Supplemental Information. 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,

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

E-mail: simone.melchionna@roma1.infn.it; jerome.henin@ibpc.fr; fabio.sterpone@ibpc.fr

[*]To whom correspondence should be addressed

[†]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

The ensemble of structural water in the X-ray structures

Table 1: Location of internal water in the X-ray structures of the EF-Tu and EF-1α proteins that are retrieved by the slow hydration dynamics in the MD simulations. Residues are indicated according to the numbering of our systems. To recover the numbering of the Pdb a shift of +7 a dn +3 should be done for the $\mathcal{M}$ and $\mathcal{H}$ proteins, respectively. For the HB partners we indicate the residue name and index, the chemical group location, i.e. backbone (bb) or side-chain (sc), and the connectivity to other water molecules (w).

water in 1EFC ($\mathcal{M}$)	location
$w_{X-ray}6018$	$S_{166}^{N^{bb}},^{O^{bb}}, W_{177}^{O^{bb}}$
w_{X-ray}^{6027}	$H_{12}^{N^{bb}}, H_{15}^{N^{sc}}, T_{108}^{O^{sc}}, V_{97}^{O^{bb}}$
w_{X-ray}^{6123}	$G_{11}^{O^{bb}}, Q_{107}^{O^{sc}}, w$
water in 1SQK ($\mathcal{H}$)	**location**
w_{X-ray}^{6564}	$H_{11}^{N^{bb}}, V_{111}^{O^{bb}}, T_{129}^{O^{sc}}$
w_{X-ray}^{6629}	$H_{14}^{N^{sc}}, Y_{119}^{O^{bb}}, T_{129}^{O^{sc}}$
w_{X-ray}^{6123}	$G_{11}^{O^{bb}}, Q_{107}^{O^{sc}}, w$

The ensemble of long residence water molecules

The location, hydrogen bond connectivity and volume of the water molecules selected for the free energy calculations are reported in table 2 and 3 for the mesophilic ($\mathcal{M}$) and hyperthermophilic ($\mathcal{H}$) domains and for the three selected configurations, A, B, and C.

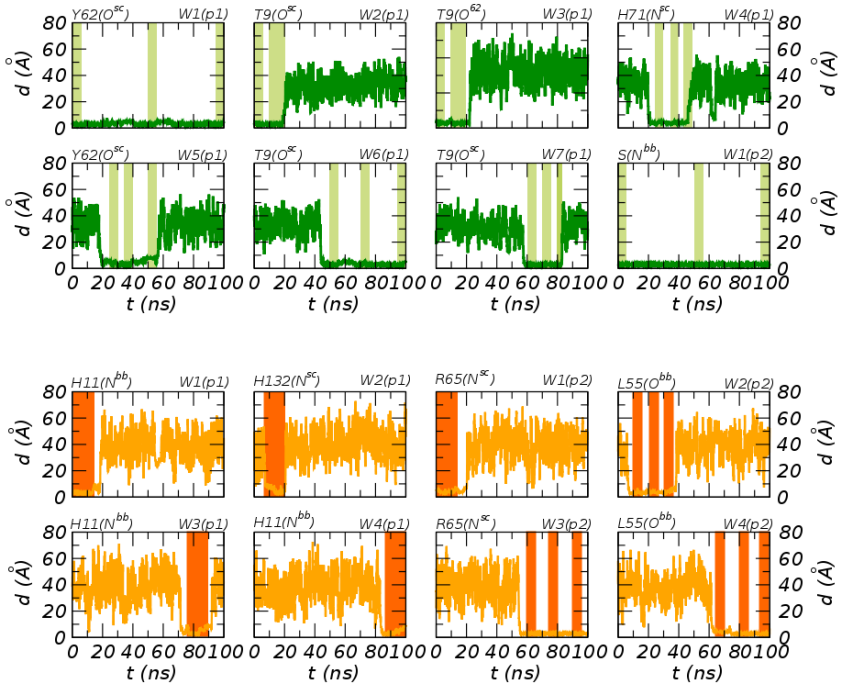

Figure 1: Time evolution of the distances between the selected molecule W_i and a reference atom in the cavity of the protein. The tree representative configurations A_i, B_i, C_i of the molecule W_i were extracted are at time intervals indicated by the shaded area. The protein cavities are indicated by p_1 and p_2. The label of the reference atom is reported on the left corner of each plot.

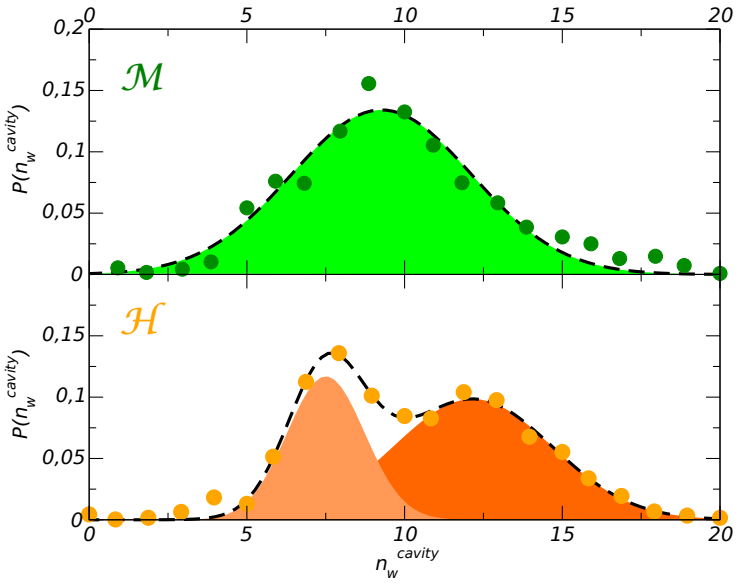

Figure 2: Frequency distribution of the number of long residence water simultaneously present in the protein matrix. The kinetic cut-off is $\tau_r > 5ns$.

Hydration free energy

In order to estimate the total free energy contribution of internal water an alternative formulation for the mean occupancy n_i can be used in eq. 8 in the main text.

$$n_i = \frac{1}{1 + \exp(\beta \Delta G_i - \mu)} \tag{1}$$

In principle, μ represents the bulk chemical potential of water, and should be zero as it is already accounted for in the local hydration free energies. Yet in this case, due to partial overlap between the sites, a direct estimate of occupancies based on our free energies would include some double counting, and hence overestimate the total hydration number of the protein. Since this overlap is impossible to quantify directly, we introduce a non-zero value of μ as an empirical free energy correction, whose role is to ensure that the total predicted occupancy matches

the average number of bound water molecules measured in long simulations. This alternative formulation gives exactly the same estimate for $\Delta\Delta G_{hyd}$: 1.3 kcal/mol in favor of $\mathcal{H}$.

Error estimate ΔG_{hyd}

Propagating the error estimates for local water binding free energies into our estimator for ΔG_{hyd} (and neglecting the variance of n^i), we find:

$$\sigma^2(\Delta G_{hyd}) = \sum_{i=1}^{8} n_i^2 \frac{\sum_{k=1}^{3} \left(\exp(-\beta \Delta G_i^k)\sigma(\Delta G_i^k)\right)^2}{\left(\sum_{k=1}^{3} \exp(-\beta \Delta G_i^k)\right)^2} \tag{2}$$

Stability curves

In order to evaluate the effect of the stability gain caused by internal hydration we have considered two model systems that have been characterised systematically from the thermodynamic point of view. The first one is a pair of proteins similar to the ones studied in this work, namely we selected the Elongation factor 1α from rabbit (eEF1a) and from the thermophilic bacterium *Thermus thermophilius* (EF1a)[2]. The second model is a pair of homologues belonging to the CheY family[3]. From the thermodynamic data that characterise the unfolding we have reconstructed the stability curve for the proteins, see Figure S.3. The stability curve represents the change in the unfolding free energy as a function of temperature and according to Hawley's equation[4] it depends on the measured melting temperature T_m, the unfolding enthalpy at T_m, H_m, and the unfolding heat capacity, ΔC_p. Pressure dependence is not considered here.

$$\Delta G^{unf}(T) = -\Delta C_P \left[T\left(\ln\left(\frac{T}{T_{\mathrm{m}}}\right) - 1\right) + T_{\mathrm{m}} \right] + \Delta H_{\mathrm{m}} \left(\frac{T_{\mathrm{m}} - T}{T_{\mathrm{m}}}\right) \tag{3}$$

The parabola have been upshifted by a constant quantity $\Delta\Delta G_{hyd} = 1.3$ kcal/mol that represents the stability gain due to internal hydration as evaluated comparing the mesophilic and the hyperthermophilic G-domains, see main text. As the plots show, the up-shift increases the melting temperature by about 6 K. We also note that according to a recent study, in about the 70% of the homologous protein pairs characterised thermodynamically, thermal stability of (hyper)thermophiles is also achieved by a smaller ΔC_p of unfolding, or in other words by a broadening of the stability curve. This implies that when up-shift is combined with a decrease in ΔC_p, the effect on the melting temperature is even higher.

	A			B			C		
water	$\cdots$HB	n_w	v_w [Å^3]	$\cdots$HB	n_w	v_w [Å^3]	$\cdots$HB	n_w	v_w [Å^3]
$w_1(P1)$	$Y_{62}^{OH*^{sc}}$,$2w$	3	34.7	$V_7^{O^{bb}}$,$Y_{62}^{OH*^{sc}}$,$H_{71}^{NH*^{sc}}$	2	35.7	$V_7^{O^{bb}}$,$T_9^{O*H^{sc}}$,$H_{71}^{NH*^{sc}}$	1	31.7
$w_2(P1)$	$G_8^{O^{bb}}$,$T_9^{O*H^{sc}}$,w	4	28.9	$T_9^{OH*^{sc}}$,$K_{17}^{O^{bb}}$,$T_{21}^{O*H^{sc}}$	5	32.1	$T_9^{OH*^{sc}}$,$K_{17}^{O^{bb}}$,$T_{21}^{O*H^{sc}}$	5	37.4
$w_3(P1)$	$T_{21}^{O^{bb}}$,$H_{71}^{N^{sc}}$,$2w$	5	38.2	$G_8^{O^{bb}}$,$T_{21}^{O*H^{sc}}$,$H_{71}^{N^{sc}}$,w	3	31.4	$G_8^{O^{bb}}$,$T_9^{O*H^{sc}}$	2	30.9
$w_4(P1)$	$T_9^{OH*^{sc}}$,$K_{17}^{O^{bb}}$,$T_{21}^{O*H^{sc}}$,$H_{71}^{N*^{sc}}$	6	31.6	$Y_{62}^{O*H^{sc}}$,$H_{71}^{NH*^{sc}}$,$2w$	3	36.3	$T_9^{OH*^{sc}}$,$T_{21}^{O*H^{sc}}$,w	3	36.4
$w_5(P1)$	$G_8^{O^{bb}}$,$T_9^{O*H^{sc}}$,w	2	32.9	$Y_{62}^{OH*^{sc}}$,$A_{70}^{O^{bb}}$,$2w$	3	28.4	$T_9^{OH*^{sc}}$,$K_{17}^{O^{bb}}$,$T_{21}^{O*H^{sc}}$	5	34.7
$w_6(P1)$	$T_9^{O*H^{sc}}$,w	2	43.9	$V_7^{O^{bb}}$,$T_9^{O*H^{sc}}$,$H_{71}^{NH*^{sc}}$,w	4	25.0	$T_9^{OH*^{sc}}$,$T_{21}^{O*H^{sc}}$	3	40.0
$w_7(P1)$	$T_9^{OH*^{sc}}$,$K_{17}^{O^{bb}}$,$T_{21}^{O*H^{sc}}$	3	26.9	$T_9^{OH*^{sc}}$,$K_{17}^{O^{bb}}$,$T_{21}^{O*H^{sc}}$,w	5	29.2	$T_9^{OH*^{sc}}$,$K_{17}^{O^{bb}}$,$T_{21}^{O*H^{sc}}$	4	36.3
$w_1(P2)$	$S_{166}^{N^{bb}}$,$W_{177}^{O^{bb}}$	0	31.8	$S_{166}^{N^{bb}}$,$^{O^{bb}}$,$W_{177}^{O^{bb}}$	0	30.9	$S_{166}^{N^{bb}}$,$^{O^{bb}}$,$W_{177}^{O^{bb}}$	1	32.1

Table 2: Characterization of the selected states of the mesophilic system for the free energy calculations. For each selected water molecule $w_i(pj)$ we report the hydrogen bond partners ($\cdots HB$), the number of water molecules present in the surrounding environment ($r_c = 6$), the water volume estimate by voronoi tessellation of the space (v_w). For the HB partners we indicate the residue name and index, the chemical group location, i.e. backbone (bb) or side-chain (sc), the directionality, i.e. the location of $*$ indicates the acceptor or the donor group involved in the HB, and the connectivity to other water molecules (w).

	A			B			C		
water	$\cdots$HB	n_w	v_w [Å^3]	$\cdots$HB	n_w	v_w [Å^3]	$\cdots$HB	n_w	v_w [Å^3]
$w_1(p1)$	$V_{111}^{O^{bb}}$,$H_{14}^{NH*^{sc}}$,w	5	30.9	$H_{11}^{NH*^{bb}}$,$H_{14}^{NH*^{sc}}$,$Q_{128}^{O^{sc}}$,$H_{132}^{N*^{sc}}$	4	30.8	$H_{11}^{NH*^{sc}}$,$V_{111}^{O^{bb}}$,w	3	27.0
$w_2(p1)$	$Q_{128}^{O^{bb}}$,$2w$	7	27.3	$H_{11}^{NH*^{bb}}$,$H_{132}^{N*^{sc}}$,w	6	32.0	$H_{11}^{NH*^{bb}}$,$H_{14}^{NH*^{sc}}$,$V_{111}^{O^{bb}}$,w	7	31.4
$w_3(p1)$	$V_{111}^{O^{bb}}$,$2w$	6	38.9	$H_{11}^{NH*^{bb}}$,$V_{111}^{O^{bb}}$,$Q_{128}^{O^{sc}}$	3	32.5	$V_{111}^{O^{bb}}$,$H_{132}^{N*^{sc}}$,w	3	31.1
$w_4(p1)$	$H_{11}^{NH*^{bb}}$,$H_{14}^{N*H^{sc}}$,$Q_{128}^{O^{sc}}$	5	27.7	$H_{11}^{NH*^{bb}}$,$V_{111}^{O^{bb}}$, $Q_{128}^{O^{sc}}$	4	31.6	$V_{111}^{O^{bb}}$,$2w$	4	33.0
$w_1(p2)$	$D_{31}^{COO*^{sc}}$,$T_{34}^{O*H^{sc}}$,w	5	28.2	$E_{62}^{O*^{bb}}$,$2R_{65}^{NH_2*^{sc}}$,w	4	31.4	$D_{31}^{COO*^{sc}}$,$T_{34}^{O*H^{sc}}$,w	5	30.2
$w_2(p2)$	$L_{55}^{O*^{bb}}$,$R_{65}^{NH_2*^{sc}}$,$2w$	4	29.4	$D_{31}^{COO*^{sc}}$,$L_{55}^{O*^{bb}}$,$R_{65}^{NH_2*^{sc}}$	1	33.1	$D_{31}^{COO*^{sc}}$,$L_{55}^{O*^{bb}}$,$R_{65}^{NH_2*^{sc}}$	3	31.6
$w_3(p2)$	$D_{31}^{COO*^{sc}}$,$R_{65}^{NH_2*^{sc}}$,$R_{65}^{NH*^{sc}}$,$F_{73}^{\pi_{sc}}$	2	35.2	$E_{62}^{O*^{bb}}$,$R_{65}^{NH_2*^{sc}}$,$R_{65}^{NH*^{sc}}$,w	4	31.1	$E_{62}^{O*^{bb}}$,$2R_{65}^{NH_2*^{sc}}$,w	3	28.7
$w_4(p2)$	$D_{31}^{COO*^{sc}}$,$R_{65}^{NH_2*^{sc}}$,$2w$	5	26.8	$L_{55}^{O*^{bb}}$,$R_{65}^{NH_2*^{sc}}$,w	3	30.2	$L_{55}^{O*^{bb}}$,$R_{65}^{NH_2*^{sc}}$,$2w$	3	35.2

Table 3: Characterization of the selected states of the hyperthermophilic system for the free energy calculations. Legends as in table 2.

Table 4: Free energy of hydration for the water molecules in the two pockets of the mesophilic protein and for three independent configurations A, B, and C. Errors estimated for the unidirectional transformation are reported in parentheses[1]. At T=300 K, the free energy to remove a molecule from the bulk water system is $\Delta G_{wat} = 6.2(0.1)$ kcal/mol.

	ΔG_{hyd} (kcal/mol)		
water	A	B	C
w_1^{P1}	-4.9(0.3)	-4.2(0.6)	-4.6(0.2)
w_2^{P1}	2.7(0.2)	-2.1(0.3)	-0.9(0.3)
w_3^{P1}	0.2(0.2)	-1.5(0.3)	-2.9(0.3)
w_4^{P1}	-3.3(0.4)	1.9(0.3)	1.1(0.3)
w_5^{P1}	-2.2(0.2)	-0.5(0.2)	-4.5(0.3)
w_6^{P1}	-1.5(0.3)	-1.4(0.3)	-0.4(0.2)
w_7^{P1}	-1.2(0.2)	-1.8(0.4)	-1.5(0.1)
w_1^{P2}	-9.2(0.3)	-3.5(0.4)	-3.7(0.2)

Table 5: Free energy of hydration for the water molecules in the two pockets of the hyperthermophilic protein and for three independent configurations A, B, and C. Errors estimated for the unidirectional transformation are reported in parentheses.

	ΔG_{hyd} (kcal/mol)		
water	A	B	C
w_1^{P1}	-1.9(0.5)	-1.8(0.3)	-0.6(0.3)
w_2^{P1}	0.1(0.3)	-1.9(0.5)	-2.4(0.3)
w_3^{P1}	-0.5(0.4)	-1.3(0.2)	-2.1(0.3)
w_4^{P1}	-1.6(0.3)	-4.9(0.2)	-0.2(0.3)
w_1^{P2}	-0.7(0.3)	-5.0(0.6)	-3.0(0.6)
w_2^{P2}	-3.4(0.4)	-5.9(0.6)	-3.8(0.3)
w_3^{P2}	-3.7(0.3)	-1.8(0.3)	-7.2(0.3)
w_4^{P2}	-3.6(0.3)	-1.8(0.3)	-1.6(0.4)

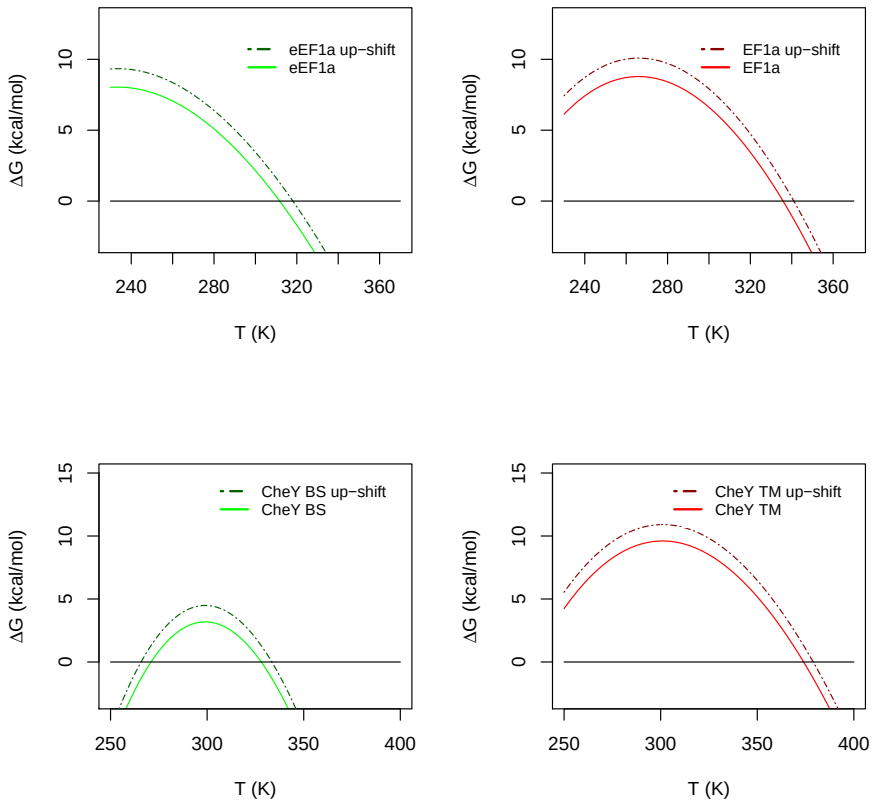

Figure 3: Top panels. Stability curve ($\Delta^{unf}G(T)$) as a function of temperature for the EF-1α proteins from rabbit (left panel, green) and the *T. thermophilus* (right panel, red). Bottom panels. Stability curve for CheY proteins from the mesophilic *B. subtilus* (left panel, green) and from the hyperthermophilic *T. marititima* (right panel, red) organisms. The shifted parabola are reported in each panel in dark green and dark red dashed lines, for the mesophilic and (hyper)thermophilic specie, respectively.

References

(1) Liu, P.; Dehez, F.; Cai, W.; Chipot, C. *Journal of Chemical Theory and Computation* **2012**, *8*, 2606–2616.

(2) Budkevich, T. V.; Timchenko, A. A.; Tiktopulo, E. I.; Negrutskii, B. S.; Shalak, V. F.; Petrushenko, Z. M.; Aksenov, V. L.; Willumeit, R.; Kohlbrecher, J.; Serdyuk, I. N.; El'skaya, A. V. *Biochemistry* **2002**, *41*, 15342–15349.

(3) Deutschman, W. A.; Dahlquist, F. W. *Biochemistry* **2001**, *40*.

(4) Hawley, S. A. *Biochemistry* **1971**, *10*, 2436–2442.