

The Multiple Roles of Waters in Protein Solvation

Adam Hospital^{1,2}, Michela Candotti^{1,2}, Josep Lluís Gelpí^{2,3,4*} and

Modesto Orozco^{1,2,3,*}

ABSTRACT

Extensive molecular dynamics (MD) simulations have been used to characterize the multiple roles of water in solvating different types of proteins under different environmental conditions. We analyzed a small set of proteins, representative of the most prevalent meta-folds under native conditions, in the presence of crowding agents, and at high temperature with or without high concentration of urea. We considered also a protein in the unfolded state as characterized by NMR and atomistic MD simulations. Our results outline the main characteristics of the hydration environment of proteins and illustrate the dramatic plasticity of water, and its chameleonic ability to stabilize proteins under a variety of conditions.

¹ Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, Barcelona, 08028, Spain

² Joint BSC-IRB Research Program in Computational Biology, Baldiri Reixac 10-12, Barcelona, 08028, Spain.

³ Department of Biochemistry and Molecular Biomedicine, University of Barcelona, Barcelona, 08028, Spain.

⁴ Barcelona Supercomputing Center (BSC), Jordi Girona 29, Barcelona, 08034, Spain

* Correspondence to Modesto Orozco modesto.orozco@irbbarcelona.org or Josep Lluís Gelpí gelpi@ub.edu, Phone nº +34 93 403 71 55

INTRODUCTION

Water is the main solvent of life and the responsible of defining the native structure of proteins. More than 40 years¹ of intense experimental research have outlined key characteristics of strongly bound water molecules²⁻⁵, but the description of the hydration atmosphere generated by mobile and fluctuating water molecules remained more challenging for experimental techniques^{1, 6-9}, fueling the use of theoretical approaches, particularly molecular dynamics (MD)^{10, 11} as an alternative and often complementary source of information on the average characteristics of protein hydration (as examples see¹²⁻¹⁶). Thanks to the combination of spectroscopic and MD techniques some aspects of protein hydration have been described, such as the lack of correlation between strong-interacting sites and water residence times¹⁷, the dynamics of ultra-slow water molecules trapped in protein cavities or channels¹⁷⁻²⁰, or the robustness of the protein hydration to changes in the aminoacidic composition of the proteins^{21, 22}.

The amount of information on the impact of environmental changes in protein hydration is more limited, and many aspects of the impact of co-solvents or of the increase in temperature remain unclear. It is generally accepted that perturbation of the protein by the presence of co-solvents induces certain changes in the structure and dynamics of waters around proteins. For example, it has been suggested^{20, 23-25} that inert crowding leads to a reduction in the dynamics of water around the protein, mimicking perhaps the situation found in the cellular cytoplasm²⁰. Denaturing solvents such as urea are suggested to affect the hydration on the surface of proteins, especially around apolar groups²⁰, but as the presence of denaturing is linked to unfolding it has been difficult to distinguish between co-solvent and unfolding effects²⁶.

Most work on protein hydration has focused on well-folded proteins and the works focused on the study of the solvent environment in unfolded or intrinsically disordered proteins are rare^{20, 22}, which maintains open the debate on the nature of hydration around hydrophobic residues^{20, 22} and the potential existence of highly ordered clathrate-like structures in unfolded proteins. The difficulties to uncouple protein and water dynamics in unfolded proteins might be on the basis of the difficulties for both theory and simulation to trace hydration in proteins without a well-defined structure. In summary, decades of work have provided a significant

amount of fragmentary information on protein hydration, but crucial aspects, such as the dynamic nature of the inner hydration layer, the plasticity of water to changes in temperature, presence of co-solvents, or to complete unfolding still remain controversial^{7, 9, 20, 27}. In other words, despite all the work done, we still lack a full understanding of water plasticity.

We present here a wide theoretical study on the protein hydration. By using atomistic MD simulations on a reduced set of highly representative proteins, we determine how hydration changes depending on protein fold, and on changes in environment such as increasing the temperature, adding urea or crowding agents. Additionally, we explore the nature of hydration in one of the few proteins, Ubiquitin, whose unfolded ensemble has been characterized²⁸ by a combination of NMR data and extended MD simulations. In all the cases, to isolate as much as possible, water dynamics from protein dynamics, MD simulations are extended for long periods, but the analysis of waters is performed in relatively small time windows, where the water movement is expected to be prevalent over slow protein rearrangements.

Overall, our results provide an accurate (static and dynamic) picture of hydration of different folds under different conditions, and illustrate the high dynamics and plasticity of water, which respond to a variety of conditions keeping always a good ability to keep solvated the protein. Static and especially dynamic properties of waters change very significantly as response of changes in temperature or the presence of co-solvents. Furthermore, water properties are not very dependent on the equilibrium folding of proteins, but are largely dependent on whether the protein is folded or denatured. Clearly, water is not an infinite continuum media which remain unaltered by the presence of proteins, but a plastic and complex solvent affecting protein structure and dynamics that in turn is largely affected by the presence of the macromolecule.

METHODS

Simulated systems and simulation details. We first explored the dependence of hydration on protein fold using three proteins representative of the most frequents meta-folds according to SCOP classification^{16, 29, 30}: PDB 1CQY (Starch binding domain of *Bacillus cereus*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

beta-amylase; Hydrolase) representing “*all β proteins*”, PDB 1OPC (Ompr DNA-binding domain, *Escherichia Coli*; Transcription Regulation) representing “*all α proteins*”, and PDB 1KTE (Thioltransferase; Electric transport) representing “ *α and β proteins*”. Trajectories for these three proteins were taken from our Model Database¹⁶, where they were neutralized by adding Na^+ and Cl^- , thermalized and equilibrated at room temperature³¹, and further subjected to 100 ns of MD simulations at constant pressure (1 atm) and temperature (300 K). Inspection of trajectories (data not shown) demonstrate that proteins were fully equilibrated after such simulation period, allowing then to concentrate on the much faster solvent dynamics, which was followed in much shorter windows (1 ns), to disconnect, as much as possible, solvent and macromolecule dynamics. Truncated octahedral simulation boxes with 5,939 (1KTE), 7,068 (1OPC), and 9,086 (1CQY) water molecules were used (boxes were created to guarantee that distance from any protein atom to the boundaries of the box was always greater than 12Å). Particle Mesh Ewald (PME) and periodic boundary conditions were used to represent long-range electrostatic effects³². All bonds linking hydrogens were frozen using SHAKE³³, which allowed us to use 2 fs time scale for integration of Newton equations of motion. Analysis performed here correspond to simulations carried out with the latest modification of Amber parm99 force-field (P99-SBldn³⁴) and the TIP3P³⁵ water model.

The last structures from the room temperature simulation were used as starting point for simulations at high-temperature. The temperature of the solvent was gradually increased (10 ns window) from 300 to 368K maintaining the protein coordinates constant. At the end of the heating process protein velocities were reassigned to those before heating (to guarantee a smooth transition to high temperature) and the final systems were then thermalized for 10 ns (without any further restraint). Under these conditions proteins should slowly depart from native fold²⁹, but after 20 ns the effect of temperature in protein structure should be still small²⁹. In any case to avoid mixing protein and water dynamics, solvation data was taken from the 1 ns simulation after the 20 ns of heating and equilibration. With a similar philosophy, an equivalent protocol was used to perform simulations for the same set of proteins at high temperature, but in the presence of 8 M urea.

In order to analyze crowding effect on protein hydration we collected trajectories of the three proteins in aqueous solution of PEG²³. These systems were created by immersing the

aqueous-equilibrated proteins in a pre-equilibrated octahedral box containing water and PEG(500) molecules at a concentration of 200g/L²³. Solvent was relaxed for 50 ns keeping protein fixed at its aqueous conformation, followed by a global thermalization and equilibration for 50 more ns, without any restraint before data collection. As crowding agents tend to stabilize proteins, no major contamination of protein and solvent dynamics is expected in these simulations, but for consistency with the rest of simulations data is collected using the last 1 ns of trajectories. Details of simulations are as before, using modified TrapPPE-UA parameters described by Fischer and colleagues for PEG molecules³⁶.

Finally, to complete our study we analyzed the solvent environment in one of the few proteins for which the unfolded ensemble has been characterized by a combination of NMR and MD simulations: Ubiquitin²⁸. We took 10 of the most representative structures in the collected ensemble transferring them from denaturing conditions to aqueous solution as described elsewhere²⁸. Simulation parameters used were taken from our MoDEL database¹⁶ as before: Na⁺ and Cl⁻ to neutralize the system, truncated octahedral simulation boxes (12Å from the protein), PME and periodic boundary conditions³², SHAKE³³ to freeze all bonds linking hydrogens (2 fs time scale), and NPT ensemble with the Amber P99-SBldn³⁴ forcefield and the TIP3P³⁵ water model. As re-folding happens very fast in aqueous solution, we equilibrated water for 10 ns keeping the protein conformation fixed, followed by thermalizing and equilibrating the system for just 5 ns, before 1 ns of data collection. With these windows the contamination of protein dynamics on solvent dynamics is moderated as shown in Suppl. Figure S1, indicating that we are collecting “bona fide” dynamics of water under protein unfolding conditions.

Amber 14 MD package³⁷ was used to run all simulations except for the crowded environment, where trajectories were produced using GROMACS v4.6.5 MD package^{23, 38}.

Analysis. The mining of trajectories was performed using analysis tools in GROMACS³⁸, AMBER^{37, 39} and our MDWeb³¹, FlexServ⁴⁰ and MODEL¹⁶ servers, as well as “in house” scripts. Main analysis set include:

Radial Distribution Function (RDF): They were used to obtain a static, distance-dependent picture of protein solvation:

$$g_{PS}(r) = \frac{1}{\rho 4\pi r^2 dr} \sum_S \langle \delta(r - |r_S - r_P|) \rangle \quad , \quad (1)$$

being ρ the average water density. The summatory corresponds to the number of solvent molecules within a spherical shell of radii between r and $r + dr$, measured from the reference site P in the protein (only exposed ($ASA > 30 \text{ \AA}^2$) O and N atoms were considered as P-sites). Index S denotes either the water hydrogen or the water oxygen atoms⁴¹. To avoid bias in the analysis related to the different molarity of water in the presence or absence of co-solvents, or at different temperatures, the density normalized RDF ($_{dn}RDF$) was determined by normalizing to the unity the RDFs at the largest distance measured (8 \AA).

Water-protein interactions and dynamics: The number of contacts, the type of residues attached to, the water surface-residence times and hydrogen bond behavior of the water molecules at the surface of the protein were analyzed considering that a water molecule was on the surface of the protein when the shortest protein-water distance was smaller than 5 \AA . For determining protein-water hydrogen bonds conformation, only the distance between the oxygen atom of the water molecule and the nearest protein atom (either heavy or hydrogen) was considered, which allowed us to distinguish between waters acting as hydrogen bond (H-bond) donor or acceptor.

Water Diffusion: The isotropic diffusion coefficient of water was computed from the time-dependent mean square displacements (MSD):

$$6Dt = \langle |\vec{r}(t) - \vec{r}(0)|^2 \rangle \quad t \rightarrow \infty, \quad (2)$$

where $\vec{r}(t)$ is the position vector of the solvent molecule at time t , and the brackets indicate that the average is taken over both the time origins and solvent molecules. As TIP3P usually overestimates the diffusion coefficient of water, all the values obtained were also scaled by $1/2.56$, a factor that corrects the diffusion coefficient of TIP3P in a box of pure water to the experimental diffusion of waters.

Residue preferences of water. In order to obtain an estimate of the preference of water for different residues, independent of their accessibility, we computed:

$$Pref_{resType} = \frac{\sum_{i=1}^n \left(\frac{\sum_{j=1}^m numSnapshotsBound_j}{ASA_i} \right)}{N * n}, \quad (3)$$

where $i=1..n$ are all the residues of type *resType* (for example ARG) for the analyzed system (the three folded proteins in a particular solvent environment or the 10 different unfolded conformations of ubiquitin); $j=1..m$ are all the atoms of a given residue; ASA_i is the Accessible Surface Area (absolute value computed with NACCESS⁴² program) of the corresponding residue side chain or backbone. Note that we only consider exposed residues ($ASA > 10\text{\AA}^2$); $numSnapshotsBound_j$ is the number of snapshots where atom j is in contact with a particular water molecule (being j the nearest protein atom for a specific water molecule). Note that we only count waters attached to the same protein residue for at least 10 snapshots (10ps). Finally, N is the total number of snapshots. Values computed are in $contacts/ps*\text{\AA}^2$. In order to compare values for different solvent environments, values corresponding to a particular set (e.g. urea) are converted to percentages.

RESULTS AND DISCUSSION

The average hydration atmosphere in proteins was first explored from the analysis of the density-normalized protein-water radial distribution functions ($_{dn}rdfs$; see Methods, and Figure 1) of protein exposed residues. All the $_{dn}rdfs$ plots show a major peak corresponding to direct water-protein contacts (distance around 3 Å), a secondary peak located at around 4.5 Å, and a very shallow remote peak around 7 Å indicating additional hydration layers. Similar $_{dn}rdfs$ plots were found for the three folded proteins, supporting previous claims by other authors²¹ on the robustness of average protein hydration on the nature of the protein fold. This similarity allowed us to combine the three independent plots in a single figure representing the average hydration around folded proteins (Figure 1a). When the temperature is increased (without adding co-solvents) a significant loss of water in the vicinities of the folded proteins occurs (reduction in both the first and second peaks in the $_{dn}rdfs$). As our distributions are normalized these changes indicate a migration of waters from vicinities of the protein to bulk. The addition of urea produces a concentration of water in the vicinities of the protein (compared with simulation in

pure water at the same temperature), as clearly seen in the first peak (increased) and second peak (decreased) of dn^2rdf (Figure 1a). A similar effect of attracting water to the protein surface is evident in PEG, supporting previous spectroscopy-derived claims on the ability of PEG to modify the hydration pattern of proteins²⁵ (see Figure 1a). Finally, it is worth noting that no dramatic differences are found when dn^2rdf s of folded and unfolded proteins (Figure 1a and b) are compared (note that Figure 1b shows average dn^2rdf obtained by the entire ensemble derived from the 10 independent trajectories), which supports the idea that, at least from a static point of view, the static picture of hydration around folded and unfolded proteins might not be so different^{9,20}.

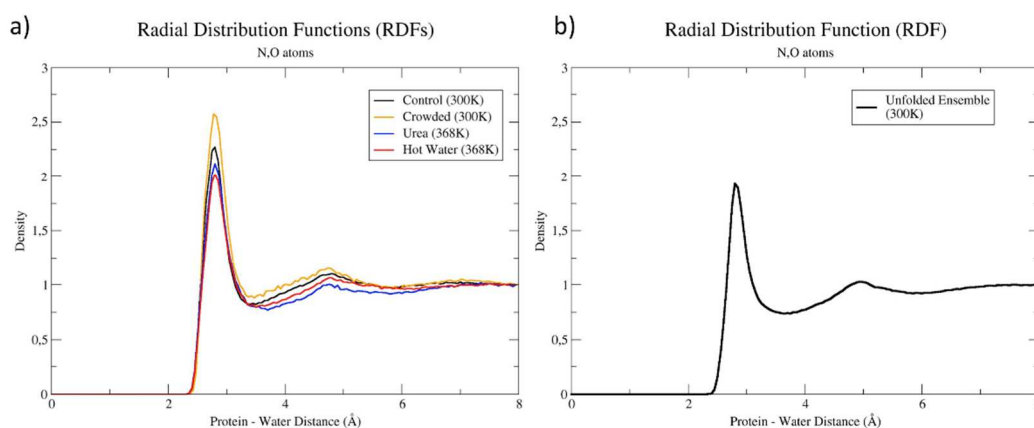


Figure 1. Radial Distribution Functions (RDFs) for exposed Nitrogen and Oxygen atoms of (a) the three folded proteins in the different environments and (b) the unfolded ensemble of ubiquitin. Note that values represented correspond to the average distribution obtained from combining individual dn^2rdf for the three folded proteins, or from the 10 independent trajectories of the unfolded ubiquitin.

An alternative picture of the static water distribution around proteins is obtained by inspecting water molecules that are sensitive to the presence of the protein, i.e. those interacting at least once with the protein in the simulation window. For most analysis purposes we classified those water molecules “feeling” the protein in three categories: i) those that are in very close contact with the protein surface (more than 80% of time in direct contact with the protein in the simulation window; we named them *surface waters*), ii) those that are in frequent contact with the protein (between 20 and 80% of time in direct contact; we named them *exchange waters*) and iii) those contacting the protein less than 20% of the time (we named them *bulk waters*). Note that this classification is arbitrary (longer the collection window greater the number of “bulk” waters), but in terms of comparison between the different environments the results are very

robust (see Suppl. Figures S2-S3). Simulations of folded proteins reveal that only a very small fraction of waters (0.3-0.5%) remains permanently bound to the protein in the collection window (0.5 ns for all these analyses), in agreement with previous claims on the low prevalence of long-living water molecules^{9,18}. These permanently bound water molecules are trapped inside cavities (Suppl. Figure S4). In fact for folded proteins we found one “permanently bound” water every 10-21 residues, consistent with Williams’ estimates⁴³, and with claims by different authors on the geometrical requirements for the placement of long-residence waters¹⁷. Exchange waters (those visiting frequently the first hydration shell) are quite frequent, around 23% of water molecules in the simulation box (see Figure 2). No dramatic changes in the ratio between *exchange/bulk* waters are detected depending on the type of residue. The nature of the fold introduces however small, but not negligible changes, with the β -protein showing (as expected) the lowest percentage of exchange waters (Figure 2). Finally, folding does not introduce dramatic changes in the percentage of waters visiting with a certain frequency the protein surface (*exchange water*), but eliminate tightly bound waters (*surface waters*) and modify in a non-negligible way the relative hydrophilicity of residues compared to folded proteins (Figure 3).

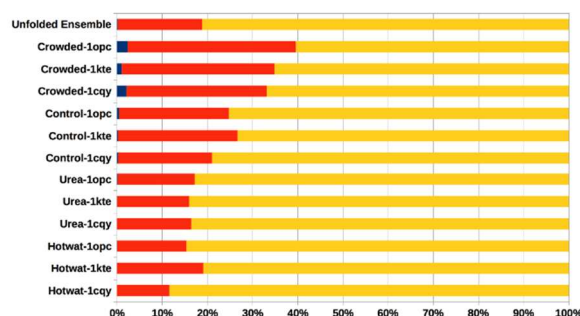


Figure 2. Percentage of Surface (blue), Exchange (red) and Bulk (yellow) water molecules found in the different environments. Data obtained with other analysis windows are shown in Suppl. Figure S2 and S3, if longer windows are used the percentage vary (longer the window, smaller the number of surface waters), but in relative terms results are equivalent.

The increase in the temperature produces the practical disappearance of the tightly bound water molecules, and a significant reduction (from 23 to around 15%; Figure 2) in *exchange waters*. The changes in the relative hydrophilicity of residues occurring upon heating are complex (Figure 3), and probably related to the stochastic effects related to the exposure of apolar atoms as a consequence of the increased thermal motion. Urea leads, in average (Figure

2), to a small increase in the number of tightly bound and frequently interacting water molecules, compared with the simulation in pure water at the same temperature, which can be attributed mostly (see Figure 3) to a greater relative hydrophilicity of charged residues. Finally, PEG leads to a very significant increase in the percentage of water molecules that are tightly bound to the protein (*surface waters* increase from 0.3-0.5% in pure water to 2-3 % in the presence of PEG; see Figure 2), and of those visiting frequently the first hydration shell (*exchange waters* increase from 23 to 34%), confirming the ability of PEG to favor recruitment of waters to the protein surface.

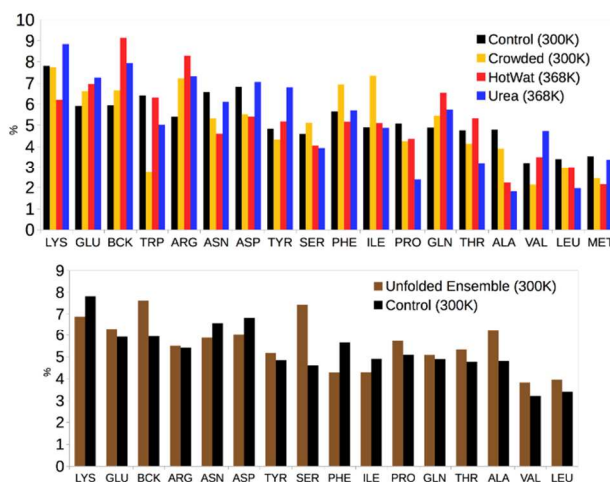


Figure 3. Protein residues preference (in %) for the water molecules in close contact to the surface (values are corrected considering the ASA of the residue as described in Methods). Top plot represents the different environments (urea, crowded and hot water) versus the control (water at T=300K). Bottom plot represents the unfolded average ensemble versus the folded state (300K).

A long standing question in the field^{27, 44, 45} is whether waters interacting with proteins are acting as hydrogen bond donors (HB-donor), or as hydrogen bond acceptors (HB-acceptors). Results in Figure 4 demonstrate that in pure aqueous simulations the water molecules located in the vicinities of folded proteins act 3-5 times more as H-bond donor (distance water(O)-protein around 2.5 Å) than as H-bond acceptor (distance water(O)-protein around 2.0 Å), in agreement with IR-findings by Bolhuis and coworkers⁴⁶. However, if the analysis is focused in long-remaining contacts (defined as the longest of all the different contacts of a water molecule with protein surface atoms in a particular time window), the situation reverses (Figure 4, inset) and

water appears mostly as H-bond acceptor, where water binds to backbone amide group placed in narrow cavities of the protein. In other words, despite the number of acidic and basic residues is similar, there is a curious asymmetry in the role of water as acceptor/donor, which seems to depend on the intensity in which waters interact with the protein. These findings agree with a previous work by Sterpone and coworkers²⁷ which indicate that waters tightly bound to lysozyme might display different H-bond donor/acceptor capabilities than those exhibiting looser contacts with the protein.

The increase in the temperature (Figure 4) leads, as discussed above, to a significant reduction in the number of water molecules visiting the vicinities of the protein, but the few remaining waters display good H-bonds with accessible protein residues. The presence of urea leads to an increase in the acceptor capabilities of water when the entire universe of interacting water molecules is considered, but to the reverse situation when trapped waters are considered. Finally, the inert crowder considered here (PEG) decreases the number of acceptor water molecules, while in parallel increases the donor capabilities of those interacting with protein residues, irrespectively of the strength of the interaction (Figure 4). Very interestingly, the differential analysis of H-bond donor/acceptor capabilities of water shows quite different results for the unfolded ensemble, with water molecules remaining at an increased distance (in average) from the protein atoms. This leads to a decrease of long-standing H-bond contacts, and to the disappearance of the bimodal distribution of H-bond donor/acceptor capabilities found for folded proteins (Figure 4). This probably reflects the fact that unfolded proteins expose hydrophobic residues, which means that a big portion of the surface water molecules is travelling along the surface (Figure 4).

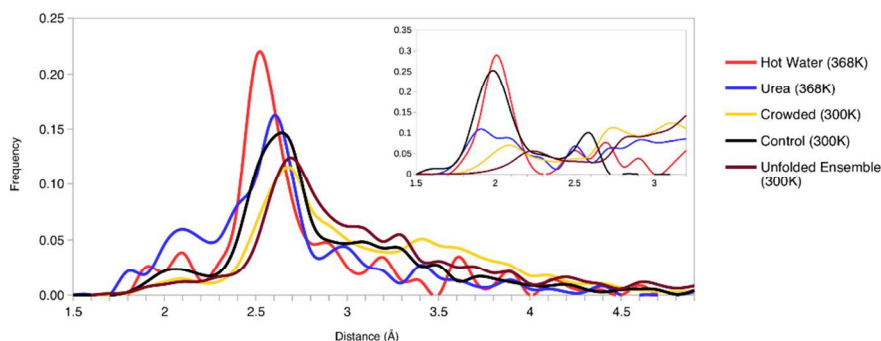


Figure 4. Distribution of protein-water oxygen distances (\AA) (distance from the water oxygen to the nearest atom (hydrogen or heavy) of the protein surface) for all the contacts formed by surface waters divided in the different environments (note that the Figure is normalized, so it is not sensitive to the total number of surface waters). The inset shows the same distribution for the reduced set of long-remaining contacts. Values for the unfolded ensemble are averaged from the 10 different simulations. Note that the number of acidic and basic residues for the proteins studied is similar, and that numbers are quite robust to the type of protein fold.

In summary, static analysis of the hydration atmosphere around proteins highlight the ability of water to adapt to the conformational state of the protein and to modify its interactions with the protein depending on the presence of co-solvents and on the temperature.

Dynamic properties of protein hydration. One of the greatest advantages of using MD simulations is that they produce direct dynamic information on the solvent atmosphere, which can be used to answer several long-standing questions in the field. For example: whether waters on the surface of folded proteins are fixed to a single residue (as suggested from the inspection of X-Ray structures) or are sliding over the protein surface. Results reported in Figure 5 demonstrate that water molecules expending long periods of time in close contact with the protein typically visit many residues, suggesting that they are sliding over the surface generating a dynamic solvent atmosphere. Water molecules showing long residence times on the surface of the protein are uncommon for any protein fold, and water molecules continuously bound to a single residue are extremely rare (see Figure 5), as was in fact anticipated from static measures (see above). In summary, the concept of structural waters as defined from X-Ray crystallography does not fit well with the dynamic picture of solvation derived from our simulations where, as discussed by others (for example^{7,9,17-19}), fast exchange of protein-interacting water is a universal trend.

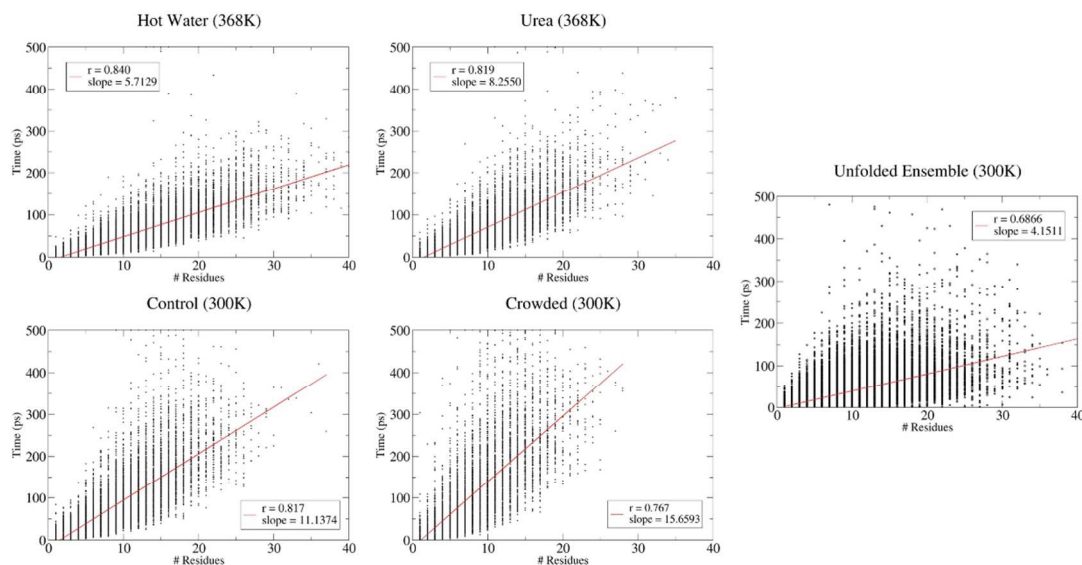


Figure 5. Correlation between total time (within the simulation window) spent for a water molecule in close contact to the protein surface against the number of different protein residues it attaches to. As no dramatic differences were found depending on protein fold, values shown here correspond to averages for the three folded proteins. Values for the ensemble are also averaged, which considering the larger variability of the structures generates a slightly larger noise in the distributions. No data is shown for waters showing longer residence time in the surface, as the number is small and values are quite noisy.

Another question that can be easily answered from MD simulations is how different are the dynamic properties of water molecules in close contact with the protein, to those in pure solvent. This information can be gained from the study of the (isotropic-translational) diffusion coefficients, which can be easily derived from the MD ensembles (see Methods). Results are displayed in Table 1, where we included also values scaled to correct the known tendency of the TIP3P model to maximize diffusion⁴⁷. Note that the scaling does not affect the comparison of water mobility in different environments, but helps in determining apparent temperatures of solvent (see below). Table 1 demonstrates that the presence of the protein slows down the movement of waters, even of those rarely visiting (<20% of time) the protein surface (Diffusion coefficient $2.1 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ compared with $2.3 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ for pure TIP3P water simulations⁴⁷). When the analysis is repeated considering *exchange waters* (i.e. those exploring more than 20% of the time the protein surface) the effect of protein is more evident, as noted in a reduction of the (unscaled) diffusion coefficient to $1.9 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$, and if the analysis is performed considering water molecules expending more than 40% of the time on the surface of the protein

the diffusion coefficient reduces at values around $1.3 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ (values below $1.0 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ are found for the few waters with very long residence times in the surface). These changes mirror the variation in instantaneous velocities (Table 1), which significantly reduce when waters are on the protein surface (some individual waters reduce their instantaneous velocity 2-3 times with respect to solvent; in perfect agreement with independent experimental estimates^{9, 27}). If scaled-TIP3P diffusion coefficients are used in conjunction with experimental curves of the dependence of diffusion of water with the temperature⁴⁸ we can conclude that water molecules on the surface of proteins behave individually as water close to freezing point (around 5°C), moving around 20-30% slower than normal waters and diffusing less efficiently, while waters in cavities show a behavior typical of “supercooled waters”⁷. The dependence of the results on the protein fold is small (in general waters in the β -protein diffuse faster than waters in α -containing proteins), but if the analysis is done for unfolded conformations significant changes emerge: waters move significantly faster, even this not always imply a large diffusion coefficient (Table 1), suggesting that water molecules around unfolded protein have a different dynamic pattern to that found for the water molecules around folded proteins. This result agrees well with MRD-derived suggestion by Qvist et al.²², and warns against the idea of the existence of clathrates in the surroundings of exposed apolar residues^{7, 9}.

		Bulk Water	>20% in Surface (Exchange waters)	>40% in Surface
		$Diff^{TIP3P} / Diff^{Scaled} / Velocity$	$Diff^{TIP3P} / Diff^{Scaled} / Velocity$	$Diff^{TIP3P} / Diff^{Scaled} / Velocity$
Control (300K)	1cgy	5.5982 / 2.1868 / 1.7727	4.9597 / 1.9374 / 1.7228	3.4785 / 1.3588 / 1.2398
	1kte	5.1212 / 2.0005 / 1.7593	4.6233 / 1.8060 / 1.7138	3.3720 / 1.3172 / 1.2554
	1opc	5.3089 / 2.0738 / 1.7643	4.8099 / 1.8789 / 1.7149	3.2010 / 1.2504 / 1.2800
Hot Water (368K)	1cgy	11.2384 / 4.3900 / 2.3819	10.2530 / 4.0051 / 2.3326	4.8734 / 1.9037 / 1.2691
	1kte	10.6360 / 4.1547 / 2.3673	10.4097 / 4.0663 / 2.3246	7.2243 / 2.8220 / 1.4569
	1opc	11.1626 / 4.3604 / 2.3684	9.8455 / 3.8459 / 2.3231	8.5066 / 3.3229 / 1.2825
Urea	1cgy	7.2980 / 2.8508 / 2.1018	6.8339 / 2.6695 / 2.0508	4.4239 / 1.7281 / 1.2948

(368K)	1kte	7.2693 / 2.8396 / 2.0961	6.6196 / 2.5858 / 2.0381	4.7027 / 1.8370 / 1.3099
	1opc	7.1275 / 2.7842 / 2.0948	6.6160 / 2.5844 / 2.0358	5.1166 / 1.9987 / 1.5667
Crowded (300K)	1cgy	2.5341 / 0.9899 / 1.5168	2.2458 / 0.8773 / 1.4667	1.5398 / 0.6015 / 1.1969
	1kte	2.6721 / 1.0438 / 1.5333	2.4785 / 0.9682 / 1.5038	1.6611 / 0.6489 / 1.2170
	1opc	2.7440 / 1.0719 / 1.5198	2.2888 / 0.8941 / 1.4908	1.4062 / 0.5493 / 1.1905
Unfolded (300K)	AVG	5.7702 / 2.2540 / 1.9506	5.5116 / 2.1530 / 1.9235	3.1121 / 1.2157 / 1.7548

Table 1. Diffusion coefficients and water velocities for the different environments. Water molecules are divided in different groups according to the time they spent on the surface of the protein (data is presented separately for *bulk waters*, *exchange waters* and a subset of exchange waters expending more than 40% of the time in contact with the surface). Diff^{TIP3P} corresponds to diffusion values obtained directly from MD simulations; Diff^{Scaled} corresponds to diffusion values corrected to their experimental diffusion range (see analysis). Diffusion values are in $\times 10^{-9} \text{ m}^2/\text{s}$, velocities are in $\text{\AA} / \text{ps}$. Values for the unfolded conformations are averaged from the 10 different simulations. Large variability in the results obtained for water molecules in contact with the protein more than 40% of the time in simulations at high temperature is related to the statistical noise derived from the small number of water molecules in this subset.

The increase in the temperature of the simulation leads to a very significant enlargement of the diffusion coefficient although the relative cooling of waters found on the protein surface is still visible (Table 1). Urea reduces the diffusion coefficient (with respect to simulations in water at the same temperature), and at a lower extend, the instantaneous water velocities, irrespectively of whether they are close or far from the protein surface. Finally, the effect of crowding is dramatic, with waters showing diffusion coefficients below those expected for ice (Table 1 and reference⁴⁸) even when only “bulk” waters are considered. Results suggest that the large viscosity of PEG solutions makes water molecules to display dynamic properties very far from those expected at the simulation temperature in pure aqueous solution. These results, which confirm previous experimental claims on the extreme “cooling effect” of crowders²⁴ suggest that “crowding” physiological conditions might imply ultra-slow water molecules²⁰, but caution is needed in extrapolating PEG results to physiological conditions²³ as water molecules in PEG solutions might behave in a very different way than those in biologically relevant environments.

The diffusion coefficient is a very useful general descriptor of water mobility, but its isotropic nature makes it sub-optimal to characterize the movements of water molecules on the protein surface. We extended then the analysis of water mobility by looking at the ability of water molecules to visit protein residues taken from the slope of the regression line between the total times of water in the surface versus the number of visited residues (Figure 5). For folded proteins waters sample approximately 1 residue every 11 ps, while at high temperature the sampling increase up to 1 residue every 6 ps. The presence of urea reduces the residue sampling velocity to around 30% (comparing always simulations at equal temperature), while crowding freezes the surface waters to an average speed of 16 ps x residue. Finally, as expected from previous analysis, water molecules over the surface of unfolded conformations differ very significantly to those in folded proteins. As Figure 5 shows, the number of long-residence water molecules is reduced when protein is unfolded, and the speed of sampling surface residues is dramatically increased (waters explore in average 1 residue every 4 ps). This high water mobility on the surface is originated by the prevalence of exposed hydrophobic residues with weak interactions with water, arguing again against the chlatrate model. This scenario, with water acting as lubricant, might facilitate the fast protein movements required for the structural reorganization and hydrophobic collapse required for protein folding.

CONCLUSIONS

We present a systematic study on the nature of water around different proteins at physiological and not physiological conditions. The study faced the technical problems derived from the coupling between protein and solvent dynamics, which can be especially severe in the case of folded proteins in unfolding conditions, or denatured proteins in folding conditions. We alleviated this problem by taking water ensembles from short simulation window times, where water dynamics is expected to equilibrate, but protein movements are expected to be reduced. We found that water is an extremely chameleonic molecule able to change its interacting patterns and dynamic properties depending on the folded state of the protein, and on the environmental conditions. For example, water properties are very different when the protein is folded and water

can slowly slide across a highly polar surface making strong hydrogen bond contacts, from those existing when the protein is unfolded, and water moves quickly for short periods of time on a heterogeneous surface where hydrophobic residues are exposed, disfavoring the formation of strong hydrogen bond contacts, and provoking a certain effective dehydration of the surface. The presence of co-solvents induces dramatic changes in the properties of water molecules around the proteins, which cannot be always explained by the reduction in the effective water molarity. The two co-solvent molecules considered here favor the migration of water molecules from bulk to the protein surface and modify the ratio between H-bond donor and acceptor roles of water, but more importantly, induce changes in the mobility of water. A dramatic effect is found for PEG500 which freezes the mobility of water to values below those expected for ice. This finding warns on the suitability of pure aqueous solutions as a mimic of physiological conditions, but also warns on the use of highly viscous co-solvents to simulate physical effects such as unfolding or crowding in physiological conditions. Overall our results suggest that different water environments around proteins might exist depending on the folded state of the protein and the presence of co-solvents.

Supporting Information: Root Mean Square deviation (RMSd) and Radius of Gyration (R_{gyr}) values for the 10 unfolded Ubiquitin simulations; Percentage of surface, exchange and bulk water molecules found in the different environments in different time windows: 500ps, 1ns, 2ns and 5ns; Decrease in the percentage of surface water molecules looking at different time windows: 500ps, 1ns, 2ns and 5ns; Trapped water molecules found in surface and internal cavities.

ACKNOWLEDGMENTS

The authors thank INB (Instituto Nacional de Bioinformática) for support in this project. Additional funding to this project comes from H2020 (BioExcel GA 675728, Elixir-Excelerate GA 676559) and the Spanish MINECO (BIO2015/64802-R) funding agencies. IRB Barcelona is recipient of a Severo Ochoa Award of Excellence from MINECO (Government of Spain, GA SEV-2015-050). MO is an ICREA Academia researcher.

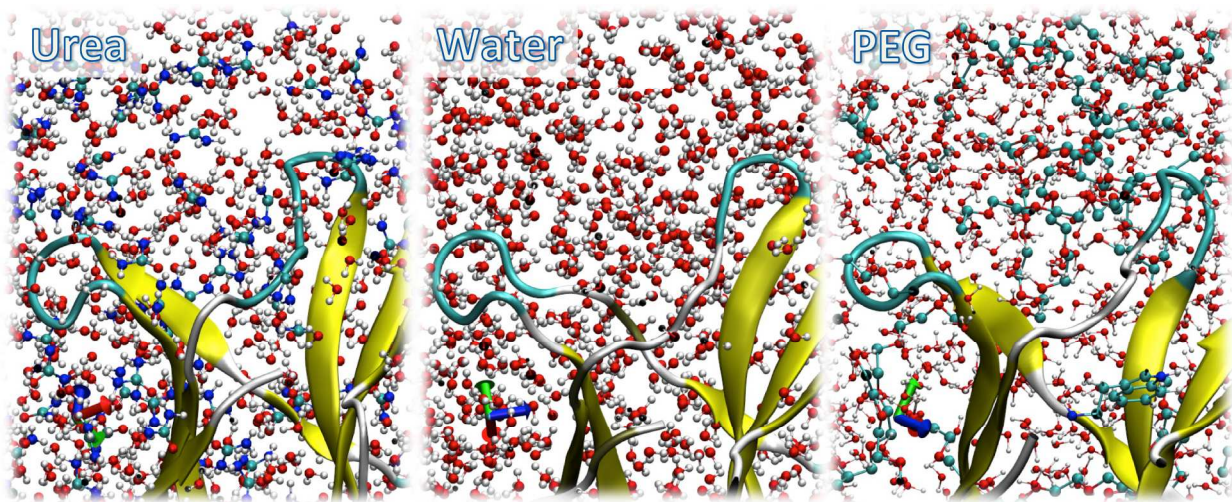
REFERENCES

1. Kuntz, I. D.; Kauzmann, W. Hydration of Proteins and Polypeptides. *Adv. Protein Chem.* **1974**, *28*, 239-345.
2. Quiocho, F. A.; Wilson, D. K.; Vyas, N. K. Substrate Specificity and Affinity of a Protein Modulated by Bound Water Molecules. *Nature* **1989**, *340*, 404-7.
3. Gerwert, K.; Freier, E.; Wolf, S. The Role of Protein-Bound Water Molecules in Microbial Rhodopsins. *Biochim. Biophys. Acta* **2014**, *1837*, 606-13.
4. Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. Bound Water Molecules and Conformational Stabilization Help Mediate an Antigen-Antibody Association. *Proc. Natl. Acad. Sci. U S A* **1994**, *91*, 1089-93.
5. Smith, K. J.; Reid, S. W.; Harlos, K.; McMichael, A. J.; Stuart, D. I.; Bell, J. I.; Jones, E. Y. Bound Water Structure and Polymorphic Amino Acids Act Together to Allow the Binding of Different Peptides to MHC Class I HLA-B53. *Immunity* **1996**, *4*, 215-28.
6. Bagchi, B. Water Dynamics in The Hydration Layer Around Proteins and Micelles. *Chem. Rev.* **2005**, *105*, 3197-219.
7. Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008**, *108*, 74-108.
8. Fogarty, A. C.; Duboué-Dijon, E.; Sterpone, F.; Hynes, J. T.; Laage, D. Biomolecular Hydration Dynamics: A Jump Model Perspective. *Chem. Soc. Rev.* **2013**, *42*, 5672-83.
9. Halle, B. Protein Hydration Dynamics in Solution: A Critical Survey. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2004**, *359*, 1207-23.
10. Brooks, C. L.; Karplus, M.; Pettitt, B. M. *Proteins : A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*. Wiley: New York ; Chichester, 1988.
11. McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585-90.
12. Bizzarri, A. R.; Cannistraro, S. Molecular Dynamics of Water at the Protein-Solvent Interface. *J. Phys. Chem. B* **2002**, *106*, 6617-6633.
13. Pal, S.; Bandyopadhyay, S. Importance of Protein Conformational Motions and Electrostatic Anchoring Sites on the Dynamics and Hydrogen Bond Properties of Hydration Water. *Langmuir* **2013**, *29*, 1162-73.
14. Pizzitutti, F.; Marchi, M.; Sterpone, F.; Rossky, P. J. How Protein Surfaces Induce Anomalous Dynamics of Hydration Water. *J. Phys. Chem. B* **2007**, *111*, 7584-90.
15. Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpí, J. L.; Orozco, M. A Consensus View of Protein Dynamics. *Proc. Natl. Acad. Sci. U S A* **2007**, *104*, 796-801.
16. Meyer, T.; D'Abramo, M.; Hospital, A.; Rueda, M.; Ferrer-Costa, C.; Pérez, A.; Carrillo, O.; Camps, J.; Fenollosa, C.; Repchevsky, D., et al. MoDEL (Molecular Dynamics Extended Library): A Database of Atomistic Molecular Dynamics Trajectories. *Structure* **2010**, *18*, 1399-409.
17. Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys. J.* **2000**, *79*, 2966-2974.
18. García, A. E.; Hummer, G. Water Penetration and Escape in Proteins. *Proteins* **2000**, *38*, 261-72.
19. Kaieda, S.; Halle, B. Internal Water and Microsecond Dynamics in Myoglobin. *J. Phys. Chem. B* **2013**, *117*, 14676-87.
20. Ball, P. Water as a Biomolecule. *ChemPhysChem* **2008**, *9*, 2677-85.
21. Fogarty, A. C.; Laage, D. Water Dynamics in Protein Hydration Shells: The Molecular Origins of the Dynamical Perturbation. *J. Phys. Chem. B* **2014**, *118*, 7715-29.
22. Qvist, J.; Ortega, G.; Tadeo, X.; Millet, O.; Halle, B. Hydration Dynamics of a Halophilic Protein in Folded and Unfolded States. *J. Phys. Chem. B* **2012**, *116*, 3436-44.

23. Candotti, M.; Orozco, M. The Differential Response of Proteins to Macromolecular Crowding. *PLoS Comput. Biol.* **2016**, *12*, e1005040.
24. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J. Crowding Induced Collective Hydration of Biological Macromolecules Over Extended Distances. *J. Am. Chem. Soc.* **2014**, *136*, 188-94.
25. Samanta, N.; Luong, T. Q.; Das Mahanta, D.; Mitra, R. K.; Havenith, M. Effect of Short Chain Poly(ethylene glycol)s on the Hydration Structure and Dynamics around Human Serum Albumin. *Langmuir* **2016**, *32*, 831-7.
26. Modig, K.; Kurian, E.; Prendergast, F. G.; Halle, B. Water and Urea Interactions with the Native and Unfolded Forms of a Beta-Barrel Protein. *Protein Sci.* **2003**, *12*, 2768-81.
27. Sterpone, F.; Stirnemann, G.; Laage, D. Magnitude and Molecular Origin of Water Slowdown Next to a Protein. *J. Am. Chem. Soc.* **2012**, *134*, 4116-9.
28. Candotti, M.; Esteban-Martín, S.; Salvatella, X.; Orozco, M. Toward an Atomistic Description of the Urea-Denatured State of Proteins. *Proc. Natl. Acad. Sci. U S A* **2013**, *110*, 5933-8.
29. Candotti, M.; Pérez, A.; Ferrer-Costa, C.; Rueda, M.; Meyer, T.; Gelpí, J. L.; Orozco, M. Exploring Early Stages of the Chemical Unfolding of Proteins at the Proteome Scale. *PLoS Comput. Biol.* **2013**, *9*, e1003393.
30. Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: A Structural Classification of Proteins Database for The Investigation of Sequences and Structures. *J. Mol. Biol.* **1995**, *247*, 536-40.
31. Hospital, A.; Andrio, P.; Fenollosa, C.; Cicin-Sain, D.; Orozco, M.; Gelpí, J. L. MDWeb And MDMoby: An Integrated Web-Based Platform for Molecular Dynamics Simulations. *Bioinformatics* **2012**, *28*, 1278-9.
32. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
33. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
34. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99sb Protein Force Field. *Proteins* **2010**, *78*, 1950-8.
35. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.
36. Fischer, J.; Paschek, D.; Geiger, A.; Sadowski, G. Modeling of Aqueous Poly(Oxyethylene) Solutions: 1. Atomistic Simulations. *J. Phys. Chem. B* **2008**, *112*, 2388-98.
37. Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; T.E. Cheatham, I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W., et al. *AMBER 2016*, University of California, San Francisco, 2016.
38. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
39. Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084-95.
40. Camps, J.; Carrillo, O.; Emperador, A.; Orellana, L.; Hospital, A.; Rueda, M.; Cicin-Sain, D.; D'Abramo, M.; Gelpí, J. L.; Orozco, M. FlexServ: an Integrated Tool for the Analysis of Protein Flexibility. *Bioinformatics* **2009**, *25*, 1709-10.
41. Schröder, C.; Rudas, T.; Boresch, S.; Steinhauser, O. Simulation Studies of the Protein-Water Interface. I. Properties at the Molecular Resolution. *J. Chem. Phys.* **2006**, *124*, 234907.
42. Hubbard, S. J. a. T., J.M. *NACCESS Computer Program*, University College London: Department of Biochemistry and Molecular Biology, 1993.

43. Williams, M. A.; Goodfellow, J. M.; Thornton, J. M. Buried Waters and Internal Cavities in Monomeric Proteins. *Prot. Sci.* **1994**, *3*, 1224-1235.
44. Juranić, N.; Likić, V. A.; Prendergast, F. G.; Macura, S. Protein–Solvent Hydrogen Bonding Studied by NMR $^1\text{JNC}'$ Coupling Constant Determination and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1996**, *118*, 7859-7860.
45. Petukhov, M.; Rychkov, G.; Firsov, L.; Serrano, L. H-bonding in Protein Hydration Revisited. *Protein Sci.* **2004**, *13*, 2120-2129.
46. Brotzakis, Z. F.; Groot, C. C.; Brandeburgo, W. H.; Bakker, H. J.; Bolhuis, P. G. Dynamics of Hydration Water around Native and Misfolded α -Lactalbumin. *J. Phys. Chem. B* **2016**, *120*, 4756-66.
47. Hertz, H. In *Water: A Comprehensive Treatise*, Vol. 3; F. Franks, ed. Plenum Press, New York: 1973.
48. Mills, R. Self-diffusion in Normal And Heavy Water in The Range 1-45°. *J. Phys. Chem.* **1973**, *77*, 685-688.

TOC GRAPHIC



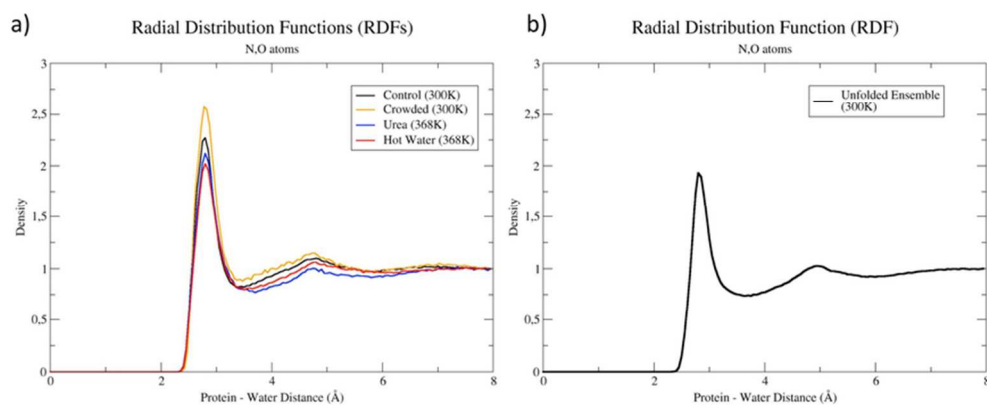


Figure 1. Radial Distribution Functions (RDFs) for exposed Nitrogen and Oxygen atoms of (a) the three folded proteins in the different environments and (b) the unfolded ensemble of ubiquitin. Note that values represented correspond to the average distribution obtained from combining individual dnrdf for the three folded proteins, or from the 10 independent trajectories of the unfolded ubiquitin.

Figure 1
74x31mm (300 x 300 DPI)

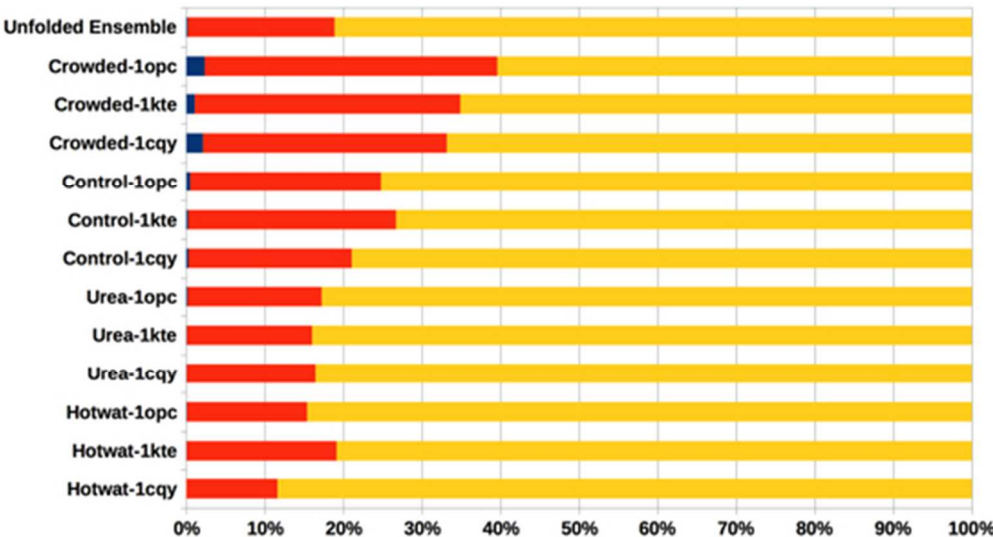


Figure 2. Percentage of Surface (blue), Exchange (red) and Bulk (yellow) water molecules found in the different environments. Data obtained with other analysis windows are shown in Suppl. Figure S2 and S3, if longer windows are used the percentage vary (longer the window, smaller the number of surface waters), but in relative terms results are equivalent.

Figure 2
46x26mm (300 x 300 DPI)

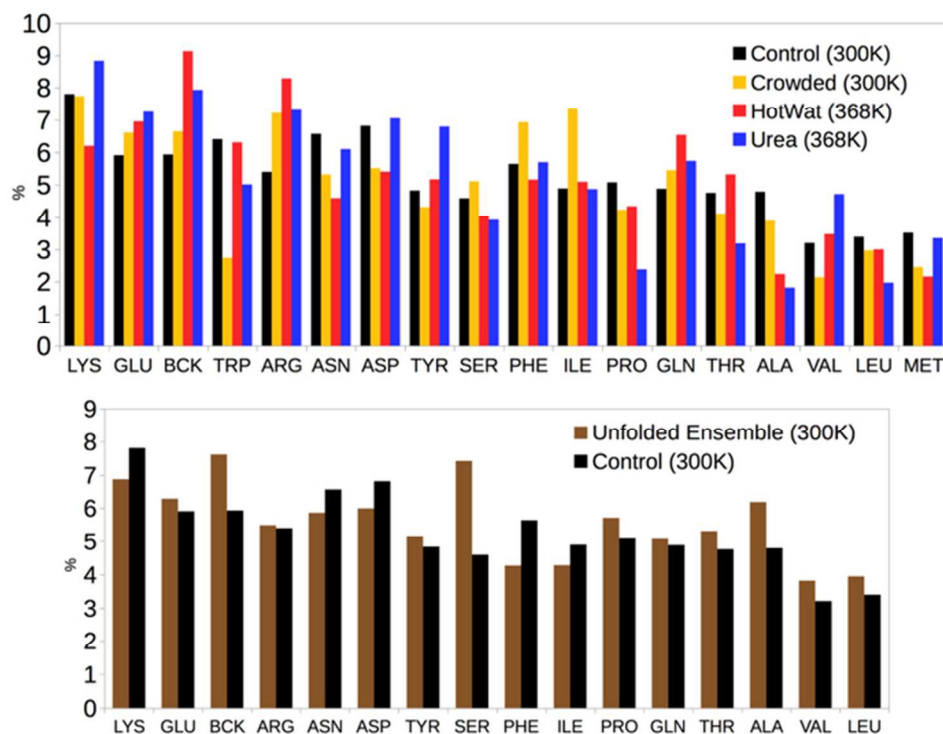


Figure 3. Protein residues preference (in %) for the water molecules in close contact to the surface (values are corrected considering the ASA of the residue as described in Methods). Top plot represents the different environments (urea, crowded and hot water) versus the control (water at $T=300\text{K}$). Bottom plot represents the unfolded average ensemble versus the folded state (300K).

Figure 3

61x46mm (300 x 300 DPI)

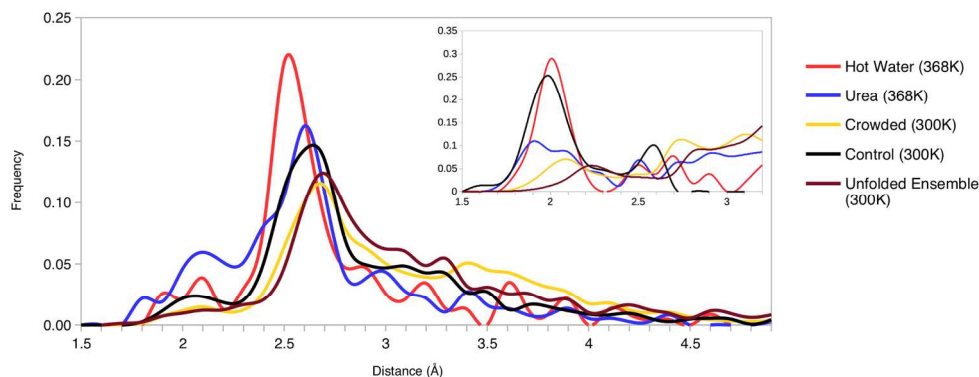


Figure 4. Distribution of protein-water oxygen distances (\AA) (distance from the water oxygen to the nearest atom (hydrogen or heavy) of the protein surface) for all the contacts formed by surface waters divided in the different environments (note that the Figure is normalized, so it is not sensitive to the total number of surface waters). The inset shows the same distribution for the reduced set of long-remaining contacts. Values for the unfolded ensemble are averaged from the 10 different simulations. Note that the number of acidic and basic residues for the proteins studied is similar, and that numbers are quite robust to the type of protein fold.

Figure 4
177x72mm (300 x 300 DPI)

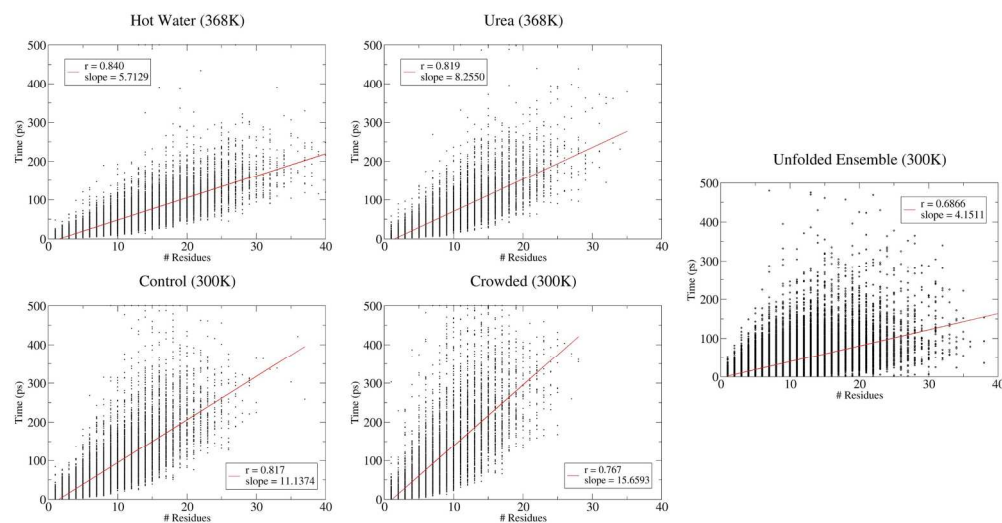


Figure 5. Correlation between total time (within the simulation window) spent for a water molecule in close contact to the protein surface against the number of different protein residues it attaches to. As no dramatic differences were found depending on protein fold, values shown here correspond to averages for the three folded proteins. Values for the ensemble are also averaged, which considering the larger variability of the structures generates a slightly larger noise in the distributions. No data is shown for waters showing longer residence time in the surface, as the number is small and values are quite noisy.

Figure 5
177x95mm (300 x 300 DPI)