Protein unfolding and aggregation near an hydrophobic interface

Author: David March Pons
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Advisor: Giancarlo Franzese
(Dated: June 2020)

Abstract: The behaviour of proteins in a crowded environment is a relevant matter of study for biological and medical purposes. Previous results show that, when the protein concentration increases, the proteins unfold and, at higher concentration, aggregate. Here we study if the presence of a hydrophobic surface affects this sequence of events. To this goal, we simulate a coarse-grained model of a particular sequence of amino acids, with a known native structure, near an ideal hydrophobic surface in aqueous environment. We discuss how the protein folding and the aggregation depends on the temperature and the protein concentration. We find that the sequence of events of unfolding and aggregation is not affected by the hydrophobic interface for this specific sequence of amino acids. This work opens the way for further systematic studies on this topic, with possible relevant implications in biotechnology.

I. INTRODUCTION

Proteins are large biomolecules formed by hydrated chains of amino acid residues, which play an essential role in many biological phenomena taking place in the human body. Depending on changes in temperature T or pressure P a protein may change its conformation, going from its native folded state to a denatured unfolded state, losing its activity. Recent studies have shown that water takes a significant role in these processes, because, by varying T and P, water contributions to the enthalpy and the energy of the system rule the protein denaturation [1]. The effect is particularly evident in the design of artificial proteins [2]. Protein concentration is another important matter of study in the biology and medicine because it is related to protein aggregation, as, for example, in Alzheimer and Parkinson diseases [3], or in other biological activities [4]. Recent computational studies about the unfolding, stability and aggregation of proteins as a function of their concentration, show that proteins tend to fold to their native structure as long as a threshold concentration is not reached [5]. Above this threshold concentration, proteins unfold. Only at higher concentrations they aggregate. The study establishes three clearly identifiable states, folded, unfolded and aggregated, separated by concentration thresholds [5]. Here, adopting the model used in [5] we present a preliminary study on the protein folding/unfolding and aggregation near a hydrophobic interface. Preliminary results show that such an interface destabilizes the protein, moving toward lower T the region where the protein folds. Our goal is to understand if the interface affects the unfolding and aggregation for a particular sequence, and also if the protein adsorbs to the interface. These results could shed light on biological mechanisms at the nanoscale.

*Electronic address: dmarchp@gmail.com

II. MODEL

A. Franzese-Stanley water model

We adopt a coarse-grain representation of the water molecules, partitioning a volume V into a fixed number N of cells, each one with volume $v \equiv N/V \geq v_0$, being v_0 the water excluded volume. For sake of simplicity, we will consider here the simplified case of the projection into two dimensions (2D) of a water monolayer with height $h \simeq 0.5$ nm. Preliminary data show no qualitative differences for the model in 3D. We fix T and P of the system, leaving $r \equiv \sqrt{v/h}$ free to change, with $r \geq r_0$ average distance between first neighbour water molecules. In its general formulation, the model is able to describe all the fluid phases of water [6]. However, here we focus only on its liquid phase. In this case, the Hamiltoninan describing the interaction of the bulk water is

$$\mathcal{H} \equiv \sum_{ij} \mathcal{U}(r_{ij}) - JN_{HB}^{(b)} - J_\sigma N_{coop}^{(b)}. \qquad (1)$$

The first term, summed over all the molecules i and j at oxygen-oxygen distance r_{ij}, accounts for the Van der Waals attraction and the repulsive forces due to Pauli's exclusion principle, and is expressed as a double-truncated Lennard-Jones potential,

$$\mathcal{U}(r) \equiv 4\epsilon \left[\left(\frac{r_0}{r}\right)^{12} - \left(\frac{r_0}{r}\right)^{6} \right], \quad \text{if } r_0 < r < 6r_0,$$

$\mathcal{U} \equiv \infty$ for $r \leq r_0$ and $\mathcal{U} \equiv 0$ for $r \geq 6r_0$, where we use ϵ as our energy scale. The second term of the Hamiltonian represents the directional (covalent) contribution to the formation of water-water hydrogen bonds (HBs). Assuming that each molecule i can form up to four HBs, the number of possible molecular conformations is discretized by the introduction of four bonding variables $\sigma_{ij} = 1...q$, one for each neighbor molecule j. A HB will be formed if the angle between the orientations of

the OH vector of a molecule and the OO vector with the neighbor molecule do not exceed $\pm 30^o$. Therefore, only 1/6 of all the possible orientations $[0^o,\ 360^o]$ are associated with a HB. Thus, we fix $q = 6$ and a HB will be formed only if $\sigma_{ij} = \sigma_{ji}$, with characteristic energy J, correctly accounting for the entropy loss associated to a HB formation. The total number of bulk HBs is $N_{HB}^{(B)} = \sum_{ij} \delta_{\sigma_{ij},\sigma_{ji}}$, where $\delta_{a,b} = 1$ if $a = b$, 0 otherwise. Finally, the third term corresponds to the cooperative interaction of the HBs, emerging from quantum many body interactions, which leads to an ordered, low density tetrahedral configuration in bulk. This phenomenon is modelled as an effective interaction between each of the six different pairs of the four variables σ_{ij} of a molecule i, coupled by an energy J_σ. $N_{coop}^{(b)} \equiv \sum_{ikl} \delta_{\sigma_{ik},\sigma_{il}}$ is the sum over the pair of bonding indices that cooperatively acquire the same value in each molecule i. By taking $J_\sigma \ll J$ we guarantee that the term plays a role only when the HBs are formed. Finally, the total volume, and hence the density field, depend on the HB formation, as $V^{(b)} = Nv + N_{HB}^{(b)} v_{HB}^{(b)}$, where $v_{HB}^{(b)}$ is a fraction of v_0. This relation accounts, on average, for the local decrease of density due to the tetrahedral HB network. The valuers of the model's parameters are given at the end of the next section.

B. Protein and interface model

Following the coarse-grain representation for the water molecules we adopt a coarse-grained lattice representation for the proteins, depicted as self-avoiding heteropolymers composed of 36 amino acids. The interface is taken as a spatially fixed hydrophobic homopolymer. For sake of simplicity, both the protein residues and the interface section have the same size as the water molecules (one cell) [5]. The residues interact through a nearest neighbour potential given by the Miyazawa-Jerningan interaction matrix [5]. The HB formation in the protein/interface hydration shell depends on the hydrophobic (Φ) or hydrophilic (ζ) nature of the hydrated amino acids. Depending if two water molecules, forming a HB, are near two hydrophobic amino acids, two hydrophilic amino acids, or one of each kind (mixed, χ), the Hamiltonian is

$$\begin{aligned}\mathcal{H}_{w,w}^{(h)} \equiv &- \left[J^\Phi N_{HB}^\Phi + J^\zeta N_{HB}^\zeta + J^\chi N_{HB}^\chi\right] + \\ &- \left[J_\sigma^\Phi N_{coop}^\Phi + J_\sigma^\zeta N_{coop}^\zeta + J_\sigma^\chi N_{coop}^\chi\right]\end{aligned} \tag{2}$$

where N_α^Φ, N_α^ζ, N_α^χ ($\alpha = HB, coop$) represent the number of directional and cooperative bonds formed at a hydrophobic, hydrophilic or mixed interface. Experiments and simulations show that near a hydrophobic interface the water-water HBs are stronger [7] and increase the local water density upon pressurization. To account for these effects, the model assumes that $J^\Phi > J$, $J_\sigma^\Phi > J_\sigma$, and that the volume associated to a HB at the Φ interface decreases upon a P increment, i.e. $v_{HB}^\Phi / v_{HB,0}^\Phi \equiv 1 - k_1 P$ where $v_{HB,0}^\Phi$ is the volume increase for $P = 0$ and k_1 is a factor accounting for the compressibility of the hydrophobic hydration shell. We consider that the hydrogen bonding and the water density do not change next to a hydrophilic interface. Thus, HBs in hydrophobic hydration shell generate an extra contribution $V^\Phi \equiv N_{HB}^\Phi v_{HB}^\Phi$ to the total volume. Taking all that into consideration the model sets $J^\zeta = J$, $J_\sigma^\zeta = J_\sigma$, $J^\chi = (J^\Phi + J^\zeta)/2$, $J_\sigma^\chi = (J_\sigma^\Phi + J_\sigma^\zeta)/2$, and lastly $v_{HB}^\zeta = v_{HB}^{(b)}$, where $v_{HB}^{(b)}$ is the bulk parameter. Finally, the model assumes that the protein-water interaction energy is different depending on the residue nature, being $-\varepsilon^\Phi$ and $-\varepsilon^\zeta$, in the hydrophobic and the hydrophilic hydration shell, respectively.

C. The model's parameters

The energy parameters are, in units of 8ε, $J = 0.3$, $J_\sigma = 0.05$, $J^\Phi = 1.2$, $J_\sigma^\Phi = 0.2$, $\varepsilon^\Phi = 0$, $\varepsilon^\zeta = 0.48$. The parameters related to the volume are $k_1\varepsilon/v_0 = 1$, $v_{HB}^{(b)}/v_0 = 0.5$, and $v_{HB,0}^\Phi/v_0 = 0.5$. This parameter choice balances the water-water, water-residue and residue-residue interactions, ensuring the protein stability in the liquid phase, including ambient conditions. It also accounts for the protein and interface surface loss by taking a 2D representation instead of 3D by enhancing the interfacial interactions.

III. METHOD

We study a protein with a native state (Fig. 1, inset) whose surface is 35% hydrophobic and 65% hydrophilic, having no side completely hydrophobic. The protein is embedded into 40x40 cells with periodic boundary conditions (PBC), each cell being occupied by either a water molecule or a protein residue. The hydrophobic interface is 40 cells long and fixed into the space. Due to the PBC, the proteins are embedded between two hydrophobic interfaces. We perform Monte Carlo (MC) simulations for different concentrations of the protein at fixed P and T. Our MC simulations include (1) water moves and (2) protein moves:

1. Water moves consist in breaking/forming HBs and rescaling the total volume of the simulation box.

2. Protein moves consist in pivot moves, corner flips and crankshaft moves.

We perform simulations for various protein concentrations, in the range $c = [5\%, 27\%]$. First, we equilibrate the system for given c and T. Next, we calculate the observables for several MC steps afterwards. To study the proteins folding/unfolding and aggregation we calculate the number N_c of native contacts of each protein,

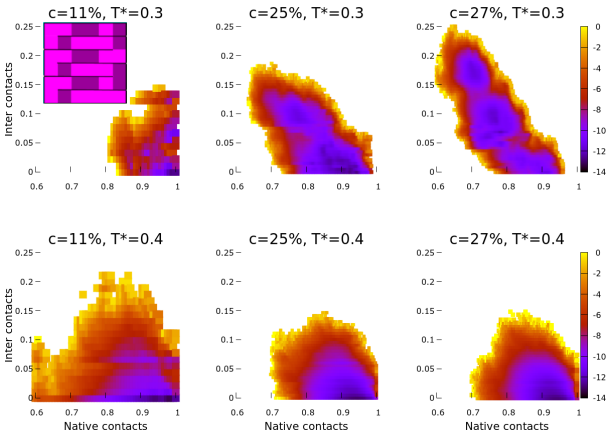
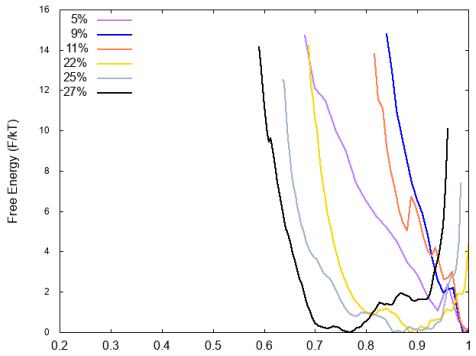
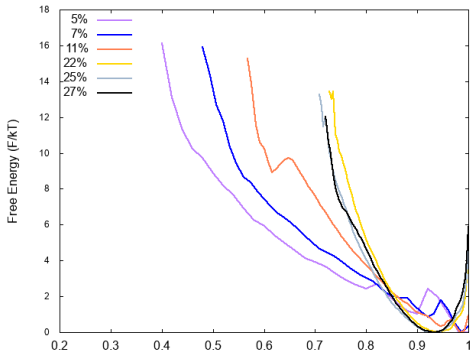


FIG. 1: Inset: Native structure of the protein; the dark/light cells are hydrophobic/hydrophilic amino acids. Main panels: Colour coded free energy F as a function of the native contacts N_c (x axis) and the inter contacts I_c (y axis) for different concentrations c and temperatures T^*. The axis are normalised by the total number of native and inter contacts, i.e., 25 and 36, respectively, times the number of proteins. The F minima correspond to the darker regions.



(a) Native contacts, $T^* = 0.3$



(b) Native contacts, $T^* = 0.4$

FIG. 2: Free energy profiles as a function of the normalized number of native contacts N_c for $T^* = 0.3$ (top) and 0.4 (bottom). Different colors correspond to different concentrations c, as indicated in the legend, with error $\pm 1\%$.The proteins are all folded when the minimum is at $N_c = 1$.

i.e., contacts in common with the native structure, the number I_c of inter contacts between different proteins or between the proteins and the interface, and the number M_c of contacts of the proteins with the interface. We compute the free energy as $F(A) \equiv -k_B T \ln P(A)$, where $P(A)$ is the normalized histogram (probability) of occurrence of a value of A, where A is N_c, I_c or M_c, all normalized by the total number of native, inter or interface contacts respectively.

IV. RESULTS

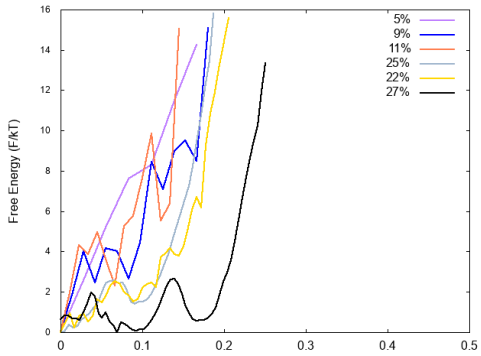
In Fig. 2 we present the free energy landscape as a function of N_c and I_c, for two different temperatures $T^* \equiv k_B T/\varepsilon = 0.3,\ 0.4$. By locating the minima in the native contacts and inter protein contacts profiles we identify three different states: the native folded state (FOL), the unfolded but not aggregated state (UNF) and the unfolded and aggregated state (AGG).

The first is given when the number of native contacts is $N_c = 1$ (Fig. 2). In this regime all the proteins have recovered their native conformation. When the minimum is displaced from $N_c = 1$, on average the proteins unfold, the smaller N_c the farther the proteins from the native state. We find that at both $T^* = 0.4$ and 0.3 the proteins unfold at large enough c.

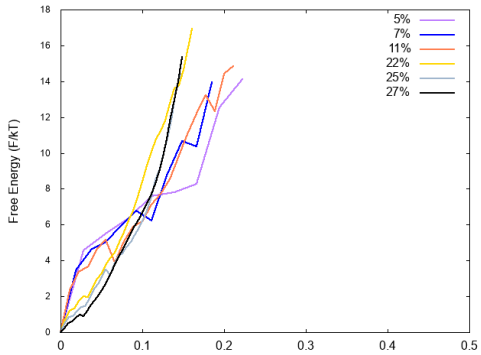
For each concentration, the proteins do not aggregate if the minimum of $F(I_c)$ is at $I_c = 0$ (Fig. 3). By inspecting Fig. 1, we observe that this occurs also for values of c at which the proteins are unfolded, for example for all the concentrations at $T^* = 0.4$. At $T^* = 0.3$ and $c = 22\%$, we observe that F has local minima at $I_c > 0$. However, only at $c = 27\%$ the global minimum of F is at $I_c > 0$ and the proteins aggregate. At this concentration, the proteins are unfolded. Hence, the aggregation occurs only if the proteins are unfolded, while unfolding is independent on aggregation.

Fig. 4 shows the free energy as a function of the interface contact points M_c. A minimum at $M_c > 0$ would correspond to proteins that on average adsorb onto the interface. However, we never find this condition. Hence, for the specific sequence we consider here and the specific temperatures and concentrations, the adsorption of the proteins would have a free energy cost that is too large for the system. This observation, does not exclude that different sequences, with more hydrophobic residues exposed to water in their native state, would not adsorb at appropriate values of temperature and concentration.

When we compare our results with the previous calculatios performed for the same protein without the hydrophobic interface [5], we find that the transition from FOL to UNF state is at similar concentrations,

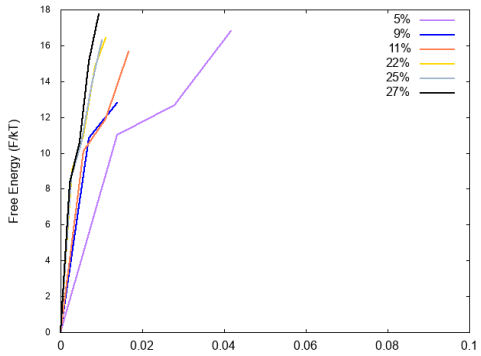


(a) Inter contacts, $T^* = 0.3$

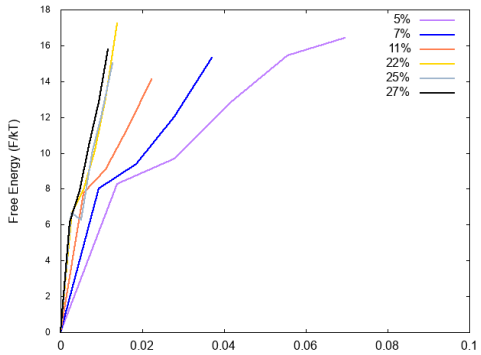


(b) Inter contacts, $T^* = 0.4$

FIG. 3: Free energy profiles as a function of the normalized number of inter contacts I_c. Colors are as in Fig. 2. The proteins are not aggregated when the minimum is at $I_c = 0$.



(a) Interface contacts, $T^* = 0.3$



(b) Interface contacts, $T^* = 0.4$

FIG. 4: Free energy profiles as a function of the normalized number of M_c. Colors are as in Fig. 2. The proteins are not adsorbed onto the interface when the minimum is at $M_c = 0$.

$c_{FOL \to UNF} = 9\%$ when $T^* = 0.3$ and $c_{FOL \to UNF} = 7\%$ when $T^* = 0.4$. As the concentration is increased this minimum displaces towards lower values of N_c.

The aggregation between proteins appears at $c_{FOL \to UNF} = 27\%$ in the case of $T^* = 0.3$, but never at $T^* = 0.4$. This is in accordance with [5], where aggregation did not appear until $c = 30\%$. Consistently, we expect aggregation also at higher concentration. Fig. 1 shows how the number of inter contacts increases as the natives contacts diminishes at large concentrations for $T^* = 0.3$, suggesting that the aggregation is ruled by unfolding. Furthermore, as shown in [8], the proteins aggregation is not influenced by the presence of other protein species. For this reason, we believe that the threshold values observed here do not vary from those without a hydrophobic interface, that could be seen at a protein of a different species.

We find no adsorption to the interface at both $T^* = 0.3$ and $T^* = 0.4$. By inspection of the free energy profiles, we find that protein make very few contacts with the interface. We believe that this result is due mainly to the restrictions imposed by the 2D system, in which the interface is reduced just to a line of points. Furthermore, the sequence we consider here is mostly hydrophilic, hence the system favors configurations where the proteins are mostly hydrated. We expect that larger hydrophobic patches on the proteins could induce larger entropy gain when the proteins adsorb onto the hydrophobic interface or aggregate among them.

We observe that temperature as a week effect on unfolding, but a large effect on aggregation. For $T^* = 0.4$ we find no aggregation (up to $c = 27\%$). We observe that the simulations for $T^* = 0.4$ reach the equilibrium more easily than those at $T^* = 0.3$, as the greater thermal fluctuations allow the system to evolve quicker. At $T^* = 0.3$ the system is trapped in metastable states for longer times than at higher T. As a consequence, the statistics for the $T^* = 0.3$ results is lower than for those at $T^* = 0.4$, increasing the statistical error for the lower T, especially at higher c.

V. CONCLUSIONS

We perform coarse-grain simulations tho study the mechanisms of folding/unfolding of proteins, competing with aggregation, in a water environment. Adopting the FS water model [1] and the computational tools used in

recent studies, we consider the effect of a hydrophobic interface when we increase the protein concentration. We find the following results:

- Increasing the concentration, folded proteins unfold. On unfolding, they do not aggregate. However, further increase of concentration at low enough temperature, leads to protein aggregation. This results is consistent with what observed in Ref. [5].
- When compared with [5], we observe that the presence of the hydrophobic interface does not affect significantly the concentration thresholds for unfolding and aggregation. This could be due to the strong water-water interactions and the small hydrophobic patches of the considered sequence. However, previous studies [8] suggest that the thresholds are not affected by the presence of proteins of other species, thus suggesting that also the presence of an interface could have a weak effect on them. Further analysis is necessary to clarify this point.
- The temperature is an important factor in protein aggregation. We find aggregation at the lower temperature we consider, but not at the higher. This could be an out-of-equilibrium effect, due to the larger probability of finding the system in a metastable state, where the proteins are aggregated, at lower T. Longer simulations, and more sofisticated analysis, will be needed to better understand this result.

Further extension of this study include (i) increasing the concentration range, (ii) studying systematically the necessary number of hydrophobic amino acids on the surface of the protein in order to observe adsorption to the interface, (iii) simulate the model in 3D.

VI. ACKNOWLEDGEMENTS

I would like to thank Dr. Valentino Bianco for his collaboration in the study, both for the technical help and his insight on the matter.

[1] Valentino Bianco and Giancarlo Franzese, *Contribution of Water to Pressure and Cold Denaturation of Proteins*, Phys. Rev, 2015, 115: 108101

[2] Valentino Bianco, Giancarlo Franzese, Christoph Dellago, and Ivan Coluzza *Role of Water in the Selection of Stable Proteins at Ambient and Extreme Thermodynamic Conditions*, Phys. Rev. X 2017, 7: 021047

[3] P. J. Knowles, M. Vendruscolo, C. M. Dobson, *The amyloid state and its association with protein misfolding diseases*, Nat. Rev. Mol. Cell Biol. 2014, 15, 384.

[4] G. G. Tartaglia, M. Vendruscolo, *Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations* Mol. BioSyst. 2009, Vol. 5, p. 1873

[5] Valentino Bianco, Giancarlo Franzese and Ivan Coluzza *In Silico Evidence That Protein Unfolding is a Precursor of Protein Aggregation*, 2020, ChemPhysChem, Vol. 21, Issue 5, p. 377

[6] Valentino Bianco and Giancarlo Franzese, *Critical behavior of a water monolayer under hydrophobic confinement*, Scientific Reports 4, 2014, 4440.

[7] Sapna Sarupria and Shekhar Garde *Quantifying Water Density Fluctuations and Compressibility of Hydration Shells of Hydrophobic Solutes and Proteins* Phys. Rev. Lett. 2009, 103: 037803

[8] Valentino Bianco, Miren Alonso-Navarro, Desire Di Silvio, Sergio Moya, Aitziber L. Cortajarena and Ivan Coluzza; *Proteins are Solitary! Pathways of Protein Folding and Aggregation in Protein Mixtures*, The Journal of Physical Chemistry Letters, 2019, Vol. 10, p. 4800.