Role of Internal Water on Protein Thermal Stability: The Case of Homologous G Domains.

Obaidur Rahaman, Maria Kalimeri, Simone Melchionna, Jérôme Hénin, Fabio Sterpone

Published in final edited form as:
J Phys Chem B. 2015 July 23; 119(29): 8939–8949. doi:10.1021/jp507571u.

On the Role of Internal Water on Protein Thermal Stability: the Case of Homologous G-domains

Obaidur Rahaman[†], **Maria Kalimeri**[†], **Simone Melchionna**[*,‡], **Jérôme Hénin**[*,†], and **Fabio Sterpone**[*,†]

[†]Laboratoire de Biochimie Théorique, IBPC, CNRS, UPR9080, Univ. Paris Diderot, Sorbonne Paris Cité, 13 rue Pierre et Marie Curie, 75005, Paris, France

[‡]CNR-IPCF, Consiglio Nazionale delle Ricerche, Physics Dept., Univ. La Sapienza, P.le A. Moro 2, 00185, Rome, Italy

Abstract

In this work we address the question whether the enhanced stability of thermophilic proteins has a direct connection with internal hydration. Our model systems are two homologues G-domains of different stability: the mesophilic G domain of the Elongation-Factor thermal unstable protein from *E. coli* and the hyperthermophilic G domain of the EF-α protein from *S. solfataricus*. Using molecular dynamics simulation at the microsecond time-scale we show that both proteins host water molecules in internal cavities and that these molecules exchange with the external solution in the nanosecond time scale. The hydration free energy of these sites evaluated via extensive calculations is found to be favourable for both systems, with the hyperthermophilic protein offering a slightly more favourable environment to host water molecules. We estimate that under ambient conditions, the free energy gain due to internal hydration is about 1.3 kcal/mol in favor of the hyperthermophilic variant. However, we also find that at the high working temperature of the hyperthermophile, the cavities are rather dehydrated, meaning that at extreme conditions other molecular factors secure the stability of the protein. Interestingly, we detect a clear correlation between the hydration of internal cavities and the protein conformational landscape. The emerging picture is that internal hydration is an effective observable to probe the conformational landscape of proteins. In the specific context of our investigation, the analysis confirms that the hyperthermophilic G-domain is characterized by multiple states and it has a more flexible structure than its mesophilic homologue.

Keywords

Thermophilic Proteins; Thermal Stability; Protein Hydration; Free-Energy; Molecular-Dynamics

Introduction

Proteins from thermophilic or hyperthermophilic organisms are stable and functional at very high temperatures, up to the boiling point of water. Several molecular factors ensure this

[*]To whom correspondence should be addressed: simone.melchionna@roma1.infn.it; jerome.henin@ibpc.fr; fabio.sterpone@ibpc.fr.

Europe PMC Funders Author Manuscripts

extreme stability and their combination results in different thermodynamic strategies for thermal adaptation[1-4]. With respect to their mesophilic homologues, (hyper)thermophiles are generally enriched in charged amino-acids and cross-linked by a larger number of hydrogen bonds (HB) and salt-bridges that contribute to stabilize the protein fold[5-7]. They have shorter loops that protect the fold from thermal excitation by reducing the extension of weak spots at the protein surface[2-4,8,9]. Moreover, an optimized and extended hydrophobic packing in the protein core or at domain interfaces enhances the intramolecular cohesive forces[10].

These factors depend, directly or indirectly, on the coupling between the protein and surrounding water. For example it has been demonstrated that the distribution of charged amino-acids at the protein surface and in contact with water can be effectively optimized to increase thermal stability[11,12]. Moreover, the stabilizing contribution of salt-bridges partially buried in the protein matrix depends on the local polarity[5,13,14] while the presence of internal water molecules could alleviate the associated desolvation penalty[15,16]. In this regard, it was proposed that (hyper)thermophiles could benefit from dense ion-pairing because of a higher internal dielectric constant[17]. Finally, the presence of water molecules inside internal cavities or superficial pockets also improves the molecular packing and extends the HB connectivity with potential effects on local kinetic stability[18]. In this regard it is worth noting that many thermophiles are also resistant to high pressure[2], in fact internal hydration could reduce the volume of internal voids which are thought to trigger pressure unfolding[19] and at the same time make intramolecular HB patterns resistant to temperature and pressure stresses[20].

While protein hydration has been the object of intense and debated research[21-25] so far only sporadic investigations tried to link the interactions between protein and water to thermal stability, or to extreme adaptation in general[26-32]. In a previous computational study of two homologous G-domains of different stability, it was observed that the thermal stability of the percolating water HB network enveloping the protein surface correlates to the different stabilities of the two molecules[27]. Due to the different polarities and morphologies, a strong coupling between the protein surface and water was invoked for the hyperthermophilic domain[33,34].

The impact of the surface composition on the organization of the hydration layer was recently highlighted for two orthologous malate dehydrogenase (*MDH*) proteins, one being halophilic and the other thermophilic and non-halophilic, whose X-ray structures were resolved at high resolution[31]. At cryogenic temperature, the thermophilic *MDH* was found to be covered by extended networks of water clusters, mostly in pentagonal configurations, and showed a remarkably high level of hydration. These findings are in line with those reported for the hyperthermophilic b-glicosidase protein[35] and question the generality of the conclusion from a previous investigation according to which thermophilic proteins are less hydrated[26].

On the other hand, as shown in a more recent work on the homologous G-domains[29], the dynamics of individual water molecules in the hydration layer was found to be unaffected by the chemical composition of the proteins, since water relaxation mainly depends on an excluded volume effect[36]. This result, later verified by other examples[25], agrees with the

results from NMR experiments[32] reporting a similar retardation factor for the hydration water of both an halophilic and a non-halophilic peptide of markedly different amino-acid composition.

Hydration dynamics provides only kinetic insights on the protein/water coupling, however an understanding of water's contribution to protein stability in extreme conditions needs to be quantified from a thermodynamic point of view. For instance, an enhanced structuring of water at the surface of the protein fold could explain the systematic lower heat capacity of unfolding, ΔC_p, measured for (hyper)thermophiles with respect to mesophiles[26,37].

Another interesting question concerns the contribution from buried water to the overall stability gap between homologues. A seminal attempt to shed light on such issue was done by Yin, Hummer and Rasaiah[30] who used computational means to estimate the stability of water clusters in the nonpolar cavities of the tetrabrachion protein from the hyperthermophilic *Staphylothermus marinus*, whose optimal growth temperature is 365 K. This coiled coil assembly, made up of four helices, is characterized by a sequence of cavities located along the protein axis and, according to the crystallographic structure resolved at low temperature (T = 100 K), they are filled by water. Yin and coworkers found that at ambient temperature the largest cavity is preferentially hydrated with clusters comprising 4 to 9 molecules, being a necessary threshold to create a convenient environment for the water HB connectivity[22,38]. By increasing temperature, the cavity dehydrates and the drying process marks the onset of denaturation. For a very similar system, the SNARE protein, the presence of internal water is instead destabilizing[39]. In the very stable neutral variant (melting temperature T_m = 365 K) the four helices are tightly bound and no buried water is visible in the structure even if long residence molecules were detected by molecular dynamics simulations[40]. On the contrary, in the less stable protein from *S. cerevisiae* (T_m = 328 K) three internal cavities are filled by water and when these cavities are removed by punctual mutations, the stability increases as monitored by a shift of T_m by about 6 K[39].

Estimating the increase in stability arising from internal water molecules, by performing single point mutations that promote/inhibit water filled cavities is, however, ambiguous. In fact, the different intramolecular interactions, like the HB propensity and the conformational reorganization, caused by the change of the amino-acid also contribute to the overall thermodynamical shift. In some cases, creating a host environment for water is stabilizing, as reported for BPTI[41], subtilisin[42], lysozyme[43] and lipases[18], but for other proteins the reverse applies, as reported for iso-1-cytochrome[44], protease[45] and lysozyme T4 mutants[46].

Computer simulations, based on modern all-atom force fields, provide a unique tool to explore the specific role of water inside cavities and pockets[22,47-52]. Free energy calculations and MD simulations successfully predict the hydration state of pockets of different chemical natures (polar vs nonpolar)[22,50,51,53], help to understand the relation between internal hydration and function as in the case of charge separation processes[54-57], ligand binding[58-60], allostery[61,62] and can be used to estimate the contribution to stability[50,63].

In this work we continue our investigation of the molecular mechanisms at the origin of extreme thermal stability by comparing the behavior of a representative pair of mesophilic/hyperthermophilic homologues: the G-domains of the Elongation Factor thermo unstable (EF-Tu) protein from the mesophilic *Escherichia coli* bacterium and EF-1α protein from the hyperthermophilic *Sulfolobus solfataricus* archaeon. The experimental melting temperatures of the two domains differ by about 40 K[64,65] and are qualitatively well reproduced *in silico*. By using extensive MD simulations at high temperature (T = 360 K) we previously probed the early steps of unfolding of the mesophilic domain. Conversely, in the same timescale and temperature range, the hyperthermophilic protein was found to remain stable[9]. Interestingly, for the folded states at ambient conditions, we noticed an extended hydration of internal cavities within the proteins. Buried water molecules were suggested to play a major role in regulating stability[29]. Following this idea, we present here a precise study of both proteins' internal hydration, and determine the contributions to their different stabilities arising from water molecules buried within the two domains.

Methods

System and MD simulations

The G-domain of the mesophilic EF-Tu (*E. coli*), PDB code 1EFC[66]) and hyperthermophilic EF-1α (*Sulfolobus solfataricus*, PDB code 1SKQ[67]) proteins were simulated in water solution and at ambient temperature for a time of 0.6 μs for each protein. The mesophilic domain ($\mathscr{M}$, 196 a.a.) was solvated with 7440 and the hyperthermophilic domain ($\mathscr{H}$, 226 a.a.) with 10673 water molecules. Counter-ions were added to neutralize the systems. In order to explore the behavior at high temperature, multiple simulations were carried out at T=360 K, see[9]. In particular, for the mesophilic protein, we collected 10 independent runs of length between 400 ns and 1.5 μs, each probing the early steps of the unfolding. For the hyperthermophilic domain, two independent simulations of 1 μs and 0.7 μs were used to probe its stability at high temperature.

To perform extensive simulations, we employed the NAMD parallel software[68]. The two proteins were modeled using the chamm22[69] force field and water was represented by the charmm-TIP3P model. The system preparation and simulation protocol are detailed in our previous work[9,27,29].

Hydration dynamics

The exchange dynamics of water molecules between the protein hydration shell and the bulk solution was analyzed by computing the survival probability function $N_w(t)$. To this aim we first defined the hydration shell by detecting water oxygen atoms within 4.5 Å of any heavy atom of the protein. The survival probability $N_w(t)$[70,71] counts how many water molecules, initially located within the hydration shell, remain in the shell for the subsequent time span t. This is formally defined as:

$$N_w(t) = \frac{1}{N_t}\sum_{n=1}^{N_t}\sum_j P_j(t_n, t). \quad (1)$$

where the indicator function $P_j(t_n, t)$ takes the value 1 if the j-th water is found continously in the hydration shell between times t_n and $t_n + t$, or zero otherwise. N_t is the number of simulation time-frames of duration t. By computing $N_w(t)$ we also measured the distribution of the continuous residence times τ_r of water in the hydration shell. Since we are interested in long time behavior, the analysis was performed on reduced trajectories with a time resolution of 20 ps.

Free energy calculations

Free energy differences were computed to estimate the water contribution to protein stability, measured on a limited number of molecules that hydrate the interior of the proteins. These water molecules were extracted from the ensemble of water molecules that were found within the hydration shell with a residence time $\tau_r > 5$ ns. The frequency distributions of the contact of these molecules with individual amino acids of the proteins allowed to detect their most probable locations. Two particularly hydrated cavities were found in each protein. For a short stretch of the trajectory (100 ns), water molecules inside these pockets were ranked depending on their cumulative residence time, and among those having the larger total residence times, we selected a small subset for our calculations. For each selected molecule we extracted three representative configurations separated in time by at least 5 ns one from another. In order to characterize the localization and binding dynamics of the water molecules, Figure S1 of the Supplemental Information (SI) reports the time evolution of distances between water oxygens and a reference atom in each pocket.

For each site, a local hydration free energy, or equivalently, a local excess chemical potential for water was calculated. Starting from each representative configuration, an equilibration step was first conducted in the NPT ensemble. During the equilibration, all protein atoms and the water molecule of interest were restrained around their initial positions by a harmonic potential.

Free energy calculations were carried out following the framework used by Olano and Rick[50] to study hydration free energy of internal cavities in BPTI and barnase, which we briefly describe below. The excess free energy of hydration is obtained from a virtual thermodynamic cycle and therefore results from two independent components, the free energy required to remove a water molecule from the bulk solution (ΔG_{wat}) and the free energy required to add the water molecule to the protein cavity (ΔG_{prot}). Both were estimated via the free energy perturbation (FEP) method[72].

At first a water molecule in the bulk liquid is decoupled from the rest of the system:

$$(N+1)\,\mathrm{H_2O} \xrightarrow{\Delta G_{wat}} N\mathrm{H_2O} + \mathrm{H_2O}_{ni} \quad (2)$$

In the second step, the non-interacting water (H_2O_{ni}), once localized inside the protein's empty cavity (EC), is coupled to the rest of the system. This step is then further decomposed into a localization process and a coupling process:

$$\mathrm{H_2O}_{ni} + \mathrm{protein}\,(EC) \xrightarrow{\Delta G_{loc}} \mathrm{protein}\left(\mathrm{H_2O}_{ni}^{loc}\right) \quad (3)$$

$$\text{protein}\left(\mathrm{H_2O}_{ni}^{loc}\right) \xrightarrow{\Delta G_{inter}} \text{protein}\,(\mathrm{H_2O}) \quad (4)$$

where $\mathrm{H_2O}_{ni}^{loc}$ refers to a non-interacting water molecule which is restrained inside the protein cavity. Localization inside the cavity is enforced by a harmonic potential added to the Hamiltonian of the system. This potential confines the motion of the water molecule in its reference site inside the cavity of interest. Since we were interested in calculating the hydration free energy for many sites inside the cavities, we designed an *ad hoc* strategy to implement the restraint with respect to these positions.

For this purpose, for each internal water we identified the carbon atoms of the protein that, in the equilibrated configuration, lie within 5 Å of the oxygen atom of the water molecule. Among them, all possible combinations of five carbons were considered and the distance d_0 between their center of mass and the oxygen of the water molecule was measured. Finally the combination of the five carbons that produced the shortest distance (usually less than 0.1 Å) was selected. The center of mass of the carbons was then defined as the reference site

and used to implement the harmonic restraint, $U_{harm}(t) = \frac{1}{2} k_{harm} (d(t))^2$ with $d(t)$ the time-dependent distance between the water oxygen and the carbons' center of mass (k_{harm} = 10 kcal/(mol Å^2)). The value of the constant was chosen such that it reproduces the fluctuations of the oxygen around the reference site in the unbiased trajectory. This procedure avoids the definition of a static reference point in space and thus the need of fitting the protein structure to a reference configuration during the dynamics[50].

For the non-interacting molecule the work of localizing due to U_{harm} is associated to a free energy contribution:

$$\Delta G_{\mathrm{loc}} = -k_B T \ln\left[\rho (2\pi k_B T / k_{harm})^{3/2}\right] \quad (5)$$

where k_B is the Boltzmann's constant, T the temperature, and ρ the bulk density of water.

The local binding free energy of a water molecule in a given site was calculated by adding all the components:

$$\Delta G = \Delta G_{wat} + \Delta G_{prot} = \Delta G_{wat} + \Delta G_{loc} + \Delta G_{inter} \quad (6)$$

In our implementation, the free energy ΔG_{inter} is given by two terms. The first one measures the coupling between the water molecule and the environment in the presence of the harmonic restraint and is calculated using the FEP method as implemented in NAMD, while the second one removes the contribution from the harmonic restraint in the interacting system (see below). A coupling parameter λ taking values in the [0, 1] interval was used to gradually switch off first the electrostatic and then the VdW interactions of the water molecule in the cavity; by this annihilation process we accessed the quantity $-\Delta G_{inter}$. The FEP procedure consisted of 11 windows along the coupling parameter λ. For each value of λ, the system was equilibrated for 500 ps, then statistics were accumulated for 2 ns. The

simulations were performed in the NPT ensemble, and by using the same protocol as in our previous work[9]. In a similar way, we estimated ΔG_{wat} by simulating a system of pure water of 512 molecules. We obtained ΔG_{wat} = 6.2 ± 0.1 kcal/mol at T = 300 K, and ΔG_{wat} = 5.6 ± 0.1 kcal/mol at T = 360 K. While in ref.[50] the harmonic restraint is progressively removed along the coupling process so as to disappear completely for a fully interacting water molecule, in our implementation this term was removed only in a separate step and its contribution to ΔG_{inter} was estimated via an independent thermodynamic integration. In this procedure the harmonic constant k_{harm} was progressively scaled to zero (11 steps), and the free energy contribution of the restraint was evaluated by computing $\int dk \langle \partial U_{harm} / \partial k_{harm} \rangle_k$ along the path[72]. The ensemble average $\langle \ldots \rangle_k$ was estimated for a given value of the scaled constant $k_{harm} = k$.

We also point out that during the annihilation process, when all the interactions between the water molecule and the rest of the system are scaled down, other molecules may overlap with the annihilated one, thus occupying the reference site. In order to avoid artifacts arising from such overlaps, we added a harmonic repulsive wall between the tagged water molecule and those nearby. We verified that without this repulsive wall the overlaps occur rarely and only in the last windows of the FEP decoupling procedure; the effect on the estimated free energy is generally comparable to the statistical error. The harmonic restraint as well as the repulsive wall were implemented using the Collective Variables Module available in NAMD[73].

Clustering

In order to correlate the protein conformational fluctuations and the change of the internal hydration we have performed a cluster analysis of the trajectories sampled by the two systems at ambient temperature[9]. For the clustering procedure we used as a collective variable the fraction of native contacts Q defined below.

The number of native contacts l_i' for a given C_α site is the number of C_α atoms located within a cut-off distance of 8 Å in the crystallographic structure. Thus, the fraction of native contacts, referring to the whole molecule, is defined as

$$Q(t) = \frac{1}{N_{C_\alpha}} \sum_{i=1}^{N_{C_\alpha}} \frac{l_i(t)}{l_i'} \quad (7)$$

where N_{C_α} is the number of C_α atoms, having l_i' native C_α contacts in the reference state and $l_i(t)$ of them appearing also at time $t \left(l_i(t) \leq l_i' \right)$.

The clustering was done using the well-established *leader* algorithm[74] with the following procedure. Each snapshot was represented by a vector of length equal to the number of C_α atoms (N_{C_α}), whose i–th component of the vector is the quantity $l_i(t)/l_i'$. Then, the distance between two snapshots was defined as

$$d\left(t,t'\right)=\sqrt{\frac{1}{N_{C\alpha}}\sum_{i=1}^{N_{C\alpha}}\left(L_i\left(t\right)-L_i\left(t'\right)\right)^2},\quad \text{with}\quad L_i\left(t\right)=\frac{l_i(t)}{l_i'} \tag{8}$$

The output of the clustering was mapped onto a network that was drawn with the use of a forced-based algorithm of the Gephi software[75].

Results

The exchange kinetics of internal water molecules

As a preliminary investigation we detected the water molecules that penetrate the internal cavities of the two proteins. In many crystallographic structures the positions of structural water molecules are resolved together with the protein. In the structure of the whole EF-Tu[66] and EF-1α[67] proteins, a few molecules are visible in the interior of each G-domain. However, these structures were resolved in the presence of substrates which possibly make inaccessible alternative sites. Moreover, according to several computational studies[49,51] when the protein is thermalized at ambient conditions locations alternative to the crystallographically determined sites are also sampled.

Exchange of these molecules with the external bulk solution occurs, as probed by both NMR[23,76] and simulations[49,77,78], on the nanosecond, or even longer timescale. In this work, we used this lower bound on the exchange time to fully map the solvent accessibility of the proteins' interior. As already probed for other globular proteins[49,77,78], for these two G-domains, the probability that a water molecule persists in the hydration shell as a function of time, $N_w(t)$, extends with a long tail in the nanoseconds range, as shown in the insets of 1. When performing an exponential fit of this long tail (t> 5 ns) we recover average characteristic times of $\langle \tau_{slow} \rangle$ 41 ns and 16 ns for $\mathscr{M}$ and $\mathscr{H}$, respectively. The dynamics concerns about 1% of the water molecules in the shell, the so-called structural water, mainly buried in internal cavities or superficial clefts. The distributions of the residence times τ_r for this ensemble of molecules is reported in the main plot of 1. The distributions, as commonly observed for other proteins[77], decay according to a power law, t^{-a}, with $a \sim 2.6$. This value is close to what is reported for cytochrome c (2.5)[77] and to the assumed distribution of reorientational times to interpret NMR experiments[79]. In the simulation of the $\mathscr{M}$ domain, one water molecule gets trapped in the protein matrix for about 80% of the simulation time, ~500 ns, causing the longest value of τ_{slow} for this system. In $\mathscr{H}$ the longest residence time is somewhat shorter, ~250 ns.

Internal cavities sampled by water

In order to map the internal cavities of the proteins we restricted our attention to those water molecules that have a residence time τ_r longer than 5 ns. The frequency distribution of their positions along the protein sequence is reported in 2. Two main pockets are found in each system, and their locations are indicated by the colored bars at the top of the x-axis. In the mesophilic domain the largest pocket, marked by the magenta bars, is located in the protein

Europe PMC Funders Author Manuscripts

core. There a cluster of molecules is confined by a hydrophobic wall and a polar surface that offers several anchoring sites for hydrogen bonds. The pocket is accessible via a gateway near the P loop, a key region in the GTP binding site. Around this opening two buried water molecules are visible in the X-ray structure, see SI Table S1. The second pocket (violet bars) hosts a single isolated molecule. In our simulation, a single molecule penetrates into this pocket after about 100 ns, and gets trapped for the remainder of the simulation time. The water is caged by three hydrogen bonds formed with the residues S166 (backbone CO and NH groups) and W177 (backbone CO group). Interestingly, the observed binding matches the location of a structural water visible in the X-ray structure (SI Table S1).

In the hyperthermophilic protein, both pockets are filled with clusters of molecules. The first pocket (maroon bars) is found in proximity of the P loop. A portion of the pocket's surface offers several HB pinning sites, mainly backbone groups, while the other part has a hydrophobic character. In the crystallographic structure three molecules are resolved in this pocket, see Supplemental Information Table S1. The second pocket (red bars) is created by a structural specific motif of the $\mathscr{H}$protein: the two small helices α^1 [E32-L45] and α^2 [E48-E63]. In our previous investigation we have shown that these structures confer stability to this region of the protein and prevent the local unfolding at high temperature that instead occurs in the $\mathscr{M}$domain[9]. The walls of this cavity are cross linked by ionic interactions, i.e. the highly stable ion-pairs R65-D31 and R58-E62, and these charged amino-acids create a highly polar environment. The fluctuations of these links are mediated by the presence of water that can form molecular bridges between the ionic groups (SI Table S3). We stress that in our simulations the ionic groups were all considered charged, however it is widely accepted that it is quite difficult to determine the protonation state of such buried charged groups, that in principle could also be influenced by the presence of water.

Local hydration free energies

We now discuss the contribution to protein stability stemming from these hydrated cavities. From the ensemble of long residence water molecules, we singled out a restricted set of molecules representative of the cavity hydration and with the longest residence times ($\tau_r >$ 15 ns). For practical reasons, this filtering was done considering a short stretch of the trajectory (100 ns). For each one of these molecules three different configurations, *A*, *B*, and *C*, separated from each other by at least 5 ns, were extracted from the trajectory and used to initiate the free energy calculations. This large ensemble allows to estimate the variability induced by the presence of multiple states of the pockets' hydration, including i) changes in the internal location of the water molecule, ii) changes of the shape of the hosting cavity, and iii) long range effects due to the global fluctuations of the protein. The values of the hydration free energy, ΔG_i^k, and of the associated statistical error, are reported in Supplementary Tables S4 and S5 for all the pockets and conformers of the $\mathscr{M}$and $\mathscr{H}$ domains, respectively.

First we note that the hydration free energy is favorable for almost all the selected internal locations. However, we also remark the large variability of the ΔG_i^k, with a single molecule experiencing different favorable interactions during its confinement in the pocket. This variability can be as high as 4 kcal/mol. In both proteins, some locations provide large

contributions to protein stability, for instance −9.0 ± 0.3 ($\mathscr{M}$, w_1^{P2}) and −7.0 ± 0.3 kcal/mol ($\mathscr{H}$, w_3^{P2}). In both these cases, we observe high HB connectivity and strong packing (see data in SI Tables S2 and S3). When focusing on the average value associated to each pocket, we obtain $\langle \Delta G_i \rangle^{P1}$ = −1.6 ± 0.3 and $\langle \Delta G_i \rangle^{P2}$ = −5.5 ± 0.4 kcal/mol for the $\mathscr{M}$ domain and $\langle \Delta G_i \rangle^{P1}$ = −1.5 ± 0.3 and $\langle \Delta G_i \rangle^{P2}$ = −3.6 ± 0.5 kcal/mol for the $\mathscr{H}$ domain. The values are close to previous estimates of the insertion free energy for water into the cavities of other proteins as the bacteriorhodopsin[48], BPTI[50], and kinase A[80].

The distribution of the free energies are plotted in 3. The vertical lines indicate the averages, $\langle \Delta G_i \rangle$ = −2.1 ± 0.3 kcal/mol and $\langle \Delta G_i \rangle$ = −2.6 ± 0.4 kcal/mol for $\mathscr{M}$ and $\mathscr{H}$, respectively. For the $\mathscr{M}$ domain the distribution is shaped by the contributions from water hydrating the larger pocket (P2), on the contrary for the $\mathscr{H}$ protein, the distribution reflects the different contribution from the two pockets.

The calculated ΔG_i^k is the free energy contribution to the protein's stability from a single buried molecule localized in a specific site *i* and calculated for different states *k* of the local environment, including the presence of surrounding water molecules having different residence times. According to the analysis presented in[48,53], our values correspond to the insertion free energy of a molecule in the presence of $n - 1$ others. The ensemble of conformers A^i, B^i, C^i helps us to sample several initial conditions of the hydrated cavities, i.e. different values of $n - 1$ (see SI Tables S2 and S3).

Global free energy of internal hydration

As discussed earlier, both proteins host several long residence water molecules at the same time. Therefore, in order to evaluate the total stabilization due to internal water molecules, we need to estimate their cumulative contribution, as represented schematically in 4. Hence, we determine a global free energy of internal hydration for both proteins based on our local estimates of water binding free energy.

For a given starting configuration *k*, the calculations described above provide local hydration free energies ΔG_i^k for each of 8 identified binding sites. Stability of an internal water molecule depends on many degrees of freedom, belonging to the protein as well as neighboring water. Direct calculation of well-converged free energy estimates could therefore be prohibitively long. We choose to obtain each binding free energy by combining data from three independent, local simulations (starting from different configurations). Each simulation is considered to sample one representative region of phase space, hence, their results are combined as if merging samples within the exponential average of FEP:

$$\exp\left(-\beta \Delta G_i\right) = \frac{1}{3}\sum_{k=1}^{3} \exp\left(-\beta \Delta G_i^k\right) \quad (9)$$

We assume additivity, meaning that the global free energy of internal hydration is estimated as the sum of local free energies weighted by the occupancy n_i of each site:

$$\Delta G_{hyd}=\sum_{i=1}^{8}-\frac{n_i}{\beta}\log\left(\frac{1}{3}\sum_{k=1}^{3}\exp\left(-\beta\Delta G_i^k\right)\right) \quad (10)$$

In turn, the mean occupancy of each site is obtained from the binding free energies as:

$$n_i=\frac{C}{1+\exp\left(\beta\Delta G_i\right)} \quad (11)$$

C is a factor common to all sites in one protein, and is fitted so that the total predicted occupancy matches the average number of bound water molecules measured in long simulations, in other words:

$$\sum_{i=1}^{8}n_i=\langle n_w\rangle \quad (12)$$

We find the optimal value of *C* to be 0.6 for both the mesophile and thermophile.

The term n_w can be calculated by scanning the trajectories and counting how many long-residence molecules are present at the same time in the interior of the proteins. Considering the molecules with residence time $\tau_r > 15$ ns we obtain $\langle n_w\rangle = 4.7$ for both proteins and for a less strict condition ($\tau_r > 5$ ns) $\langle n_w\rangle = 9.8$ and $\langle n_w\rangle = 10.4$ for the $\mathscr{M}$ and $\mathscr{H}$ proteins, respectively.

Within this framework and counting $\langle n_w\rangle$ with the stricter condition ($\tau_r > 15$ ns), we find $\Delta G_{hyd} = -15.7 \pm 0.5$ kcal/mol for the mesophile, and -17.0 ± 0.6 kcal/mol for the hyperthermophile, meaning that the relative stabilization of the $\mathscr{H}$ protein with respect to $\mathscr{M}$ by internal hydration is $\Delta\Delta G_{hyd} = 1.3 \pm 0.5$ kcal/mol. By releasing condition (12) on the total occupancy, the relative stabilisation of the $\mathscr{H}$ protein with respect to $\mathscr{M}$ is slightly larger, $\Delta\Delta G_{hyd} = 2.5 \pm 0.9$ kcal/mol. The same numerical results are obtained with an alternative formulation of the problem that we report in SI. The obtained value for the relative stability $\Delta\Delta G_{hyd}$ is a small but statistically meaningful number, and its possible effect on protein thermal stability is discussed below.

Unfortunately, for our EF-Tu and -1α G-domains, no thermodynamic values are available for the unfolding free energy $\Delta G^{unf}(T)$. The magnitude of the stability gap between homologues, i.e. $\Delta\Delta G^{unf}(T)$ at ambient conditions, can vary substantially depending on the pair. In some cases this gap can be as high as 20 kcal/mol but for many pairs its just on the order of a few kcal/mol[81,82]. In order to quantify the possible effect of the stability gain due to the internal hydration ($\Delta\Delta G_{hyd}$) we have considered two model systems. The first one is represented by a pair of homologues close to our proteins, namely the EF-1α from rabbit (mesophile) and from the thermophilic bacterium *T. thermophilus*[83]. The second is a pair of mesophilic/hyperthermophilic proteins belonging to CheY family. For these pairs the available thermodynamic data[83,84] allow to estimate the unfolding free energy as a function of temperature and hence construct the stability curve, $\Delta G^{unf}(T)$. For each protein we

Europe PMC Funders Author Manuscripts

upshifted the stability curve by $\Delta\Delta G_{hyd}$ = 1.3 kcal/mol (see above) and as expected the melting temperature of the proteins increases, ΔT_m is about 6 K. Depending on the value of the specific heat of unfolding ΔC_p^{unf}, that controls the curvature of the stability curve, the shift of the melting temperature could be even larger. Thus, even a small contribution as that arising from internal hydration can result in an enhanced thermal stability. The details of this qualitative modelling are provided in SI, see Fig. S.3.

Alternative strategies to estimate the free energy contribution to stability stemming from the hydration of large cavities have been presented before and reporting for non-polar cavities and channels cooperative effects[22,48,53]. For example in hydrophobic pockets of large volumes favorable hydration is found to take place only above a given threshold: about 4/5 water molecules should be hosted to compensate the loss of HBs due to the cavity nature[22]. Still, for cavities offering HB acceptor/donor sites, the size and cooperative effects are expected to be less important; for example in the case of a protein kinase, the progressive hydration of an internal cavity is associated to a stable free energy contribution of about −3 to −2 kcal/mol per added molecule[80]. A detailed analysis of the problem is reserved for future work. Obviously, our conclusions also depend on the reconstructed distributions $p(\Delta G_i)$, here assumed to be representative of the cavities' hydration, and on the residence time threshold used to determine the number of buried water molecules, $\langle n_w \rangle$.

While hereby we have considered only the contribution from internal hydration, we stress that a more complete view of the role of hydration on protein thermal stability would require to account for the changes in the solvatation of the external surface as well. This can be only approached using specific modelling techniques, as those explored recently by Harano[85] and Ham[86].

Hydration and global protein flexibility

The number of internal waters fluctuates over time and the penetration/escape dynamics correlates with the breathing soft-modes of the protein that visits different conformational states[78,87]. For our proteins, the fluctuations are rather high (~ 40%) and have characteristic oscillation period ranging from tenths to one hundred nanoseconds. As a consequence the internal water contribution to protein stability varies substantially with the proteins' conformational fluctuations.

The distribution of the number of long residence water molecule $p(n_w)$, is reported in 5 (top panel) for the residence time threshold τ_r > 15 ns. While for the $\mathscr{M}$ domain the fluctuations of n_w produce an unimodal distribution, for $\mathscr{H}$ the distribution is bi-modal. This two-state regime relates to the simultaneous hydration dynamics of the two cavities P1 and P2, while in $\mathscr{M}$ only the hydration of pocket P2 varies substantially in time. Interestingly, we note that despite the qualitative differences, the hydrations dynamics of the internal cavities result in very similar overall compressibilities of the two proteins[9], $\beta_a \simeq 9 \cdot 10^{-5}\, MPa^{-1}$. Hence, internal hydration has similar effects on the global flexibility of the two proteins.

In the bottom panel of the figure we provide a more direct mapping of the internal hydration on the conformational landscape, represented as a network of states. In the specific case the network is obtained by clustering the trajectories and using the native contacts order

parameter, Q[9]. The details of the calculation are provided in section Method. The size of each cluster is proportional to its occupation and the color scale reflects the average number of long residence water molecules computed for the configurations in each cluster.

For the $\mathscr{H}$ domain, two zones of different color appear well separated kinetically; we do also note, however, that some basins that have similar Q values with rather different internal hydration are kinetically connected. For the mesophilic $\mathscr{M}$ protein, the unimodal distribution of the internal hydration correlates with a compact conformational landscape. The correlation between the fluctuations of the internal hydration and the presence of multiple conformational states measured by conformational order parameters, such as the native contacts Q, supports the notion proposed some years ago[76], that the exchange dynamics of internal hydration could be used to probe roughness of the conformational landscape of proteins.

This gives quantitative support to the qualitative sketch of 4, where the free energy landscape of the protein is projected along a generic conformation reaction coordinate. The filled blue zones represent the internal water free energy contribution to stability as it varies along the conformational landscape of the folded state. The thermophile exhibits a broader variety of internal hydration states, some of which confer it additional stability. The tool we have used to investigate the correlation among the dynamics of internal hydration and the protein conformational landscape could result effective to explore the mechanism of unfolding caused by high temperature or pressure and to characterise the kinetic heterogeneity of these processes that have been related to a different transient hydration of internal cavities[88].

Stability at high temperature

We address here the key issue of high temperature stability. For the same set of water molecules we evaluated the hydration free energy at T = 360 K. The calculations were performed starting from the same protein configurations used at T = 300 K, thus we assumed that the temperature excitation does not distort these representative conformations of the folded state. At high temperature, water binding free energy decreases substantially for most sites, in both proteins; water transfer into some of the protein sites is even unfavorable (6). As a result, there is no significant difference in overall free energy of internal hydration, yielding no clear contribution to the stability gap. The average $\langle \Delta G_i \rangle$ is −1.8 and −1.6 kcal/mol, for $\mathscr{M}$ and $\mathscr{H}$. A comparable decrease of the favorable hydration was reported also for nonpolar cavities, and in a less extended temperature window for the favorable hydration of BPTI cavity[50].

In the high temperature regime, it is difficult to estimate the overall contribution from internal water to the stability gap between the two proteins because the mesophilic homologue starts to unfold. The early step of unfolding was monitored in a set of independent simulations and occurred within a characteristic time of 100 to 300 ns[9]. We have estimated the number of long residence waters residing simultaneously in the protein matrix for each of these runs by restraining the calculation to the part of the trajectory before the unfolding event. For the hyperthermophilic domain two simulations of length 0.7 and 1 μs were used. Using a strict threshold (τ_r > 9.6 ns) we obtain $\langle n_w \rangle$ = 2.6 and $\langle n_w \rangle$ = 2 for $\mathscr{M}$

and $\mathscr{H}$, respectively while for a less strict one (τ_r > 3.8 ns), the numbers are $\langle n_w \rangle$=3.6 vs $\langle n_w \rangle$ = 4.6 for $\mathscr{M}$ and $\mathscr{H}$, respectively [1]. When estimating the ΔG_{hyd} via eq. 10 the contribution of internal hydration to the different stabilities ($\Delta\Delta G_{hyd}$) can be negligible (0.7 kcal/mol) or even in favour of the mesophile (1.7 kcal/mol) depending on the values of $\langle n_w \rangle$. Therefore our results are not conclusive and not robust for this high temperature regime.

Alternatively, one could investigate the contribution of internal water hydrating the cavity P2 to the kinetic stability of the specific region a^1 [E32-L45] and a^2 [E48-E63] of the $\mathscr{H}$ domain. Here, water could link these key secondary structures and indirectly favor the stability of the local ion-pairs. On the other hand, in the mesophilic homologue $\mathscr{M}$, the similar region represents the weakest point of the protein structure where unfolding early steps take place[9]. How transient internal hydration couples to the unfolding of this region and the associated kinetic barriers will be the object of a separate investigation.

Conclusion

In this work we investigated the internal hydration of two homologous proteins and estimated its contribution to the proteins' different thermal stabilities. Namely, we compared the mesophilic EF-Tu and the hyperthermophilic EF-1α G-domains. Two internal cavities were found to be hydrated in both systems. The location of one of them is common for the two domains with an access near the P-loop at the catalytic site. The other ones are peripheral and domain specific.

In both systems, the per molecule hydration free energy was found to be favorable for all the sites and the protein configurations.

At ambient conditions, the free energy of internal hydration for the hyperthermophilic domain is more favorable than for the mesophilic one by 1.3 kcal/mol, and resulting in a sizeable contribution to the stability gap. At the high working temperature of the hyperthermophile (T = 365 K), the internal hydration free energy is systematically less favorable when compared to ambient condition, giving probably a negligible contribution to the stability gap between the two systems.

We also found a very clear correlation between the conformational fluctuations of the proteins, represented as a connected network of conformational states, and the fluctuations of the number of the long residence waters wetting the interior of the proteins. While the mesophilic protein is characterized by unimodal fluctuations around a single conformational or hydration state, the hyperthermophilic domain visits multiple states for both the protein conformation and the internal hydration. Again, this finding questions the general notion according to which thermophilic proteins are more rigid objects than their mesophilic counterparts.

[1]The number of long-residence water molecules was estimated by considering the temperature scaling on the time threshold and assuming the escaping process rate-limited by a single barrier. The lower bound at T=360 K is therefore τ_r > 9.6 ns and the weaker threshold is about 3.8*ns*.

Our investigation shows that thermal stabilization can be made possible by enhanced internal hydration. However, the thermodynamic gain results from different factors, such as the overall number of structural water and how the local internal polarity is configured. By *in silico* approaches one can envision a way to enhance thermal stability by using internal hydration and by targeting such factors selectively. One strategy is to tune the internal polarity of a cavity so as to increase the free energy gain. Alternatively, de novo new cavities can be created, possibly by creating pockets near the protein surface, to increase the number of structural waters that contribute to stability. Internal hydration can also be used to kinetically stabilize specific regions of the protein. However, it must be pointed out that the creation of cavities and their hydrated or dry states can result in a complex thermodynamic and kinetic response of the proteins, as pointed out in the context of pressure induced unfolding[19,88]. Moreover, these changes however must be effective in a broad range of temperatures. A more systematic investigation on the internal and the total hydration of homologues of different protein families is reserved for future work and can benefit from recent methodological development for treating solvation contribution to protein stability[85,86].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no.258748. Part of this work was performed using HPC resources from GENCI [CINES and TGCC] (Grant 2012 c2012086818 and 2013 x201376818) and from CINECA (ISCRA Grant FLOWPROT). We acknowledge the financial support for infrastructures from ANR-11-LABX-0011-01.

References

(1). Nojima H, Ikai A, Oshima T, Noda H. Reversible Thermal Unfolding of Thermostable Phosphoglycerate Kinase. Thermostability Associated With Mean Zero Enthalpy Change. J. Mol. Biol. 1977; 116:429–442. [PubMed: 338921]

(2). Vieille C, Zeikus GJ. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. Microbiol. Mol. Biol. Rev. 2001; 65:1–43. [PubMed: 11238984]

(3). Kumar S, Nussinov R. How do Thermophilic Proteins Deal with Heat? Cell. Mol. Life Sci. 2001; 58:1216–1233. [PubMed: 11577980]

(4). Sterpone F, Melchionna S. Thermophilic Proteins: Insight and Perspective from In Silico Experiments. Chem. Soc. Rev. 2012; 41:1665–1676. [PubMed: 21975514]

(5). Kumar S, Ma B, Tsai CJ, Nussinov R. Electrostatic Strengths of Salt Bridges in Thermophilic and Mesophilic Glutamate Dehydrogenase Monomers. Proteins. 2000; 38:368–83. [PubMed: 10707024]

(6). Karshiko A, Ladenstein R. Ion Pairs and the Thermotolerance of Proteins from Hyperthermophiles: a "Traffic Rule" for Hot Roads. Trends Biochem. Sci. 2001; 26:550–6. [PubMed: 11551792]

(7). Chan C-H, Yu T-H, Wong K-B. Stabilizing Salt-Bridge Enhances Protein Thermostability by Reducing the Heat Capacity Change of Unfolding. PLoS ONE. 2011; 6:e21624. [PubMed: 21720566]

(8). Balasco N, Esposito L, Simone AD, Vitagliano L. Role of Loops Connecting Secondary Structure Elements in the Stabilization of Proteins Isolated from Thermophilic Organisms. Protein Sci. 2013; 22:1016–1023. [PubMed: 23661276]

(9). Kalimeri M, Rahaman O, Melchionna S, Sterpone F. How Conformational Flexibility Stabilizes the Hyperthermophilic Elongation Factor G-Domain. J. Phys. Chem. B. 2013; 117:13775–13785. [PubMed: 24087838]

(10). Rathi PC, Höffken HW, Gohlke H. Quality Matters: Extension of Clusters of Residues with Good Hydrophobic Contacts Stabilize (Hyper)Thermophilic Proteins. J. Chem. Inf. Model. 2014; 54:355–361. [PubMed: 24437522]

(11). Sanchez-Ruiz JM, Makhatadze GI. To Charge or not to Charge? Trends Biotechnol. 2001; 19:132–5. [PubMed: 11250029]

(12). Gribenko AV, Patel MM, Liu J, McCallum SA, Wang C, Makhatadze GI. Rational Stabilization of Enzymes by Computational Redesign of Surface Charge-Charge Interactions. Proc. Natl. Acad. Sci. USA. 2009; 106:2601–6. [PubMed: 19196981]

(13). Xiao L, Honig B. Electrostatic Contributions to the Stability of Hyperthermophilic Proteins. J. Mol. Biol. 1999; 289:1435–44. [PubMed: 10373377]

(14). Gao J, Bosco D, Powers E, Kelly J. Localized Thermodynamic Coupling between Hydrogen Bonding and Microenvironment Polarity Substantially Stabilizes Proteins. Nat. Struct. Mol. Biol. 2009; 16:684–691. [PubMed: 19525973]

(15). Fitch C, Karp D, Lee K, Stites W, Lattmanand EE, Garcia-Moreno E. Experimental pKa Values of Buried Residues: Analysis with Continuum Methods and Role of Water Penetration. Biophys. J. 2002; 82:3289–3304. [PubMed: 12023252]

(16). Salari R, Chong LT. Effects of High Temperature on Desolvation Costs of Salt Bridges Across Protein Binding Interfaces: Similarities and Differences between Implicit and Explicit Solvent Models. J. Phys. Chem. B. 2012; 116:2561–2567. [PubMed: 22300130]

(17). Dominy BN, Minoux H, Brooks CL 3rd. An Electrostatic Basis for the Stability of Thermophilic Proteins. Proteins. 2004; 57:128–41. [PubMed: 15326599]

(18). Ahmad S, Kamal MZ, Sankaranarayanan R, Rao NM. Thermostable Bacillus Subtilis Lipases: In Vitro Evolution and Structural Insight. J. Mol. Biol. 2008; 381:324–40. [PubMed: 18599073]

(19). Roche J, Caro JA, Norberto DR, Barthe P, Roumestand C, Schlessman JL, Garcia AE, García-Moreno BE, Royer CA. Cavities Determine the Pressure Unfolding of Proteins. Proc. Natl. Acad. Sci. USA. 2012; 109:6945–6950. [PubMed: 22496593]

(20). Nisius L, Grzesiek S. Key Stabilizing Elements of Protein Structure Identified through Pressure and Temperature Perturbation of its Hydrogen Bond Network. Nat. Chem. 2012; 4:711–717. [PubMed: 22914191]

(21). Levy Y, Onuchic JN. Water Mediation in Protein Folding and Molecular Recognition. Annu. Rev. Biophys. Biomol. Struct. 2006; 35:389–415. [PubMed: 16689642]

(22). Rasaiah JC, Garde S, Hummer G. Water in Nonpolar Confinement: from Nanotubes to Proteins and beyond. Annu. Rev. Phys. Chem. 2008; 59:713–40. [PubMed: 18092942]

(23). Halle B. Protein Hydration Dynamics in Solution: a Critical Survey. Philos. Trans. R. Soc. Lond. B Biol. Sci. 2004; 359:1207–23. [PubMed: 15306377]

(24). Ball P. Water as an Active Constituent in Cell Biology. Chem. Rev. 2008; 108:74–108. [PubMed: 18095715]

(25). Fogarty AC, Duboue-Dijon E, Sterpone F, Hynes JT, Laage D. Biomolecular Hydration Dynamics: a Jump Model Perspective. Chem. Soc. Rev. 2013; 42:5672–5683. [PubMed: 23612685]

(26). Pechkova E, Sivozhelezov V, Nicolini C. Protein Thermal Stability: the Role of Protein Structure and Aqueous Environment. Arch. Biochem. Biophys. 2007; 466:40–8. [PubMed: 17765863]

(27). Sterpone F, Bertonati C, Briganti G, Melchionna S. Key Role of Proximal Water in Regulating Thermostable Proteins. J. Phys. Chem. B. 2009; 113:131–7. [PubMed: 19072709]

(28). Paredes D, Watters K, Pitman D, Bystro C, Dordick J. Comparative Void-Volume Analysis of Psychrophilic and Mesophilic Enzymes: Structural Bioinformatics of Psychrophilic Enzymes Reveals Sources of Core Flexibility. BMC Struct. Biol. 2011; 11:42. [PubMed: 22013889]

(29). Rahaman O, Melchionna S, Laage D, Sterpone F. The Effect of Protein Composition on Hydration Dynamics. Phys. Chem. Chem. Phys. 2013; 15:3570–3576. [PubMed: 23381660]

(30). Yin H, Hummer G, Rasaiah JC. Metastable Water Clusters in the Nonpolar Cavities of the Thermostable Protein Tetrabrachion. J. Am. Chem. Soc. 2007; 129:7369–77. [PubMed: 17508748]

(31). Talon R, Coquelle N, Madern D, Girard E. An Experimental Point of View on Hydration/Solvation in Halophilic Proteins. Frontiers in Microbiology. 2014; 5

(32). 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; 116:3436–3444. [PubMed: 22329545]

(33). Melchionna S, Sinibaldi R, Briganti G. Explanation of the Stability of Thermophilic Proteins Based on Unique Micromorphology. Biophys. J. 2006; 90:4204–12. [PubMed: 16533850]

(34). Sterpone F, Bertonati C, Melchionna S. Water Around Thermophilic Proteins: The Role of Charged and Apolar Atoms. J. Phys.: Condens. Matter. 2010; 22:284113. [PubMed: 21399285]

(35). Chi YI, Martinez-Cruz LA, Jancarik J, Swanson RV, Robertson DE, Kim SH. Crystal Structure of the Beta-Glycosidase from the Hyperthermophile Thermosphaera Aggregans: Insights into its Activity and Thermostability. FEBS Lett. 1999; 445:375–383. [PubMed: 10094493]

(36). Sterpone F, Stirnemann G, Laage D. Magnitude and Molecular Origin of Water Slow-down Next to a Protein. J. Am. Chem. Soc. 2012; 134:4116–4119. [PubMed: 22335572]

(37). Ninad VP, Sharp KA. Heat Capacity in Proteins. Annu. Rev. Phys. Chem. 2005; 56:521–548. [PubMed: 15796710]

(38). Liu L, Quillin ML, Matthews BW. Use of Experimental Crystallographic Phases to Examine the Hydration of Polar and Nonpolar Cavities in T4 Lysozyme. Proc. Natl. Acad. Sci. USA. 2008; 105:14406–14411. [PubMed: 18780783]

(39). Strop P, Kaiser SE, Vrljic M, Brunger AT. The Structure of the Yeast Plasma Membrane SNARE Complex Reveals Destabilizing Water-Filled Cavities. J. Biol. Chem. 2008; 283:1113–9. [PubMed: 17956869]

(40). Durrieu M-P, Lavery R, Baaden M. Interactions between Neuronal Fusion Proteins Explored by Molecular Dynamics. Biophys. J. 2008; 94:3436–3446. [PubMed: 18212009]

(41). Berndt K, Beunink J, Schroder W, Wuthrich K. Designed Replacement of an Internal Hydration Water Molecule in BPTI: Structural and Functional Implications of a Glycine-to-Serine Mutation. Biochemistry. 1993; 32:4564–4570. [PubMed: 7683491]

(42). Pedersen JT, Olsen OH, Betzel C, Eschenburg S, Branner S, Hastrup S. Cavity Mutants of Savinase™: Crystal Structures and Differential Scanning Calorimetry Experiments Give Hints of the Function of the Buried Water Molecules in Subtilisins. J. Mol. Biol. 1994; 242:193–202. [PubMed: 8089841]

(43). Takano K, Funahashi J, Yamagata Y, Fujii S, Yutani K. Contribution of Water Molecules in the Interior of a Protein to the Conformational Stability. J. Mol. Biol. 1997; 274:132–142. [PubMed: 9398521]

(44). Hickey DR, Berghuis AM, Lafond G, Jaeger JA, Cardillo TS, McLendon D, Das G, Sherman F, Brayer GD, McLendon G. Enhanced Thermodynamic Stabilities of Yeast Iso-1-Cytochromes C with Amino Acid Replacements at Positions 52 and 102. J. Biol. Chem. 1991; 266:11686–11694. [PubMed: 1646814]

(45). Vriend G, Berendsen H, van der Zee J, van den Burg B, Venema G, Eijsink V. Stabilization of the Neutral Protease of Bacillus Stearothermophilus by Removal of a Buried Water Molecule. Protein Eng. 1991; 4:941–945. [PubMed: 1817257]

(46). Xu J, Baase W, Quillin M, Baldwin E, Matthews B. Structural and Thermodynamic Analysis of the Binding of Solvent at Internal Sites in T4 Lysozyme. Protein Sci. 2001; 10:1067–1078. [PubMed: 11316887]

(47). Wade RC, Mazor MH, McCammon JA, Quiocho FA. A Molecular Dynamics Study of Thermodynamic and Structural Aspects of the Hydration of Cavities in Proteins. Biopolymers. 1991; 31:919–931. [PubMed: 1782354]

(48). Roux B, Nina M, Pomes R, Smith J. Thermodynamic Stability of Water Molecules in the Bacteriorhodopsin Proton Channel: a Molecular Dynamics Free Energy Perturbation Study. Biophys. J. 1996; 71:670–681. [PubMed: 8842206]

(49). Sterpone F, Ceccarelli M, Marchi M. Dynamics of Hydration in Hen Egg White Lysozyme. J. Mol. Biol. 2001; 311:409–19. [PubMed: 11478869]

(50). Olano LR, Rick SW. Hydration Free Energies and Entropies for Water in Protein Interiors. J. Am. Chem. Soc. 2004; 126:7991–8000. [PubMed: 15212549]

(51). Damjanovi A, Schlessman J, Fitch C, Garcia A, Garcia-Moreno E. Role of Flexibility and Polarity as Determinants of the Hydration of Internal Cavities and Pockets in Proteins. Biophys. J. 2007; 93:2791–2804. [PubMed: 17604315]

(52). Chakrabarty S, Warshel A. Capturing the Energetics of Water Insertion in Biological Systems: The Water Flooding Approach. Proteins. 2013; 81:93–106. [PubMed: 22911614]

(53). Oikawa M, Yonetani Y. Molecular Dynamics Free Energy Calculations to Assess the Possibility of Water Existence in Protein Nonpolar Cavities. Biophys. J. 2010; 98:2974–2983. [PubMed: 20550910]

(54). Baudry J, Tajkhorshid E, Molnar F, Phillips J, Schulten K. Molecular Dynamics Study of Bacteriorhodopsin and the Purple Membrane. J. Phys. Chem. B. 2001; 105:905–918.

(55). Tashiro M, Stuchebrukhov AA. Thermodynamic Properties of Internal Water Molecules in the Hydrophobic Cavity around the Catalytic Center of Cytochrome c Oxidase. J. Phys. Chem. B. 2005; 109:1015–1022. [PubMed: 16866474]

(56). Pisliakov AV, Sharma PK, Chu ZT, Haranczyk M, Warshel A. Electrostatic Basis for the Unidirectionality of the Primary Proton Transfer in Cytochrome C Oxidase. Proc. Natl. Acad. Sci. USA. 2008; 105:7726–7731. [PubMed: 18509049]

(57). Lee HJ, Svahn E, Swanson JMJ, Lepp H, Voth GA, Brzezinski P, Gennis RB. Intricate Role of Water in Proton Transport through Cytochrome c Oxidase. J. Am. Chem. Soc. 2010; 132:16225–16239. [PubMed: 20964330]

(58). Helms V, Wade R. Thermodynamics of Water Mediating Protein-Ligand Interactions in Cytochrome P450cam: a Molecular Dynamics Study. Biophys. J. 1995; 69:810–824. [PubMed: 8519982]

(59). Li Z, Lazaridis T. Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease. J. Am. Chem. Soc. 2003; 125:6636–6637. [PubMed: 12769565]

(60). Baron R, Setny P, McCammon JA. Water in Cavity-Ligand Recognition. J. Am. Chem. Soc. 2010; 132:12091–12097. [PubMed: 20695475]

(61). Szep S, Park S, Boder ET, Van Duyne GD, Saven JG. Structural Coupling between FKBP12 and Buried Water. Proteins. 2009; 74:603–611. [PubMed: 18704951]

(62). Prakash P, Sayyed-Ahmad A, Gorfe AA. The Role of Conserved Waters in Conformational Transitions of Q61H K-ras. PLoS Comput. Biol. 2012; 8:e1002394. [PubMed: 22359497]

(63). De Simone A, Dodson GG, Verma CS, Zagari A, Fraternali F. Prion and Water: Tight and Dynamical Hydration Sites Have a Key Role in Structural Stability. Proc. Natl. Acad. Sci. USA. 2005; 102:7535–7540. [PubMed: 15894615]

(64). Jensen M, Cool R, Mortensen K, Clark B, Parmeggiani A. Structure-function Relationships of Elongation Factor Tu. Isolation and Activity of the Guanine-nucleotide-binding Gomain. Eur. J. Biochm. 1989; 182:247–255.

(65). Bocchini V, Adinolfi BS, Arcari P, Arcucci A, Russo AD, Vendittis ED, Ianniciello G, M. Masullo M, Raimo G. Protein Engineering on Enzymes of the Peptide Elongation Cycle in Sulfolobus Solfataricus. Biochimie. 1998; 80:895–898. [PubMed: 9893948]

(66). Song H, Parsons MR, Rowsell S, Leonard G, Phillips SE. Crystal Structure of Intact Elongation Factor EF-Tu from Escherichia coli in GDP Conformation at 2.05 Å Resolution. J. Mol. Biol. 1999; 285:1245–1256. [PubMed: 9918724]

(67). Vitagliano L, Ruggiero A, Masullo M, Cantiello P, Arcari P, Zagari A. The Crystal Structure of Sulfolobus Solfataricus Elongation Factor 1a in Complex with Magnesium and GDP. Biochemistry. 2004; 43:6630–6636. [PubMed: 15157096]

(68). Phillips JC, Braunand R, Wei W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable Molecular Dynamics with NAMD. J. Comp. Chem. 2005; 26:1781–1802. [PubMed: 16222654]

(69). MacKerell AD, Feig M, Brooks CL III. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-phase Quantum Mechanics in Reproducing Protein

Conformational Distributions in Molecular Dynamics Simulations. J. Comput. Chem. 2004; 25:1400–1415. [PubMed: 15185334]

(70). Impey RW, Madden PA, McDonald IR. Hydration and Mobility of Ions in Solution. J. Phys. Chem. 1983; 87:5071–5083.

(71). Marchi M, Sterpone F, Ceccarelli M. Water rotational relaxation and diffusion in hydrated lysozyme. J. Am. Chem. Soc. 2002; 124:6787–91. [PubMed: 12047201]

(72). Tuckerman, M. Statistical Mechanics. Theory and Molecular Simulation. Oxford University Press; 2010.

(73). Fiorin G, Klein ML, Hénin J. Using Collective Variables to Drive Molecular Dynamics Simulations. Mol. Phys. 2013; 111:3345–3362.

(74). Hartigan, J. Clustering Algorithms. Wiley; New York: 1975.

(75). Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. 2009.

(76). Denisov VP, Peters J, Hörlein HD, Halle B. Using Buried Water Molecules to Explore the Energy Landscape of Proteins. Nat. Struct. Mol. Biol. 1996; 3:505–509.

(77). Garcia AE, Hummer G. Water Penetration and Escape in Proteins. Proteins. 2000; 38:261–72. [PubMed: 10713987]

(78). Persson F, Halle B. Transient Access to the Protein Interior: Simulation versus NMR. J. Am. Chem. Soc. 2013; 135:8735–8748. [PubMed: 23675835]

(79). Mattea C, Qvist J, Halle B. Dynamics at the Protein-Water Interface from 17O Spin Relaxation in Deeply Supercooled Solutions. Biophys. J. 2008; 95:2951–63. [PubMed: 18586840]

(80). Knight JDR, Hamelberg D, McCammon JA, Kothary R. The Role of Conserved Water Molecules in the Catalytic Domain of Protein Kinases. Proteins. 2009; 76:527–535. [PubMed: 19425109]

(81). Razvi A, Scholtz JM. Lessons in Stability from Thermophilic Proteins. Protein Sci. 2006; 15:1569–1578. [PubMed: 16815912]

(82). Liu C-C, LiCata VJ. The stability of Taq DNA Polymerase Results from a Reduced Entropic Folding Penalty; Identification of other Thermophilic Proteins with Similar Folding Thermodynamics. Proteins. 2013 n/a–n/a.

(83). Budkevich TV, Timchenko AA, Tiktopulo EI, Negrutskii BS, Shalak VF, Petrushenko ZM, Aksenov VL, Willumeit R, Kohlbrecher J, Serdyuk IN, et al. Extended Conformation of Mammalian Translation Elongation Factor 1A in Solution. Biochemistry. 2002; 41:15342–15349. [PubMed: 12484773]

(84). Deutschman WA, Dahlquist FW. Thermodynamic Basis for the Increased Thermostability of CheY from the Hyperthermophile Thermotoga Maritima. Biochemistry. 2001; 40

(85). Maruyama Y, Harano Y. Does Water Drive Protein Folding? Chem. Phys. Lett. 2013; 581:85–90.

(86). Chong S-H, Ham S. Protein Folding Thermodynamics: A New Computational Approach. J. Phys. Chem. B. 2014; 118:5017–5025. [PubMed: 24779395]

(87). Pizzitutti F, Marchi M, Sterpone F, Rossky PJ. How Protein Surfaces Induce Anomalous Dynamics of Hydration Water. J. Phys. Chem. B. 2007; 111:7584–90. [PubMed: 17564431]

(88). Roche J, Dellarole M, Caro JA, Norberto DR, Garcia AE, Garcia-Moreno BE, Roumestand C, Royer CA. Effect of Internal Cavities on Folding Rates and Routes Revealed by Real-Time Pressure-Jump NMR Spectroscopy. J. Am. Chem. Soc. 2013; 135:14610–14618. [PubMed: 23987660]

Europe PMC Funders Author Manuscripts

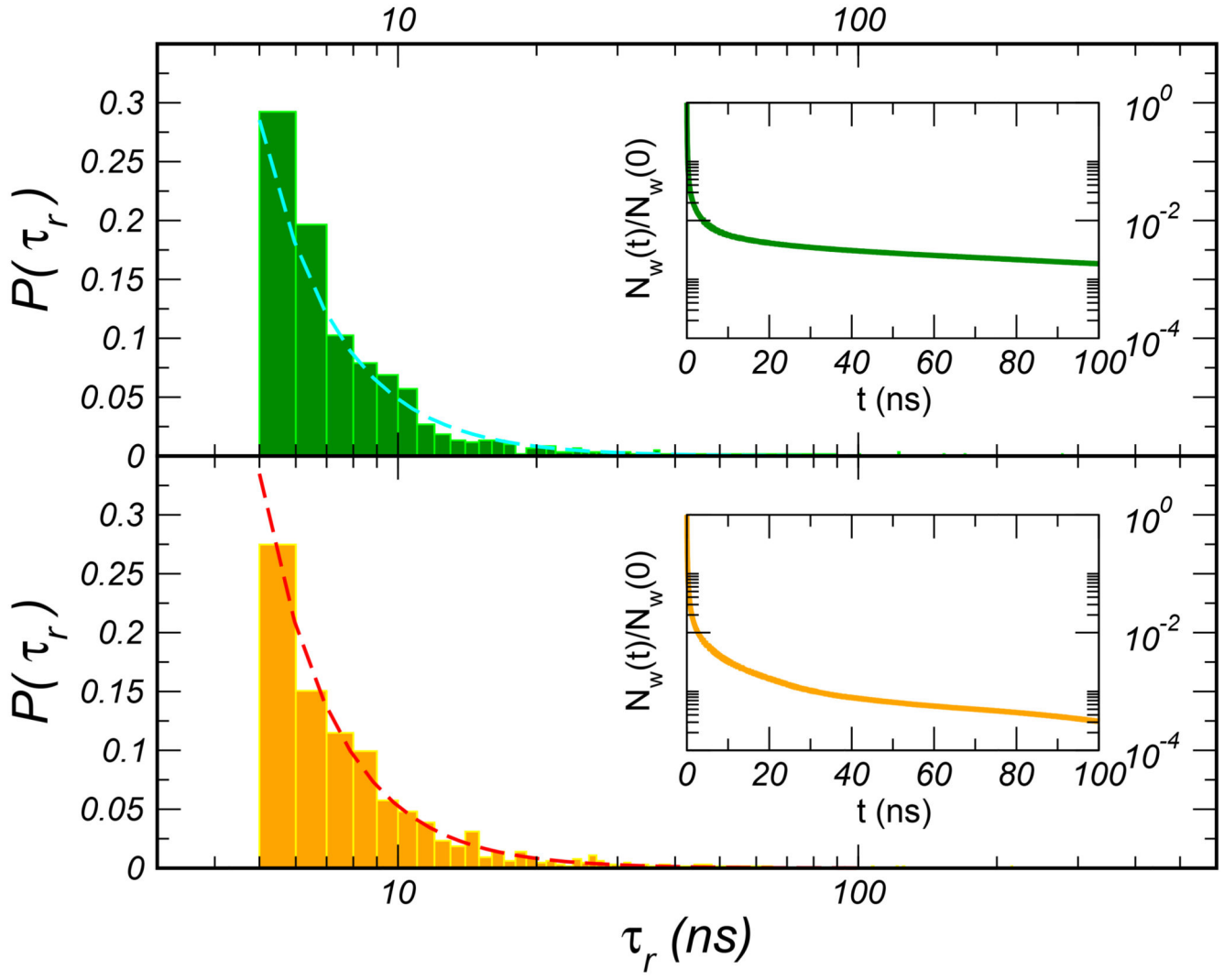

Figure 1.

Frequency distribution of the long residence times τ_r for the $\mathscr{M}$ (top) and $\mathscr{H}$ (bottom) proteins along with the power-law fits (dashed lines). In the figure insets, the fraction of water molecules that continuously stay in the hydration shell as a function of time are plotted.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

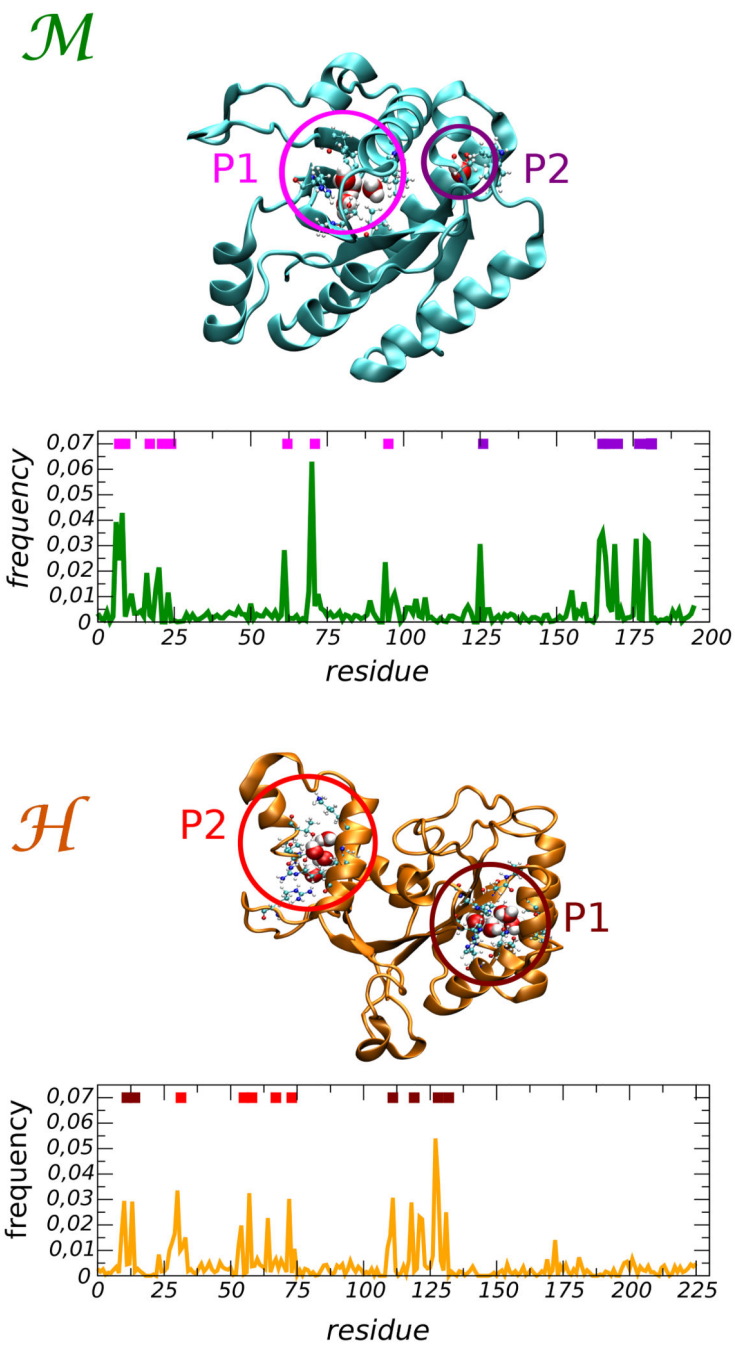

Figure 2.

Frequency distribution of long residence water molecules in the protein matrix for the $\mathscr{M}$ (top) and the $\mathscr{H}$ (bottom) domains. Two pockets are distinguished for each protein and indicated by the colored bars within the graphs. Their location in the 3D protein structure is shown at the top of each graph.

Europe PMC Funders Author Manuscripts

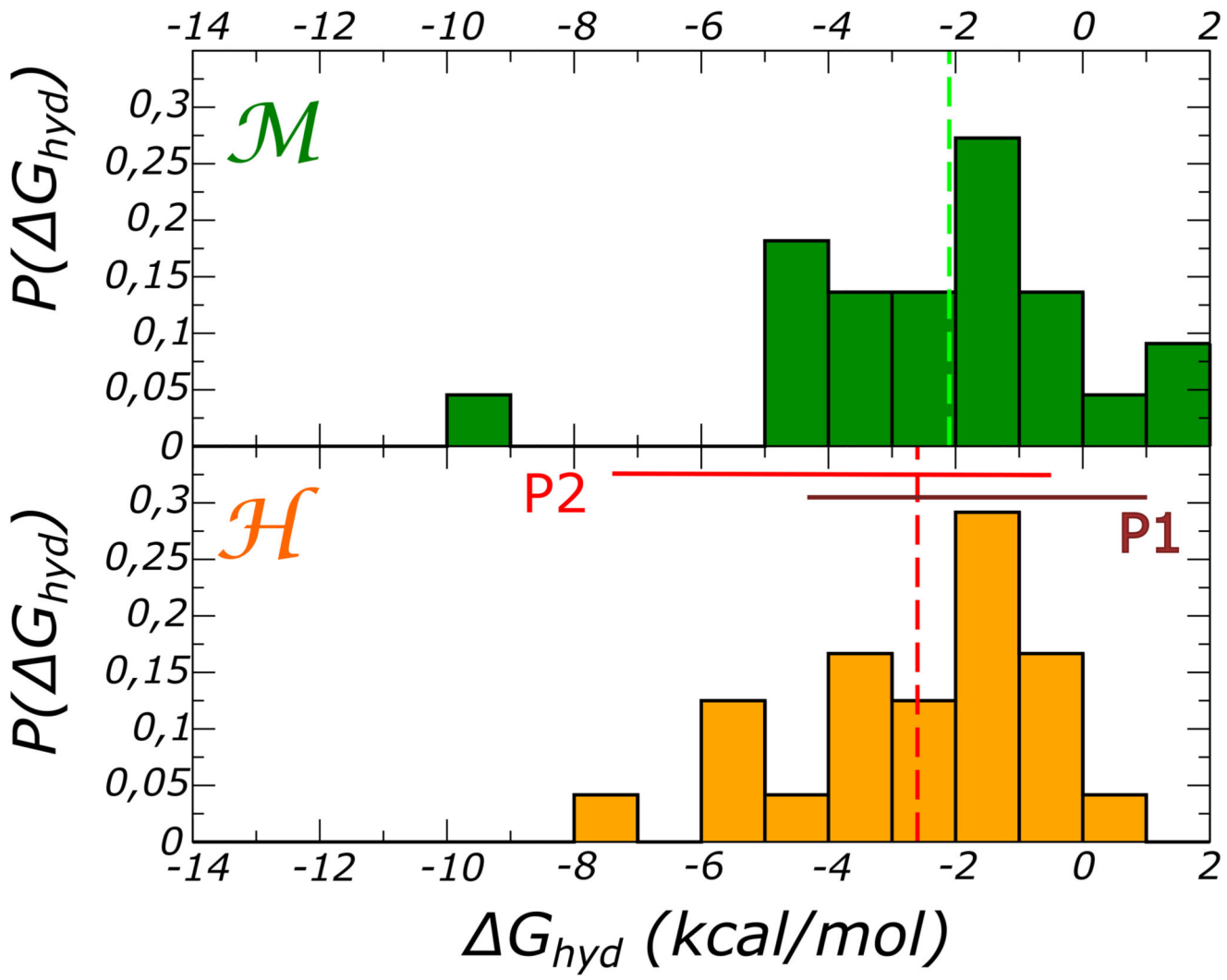

Figure 3.

Frequency distribution of the hydration free energy calculated from representative molecules hydrating the internal cavities of $\mathscr{M}$ (top) and the $\mathscr{H}$ (bottom) domains. The calculations were performed at T = 300 K. The vertical lines indicate the average value $\langle \Delta G_i \rangle$.

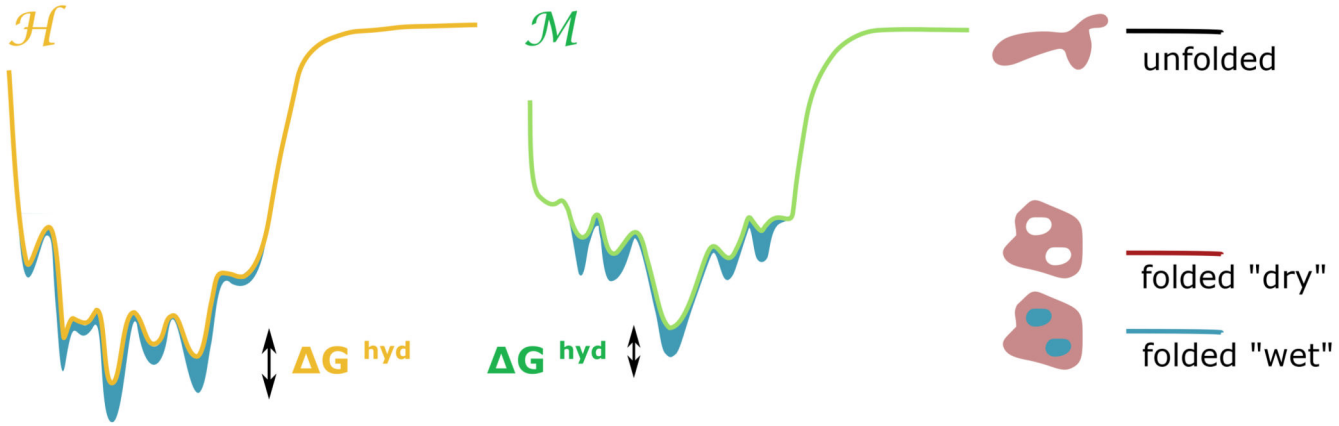

Figure 4.

Schematic view of the contribution from internal water to protein stability (blue zone) as a function of a generic conformational coordinate. The free energy landscape associated with the most stable $\mathscr{H}$ protein is represented in yellow, while that associated with the less stable $\mathscr{M}$ protein is represented in green. We assume that the unfolding free energy is decomposed in two terms, one that measures the difference between the unfolded and “dry” folded state (with the internal cavities dehydrated), and another term (blue filled zone) that measures the gain from internal hydration of the cavities. The free energy gap between the two proteins then contains a contribution from the difference in internal hydration free energy, $\Delta\Delta G_{hyd}$.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

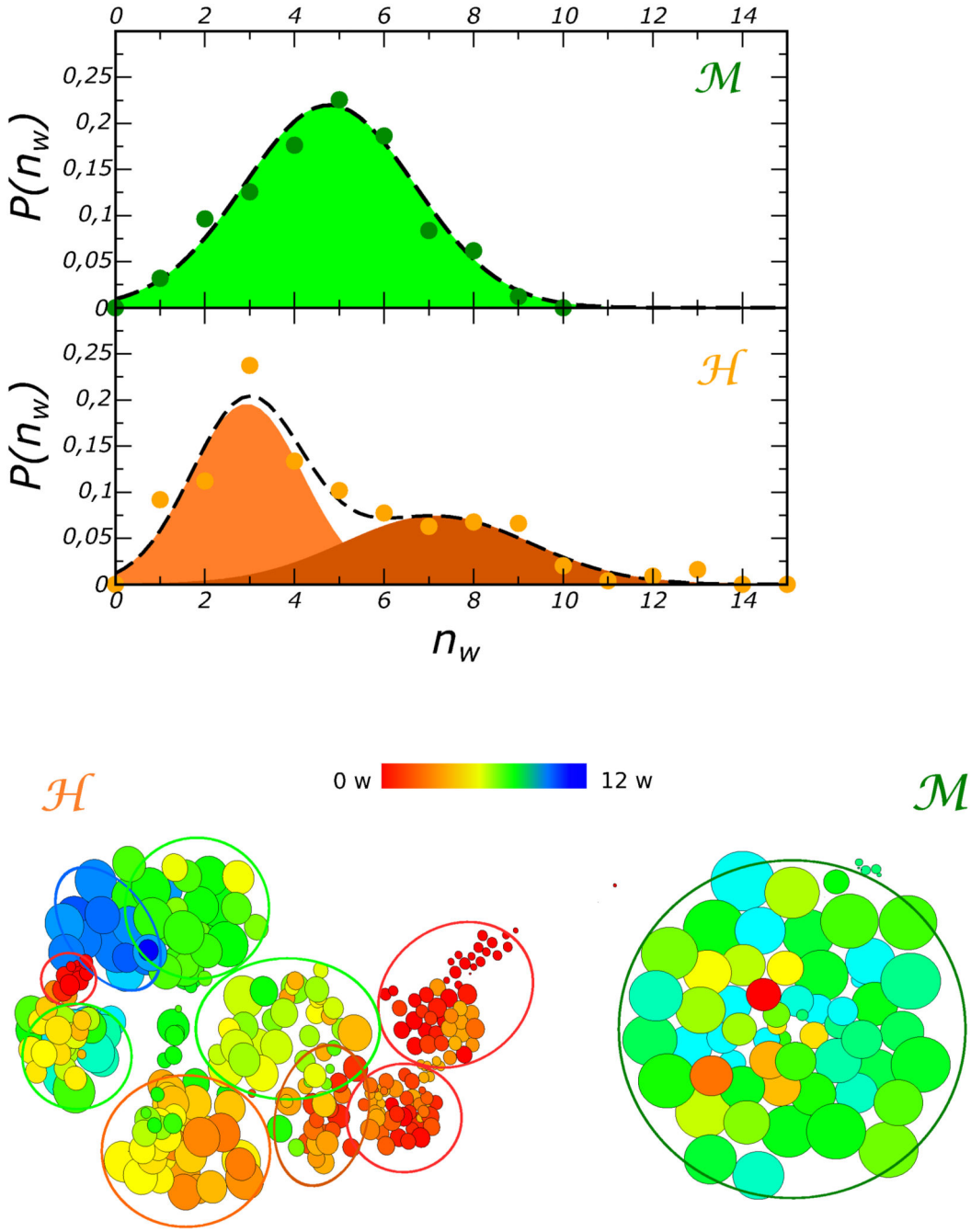

Figure 5.

Top panel. Frequency distribution of the number of long-residence water simultaneously located inside the protein matrix, n_w. For this plot the residence time threshold is $\tau_c > 15$ ns. The colored zones represent the area underneath the Gaussian functions fitting the distributions.

Bottom panel. Network representation of the conformational landscapes sampled by the two proteins. The network is obtained by clustering the MD trajectories and using the native

contacts order parameter *Q*. The color map represents the hydration states of internal sites, red indicates fully dehydration, while blue indicates maximal internal hydration.

Europe PMC Funders Author Manuscripts

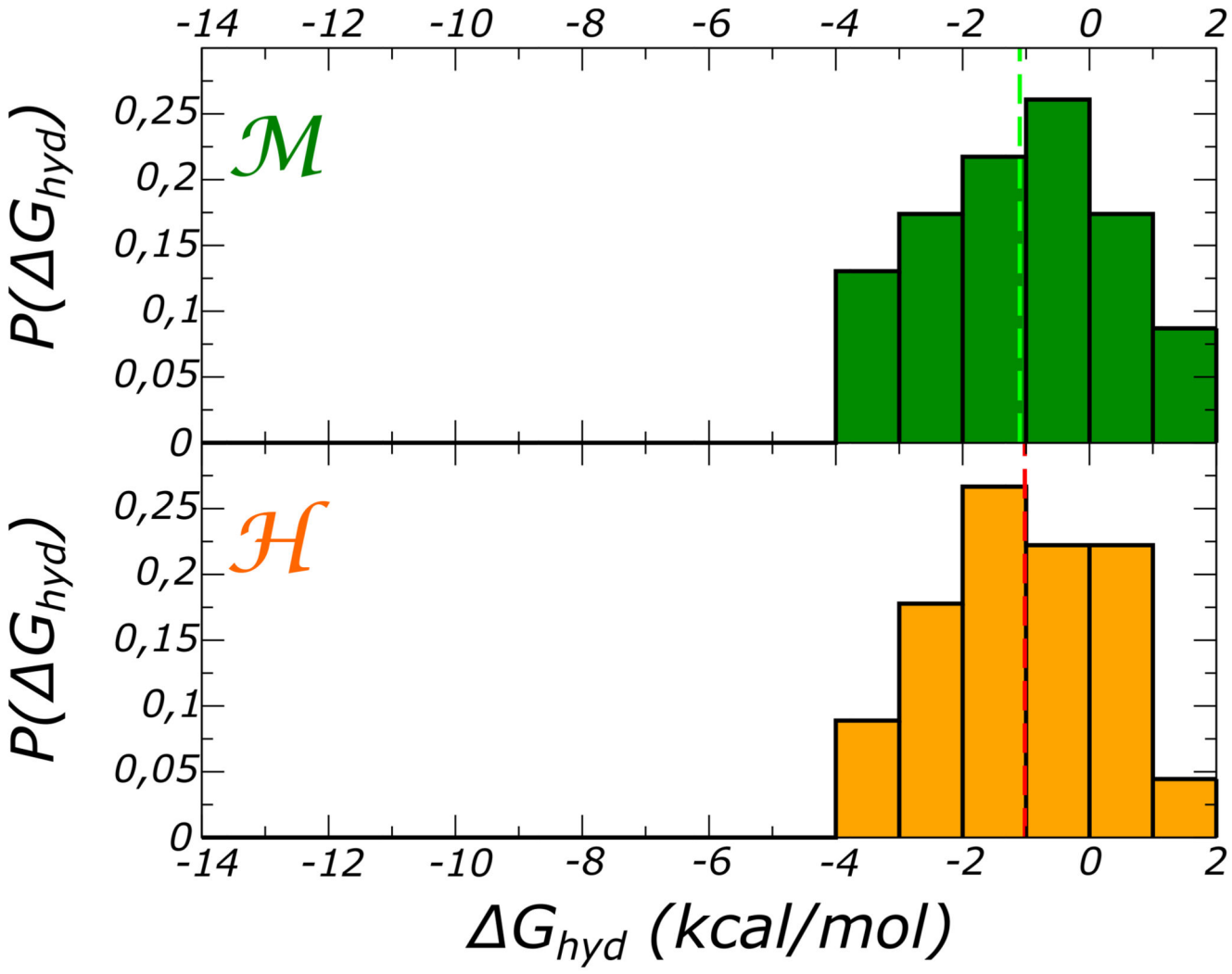

Figure 6.

Frequency distribution of the hydration free energy at T=360 K calculated for the same set of water molecules used in 3. The top panel refers to $\mathscr{M}$; the bottom one, to the $\mathscr{H}$ domain.

Europe PMC Funders Author Manuscripts