Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Supplementary Information: The structure of liquid water beyond the first hydration shells

Andres Henao,[a,b] Sebastian Busch,[c] Elvira Guàrdia,[b] Josep Lluis Tamarit,[a] and Luis Carlos Pardo*[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

1 Summary of the used Information Theory approach

In order to be able to evaluate the wealth of information contained in the trajectories of the molecules in a Molecular Dynamics simulation one can turn to information theory. It provides a framework to extract useful information about the system by looking at the different contributions to its excess entropy. The excess entropy of any liquid S_{excess}, i. e. the difference to the entropy of an ideal gas at the same average density and temperature, can be calculated from the six-dimensional correlation function $g(r,\Omega)$ as a function of the distance r and a group of five angles defining the relative position and orientation Ω. When restricted to a given length scale (a determined distance r between particles) $g(r,\Omega)$ depends only on the five angular variables, and the correlation function is then written as $g(\Omega|r)$. A good approximation for the calculation of the excess entropy is to consider only the term related to two-particle correlations, in which case the excess entropy can be separated into two terms[1]

$$S_{\text{excess}} = S_{\text{trans}}^{\text{tot}} + S_{\text{ang}}^{\text{tot}} \tag{1}$$

where $S_{\text{trans}}^{\text{tot}}$ is the total entropy connected to the particle density at a given length scale, i. e. features in $g_{\text{OO}}(r)$, and $S_{\text{ang}}^{\text{tot}}$ contains all contributions that are a function of Ω. These two contributions can be analyzed as functions of the distance r, defined as

$$S_{\text{trans}}^{\text{tot}} = \frac{1}{2} k\rho \int S_{\text{trans}}(r)\, \mathrm{d}r \tag{2}$$

$$S_{\text{trans}}(r) = -\left[g_{\text{OO}}(r)\ln g_{\text{OO}}(r) - g_{\text{OO}}(r) + 1\right] 4\pi r^2 \tag{3}$$

with the Boltzmann factor k and number density ρ, and

$$S_{\text{ang}}^{\text{tot}} = \frac{1}{2} k\rho \int 4\pi r^2 g_{\text{OO}}(r) S_{\text{ang}}(r)\, \mathrm{d}r \tag{4}$$

$$S_{\text{ang}}(r) = -\frac{1}{\Omega} \int g(\Omega|r) \ln g(\Omega|r)\, \mathrm{d}\Omega \tag{5}$$

$$= S_{\text{pos}}(r) + S_{\text{ori}}(r) + S_{\text{pos}\star\text{ori}}(r) \tag{6}$$

where $g(\Omega|r)$ denotes the five-dimensional Ω distribution function of molecules at a given value of the distance r. As shown in the equation 6, it is possible to write this quantity as a sum of three parts, S_{pos} stemming from the relative position of two molecules $(\theta_{\text{pos}}, \phi_{\text{pos}})$, S_{ori} stemming from their relative orientation $(\theta_{\text{ori}}, \phi_{\text{ori}}, \psi_{\text{ori}})$, and the cross-term $S_{\text{pos}\star\text{ori}}$.

One can perform an expansion of the correlation function, so that the entropy $S_{\text{ang}}(r)$ can be calculated as the entropy $S^{(1)}(r)$ of one-angle correlation functions $g(\alpha|r)$, corrected by the contribution of correlation functions of increasing order; $S^{(2)}(r)$ from $g(\alpha,\beta|r)$; $S^{(3)}(r)$ from $g(\alpha,\beta,\gamma|r)$ and so on. With this method, the entropy $S^{(5)}(r)$ of the five-dimensional correlation function appearing in equation 5 can be calculated as the summation of the

---

[a] Grup de Caracterització de Materials, Departament de Física, ETSEIB, Universitat Politècnica de Catalunya, Diagonal 647, E-08028 Barcelona, Catalonia, Spain; E-mail: luis.carlos.pardo@upc.edu
[b] Grup de Simulació per Ordinador en Matèria Condensada, Departament de Física, B4-B5 Campus Nord, Universitat Politècnica de Catalunya, E-08034 Barcelona, Catalonia, Spain
[c] German Engineering Materials Science Centre (GEMS) at Heinz Maier-Leibnitz Zentrum (MLZ), Helmholtz-Zentrum Geesthacht GmbH, Lichtenbergstr. 1, 85747 Garching bei München, Germany

excess entropies related to each angle, corrected by higher-order mutual information terms $I^{2\text{–}4}$ which can be written as (neglecting fourth-order terms):

$$S^{(5)}(r) = \sum S^{(1)}(r) - \sum I^{(2)}(r) + \sum I^{(3)}(r) + \mathscr{O}^{(4)} \tag{7}$$

where

$$I^{(2)}(r) = \sum S^{(1)}(r) - S^{(2)}(r) \tag{8}$$

$$I^{(3)}(r) = \sum S^{(1)}(r) - \sum S^{(2)}(r) + S^{(3)}(r) \tag{9}$$

so that for example the mutual information of two variables A and B, $I(A,B)$ of type $I^{(2)}$, is calculated as $I(A,B) = S(A) + S(B) - S(A,B)$.

This expansion is firmly based on information theory concepts[2] and it is a most powerful method to study the ordering of molecules in disordered phases.[4] Mutual information is a measure of the interdependency of N variables: high mutual information values in a two-fold correlation function imply that the two variables are strongly correlated, and therefore that a factorization is not possible ($g(\alpha,\beta) \neq g(\alpha)g(\beta)$). On the contrary a low value of mutual information implies that the variables are independent, allowing a factorization of the correlation function. If the excess entropy is expanded up to the third order it is possible to extract valuable physical information from the expansion of equation (7).

The variables determining the position of two neighboring water molecules, at a certain distance, in spherical coordinates (θ_{pos} and ϕ_{pos}) can be separated from the Euler angles related to their relative orientation (θ_{ori}, ϕ_{ori} and ψ_{ori}) (for a definition of the angles see figure 1 in the main text). Separating the contributions due to the relative position and relative orientation of two molecules one obtains, following equation (8), for the position:

$$S_{\text{pos}}(r) := S(\theta_{\text{pos}}|r) + S(\phi_{\text{pos}}|r) - I(\theta_{\text{pos}},\phi_{\text{pos}}|r) \tag{10}$$

and for the relative orientation, using equation (9)

$$\begin{aligned} S_{\text{ori}}(r) := {} & S(\theta_{\text{ori}}|r) + S(\phi_{\text{ori}}|r) + S(\psi_{\text{ori}}|r) \\ & - I(\theta_{\text{ori}},\phi_{\text{ori}}|r) - I(\theta_{\text{ori}},\psi_{\text{ori}}|r) - I(\phi_{\text{ori}},\psi_{\text{ori}}|r) \\ & + I(\theta_{\text{ori}},\phi_{\text{ori}},\psi_{\text{ori}}|r) \end{aligned} \tag{11}$$

and finally the cross-term between position and orientation in equation (6) can be written as:

$$\begin{aligned} S_{\text{pos}\star\text{ori}}(r) := {} & S_{\text{ang}}(r) - S_{\text{pos}}(r) - S_{\text{ori}}(r) \\ = {} & - \sum_{\substack{\omega_{\text{pos}} \in \theta_{\text{pos}},\phi_{\text{pos}} \\ \omega_{\text{ori}} \in \theta_{\text{ori}},\phi_{\text{ori}},\psi_{\text{ori}}}} I(\omega_{\text{pos}},\omega_{\text{ori}}) \\ & + \sum_{\substack{\omega_{\text{pos}} \in \theta_{\text{pos}},\phi_{\text{pos}} \\ \omega'_{\text{pos}} \in \theta_{\text{pos}},\phi_{\text{pos}} \\ \omega_{\text{ori}} \in \theta_{\text{ori}},\phi_{\text{ori}},\psi_{\text{ori}}}} I(\omega_{\text{pos}},\omega'_{\text{pos}},\omega_{\text{ori}}) \\ & + \sum_{\substack{\omega_{\text{pos}} \in \theta_{\text{pos}},\phi_{\text{pos}} \\ \omega_{\text{ori}} \in \theta_{\text{ori}},\phi_{\text{ori}},\psi_{\text{ori}} \\ \omega'_{\text{ori}} \in \theta_{\text{ori}},\phi_{\text{ori}},\psi_{\text{ori}}}} I(\omega_{\text{pos}},\omega_{\text{ori}},\omega'_{\text{ori}}) \\ & + \mathscr{O}^{(4)} \end{aligned} \tag{12}$$

It must be highlighted that correlation functions involving an angle between vectors α such as θ_{pos} and θ_{ori} have not been calculated using a uniform binning in angle but the cosine of the angle instead. This is necessary in order to have a featureless correlation function when two vectors are randomly distributed. If the angle were used to perform the binning, the correlation function between two vectors would have a maximum at $\alpha = 90^\circ$ and would be zero at $\alpha = 0^\circ$ and $\alpha = 180^\circ$. This would give rise to non-vanishing mutual information terms that would alter the values of the total excess entropy.

In order to perform the calculations, a coordinate system was attached to each molecule with the z axis along the dipole moment and x perpendicular to the HOH plane. θ_{ori} is therefore related to the relative orientation of two water dipoles, and ϕ_{ori} and ψ_{ori} to the rotation around this axis.[5] The analysis was carried out with the program ANGULA,[5] choosing an angular increment of 20° for angles between planes and an increment for the cosine of angles between vectors of 0.2. Excess entropies, and therefore also mutual information terms, have been rescaled so that they vanish for long distances.[6] The distance increment to produce each map has been chosen to be Δr=0.1 Å. The excess angular entropy obtained with these conditions is 10.34 cal/K/mol (86% of the total excess angular entropy) in good agreement with literature values.[3] Spatial Density Maps (SDM) were plotted using VMD.[7]

Figure 1 shows the integrated second order mutual information contributions to the total entropy, in other words: correlations between different pairs of angles. They have been calculated similarly to the total entropy in equation (2), i. e.

$$I^{(i)\text{tot}} = -\frac{1}{2}k\rho \int I^{(i)}(r)\,\mathrm{d}r \quad . \tag{13}$$

The strongest correlation between two angles is not surprisingly

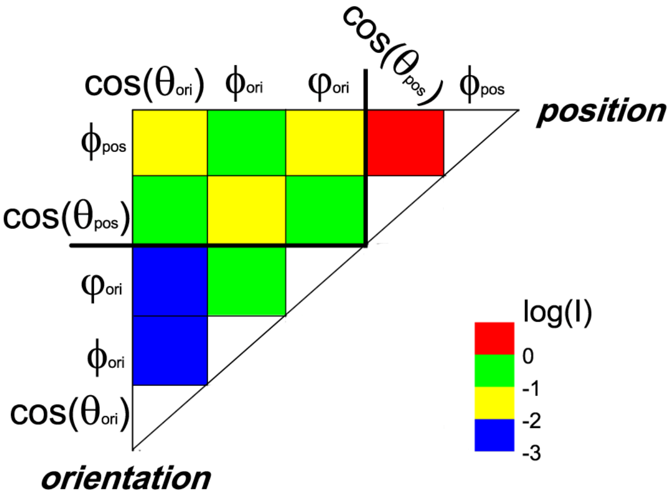

Fig. 1 (color online). Logarithm of the integrated mutual information values calculated for all possible combinations of angles or their cosine (cf. text). The angles have been grouped separating those describing the relative position of two molecules (upper right) from the orientation (lower left).

the one relating the positional angles θ_{pos} and ϕ_{pos}. Concerning the purely orientational contributions one can see that the correlation between the variables ϕ_{ori} and ψ_{ori} is stronger than the correlations between any of these angles and the one defining the dipole orientation θ_{ori}. The correlation function of the relative orientation of two molecules can therefore be naturally factorized into one term concerning the dipole orientation and another term defining the rotation around the dipolar axes: $g(\theta_{\text{ori}}, \phi_{\text{ori}}, \psi_{\text{ori}}) \sim g(\theta_{\text{ori}}) \cdot g(\phi_{\text{ori}}, \psi_{\text{ori}})$.

2 Molecular position: excess entropy, spatial density maps, and correlation with $g_{\text{OO}}(r)$

This section contains a detailed description of the positional ordering of water as a function of the distance to a central molecule. It will also be shown that changes in excess positional entropy $S_{\text{pos}}(r)$ are able to quantify where changes of structural motifs happen. In order to do so, figure 2 shows in the upper panel (a) the partial radial distribution function (PRDF) associated to water oxygens $g_{\text{OO}}(r)$ and in the lower panel (b) the positional contribution to the excess entropy $S_{\text{pos}}(r)$. It can be seen that $S_{\text{pos}}(r)$ consists of a series of more or less defined minima (regions of increased order) which are associated with structural motifs. $S_{\text{pos}}(r)$ is divided in this figure into regions delimited by maxima of $S_{\text{pos}}(r)$, except in the case of regions 2 and 3 which are not separated by a clear maximum – the reason for this is discussed below.

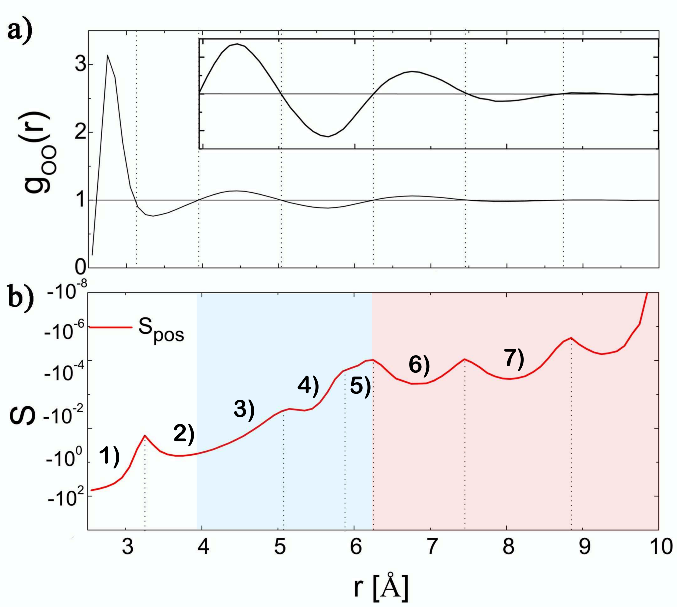

Fig. 2 (color online). Panel a) The oxygen-oxygen radial distribution function, with a zoom in the inset to show the oscillations more clearly. Panel b) Positional contribution $S_{\text{pos}}(r)$. Dotted lines show the maxima of the positional entropy and the number labels correspond to regions for which the spatial density maps are shown in figure 4.

As expected, the first peak in the Oxygen-Oxygen partial radial distribution function contains the first four molecules forming the well-known tetrahedral structure. The minimum of $g_{\text{OO}}(r)$ and the maximum of $S_{\text{pos}}(r)$ are located at the same distance of about 3.3 Å. Between regions 2 and 3 there is a node in $g_{\text{OO}}(r)$ but no maximum in the function $S_{\text{pos}}(r)$. Between regions 3 and 4 at about 5 Å, there is a node of $g_{\text{OO}}(r)$ coinciding with a maximum in the positional excess entropy. Between regions 4 and 5, there is no node of $g_{\text{OO}}(r)$ but a feature in $S_{\text{pos}}(r)$. Between regions 5 and 6, finally, there is a node of $g_{\text{OO}}(r)$ coinciding with a maximum in $S_{\text{pos}}(r)$. It can be concluded that for distances smaller than 6 Å, there is no univocal correlation between the positional entropy and the Oxygen-Oxygen PRDF, i. e. although the nodes in the PRDF are always related to a change in the structure, a node in the PRDF does not imply a change in $S_{\text{pos}}(r)$ and also the absence of a node in PRDF does not imply that there is no change in $S_{\text{pos}}(r)$.

The two following figures will make the meaning of figure 2 more accessible: First, figure 3 shows the two dimensional correlation function $g\left(\theta_{\text{pos}}, \phi_{\text{pos}}|r\right)$ (which will be called *maps* in the following) of the oxygen atoms for distance slices of $\Delta r = 0.1$ Å. The regions 1 to 7 introduced in figure 2 stretch over several maps each and are clearly marked. Second, figure 4 contains the corresponding 3D representations of the same function; the maps

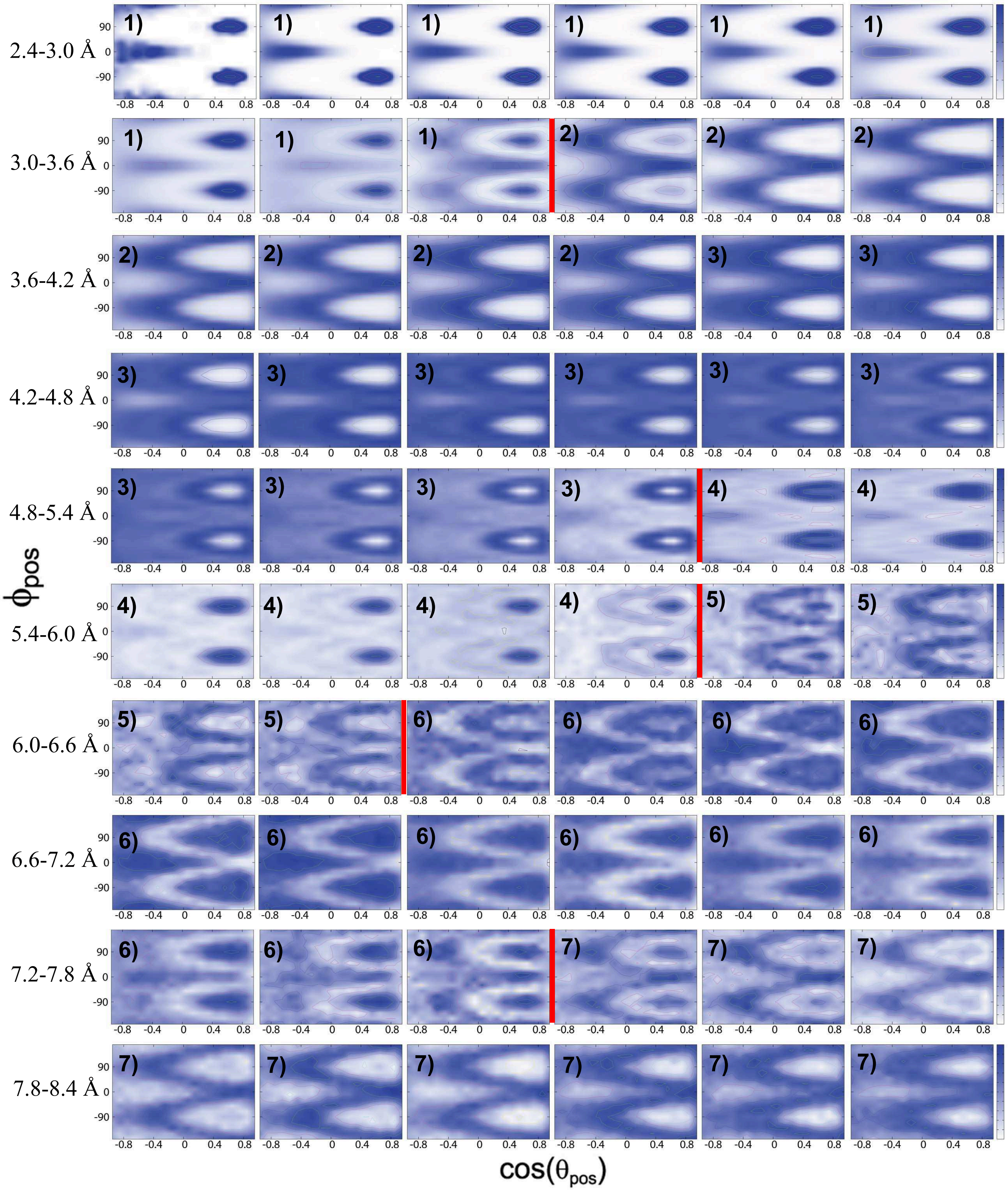

Fig. 3 (color online). Distribution functions $g\left(\theta_{\text{pos}},\phi_{\text{pos}}|r\right)$ for the oxygen atom in liquid water. 60 maps have been generated, each for a slice of 0.1 Å thickness between distances of 2.4 Å and 8.4 Å. The first map corresponds to a distance range of 2.4-2.5 Å, the second map to 2.5-2.6 Å, and so on up to a range of 8.3-8.4 Å in the 60th map. Blue colors represent probability maxima while white colours represent probability minima. Vertical red lines correspond to the limits of the regions described in figure 2.

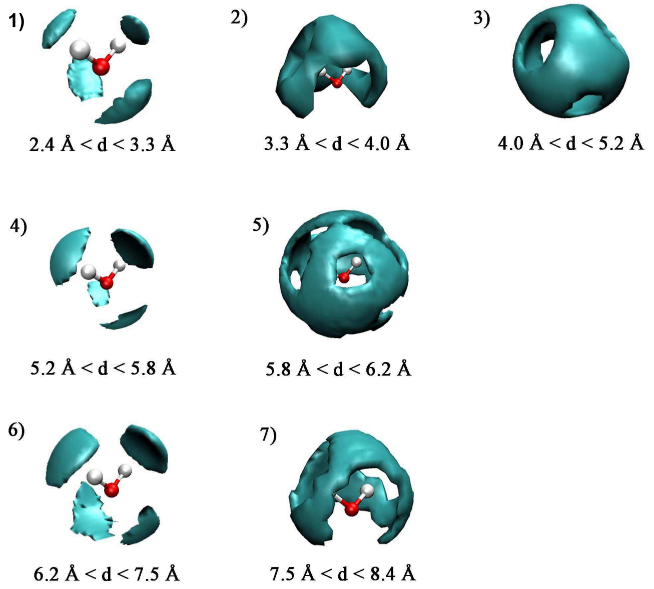

Fig. 4 (color online). Spatial density maps generated for the regions of the positional entropy $S_{\mathrm{pos}}(r)$ in figure 2.

of each region are taken together and the $\theta_{\mathrm{pos}}, \phi_{\mathrm{pos}}$ distribution is shown. It should be noted that these two figures are merely different ways of plotting the same information – the directions of high probability to find another water molecule in a given distance range from a central one. In the following, these seven regions are discussed one by one.

Regime 1 shows the first hydration shell, i. e. the first four neighbours at a distance of up to 3.3 Å (the first minimum of $g_{\mathrm{OO}}(r)$). The tetrahedral arrangement of the molecules is very obvious and well-known.

We know from the discussion above that regions 2 and 3 are not delimited by any clear maximum in $S_{\mathrm{pos}}(r)$. As it can be seen in figure 3 this is explained by the fact that is no drastic change in molecular position between these two regions, even though the Oxygen-Oxygen PRDF crosses a node. Figure 4 shows that molecules are located more or less in the same places for these two regions. Moreover, as expected, molecules are occupying the holes left by the tetrahedral structure of the first four neighbours.

The situation in regions 4 and 5 is the opposite of regions 2 and 3: There is no node in $g_{\mathrm{OO}}(r)$, but there is a change in the positional entropy: it has a small maximum in $S_{\mathrm{pos}}(r)$ delimiting two small minima: this is indicative of a change in the structure that can be seen in figure 3. In that figure, one can see that there is a fairly well defined structure in region 4 which resembles the one of the tetrahedral arrangement of the first four molecules, region 5 constitutes a less defined structure where water molecules are occupying the gaps left by the previous region.

For regions 6 and 7 (for $r > 6$ Å), contrary to what happened in the previous regions, the maxima of $S_{\mathrm{pos}}(r)$ seem to be ruled by the nodes in $g_{\mathrm{OO}}(r)$. Moreover the structural motifs giving rise to the minima of $S_{\mathrm{pos}}(r)$ in regions 6 and 7 seem to be associated with well defined structures: the high density region 6 with a tetrahedral symmetry, and the low density region 7 with molecules occupying the interstices of the tetrahedral motif. Moreover both SDMs resemble the ones of regions 1 and 2 defined above. This alternance of structures mainly determined by the location of the first molecules has been seen before for other substances.[8–10]

Summarizing, although changes in the structure seem to happen always when the Oxygen-Oxygen PRDF has a node, there is only a biunivocal correlation between density and the change of structural motifs, brought out by maxima in $S_{\mathrm{pos}}(r)$ at distances over 6 Å. For shorter distances, there is an alternance that holds for unexpected long distances.

References

1 T. Lazaridis and M. Karplus, *The Journal of Chemical Physics*, 1996, **105**, 4294–4316.

2 H. Matsuda, *Phys. Rev. E.*, 2000, **62**, 3096–3102.

3 D. J. Huggins, *J. Chem. Phys.*, 2012, **136**, 064518.

4 L. C. Pardo, A. Henao and A. Vispa, *J. Non-Cryst.*, 2015, **407**, 220–227.

5 L. C. Pardo, A. Henao, S. Busch, E. Guàrdia and J. L. Tamarit, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24479–24483.

6 R. S. *et al.*, *Bioinformatics*, 2002, **18**, (suppl 2): S231–S240.

7 W. Humphrey, A. Dalke and K. Schulten, *J. Molec. Graphics*, 1996, **14**, 33–38.

8 N. Veglio, F. Bermejo, L. Pardo, J. L. Tamarit and G. Cuello, *Physical Review E*, 2005, **72**, 031502.

9 L. Pardo, N. Veglio, F. Bermejo, J. L. Tamarit and G. Cuello, *Physical Review B*, 2005, **72**, 014206.

10 M. Rovira-Esteva, A. Murugan, L. Pardo, S. Busch, M. Ruiz-Martin, M.-S. Appavou, J. L. Tamarit, C. Smuda, T. Unruh, F. Bermejo *et al.*, *Physical Review B*, 2010, **81**, 092202.