Numerical study of water clusters at protein interfaces

Author: Andreu Puy Contreras
Advisor: Giancarlo Franzese
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Abstract: We simulate by Monte Carlo method how the protein interface in aqueous solution affects the formation of clusters whose connectivity length is related to the thermodynamic correlation length of water, following the definition of the "correlated percolation". We consider a model for a protein in a water monolayer that has been shown to reproduce the anomalies of water, including the occurrence of two specific heat maxima at low temperature and pressure in a region that approaches a liquid-liquid critical point. We show that the maxima can be characterized in terms of percolation quantities at any temperature and pressure. Furthermore, we find that the protein interface promotes the formation of water clusters, a result that might be relevant for the biological functions of the proteins.

I. INTRODUCTION

Water has many anomalies [5]. One of them is the increase of the isobaric specific heat c_p when T decreases, especially for $T < 0°C$ in the supercooled liquid phase, which is experimentally accessible down to $T \approx 150\ K$ [9]. Below this T, c_p maxima have been extrapolated from experimental data [6][7] and have been calculated by numerical simulation of water models. In these models these maxima have been shown to be related to the occurrence of a liquid-liquid phase transition [8]. A maximum in c_p is a consequence of a maximum of the fluctuation of entropy, i.e. a maximum in structural changes. Here we explore how we can characterize the structural changes of the hydrogen bond (HB) network in water using a percolation approach. In particular, we will consider a protein solution and study how the protein interface affects the structural changes.

A HB is a bond between hydrogen and an electronegative atom such as oxygen. Liquid water forms a dynamical network of HBs, with each molecule bonded to four others on average in a tetrahedral structure that changes continuously as water diffuses. Due to the quantum nature of the HB, the (many) atomistic water models with only pairwise interactions are unable to accounts properly for all the water properties, many of which are anomalous. Other models include quantum many-body interactions but are too numerically expensive for large-scale studies as those required for biological systems. Here we follow an alternative strategy that relies on the adoption of a coarse-grained water model that includes in an effective way many-body interactions for the HB. This model has been shown to describe the water anomalies [1][2][3] and is efficient for numerical studies of biological problem as complicated as the protein folding [4].

For sake of simplicity we consider here the case of a water monolayer, whose configuration can be projected on 2D. We assume a homogenous distribution of water molecules at constant pressure P, temperature T and number of molecules N, and we replace the molecule coordinates by a discretized density field with spatial resolution of $v \equiv \frac{V}{N}$, where $V(P,T)$ is the total volume and N a number larger than the number of water molecules. Defining $v_0 < v$ as the van der Waals water volume, the density field is set to 0 if $v_0/v < 0.5$ (gas-like density) and to 1 (liquid-like density) otherwise. Because we are interested in studying a liquid solution, in the following we will consider only the case in which the density field is 1 in the entire system.

The water model has an enthalpy with four terms:

1. The *Lennard-Jones interaction* $U(r)$ is a long range and isotropic interaction between molecules i and j at a distance r_{ij}. Because we fix *P, V* and *r* are continuous variables. It represents the Van der Waals forces with attraction due to fluctuating induced dipoles and repulsion (here represented as a hard-core) due to Pauli exclusion principle:
$$U(r) \equiv \begin{cases} \infty, & r_{ij} < r_0 \sim v_0^{1/3} \\ 4\varepsilon\left[\left(\frac{r_0}{r_{ij}}\right)^{12} - \left(\frac{r_0}{r_{ij}}\right)^{6}\right], & r_0 \le r_{ij} \le 3r_0 \end{cases}$$
where ε is the characteristic energy, $r_0 \equiv \sqrt{(v_0/h)}$ and *h* the monolayer height.
2. The *directional component of the HB interaction* $\mathcal{H}_{\mathrm{I}}$ is due to the formation of the HB. A HB of a water molecule i with a molecule j is described thanks to a bonding variable $\sigma_{ij} = 1, \dots, q$ of the molecule *i* and the corresponding bonding variable σ_{ji} of the molecule *j*. Here $q = 6$ because a HB can be formed between two near molecules only if the hydrogen involved deviates from the O-H-O perfect alignment by $|\Delta\theta| \le 30°$, hence only 1/6 of the possible orientations in the O-H-O plane correspond to a bonded state. The HB is formed for $\sigma_{ij} = \sigma_{ji}$, i.e. when both molecules have the correct orientation. In this way the total number of HB is $N_{HB} \equiv \sum_{\langle i,j\rangle} \delta_{\sigma_{ij},\sigma_{ji}}$ where the sum is over nearest neighbours in the discrete partition of *V*. Therefore, $\mathcal{H}_{\mathrm{I}} \equiv -JN_{HB}$, where $J > 0$, guaranties the energy and entropy decrease associated to the directional component of the HB interaction.
3. The *cooperative HB interaction* $\mathcal{H}_{\mathrm{Coop}}$ accounts for the many-body component of the HB and has a characteristic energy $J_\sigma < J$. This interaction is represented in the Hamiltonian as $\mathcal{H}_{\mathrm{Coop}} \equiv -J_\sigma N_{coop}$ where $N_{coop} \equiv \sum_i \sum_{(l,k)_i} \delta_{\sigma_{ij},\sigma_{ji}}$ is a sum over the six pairs of bonding variables of the molecule *i*. The asymmetric condition $J_\sigma < J$ guarantees that this term implies an effective interaction among the four bonded molecules in the hydration shell of molecule *i*. This term is responsible for transmitting the cooperative fluctuations along the HB network.
4. From the experiments we know that the HB tetrahedral structure of water molecules induces a decrease of density. The model accounts for this effect by including a volume increase v_{HB} per HB. Therefore the total

* Electronic address: andreupuy@hotmail.com

volume of the system is $V \equiv V_0 + N_{HB} v_{HB}$ where $V_0 \equiv v_0 N$.

In conclusion, the water model has an enthalpy

$$\mathrm{H} \equiv U(r) - (J - P v_{HB}) \sum_{\langle i,j \rangle} \delta_{\sigma_{ij},\sigma_{ji}} - J_\sigma \sum_i \sum_{(l,k)_i} \delta_{\sigma_{ij},\sigma_{ji}} + P V_0 \qquad (1)$$

The model is defined in such a way that it can represent liquid water below the melting temperature, where the (supercooled) liquid is metastable with respect to the crystal. In this region the model displays a liquid-liquid critical point at positive P_C and finite T_C [3]. For $P < P_C$ the model has two c_p maxima, weak at $T_W(P)$ and strong at $T_S(P)$, with $T_W(P) > T_S(P) \geq T_C$, that converge at $P \approx P_C$ and diverge at P_C.

Because we want to understand how the structural changes are affected by the presence of an interface, we include a hydrophobic polymer in solution as a simplified representation of a protein. Bianco et al. [4] have shown that the presence of the interface can be taken into account in the model by introducing parameters J^B, J^B_σ, v^B_{HB} for the bulk water and the corresponding parameters $J^p > J^B$ and $J^p_\sigma \geq J^B_\sigma$ and $v^p_{HB}/v^B_{HB} \equiv 1 - kP$ with $k > 0$ for the first hydration layer. As a consequence, the model can qualitatively reproduce the stability region of proteins, with folding, cold unfolding and pressure unfolding.

To describe the HB structure we use the *correlated percolation* in which two interacting bonding variables $\sigma_{\alpha\beta}$ and $\sigma_{\beta\alpha}$ belong to the same cluster with probability $p(T,P) \equiv 1 - exp(2H_{\alpha\beta}/k_B T)$ where $H_{\alpha\beta} \leq 0$ is the term in Eq. (1) that depends on $\delta_{\sigma_{\alpha\beta},\sigma_{\beta\alpha}}$ and k_B the Boltzmann constant. It can be shown that with this definition the clusters statistically coincide with the regions of correlated molecules [1]. The size s of a cluster is by definition the number of bonding variables that form the cluster. The probability that an arbitrary bonding variable belongs to the cluster of size s is

$$P_s \equiv n_s s, \qquad (2)$$

where n_s is the density of the cluster of size s, and the *mean cluster size* is

$$S \equiv \sum_s s P_s = \sum_s s^2 n_s. \qquad (3)$$

The *percolation transition* occurs at any P at $T_P(P)$, where there is an incipient infinite cluster, i.e. a cluster spanning the entire system. At $T_P(P)$, S diverges. The percolation line in the P-T plane, where S diverges, marks the threshold of the formation of the structured HB network. At the critical isobar $T_P(P_C) = T_C$ the critical temperature by construction, while for $P < P_C$ the percolation line coincides with the locus of strong c_p maxima $T_S(P)$.

II. METHOD

We perform Cluster Monte Carlo simulations with a code developed in house [1] for a system of size $L \times L$, with $L = 40$ with a fully hydrated protein of length $L^p = 30$ and $N = L \times L - L^p = 1570$ molecules of water. To check the effect of the extension of the interface, we consider four protein configurations (Fig. 1): 1 completely unfolded, 2 folded on both extremes, 3 folded only in one extreme, and 4 completely folded.

Starting from a high T configuration for water we run simulations with 10^6 Monte Carlo steps as initial equilibration, followed by 10^6 steps of production, averaging the calculation every 10^2 steps. We adopt the parameters $J^B = 0.3$, $J^B_\sigma = 0.05$, $J^p = 0.55$, $J^p_\sigma = 0.05$, $v^B_{HB} = v^p_{HB} = 0.5$. All the quantities are expressed in internal units: $[T] = \frac{4\varepsilon}{k_B}$, $[P] = \frac{4\varepsilon}{v_0}$, $[V] = v_0$, $[E] = 4\varepsilon$ and $[c_p] = k_B$, where E is an energy. We explore the system for $-0.5 \leq P \leq 0.8$ and for $0.03 \leq T \leq 0.8$. Once we achieve the equilibrium conditions, we calculate the energy, the volume, the enthalpy, the isobaric specific heat

$$c_p \equiv \left(\frac{\partial \langle H \rangle}{\partial T} \right)_P = \frac{\langle \Delta H^2 \rangle}{k_B T^2}, \qquad (4)$$

where $\langle \Delta H^2 \rangle$ is the enthalpy fluctuations, and the percolation quantities S, P_2 and P_3. The percolation quantities are calculated separately for the *interfacial* clusters, which include water from the first hydration layer, and the *bulk* clusters made only of bulk water.

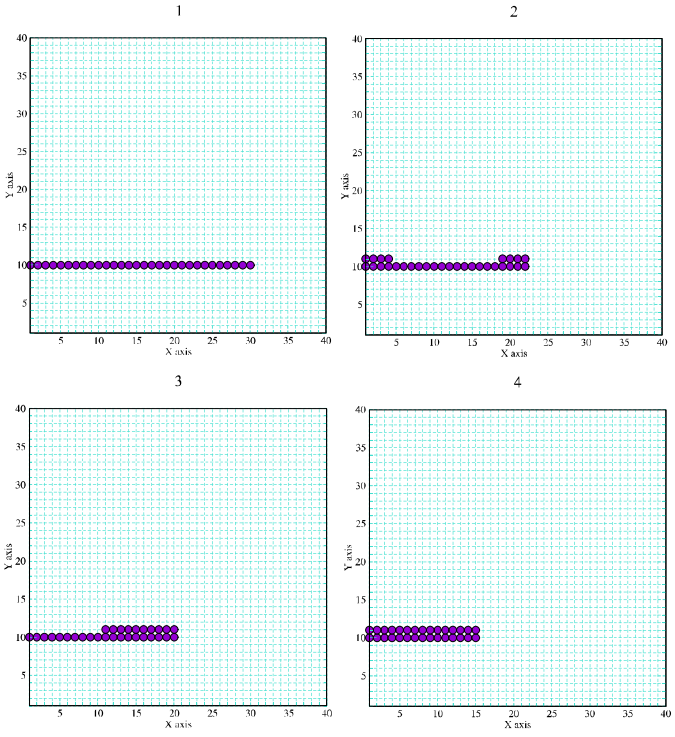

Fig. 1. The four protein configurations used in this work, labeled as 1, 2, 3 and 4.

III. RESULTS

At high P (Fig. 2a) all bulk clusters are made of one single molecule ($s = 1$) but interfacial clusters are usually larger at small T. This can be understood as a consequence of the stronger water-water interaction at the protein interface. At small P instead bulk cluster have a larger probability and cluster made of $s = 2$ or 4 favored against the others (Fig. 2b). This is due to the fact that clusters are made of HBs and HBs connect two water molecules at the time. At the interface this effect is more moderate due to the difficulties of forming tetrahedral structures.

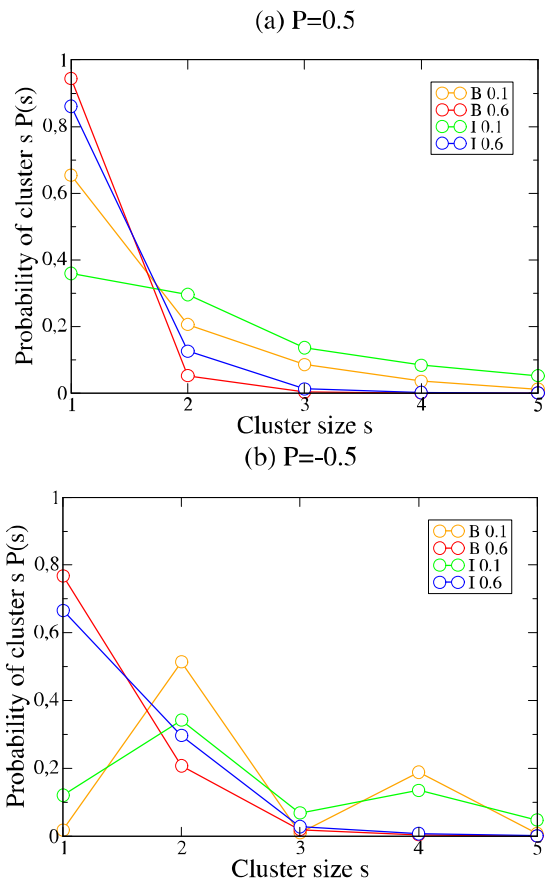

Fig. 2. Probability density P_s for (a) pressure $P = 0.5$ and (b) $P = -0.5$ for $T = 0.1$ and 0.6 and for bulk (B) and interfacial (I) water for the protein configuration 1.

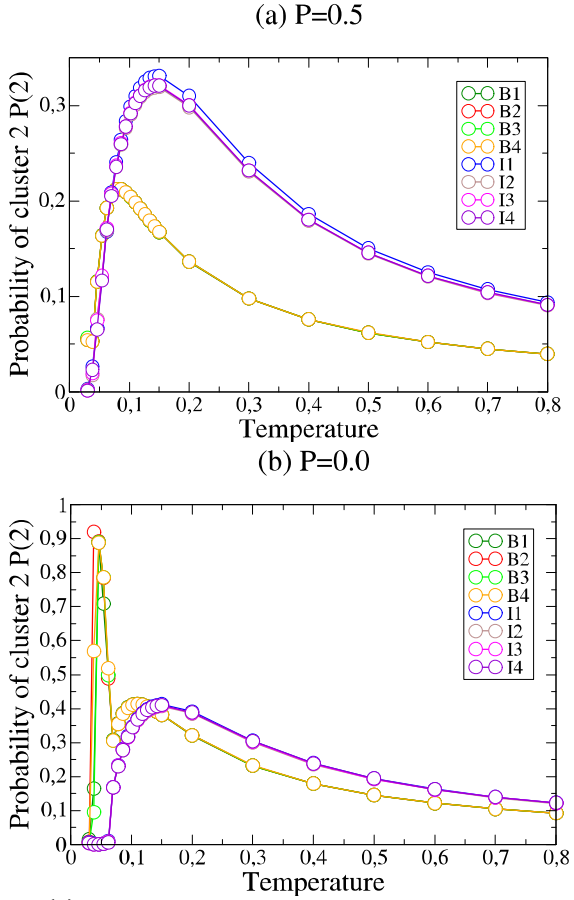

Fig. 3. (a) $P_2(T)$ for $P = 0.5$ and (b) 0.0 for bulk (B) and interfacial (I) water for all the configurations of the protein (1, 2, 3 and 4).

The probability P_2 of clusters of size 2 has maxima for high and low pressures (Fig. 3). At high P we observe again that the interface favours the cluster formation (Fig. 3a), while at low P the probability P_2 has two maxima for bulk clusters and only one for interfacial clusters (Fig. 3b). As we will discuss later, the two maxima, one weak at higher T and one strong at lower T, are related to the specific heat behaviour of the system. For $P = 0.5$ we can see too that there is a small difference on the behaviour of the configuration 1 which is higher than the others. This could be rationalized as in this case the protein is completely unfolded and there are more molecules of water on the interface than on the other cases.

The probability P_3 behaves in a way similar to P_2 at high pressure, with the interface favouring the cluster formation, (Fig. 4a) however, as observed in Fig. 2b, it is smaller than P_2 at the same P and T, showing that the clusters are preferentially made of bonding variable that belongs to different water molecules. At low pressure P_3 not only is 10 times smaller than P_2 at the same P and T, but it is also qualitatively different because it does not show two maxima for bulk clusters (Fig. 4b). Interfacial clusters in this case show a maximum and a minimum within the T interval of interest (possibly with another maximum outside this interval).

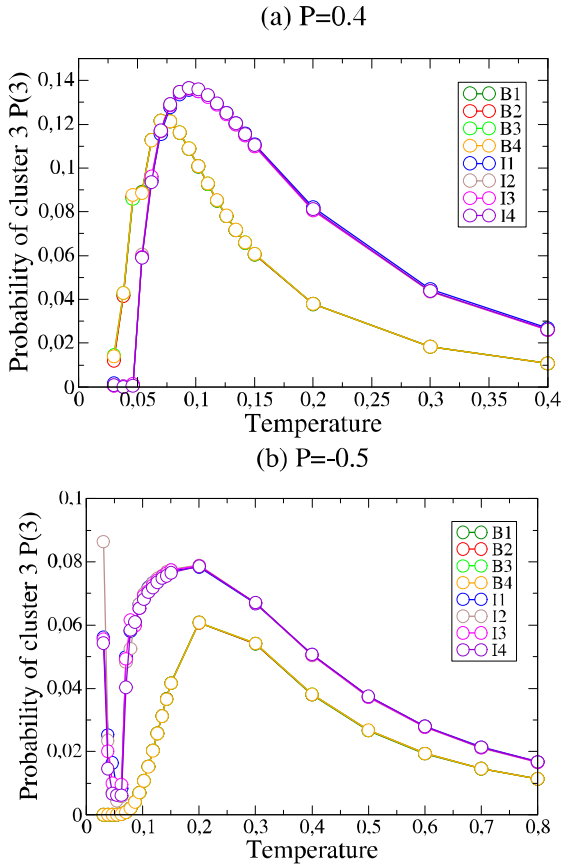

Fig. 4. (A) $P_3(T)$ for $P = 0.4$ and (b) -0.5 for bulk (B) and interfacial (I) water for all the configurations of the protein (1, 2, 3 and 4).

The presence of strong maxima, at least for the bulk clusters, suggests the possibility that water is forming large clusters at low T. This is indeed confirmed by the analysis of the mean cluster size S. At high P we find that S has a moderate maximum for the interfacial clusters but only a monotonic decrease for bulk clusters (Fig. 5a). This behaviour is consistent with the interpretation that at high P the percolation line (marked by the bulk S) occurs at $T \approx 0$, with a small precursor at finite T for the interfacial cluster, because there is no critical phase transition for the

thermodynamics system at finite T. The situation changes drastically at low P, where S has a sharp increase for both bulk and interface clusters (Fig. 5b). These maxima mark the expected percolation transition at finite T. As we will discuss next, these maxima relate to the maxima in the specific heat of the system.

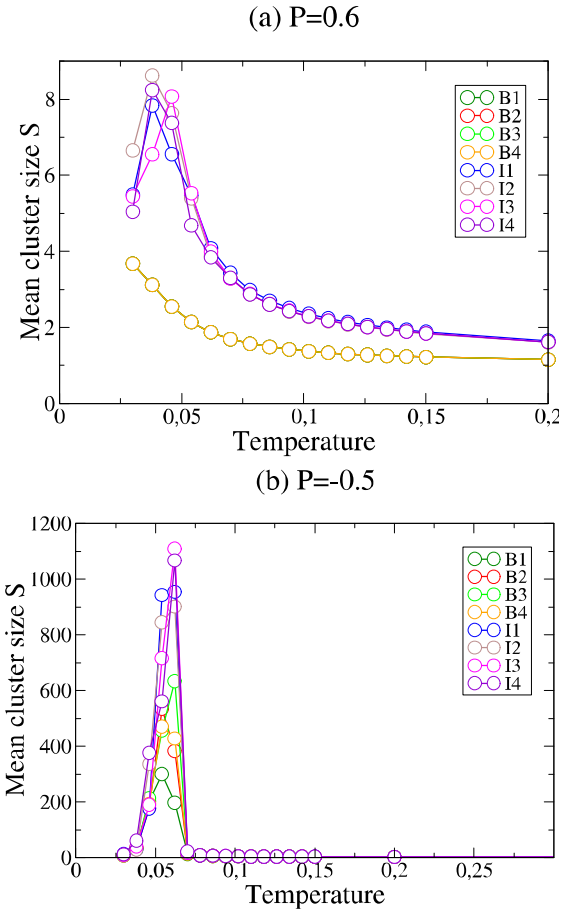

Fig. 5. (a) $S(T)$ for $P = 0.6$ and (b) -0.5 for bulk (B) and interfacial (I) water for all the configurations of the protein (1, 2, 3 and 4).

We find that the specific heat c_p has two maxima for $P \leq 0.1$ for bulk clusters, one weak at $T \approx 0.1$ and one strong for $T \leq 0.075$ (Fig. 6a). At higher P we find that the interfacial clusters grow very sharply, possibly diverging at $T = 0$. At lower P we continue to observe both maxima (Fig. 6b) and we can see (not shown) that at higher temperatures the isobaric specific heat explodes due to the liquid gas phase transition of the system. All these results are consistent with previous results [1] showing that at low P there are two c_p maxima that converge to each other by increasing P, marking the occurrence of the liquid-liquid critical point.

IV. DISCUSSION

Plotting (Fig. 7a) the loci of maxima of S and P_2 in the P-T phase diagram allows us to observe that the interface affects the weak maxima of P_2 promoting the clustering, but does not affect the maxima of S that mark the percolation line. In particular the strong maxima of bulk P_2 follow the percolation line. At high P in this and the following plots we observe that the loci of maxima have a positive slope in the P-T phase diagram mimicking the peculiar high-P change of slope of the water melting line, not calculated here.

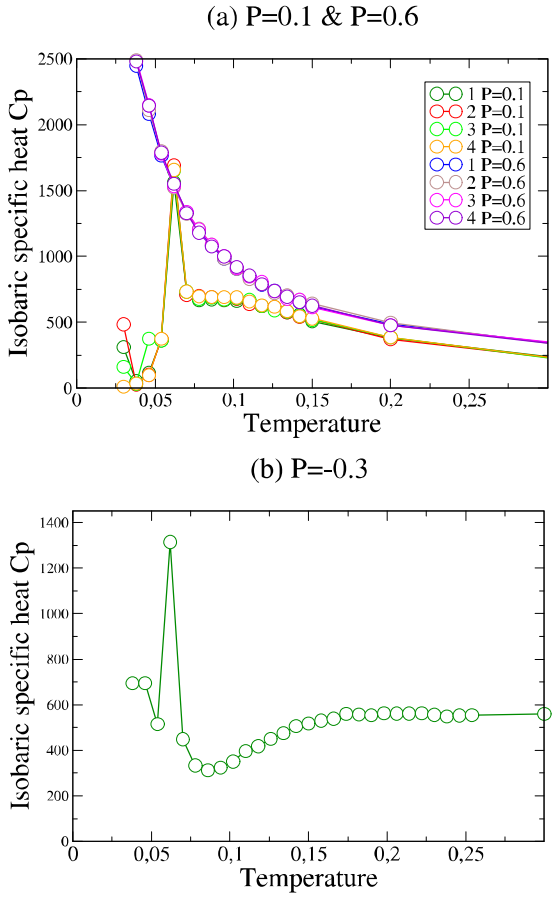

Fig. 6. (a) $c_p(T)$ for $P = 0.1, 0.6$ for all the configurations of the protein (1, 2, 3 and 4) and (b) -0.3 for only configuration 1 for bulk (B) and interfacial (I) water.

On the other hand plotting together the loci of maxima of S and P_3 we do not find strong correlations, suggesting that the P_3 does not contribute significantly to the percolating behaviour of S (Fig. 7b).

Finally plotting together the loci of maxima of c_p, weak maxima of P_2 and P_3 and maxima of S (Fig. 7c) we find that the weak maxima of c_p correlate with the weak maxima of P_2 and P_3, while strong c_p maxima are related to the strong maxima of P_2 that overlap the maxima of S.

The last figure allows us to conclude that the weak c_p maxima are associated to the formation of clusters of two water molecules with a reciprocal HB. Therefore the weak c_p maxima are associated to the fluctuations of the number N_{HB} of HB.

On the other hand, the strong maxima of c_p are associated to the cooperative (collective) behavior of the HB, characterized by the percolation transition marked by S. Hence the strong c_p maxima are associated to the fluctuations of the number N_{coop} of cooperative HB.

Previous analyses have shown that the two c_p maxima line converge at the liquid-liquid phase transition of the system [3]. We observe here that by increasing P all the bulk maxima converge near $P = 0.4$ (Fig. 7). We therefore conclude that the liquid-liquid critical point for our protein solution occurs at near $P = 0.4$ and $T = 0.05$.

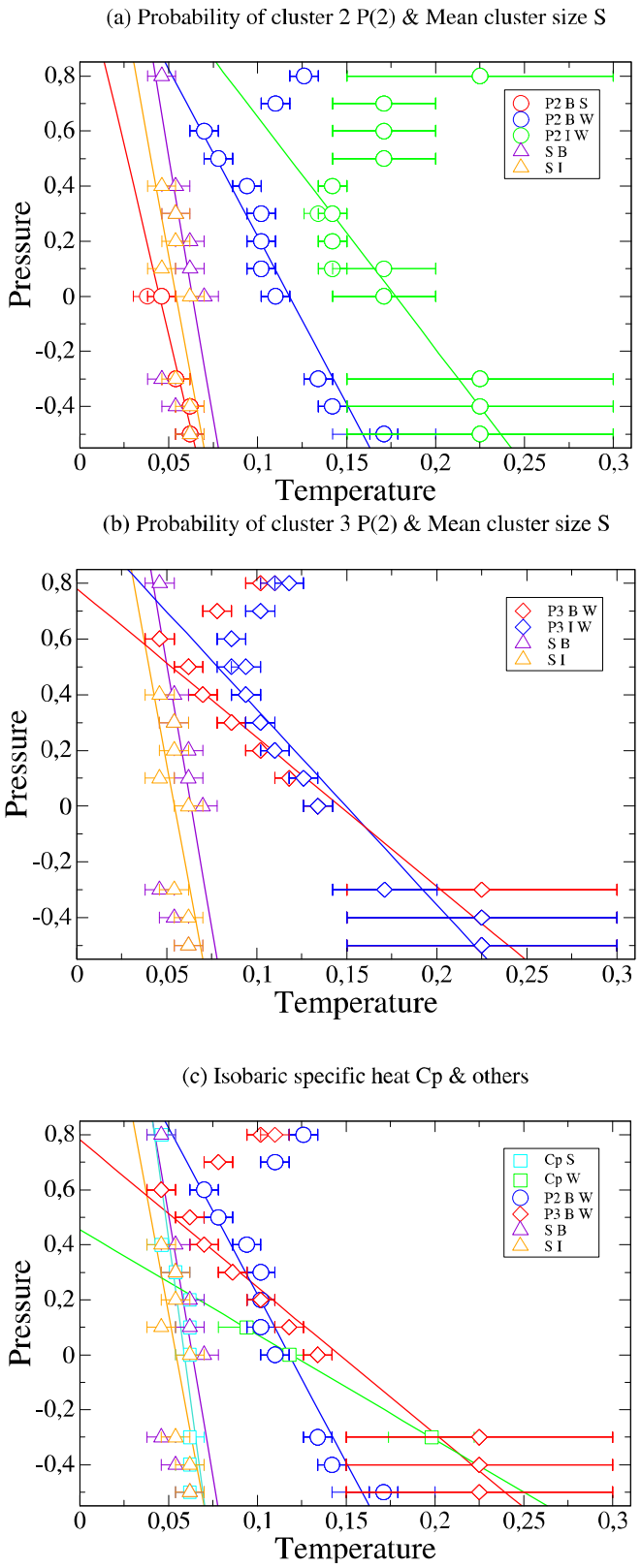

Fig. 7. (a) *P-T* diagrams for the loci of maxima of the mean cluster size S and strong (S) and weak (W) maxima of the P_2 in bulk (B) and interfacial (I) clusters. (b) As in (a) but for P_3 instead of P_2. (c) Loci of the strong and weak maxima of the isobaric specific heat c_p compared to the mean cluster size S maxima of bulk and interfacial water and with the weak maxima of bulk P_2 and P_3. Lines are linear fits of the low-P calculations. Error bars are determined by our T-resolution.

V. CONCLUSION

We perform a percolation analysis for a protein solution in a monolayer of water in the supercooled liquid region where we expect to observe two maxima for the water specific heat. We find that we can characterize the two c_p maxima. In particular, the higher T, weak, c_p maxima are associated to the formation of single HB, whose occurrence is revealed by the weak maxima of P_2 for bulk water.

The strong c_p maxima is instead the consequence of a cooperative rearrangement of the HB network as revealed by its overlap in the P-T plane with the maxima of the mean cluster size S and the strong maxima of P_2.

Because the two c_p maxima merge and diverge at the liquid-liquid critical point, our results show that we can fully characterize the liquid-liquid phase transition with the correlated percolation theory.

Our comparison of the percolation quantities calculated in the bulk with those calculated including the protein hydration shell reveals that the protein favors the formation of HB clusters at the interface. Indeed P_2 and P_3 weak maxima appear at the interface at higher temperatures than in the bulk, with an effect much more clearer for the P_2. This can be understood because the interface favors the formation of clusters due to the stronger HB interaction near hydrophobic residues.

It is worth to noticing that our results show clearly that the clusters on the protein interface are much less cooperative than those in the bulk, because P_2 at the interface does not have strong maxima, which are related to the cooperative interaction. This could be rationalized by considering that water cooperativity is enhanced by the formation of a full HB network, while near an interface water loses HB. This can be related to the fact that water near a hydrophobic surfaces forms a liquid layer even at $T < 0°C$ regardless the formation of ice at further distance.

Finally, we find that the different configurations of proteins do not affect much the behavior of the maxima on the P-T plane. This is probably due to the fact that our toy protein is completely hydrophobic, with no strong differences between the folded and the unfolded case.

Acknowledgments

I like to thank Prof. Giancarlo Franzese and Dr. Valentino Bianco for their patience and effort answering my doubts and helping me finishing the project. Special thanks too to my family and friends.

[1] V. Bianco, Statistical physics of water in hydrophobic nano-confinement and at proteins interfaces, Tesis Doctoral, Ed. Barcelona: Universitat de Barcelona, 2013.

[2] G. Franzese, H. E. Stanley, Liquid-liquid critical point in a Hamiltonian model for water: analytic solution, *J. Phys.: Conden. Matter* 14, 2201-2209 (2002).

[3] V. Bianco, G. Franzese, Critical behavior of a water monolayer under hydrophobic confinement, *Sci. Rep.* 4, 4440; DOI: 10.1038/srep04440 (2014).

[4] V. Bianco, G. Franzese, Contribution of Water to Pressure and Cold Denaturation of Proteins, *PRL* 115, 108101; DOI: 10.1103/PhysRevLett.115.108101 (2015).

[5] P. Ball, Water: Water – and enduring mystery. *Nature*, 452(7185):291-292, 2008.

[6] A. Nilsson, et al. Resonant inelastic X-ray scattering of liquid water. *J. electron. Spectrosc.* 188, 84-10 (2012).

[7] A. Taschin, P. Bartolini, R. Eramo, R. Righini, R. Torre, Evidence of two distinct local structures of water from ambient to supercooled conditions. *Nature Comm.* 4, 2401 (2013).

[8] P. H. Poole, F. Sciortino, U. Essman, H. E. Stanley, Phase-Behavior Of Metastable Water. *Nature* 360, 324 (1992).

[9] O. Mishima, H. E. Stanley, The relationship between liquid, supercooled and glassy wate. *Nature* 396, 329 (1998).