Supporting information:

Physical origin underlying the entropy loss upon hydrophobic hydration

Aljaž Godec[*] and Franci Merzel[*]

National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia

E-mail: aljaz.godec@ki.si; franci.merzel@ki.si

Probability densities and the mapping onto potential energy space

Throughout the paper as well as this section we strictly discriminate between a random variable, which is denoted by uppercase letters, and a value it can take on (lowercase letters). The random variable transformation theorem is used to map the joint configurational probability density, onto its functionally dependent joint probability density for pair potential energies. Thereby molecular positions and orientations, $\boldsymbol{\Omega}^N = (\boldsymbol{R}^N, \boldsymbol{\alpha}^N)$, are mapped onto a vector of interaction energies, $\boldsymbol{U}$:

$$p(\boldsymbol{u}) = \int \mathrm{d}\boldsymbol{\omega}^N P(\boldsymbol{\omega}^N) \delta(\boldsymbol{u} - \boldsymbol{f}(\boldsymbol{\omega}^N)), \tag{1}$$

The functional relation $\delta(\boldsymbol{u} - \boldsymbol{f}(\boldsymbol{\omega}^N))$ contains class indicators (close contact, hydrogen-bonded, etc.) and the the appropriate averaging operation accounting for the indistinguishability of pairs within a certain class.

If we denote the solute position as $\boldsymbol{R}_s$ the general functional expressions $\delta(\boldsymbol{u} - \boldsymbol{f}(\boldsymbol{\omega}^N))$ can be written as follows.

---

[*]To whom correspondence should be addressed

In the scalar case (*i.e.* density $p(U_x)$) we have:

$$\delta(u - f(\boldsymbol{\omega}^N)) \equiv \frac{\sum_{i,j \neq i} \delta(u - u(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j)) \chi_d(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \chi_{HB}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \Pi_{R_{min}, R_{max}}(|\boldsymbol{r}_i - \boldsymbol{r}_s|)}{\int d\boldsymbol{\omega}^N P(\boldsymbol{\omega}^N) \left\{ \sum_{i,j \neq i} \chi_d(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \chi_{HB}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \Pi_{R_{min}, R_{max}}(|\boldsymbol{r}_i - \boldsymbol{r}_s|) \right\}} \quad (2)$$

where $\chi_d(\boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j)$ is the indicator function for the distance cutoff for nearest neighbors

$$\chi_d(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \equiv (1 - H(|\boldsymbol{r}_i - \boldsymbol{r}_j|)), \quad (3)$$

$H(x)$ being the Heaviside function and $\chi_{HB}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j)$ is the indicator function for hydrogen bonding (which is left out in the case of non-hydrogen bonded neighbors)

$$\chi_{HB}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \equiv H(\vartheta[OH_i \cdots O_j] - \vartheta_{min}) + H(\vartheta[OH_j \cdots O_i] - \vartheta_{min}), \quad (4)$$

where $\vartheta[OH_i \cdots O_j]$ is the angle between the OH bond of the donor molecule and vector connecting the proton of the donor molecule and the acceptor oxygen. ϑ_{min} is the lower cutoff for the hydrogen bonding angle. $\Pi_{a,b}(x)$ represents the boxcar function

$$\Pi_{R_{min}, R_{max}}(|\boldsymbol{r}_i - \boldsymbol{r}_s|) \equiv \begin{cases} 0 & \text{if } |\boldsymbol{r}_i - \boldsymbol{r}_s| < R_{min}, \\ 1 & \text{if } R_{min} \leq |\boldsymbol{r}_i - \boldsymbol{r}_s| \leq R_{max}, \\ 0 & \text{if } |\boldsymbol{r}_i - \boldsymbol{r}_s| > R_{max}. \end{cases} \quad (5)$$

which localizes the tagged molecule i in a certain hydration shell (first, second shell or bulk).

In the non-scalar case $p(U)$ is a joint distribution. Then we have, for example for a joint distribution for a 3-body or 4-body density (or equivalently for a pair density of pairs), $p(U_{0x}, U_{0y})$ or $p(U_{0x}, U_{jz})$ respectively:

$$\delta(\boldsymbol{u} - \boldsymbol{f}(\boldsymbol{\omega}^N)) \equiv \frac{\sum_{ijkl} \delta(u - u^{ij}) \delta(u - u^{kl}) \chi_d^{ij} \chi_d^{kl} \chi_{HB}^{ij} \chi_{HB}^{kl} \Pi_{R_{min}, R_{max}}(d_i)(1 - \delta_{il} \delta_{jl} \delta'_{ik} \delta_{jk}) \Theta_c^{ijkl}}{\int d\boldsymbol{\omega}^N P(\boldsymbol{\omega}^N) \left\{ \sum_{ijkl} \chi_d^{ij} \chi_d^{kl} \chi_{HB}^{ij} \chi_{HB}^{kl} \Pi_{R_{min}, R_{max}}(d_i)(1 - \delta_{il} \delta_{jl} \delta'_{ik} \delta_{jk}) \Theta_c^{ijkl} \right\}}, \quad (6)$$

2

where the explicit coordinate dependences $(\mathbf{\Omega}_i, \mathbf{\Omega}_j)$ were replaced by superscripts ^{ij}, $d_i = |\mathbf{r}_i - \mathbf{r}_s|$ and δ_{ij} is the Kronecker delta. The primed Kronecker delta denotes that i may or may not be equal to k, depending on whether we take geminal or vicinal pairs of molecules, that is, 3- or 4- particle correlations (for the explicit definition of geminal and vicinal pairs see main text). For geminal pairs the primed Kronecker delta is omitted. Θ_c^{ijkl} is the connectivity indicator function and defines the topological relation between both pairs. Two pairs may be connected with a hydrogen bond or not.

Corresponding equation for higher dimensional probability densities can be constructed accordingly.

Distribution of dipolar order parameter in the case of SPC/E water

As in the case of TIP5P SPC/E water is on average also slightly more orientationally ordered in the first hydration shell of hydrophobic particles (see Figure 1) but the width of the distributions $p(D)$, which reflects the constraining of orientational degrees of freedom, remains rather unaffected. Comparing the results in bulk TIP5P and SPC/E, shown in the inset of Figure 1, we find that TIP5P is slightly more ordered.

Calculation of the orinentational distribution of the local electrostatic field

The systems under investigation all have spherical symmetry around the center of the solute. Thus, all points in the space-fixed frame (denoted by prime) with with the same radius r' are positionally equivalent. As we classify the water molecules according to their position with respect to the solute as belonging to the first and second hydration shell or the bulk, we can drop the radial component when dealing with molecules in a particular shell and simply retain the angular variables (θ', φ'),
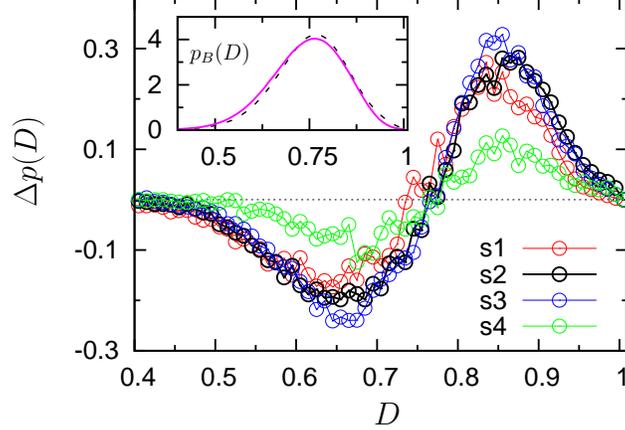
Figure 1: Difference in the distribution of the dipolar order parameter with respect to the bulk, $\Delta p(D) = p(D) - p_B(D)$, in the first hydration shell in the case of SPC/E water. The inset shows the comparison between $p(D)$ in bulk TIP5P (dashed) and in SPC/E (magenta).

where $\theta' = \arccos(\frac{z'}{r'})$ and $\varphi' = \arctan(\frac{y'}{x'})$. We choose to describe the orientational distribution of the instantaneous electrostatic filed orientation in a given point of the space-fixed frame as a distribution over the surface of a unit sphere. In order to take into account the equivalence of positions at given radius in the space-fixed frame we set the secondary local frame in the following manner. The secondary x, y and z axes are chosen to coincide with the polar, azimuthal and radial directions at (θ', φ'). Specifically, this amounts to the following basis:

$$\hat{\mathbf{x}}(\theta', \varphi') = \begin{pmatrix} \cos\varphi' \sin\theta' \\ \sin\varphi' \cos\theta' \\ -\sin\theta' \end{pmatrix}, \qquad \hat{\mathbf{y}}(\theta', \varphi') = \begin{pmatrix} -\sin\varphi' \\ \cos\varphi' \\ 0 \end{pmatrix}, \qquad \hat{\mathbf{z}}(\theta', \varphi') = \begin{pmatrix} \cos\varphi' \sin\theta' \\ \sin\varphi' \sin\theta' \\ \cos\theta' \end{pmatrix}. \tag{7}$$

The local electrostatic field at a point $\mathbf{r}' = (x', y', z')$ due to a collection of N point charges q_i with coordinates $\mathbf{r}_i''$ is $\mathbf{E}(\mathbf{r}') = \sum_{i=1}^{N} \frac{q_i(\mathbf{r}' - \mathbf{r}_i'')}{|\mathbf{r}' - \mathbf{r}_i''|^3}$. The unit Cartesian components of the field in the secondary frame at (r', θ', φ') are thus:

$$\hat{E}_x(r', \theta', \varphi') = \frac{(\mathbf{E}(\mathbf{r}') \cdot \hat{\mathbf{x}})}{|\mathbf{E}(\mathbf{r}')|}, \qquad \hat{E}_y(r', \theta', \varphi') = \frac{(\mathbf{E}(\mathbf{r}') \cdot \hat{\mathbf{y}})}{|\mathbf{E}(\mathbf{r}')|}, \qquad \hat{E}_z(r', \theta', \varphi') = \frac{(\mathbf{E}(\mathbf{r}') \cdot \hat{\mathbf{z}})}{|\mathbf{E}(\mathbf{r}')|}. \tag{8}$$

4

After a trivial transformation to spherical coordinates in the secondary frame, we may write for the orientational distribution of the electrostatic field experienced by a water molecule in shell d_i (first, second or bulk):

$$p(\theta,\varphi)[d_j] = \left\langle \frac{\sum_{i=1}^{N} \delta(\hat{E}_{\theta,i} - \theta)\delta(\hat{E}_{\varphi,i} - \varphi)\Pi_{R_{min},R_{max}}(d_i)}{\sum_{i=1}^{N} \Pi_{R_{min},R_{max}}(d_i)} \right\rangle. \tag{9}$$

$p(\theta,\varphi)$ in the second hydration shell of solute S2 and in bulk water are shown in Figure 2. Compared to those in the first hydration shell shown in the main article, which exhibit pronounced orientational ordering, they are isotropic.
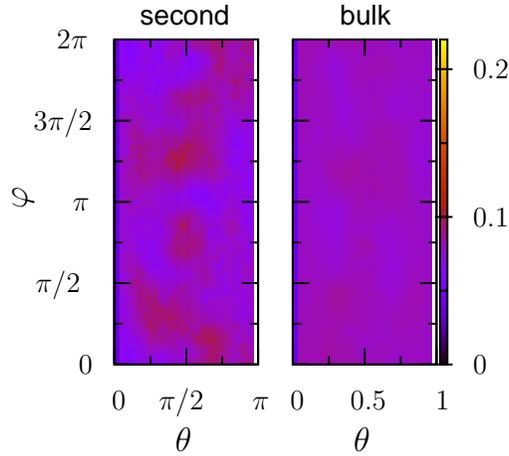
Figure 2: $p(\theta,\varphi)$ in the second hydration shell of the apolar solute S4 (left) and in bulk water.

Gibbs-Shanon and correlation entropy differences

We denote a tagged molecule with 0, its neighbors with i, j, k; $i \neq j \neq k \neq \ldots$, and the nearest neighbors of the nearest neighbors with primes. A 3-body correlation entropy of a tagged molecule and its two nearest HB neighbors is thus, for example, S_{HB-HB}^{i0-0j}, and the 4-body correlation entropy of two adjacent HB pairs is $S_{HB-HB}^{i0-jk'}$, if j is not hydrogen-bonded to 0, and $S_{HB-HB}^{i0-jk'*}$, if j is a HB neighbor of 0.
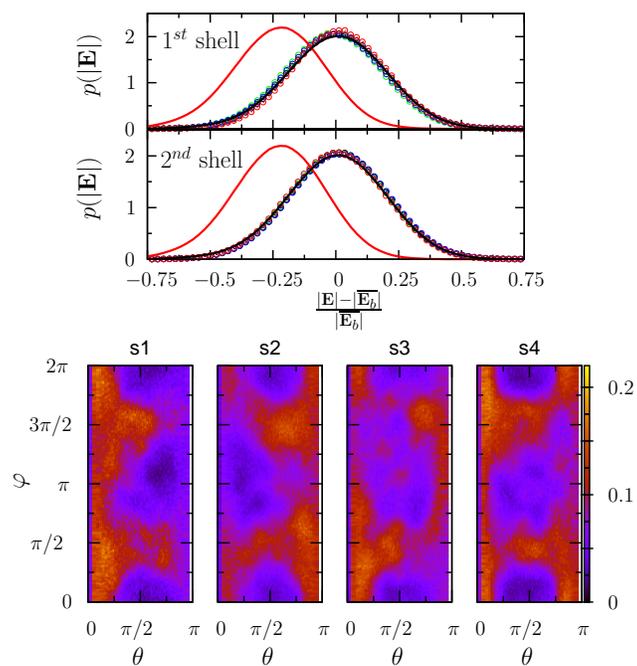
Figure 3: $p(|\boldsymbol{E}|)$ (top) and $p(\theta, \varphi)$ (bottom) in the first hydration shell of the apolar solutes in SPC/E water. Top: black lines denote distributions in bulk water and symbols represent results for water molecules around apolar solute. If one nearest neighbor water molecule is omitted from the calculation of the local field in the bulk, this leads to uncompensated fluctuations shown by the red distribution.

6

Table 1: Entropy differences per TIP5P water molecule/cluster in the first and second shell for various solute sizes. ΔS_i stands for $S^i - S^i_B$ and the asterisk denotes that molecule j is hydrogen-bonded to the tagged molecule 0.

	s1 1st	s1 2nd	s2 1st	s2 2nd	s3 1st	s3 2nd	s4 1st	s4 2nd
$\Delta S_D [10^{-3} k_B]$	-7.9	-6.4	4.5	2.0	-4.6	-6.1	5.6	3.3
$\Delta S_D / S_D^B [10^{-1}\%]$	8.6	6.9	-4.9	-2.2	5.0	6.7	-6.2	-3.6
$\Delta S_{tot-c} [10^{-2} k_B]$	-1.2	-0.3	0.4	0.2	2.6	-0.3	1.9	0.2
$\Delta S_{tot-c} / S_{tot-c}^B [\%]$	-0.5	-0.1	0.2	0.1	1.1	-0.1	0.7	0.1
$\Delta S_{HB} [10^{-2} k_B]$	-2.8	-1.1	-2.6	0.4	-2.2	0.07	-2.1	-0.5
$\Delta S_{HB} / S_{HB}^B [\%]$	-1.9	-0.7	-1.7	0.3	-1.5	0.05	-1.4	0.3
$\Delta S_{nHB} [10^{-2} k_B]$	-4.8	-0.5	-4.8	-0.1	-9.1	-0.3	-8.3	-0.9
$\Delta S_{nHB} / S_{HB}^B [\%]$	-2.6	-0.3	-2.6	-0.05	-4.9	-0.16	-4.4	-0.5
$\Delta S_{HB-HB}^{i0-0j} [10^{-3} k_B]$	5.1	1.7	4.8	1.3	6.7	1.9	5.2	1.8
$\Delta S_{HB-HB}^{i0-0j} / S_{HB-HB}^{i0-0j;B} [\%]$	94	31	91	25	126	36	98	34
$\Delta S_{HB-nHB}^{i0-0j} [10^{-3} k_B]$	6.9	2.0	6.0	1.8	7.4	2.2	6.1	1.7
$\Delta S_{HB-nHB}^{i0-0j} / S_{HB-nHB}^{i0-0j;B} [\%]$	812	235	706	212	871	256	718	200
$\Delta S_{HB-HB}^{i0-jk'} [10^{-3} k_B]$	3.9	1.2	3.8	1.1	4.8	1.3	3.8	1.1
$\Delta S_{HB-HB}^{i0-jk'} / S_{HB-HB}^{i0-jk;B} [\%]$	910	440	884	260	1120	320	884	256
$\Delta S_{HB-HB}^{i0-jk'*} [10^{-3} k_B]$	4.7	1.2	4.4	0.9	6.0	1.4	4.1	1.4
$\Delta S_{HB-HB}^{i0-jk'*} / S_{HB-HB}^{i0-jk*;B} [\%]$	72	18	68	14	92	22	63	22
$\Delta S_{nHB-nHB}^{i0-jk'} [10^{-2} k_B]$	2.2	0.6	4.3	1.1	5.6	1.3	4.1	0.9
$\Delta S_{nHB-nHB}^{i0-jk'} / S_{nHB-nHB}^{i0-jk;B} [\%]$	149	41	291	74	378	88	277	61
$\Delta S_{nHB-nHB}^{i0-jk'*} [10^{-2} k_B]$	0.9	0.3	1.3	0.3	1.6	0.4	1.2	0.3
$\Delta S_{nHB-nHB}^{i0-jk'*} / S_{nHB-nHB}^{i0-jk*;B} [\%]$	1216	405	1757	405	2162	541	1622	405
$\Delta S_{HB-nHB}^{i0-jk'} [10^{-2} k_B]$	0.7	0.2	0.9	0.2	1.2	0.3	0.9	0.2
$\Delta S_{HB-nHB}^{i0-jk'} / S_{HB-nHB}^{i0-jk;B} [\%]$	875	250	1125	250	1500	375	1125	250
$\Delta S_{HB-nHB}^{i0-jk'*} [10^{-2} k_B]$	0.8	0.2	0.9	0.2	1.2	0.3	0.9	0.3
$\Delta S_{HB-nHB}^{i0-jk'*} / S_{HB-nHB}^{i0-jk*;B} [\%]$	145	26	164	26	218	55	164	55

Table 2: Entropy differences per SPC/E water molecule/cluster in the first and second shell for various solute sizes. ΔS_i stands for $S^i - S_B^i$ and the asterisk denotes that molecule j is hydrogen-bonded to the tagged molecule 0.

	s1 1st	s1 2nd	s2 1st	s2 2nd	s3 1st	s3 2nd	s4 1st	s4 2nd
$\Delta S_D[10^{-3}k_B]$	-8.9	3.8	3.9	1.9	-4.6	-2.4	3.9	1.4
$\Delta S_D/S_D^B[10^{-1}\%]$	9.8	-4.2	-4.3	-2.1	5.1	2.7	-4.3	-1.6
$\Delta S_{tot-c}[10^{-2}k_B]$	-1.1	0.4	0.3	0.2	-0.3	-0.04	0.6	-0.08
$\Delta S_{tot-c}/S_{tot-c}^B[\%]$	-0.4	0.1	0.1	0.07	-0.1	-0.01	0.22	0.03
$\Delta S_{HB}[10^{-2}k_B]$	-1.8	-0.6	-2.1	0.4	-2.8	0.3	-2.4	-0.6
$\Delta S_{HB}/S_{HB}^B[\%]$	-1.3	-0.4	-1.5	0.3	-2.0	0.2	-1.7	0.4
$\Delta S_{nHB}[10^{-2}k_B]$	-5.8	0.8	-7.8	-0.8	-7.9	0.3	-8.9	-0.9
$\Delta S_{nHB}/S_{HB}^B[\%]$	-2.9	0.4	-3.9	-0.4	-3.9	-0.16	-4.9	-0.5
$\Delta S_{HB-HB}^{i0-0j}[10^{-3}k_B]$	4.5	1.6	3.8	0.6	3.9	0.8	3.8	0.6
$\Delta S_{HB-HB}^{i0-0j}/S_{HB-HB}^{i0-0j;B}[\%]$	85	30	71	11	74	15	72	11
$\Delta S_{HB-nHB}^{i0-0j}[10^{-3}k_B]$	10.4	4.0	8.8	1.9	9.9	2.1	10.9	2.7
$\Delta S_{HB-nHB}^{i0-0j}/S_{HB-nHB}^{i0-0j;B}[\%]$	640	246	541	116	609	129	671	166
$\Delta S_{HB-HB}^{i0-jk'}[10^{-3}k_B]$	4.6	1.3	4.1	0.9	4.8	1.3	5.1	1.1
$\Delta S_{HB-HB}^{i0-jk'}/S_{HB-HB}^{i0-jk;B}[\%]$	1264	357	1126	247	1319	357	1401	302
$\Delta S_{HB-HB}^{i0-jk'*}[10^{-3}k_B]$	2.9	1.1	2.5	0.6	4.1	1.4	3.9	1.5
$\Delta S_{HB-HB}^{i0-jk'*}/S_{HB-HB}^{i0-jk*;B}[\%]$	131	50	112	27	185	63	176	68
$\Delta S_{nHB-nHB}^{i0-jk'}[10^{-2}k_B]$	16.0	2.1	15.2	1.9	17.6	2.2	16.8	1.9
$\Delta S_{nHB-nHB}^{i0-jk'}/S_{nHB-nHB}^{i0-jk;B}[\%]$	644	85	611	76	748	89	676	89
$\Delta S_{nHB-nHB}^{i0-jk'*}[10^{-2}k_B]$	3.2	0.8	2.8	0.5	3.6	0.5	4.0	0.3
$\Delta S_{nHB-nHB}^{i0-jk'*}/S_{nHB-nHB}^{i0-jk*;B}[\%]$	1824	456	1596	285	2052	286	2280	171
$\Delta S_{HB-nHB}^{i0-jk'}[10^{-2}k_B]$	1.9	0.4	1.5	0.3	1.8	0.3	2.1	0.4
$\Delta S_{HB-nHB}^{i0-jk'}/S_{HB-nHB}^{i0-jk;B}[\%]$	3940	830	3110	622	3732	622	4354	833
$\Delta S_{HB-nHB}^{i0-jk'*}[10^{-2}k_B]$	1.2	0.3	1.0	0.2	1.6	0.3	1.5	0.3
$\Delta S_{HB-nHB}^{i0-jk'*}/S_{HB-nHB}^{i0-jk*;B}[\%]$	370	92	308	62	493	62	462	93