ increases, even for a vanishing $J_b$. The negative slope of the cold denaturation line at $P < 0$ (Fig. 3a for 70% curve) is because the larger the $|P|$, the larger is the term proportional to $N_{\text{HB}}^{(0)}$ in the free-energy balance.

Reducing $J_b$ makes the folded protein more stable also at high $T$, because the entropy term overcomes the energy term at $T$ lower than in the reference case. A similar observation holds also at high $P$, because a reduced $J_b$ implies a decrease in $N_{\text{HB}}^{(0)}$, hence a decrease in enthalpy gain associated to the exposure of the $\Phi$-interface.

On the other hand, the larger the $|P|$, the more negative is the quadratic $P$-dependent coefficient that, as mentioned above, multiplies $N_{\text{HB}}^{(0)}$ in the free energy, and the larger is the free-energy gain in exposing the $\Phi$-interface at high $T$. Hence, the hot-denaturation curve in the $P$-$T$ plane has a negative slope for $P > 0$ and a positive slope for $P < 0$. As a consequence, the ellipse describing the SR (Fig. 3a for 50% curve) becomes more elongated than in the reference case with a negatively-sloped major axis and an eccentricity that grows towards 1.

On the contrary, for increasing $J_b$, the SR is lost, due to the energetic gain associated to wetting the entire $\Phi$-interface of the protein (Fig. 3b). The $P$-dependence of the contour lines is the same as that discussed for the case with $J_b / J < 1$, hence they keep the shape but shrink.

### 3.1.2. Varying the water compressibility factor $k_1$ at the $\Phi$-interface

Decreasing the water compressibility factor $k_1$ leads to a stretching of the SR along the $P$ direction and a rotation of the ellipse axes in such a way that the main axis increases its negative slope in the $P$-$T$ plane (Fig. 4a). On the other hand, increasing $k_1$ results in a contraction of the SR along $P$ with a rotation of the main axis toward a zero slope in the $P$-$T$ plane (Fig. 4b).

These effects can be understood observing that the free energy of the system has a term $-k_1 P V_{\text{HB}}^{(0)} N_{\text{HB}}^{(0)}$. This term is associated to the fact that there is a larger water compressibility at the $\Phi$-interface, reducing the total free energy. Therefore, by decreasing $k_1$ the destabilizing effect of the increased water-compressibility is reduced and the protein gains stability in $P$ at constant $T$, while the opposite effect is achieved by increasing $k_1$. The observations about the slope of the contour lines discussed in the previous subsection apply also in this case explaining the rotation of the ellipses axes.

### 3.1.3. Varying the HB volume-increase $v_0$ at the $\Phi$-interface and $P = 0$

A decrease of $v_0^{(0)} / V_0$, with respect to the reference case, moves the SR at lower $P$, while an increase moves the SR at higher $P$ (Fig. 5). This effect can be understood observing that the free energy of the system has a term $P V_{\text{HB}}(0) N_{\text{HB}}^{(0)}$ that, at each $P$, implies a decreasing enthalpy cost for decreasing $v_0^{(0)}$ if $N_{\text{HB}}^{(0)}$ is kept constant. Hence, this term favors the unfolding at high $P$ when $v_0^{(0)}$ is small, decreasing the stability of the native state upon pressurization (Fig. 5a). The opposite occurs for increasing $v_0^{(0)}$ (Fig. 5b).

We also find that the slope of the main ellipse axis changes from positive, for small $v_{\text{HB}}^{(0)}$, to negative, for large $v_{\text{HB}}^{(0)}$. This is a consequence of the inversion of the contribution of the free-energy term $P V_{\text{HB}}^{(0)} N_{\text{HB}}^{(0)}$ when $P$ changes sign. Because a variation of $v_{\text{HB}}^{(0)}$, where the SR crosses the $P = 0$ axis, the stability contour-line changes shape as a consequence, resulting in an effective rotation of its elliptic main axis: the main axis is positive when the majority of the SR is at $P < 0$ (Fig. 5a) and is negative otherwise (Fig. 5b).

---

**Fig. 3.** Effect on the SR of changing the water-water HB directional component $J_b$ at the $\Phi$-interface. In both panels symbols with continuous lines delimit the regions with 30% (green), 40% (turquoise), 50% (red) and 70% (blue) of the protein folded. Dashed lines (with the same color code as for continuous lines) are for the reference system in Fig. 2 (Table 1) with $J_b / 4e = 0.55$. All lines are guide for eyes. (a) For $J_b / 4e = 0.20$, smaller than the reference value, the SR expands to lower $T$ and $P$ and to higher $T$ and $P$. (b) For $J_b / 4e = 0.75$, greater than the reference value, the SR shrinks.
3.1.4. Adding the quadratic $P$-dependence of $v_{\text{HB}}^{(0)}$ at the $\Phi$-interface

So far we have shown the SRs for the model with $v_{\text{HB}}^{(0)}$ linearly-dependent on $P$. This truncation of Eq. (6) implies that the model for $P < 1/k_1 \equiv P_1$ describes a system where water-water HBs at the $\Phi$-interface decrease the local density, as expected, while for larger $P$ they do the opposite. Thanks to our specific choice of parameters for the reference system, our truncation does not affect the results because for $P > P_1$ the HB probability, both in bulk and at the $\Phi$-interface, is vanishing.

However, to check how qualitatively robust are our results against this truncation of Eq. (6), we consider also the case with the quadratic $P$-dependence of $v_{\text{HB}}^{(0)}$, i.e.,

$$v_{\text{HB}}^{(0)}/v_{\text{HB}}^{(0)} = 1 - k_1 P - k_2 P^2,$$

where $k_2 > 0$ is a new parameter with units of $k_1/P$. With this new approximation of Eq. (6) results $P_1 \equiv (2/\alpha)/\sqrt{1 + \alpha - 1}$, with $\alpha \equiv 4k_2/k_1$. Therefore, $P_1$ decreases for increasing $x$.

We fix $k_1$ to the reference value, and vary $k_2$ (Fig. 6). We find that for increasing $k_2$, the SR is progressively compressed on the high-$P$ side, with minor effects on the SR $T$-range. Adding a cubic term in Eq. (6) affects the SR in a similar way (data not shown). The rational for this behavior lies in the enhanced enthalpic gain upon exposing the $\Phi$-residue to the solvent since $v_{\text{HB}}^{(0)}$ decreases faster upon approaching $P_1$ that, in turn, decreases for increasing $k_2$.

3.1.5. Adding an attractive interaction $\epsilon_{w,\Phi}$ between water and $\Phi$-residues

Here, we check how a non-zero water–hydrophobic residue interaction, $\epsilon_{w,\Phi} > 0$, would affect the SR of the hydrophobic homopolymer. Indeed, despite the common misunderstanding of “water-phobia” due to the oversimplified terminology, it is well known that a hydrophobic interface attracts water, but with an interaction that is smaller than a hydrophilic surface.

We find that by setting $\epsilon_{w,\Phi}/4\epsilon = 0.05$, smaller than bulk water–water attraction, the SR is reduced in $P$ and lightly shifted toward lower $T$ (Fig. 7). In fact, an attractive water–$\Phi$ interaction enhances the propensity of the polymer to expose the $\Phi$ residues to the solvent, resulting in a global reduction of the SR and destabilizing the folded protein.
3.1.6. Enhancing the cooperative interaction \( J_{\alpha,\beta} \) at the \( \Phi \)-interface

Lastly, in the contest of the hydrophobic protein model, we consider a different scenario. As discussed in the model description, the enthalpic gain upon cold denaturation would be consistent also with the assumption \( J_{\alpha,\beta} > J_{\alpha} \) associated to a larger cooperativity of the HBs at the \( \Phi \)-interface. Hence, to analyze this scenario, we compute the SR considering the directional component of the HB unaffected by the \( \Phi \)-interface \( J_{\Phi} = J \), while assuming an enhanced HB cooperativity at the \( \Phi \)-interface \( J_{\alpha,\beta} > J_{\alpha} \). Note that the increase of \( J_{\alpha,\beta} \), promotes the number of cooperative HBs at the \( \Phi \)-interface only once they are formed as isolated HBs (\( J_{\alpha,\beta} < J_{\alpha} \)). Our finding (Fig. 8) is consistent with a close SR, presenting cold- and pressure-denaturation.

Although not discussed here, we expect that varying the parameters \( k_1 \) and \( V_{w,0}^{(b)} \) with the current choice of \( J_{\alpha,\beta} > J_{\alpha} \) and \( J_{\Phi} = J \), would affect the SR similarly to the cases discussed in previous subsections.

3.2. Results for the polar protein model

Next we summarize the results for the polar protein model. As shown in Ref. [21], also in this case the SR recovers a close elliptic-like SR in the \( T-P \) plane (Fig. 9). In particular, despite that we reduce the value of \( J_{\Phi}/4k \) with respect to the hydrophobic protein model in Table 1, the additional residue-residue interaction \( \epsilon_{\alpha \Phi} \) and water–\( \zeta \)-residue interaction \( \epsilon_{\alpha \zeta} \) stabilize the folded state to higher \( P \) and \( T \), as can been seen by comparing Fig. 9 with Fig. 2.

### Table 1

| Parameters for the reference system of the hydrophobic protein model (Fig. 2) with which we compare the results after varying the constants at the \( \Phi \)-interface one by one. We fix \( \nu_0 = \nu_{\Phi} = v_0 = 24.4 \text{ Å}^3 \) and \( \epsilon = 5.8 \text{ kJ/mol} \). |
|---|---|---|---|---|---|---|
| \( \psi_{v_{\Phi}}/v_0 \) | \( J/4k \) | \( J_{\alpha}/4k \) | \( \psi_{v_{\Phi}}/v_0 \) | \( J_{\Phi}/4k \) | \( J_{\alpha}/4k \) | \( k_1(4k)/v_0 \) | \( k_2 = k_3 \) |
| 0.5 | 0.3 | 0.05 | 0.5 | 0.55 | 0.05 | 1 | 0 |
4. Perspective on the protein design

As we mentioned in the previous sections, the hydrated protein models discussed here simplify the dependence of the stability against unfolding on the protein sequence. In fact, in the homopolymer protein model, the sequence is reduced to a single amino acid, hence we have the alphabet $A = 1$, while in the polar protein model the alphabet size coincides, by construction, with the protein length $l$, $A = l$, because the interaction matrix has $(l^2 - l)/2$ different elements that depend on the native state configuration.

In a more realistic case we would deal with proteins composed, at most, by 20 different amino acids, irrespective of the protein length. The amino acids assemble in a linear chain, which defines the protein sequence, in such a way that the protein is capable to fold into a unique native structure. Usually, among the huge amount of possible sequences, only few are good folders for a given native structure, smoothing and funneling the free energy landscape in order to lead the open protein conformation towards the native one.

Protein design strategies allow us to identify good folding sequences for each native conformation. Different methodologies have been proposed and studied in the past years [111–125] but water properties are not explicitly accounted, apart from a few cases [111,114–116,120,122] usually referred only to ambient conditions. Despite the fact that the evolution has selected natural protein sequences capable to fold and work in extreme thermodynamic conditions (like the anti-freeze proteins or the thermophilic proteins), all the design methods are not efficient in establishing which are the key elements to predict artificial sequences stable in thermodynamic conditions far from the ambient situation.

On this important aspect our model can give a relevant insight. Indeed, following the works of Shakhnovich and Gutin [84,126] on lattice proteins, we can easily introduce an interaction matrix between the 20 amino acids—like the Miyazawa Jernigan residue-residue interaction matrix $S$ [127]—and look for the protein sequences which minimize the energy of the native structure. This scheme can be improved to account for the water properties of the surrounding water, since the protein interface affects the water-water hydrogen bonding at least in the first hydration shell. In this way, we expect to find sequences with patterns depending on the $T$ and $P$ conditions of the surrounding water. Our preliminary results show that the protein sequences designed with our explicit-water model strongly depend on the thermodynamic conditions of the aqueous environment.

5. Conclusions

In this work we have presented a protein–water model to investigate the effect of the energy and density fluctuations at the hydrophobic interface ($\Phi$) of the protein. In particular, we have considered two protein models. In the first we simplify the discussion assuming that the protein is a hydrophobic homopolymer. In the second model we consider a more realistic case, assuming that the protein has a unique native state with a hydrophilic ($\Omega$) surface and a hydrophobic core and that the hydrophilic residues polarize the surrounding water molecules. These models can be considered as representative of the disordered proteins—where the collapsed protein state is not unique [1]—and of the ordered proteins, respectively.

In both cases, we model the hydrophobic effects considering that the water–water hydrogen bonding is at the $\Phi$-interface is stronger with respect to the bulk, and that the corresponding density fluctuations are reduced upon pressurization.

Our model qualitatively reproduces the melting, the cold– and the pressure–denaturation experimentally observed in proteins. The stability region, i.e. the $T$–$P$ region where the protein attains its native state, has an elliptic–like shape in the $T$–$P$ plane, as predicted by the theory [42].
We discuss in detail how each interaction affects the stability region, showing that our findings are robust with respect to model parameters changes. Aiming at summarizing our findings, although the parameter variations result in a non-trivial modification of the protein stability region, we observe that the strength of the interfacial water-water HB compared to the bulk ones, mainly affects the T-stability range of proteins, while the compressibility of the hydrophobic hydration shell mainly regulates the P-stability range. The scenario remains substantially unvaried by changing the protein model from the oversimplified hydrophobic homopolymer to the polar protein model. Our findings put water's density and energy fluctuations in a primary role in the tunable stability protein structure and pave the way for a water-dependent design of artificial proteins, with tunable stability.

Acknowledgments

V.B acknowledges the hospitality at the Universitat de Barcelona during his visits and the support from the Austrian Science Fund (FWF) project M 2150-N36. V.B. and I.C. acknowledge the support from the Austrian Science Fund (FWF) project P 26253-N27. G.F. acknowledges the support from the ICREA Academia prize, funded by the Spanish MINECO and the European Regional Development Fund (ERDF/FEDER). V.B., I.C. and G.F. acknowledge the support from the ICREA Academia prize, funded by the Spanish MINECO and the European Regional Development Fund (ERDF/FEDER).

References

[23] H. Lesch, H. Stadlbauer, J.M. Vanderkooi, Stability dia-