Hydrophobic hydration at the level of primitive models. II: Large solutes and water restructuring

MILAN PŘEDOTA¹, IVO NEZBEDA¹* and PETER T. CUMMINGS¹,²,³

¹ Department of Chemical Engineering, and ² Departments of Chemistry and Computer Science, University of Tennessee, Knoxville, TN 37996-2200, USA
³ Chemical Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6110, USA

(Received 24 September 2001; revised version accepted 11 January 2002)

Details of structural changes that take place in water near an apolar solute have been studied by Monte Carlo simulations for hard sphere solutes of increasing size, including the limiting case of water at a hard structureless wall. Water has been modelled by two different types of extended primitive model, the four-site EPM4 model and five-site EPM5 model. Two different patterns of the orientational ordering of the water molecules around the solute as a function of its size have been found. For the EPM5 model, the structure of water and the orientation of its molecules near an apolar solute of finite diameter do not seem to be sensitive to the size of the solute, and only become more pronounced when the solute becomes a hard wall. On the other hand, the orientation ordering of the EPM4 molecules gradually changes with increasing size of the solute, and for solutes larger than approximately five times the size of the water molecule it is opposite to that near a small solute. A novel method to evaluate the excess chemical potential of large solutes has been implemented, and some thermodynamic quantities for water (distribution of hydrogen bonds and the excess chemical potential) have been computed as a function of the distance from the solute.

1. Introduction

Hydrophobic hydration (HH) is the term used at the molecular level to express structural changes that take place in bulk water when a single non-polar molecule is transferred into it [1]. Because of its importance in both science and applications, this process has been the subject of continuous research using various methods [2, 3]. The modern approach to studying HH begins with well defined realistic intermolecular potentials, and uses computer simulations with the aim of determining the reorganization of the solvent molecules around the hydrophobic solutes [4, 5]. Such simulations may provide invaluable information and make it possible to study even very fine structural details. Nonetheless, some properties (e.g. those given by higher order derivatives) remain inaccessible, and severe technical problems may be encountered when simulating systems with large solutes. From a variety of possible structural information, pair correlation functions, i.e. the spatial arrangement of water molecules around the solute, typically are evaluated, which, however, need not be sufficient for drawing conclusions about the overall behaviour of the solutions studied. Finally, establishing a direct link between the observed effects and individual physically meaningful intermolecular interactions (e.g. hydrogen bonding or dipole–dipole interaction) may be problematic for complex realistic models. In order to avoid at least some of these problems, simplified statistical mechanical models are convenient for contributing to better understanding and elucidation of hydrophobic phenomena at the molecular level.

In [6] we reported a study of HH at a very basic elementary level, considering an infinitely dilute solution of hard spheres of small and moderately large sizes in two qualitatively different simple models of water, the primitive models EPM4 and EPM5 [7]. Despite the simplicity of this model mixture, mimicking an infinitely diluted aqueous solution of apolar solutes, all the findings on the details of the orientational ordering of water molecules in the first hydration shell as well as those for the site–site correlation functions were in full agreement with the results obtained using much more complex models [4, 5]: configurations with one hydrogen bond pointing towards the bulk water and three bonds straddling the surface of the solute are preferred. In the preferred configurations, the ability of water molecules to create hydrogen bonds is only weakly restricted in comparison with bulk water.

* Author for correspondence e-mail: ivonez@icpf.cas.cz
Hard sphere solutes considered in [6] covered the range of diameters up to three times that of the water molecule. In addition to the spatial arrangement of water molecules we also determined their orientation in the first hydration shell, and found that it remained practically unchanged when the hard sphere diameter was increased, other than making orientation correlations more pronounced. However, in the case of the EPM4 model and the largest solute considered, the behaviour observed did not fit completely into this established pattern. Moreover, there are indications that water solvates large and small solutes differently [8], and that very large solutes may even induce additional effects (see, e.g. [9]). All these findings, along with experimental evidence that solubilities of apolar solutes in water may depend on their surface area [10, 11], have prompted us to focus in detail on solutions with very large solutes, including the limiting case of infinite diameter. The latter case corresponds to an inhomogeneous system, namely water at a hard planar structureless wall, which is also of interest by itself.

The purpose of this paper is to study in detail the HH for large solutes and to detect and characterize gradual changes (if there are any) in both the spatial and orientational arrangement of water molecules when the solute increases until it becomes infinitely large to form a hard structureless wall. We use again primitive water for the solvent and additional unknowns are (i) whether the primitive models maintain their capability to reproduce faithfully the structural properties of real water even in these cases, and (ii) the extent to which the orientational ordering of water molecules is model dependent. As in [6] we use Monte Carlo (MC) simulations and the subjects of the study are: (i) the spatial distribution of water molecules near the solute and the wall, i.e. the oxygen–solute correlation function and the density profile of water oxygens, respectively; (ii) the structure of the hydration shell including a preferred orientational ordering of water molecules; and (iii) selected thermodynamic properties. In this paper we focus primarily in detail on the first two items. Directly related thermodynamic properties (e.g. hydrogen bonding) are also discussed and simulation results for the chemical potential are given, but a full discussion of thermodynamics associated with the observed structural changes and its comparison with experimental data is planned for a subsequent paper [12].

2. Theoretical background and computational details

All definitions and details of the simulations were given at length in [6]. Therefore they are only briefly summarized here.

The two primitive models used, EPM4 [13] and EPM5 [14], are derived from the realistic TIP4P [15], and ST2 [16] and TIP5P [17] potential models, respectively. They are made up of a hard sphere of diameter \( \sigma_0 \) with embedded sites of two kinds (see figure 1) and incorporate two types of short range interaction: (i) repulsions between the hard cores and between like charges, and (ii) attractions corresponding to the interaction between unlike charges. These two types of interaction are considered in their simplest forms, repulsions as a hard sphere interaction,

\[
 u_{\text{hs}}(r_{12}; \sigma) = \begin{cases} 
 +\infty & \text{for } r_{12} < \sigma, \\
 0 & \text{for } r_{12} > \sigma, 
\end{cases}
\]

and attractions as a square-well interaction,

\[
 u_{\text{sw}}(r_{12}; \lambda) = -\varepsilon_{\text{HB}} & \text{for } r_{12} < \lambda, \\
 0 & \text{for } r_{12} > \lambda. 
\]

The pair potential of the EPM models assume the form

\[
 u_{\text{EPM}}(1, 2) = u_{\text{hs}}(R_{\text{OO}}; \sigma_0) + \sum_{ij} u_{\text{hs}}(|r_i^{(1)} - r_j^{(2)}|; \sigma_R) + \sum_{ij} u_{\text{sw}}(|r_i^{(1)} - r_j^{(2)}|; \lambda),
\]

where the summation in the first sum runs over all pairs of like sites, and the summation in the other sum runs over all pairs of unlike sites. \( R_{\text{OO}} \) denotes the separation between the oxygen sites on two molecules, \( r_i^{(k)} \) is the position vector of site \( i \) on molecule \( k \), \( \varepsilon_{\text{HB}} \) measures the strength of the attractive interaction and \( \lambda \) defines its range, and \( \sigma_R \) is the range of the sort range repulsive interaction between like sites. The actual arrangement of the off-centre sites of the models considered, denoted as H and M sites, is shown in figure 1. The central oxygen site (O site) accounts for the hard core repulsion. Throughout the paper we use the units such that \( \sigma_0 = \varepsilon_{\text{HB}}/k_B \equiv 1 \), where \( k_B \) is the Boltzmann constant.

We simulated solutions consisting of \( N \) water molecules (solvent) and one hard sphere of diameter \( d \) (solute) which interacts with the solvent molecules as a hard sphere with respect to the oxygen sites only, i.e. the contact distance of the solute and water molecules is \( r_c = (d + 1)/2 \). (We recall that for small solutes we simu-
lated in [6] pure water and used the virtual insertion method to compute solute–oxygen properties. In the limit of \( d \rightarrow \infty \), the hard sphere solute becomes a hard structureless wall. To mimic this limiting arrangement in simulations, the water molecules are restricted to a slab formed by two hard walls whose width \( L \) is sufficiently large to allow a homogeneous bulk phase of density \( \rho_{\text{bulk}} \) to establish in the central portion. To maintain consistency with spherical solutes of a finite diameter, the walls interact with the oxygen sites only, and the periodic boundary conditions are applied in the \( x \) and \( y \) directions.

The distribution of the oxygen sites of water around the solute is given by the conventional solute–oxygen correlation function \( g_{SO} \), defined as

\[
\rho_w g_{SO}(r) = \lim_{\Delta r \to 0} \frac{\langle N_O(r, r + \Delta r) \rangle_N}{4\pi r^2 \Delta r},
\]

where \( \rho_w = N/V \) is the number density of water, and \( \langle N_O(r, r + \Delta r) \rangle \) is the ensemble average of the number of oxygen sites in the spherical shell of thickness \( \Delta r \) centred on the solute. In the case of an inhomogeneous system a related function is the one-particle density profile of oxygen sites,

\[
\rho(z) = \lim_{\Delta z \to 0} \frac{\langle N_O(z, z + \Delta z) \rangle_N}{\Delta z^2 \xi_{xy}},
\]

where \( N_O \) denotes the number of oxygen sites in a planar slab of width \( \Delta z \) at distance \( z \) from the wall, and \( \xi_{xy} \) is the area of the basic cell in the \( x \) and \( y \) directions. Instead of the one-particle density profile we use a dimensionless one-particle correlation function

\[
g_w(z) = \rho(z)/\rho_{\text{bulk}}.
\]

This function can be compared directly with the solute–oxygen correlation function \( g_{SO} \), if the solute–oxygen separation is measured from the contact separation. Furthermore, the contact value of \( g_w \) is directly related to the properties of the bulk fluid through the contact theorem [18]:

\[
g^* = g_w(z_c) = \beta P/\rho_{\text{bulk}},
\]

where \( z_c \) denotes the distance of the water molecule from the wall at contact, and \( \beta = 1/k_B T \).

Equation (4) is formally correct but does not approximate properly the required correlation function at infinite dilution. Provided that the simulation cell is large enough, far from the solute the solvent must be homogeneous with density \( \rho_{\text{bulk}} \), which must be equal to the density of pure solvent at the given pressure \( P \). In order to guarantee the correct asymptotic behaviour of \( g_{SO} \) at infinite solution, we use throughout the paper \( g_{SO} \) defined by

\[
\rho_{\text{bulk}} g_{SO}(r) = \lim_{\Delta r \to 0} \frac{\langle N_O(r, r + \Delta r) \rangle_N}{4\pi r^2 \Delta r},
\]

which satisfies the condition \( g_{SO}(r) \to 1 \) for large \( r \).

Another quantity of interest is the coordination number \( N_c \), given by

\[
N_c = 4\pi \rho_{\text{bulk}} \int_{r}^{r_{\text{min}}} r^2 g_{SO}(r) \, dr,
\]

where the integration range goes up to the first minimum of \( g_{SO} \). For water between the walls the appropriate definition of \( N_c \) reads

\[
N_c = L_{\text{xy}} \int_{r}^{r_{\text{min}}} \rho(z) \, dz.
\]

A quantity related to \( N_c \) is the occupancy number, \( N_o \), providing the number of molecules per unit volume,

\[
N_o = \frac{N_c}{\xi_{\text{min}} L_{\text{xy}}^3},
\]

or, for the hard wall system,

\[
N_o = \frac{N_c}{\xi_{\text{min}} L_{\text{xy}}^2}.
\]

As in [6], we computed the orientational ordering of water molecules in the first hydration shell using angles \( \alpha \) and \( \gamma \) formed, respectively, by the OH vector of the water molecule or by the axis of the water molecule, and by the oxygen–solute radius vector. If the water molecules bore appropriate charges, direction \( \gamma \) would define the orientation of the dipole moment. In the case of water between the walls, the first hydration shell is defined as the slab stretching from the wall to the first minimum of \( g_w \), and angles \( \alpha \) and \( \gamma \) are measured with respect to the inward pointing normal of the wall (i.e. in the direction from the solvent towards the wall).

Selected thermodynamic properties (energy per molecule, distribution of hydrogen bonds, excess chemical potential) were computed both for the entire system and as a function of the solute–oxygen distance in co-spherical shells (slabs in the case of inhomogeneous systems). For the excess chemical potential of water we used the Widom’s particle insertion method [19],

\[
\beta \mu_{\text{ex}} = -\ln \left< \exp \left[ -\beta \Delta U \right] \right>,
\]

where \( \Delta U \) is the change in energy due to the insertion of a water molecule. The direct insertion was used also in [6] for small solutes, but for large solutes the method is not applicable. For evaluation of the excess chemical potential of the solute, \( \mu_{\text{ex}}^s \), we therefore implemented a novel method based on an exact relationship between the contact value of the solute–oxygen correlation function and a derivative of the chemical potential [20]; the method is explained in detail in §3.3.
In order to compare solutions with different solutes and assess the effect of the solute on the properties of water, a comparison must be made at the same pressure (and temperature). Nonetheless, in [6] all simulations were performed in an NVT ensemble of the same size, regardless of the solute diameter. This choice followed from the virtual insertion method used to compute properties for solutions with solutes up to \( d \leq 1.6 \). Because of consistency, the volume was kept constant even in simulations for \( d = 2 \) and \( 3 \) when the solute was actually present in the system. This was made possible by the fact that the overall pressure is affected only marginally by the presence of the solute for, approximately, \( d \leq 3 \) and large \( N \). However, as the size of the hard sphere solute is further increased, the effect of the solute upon pressure is no longer negligible. In order to avoid direct NPT simulations characterized by slow convergence, we used in this study again an NVT ensemble set up so as to yield the required pressure. This strategy required, typically, two auxiliary simulations to determine the appropriate volume \( V \) of the simulation box. Pressure \( P \) was evaluated from the standard relationship [21] using the contact values of the site–site correlation functions \( g_{ij} \), and the projections of the unit site–site vector, \( v_{ij} \), on the oxygen–oxygen vector, \( R_{OO} \):

\[
\beta P = \frac{N+1}{V} + \frac{2\pi}{3}N \frac{N-1}{p^2} (g_{OO}(\sigma_{OO})^2 \\
+ (n_{HH} + n_{MM})g_{HH}(\sigma_{HH})^2(\langle v_{RO} \rangle_{\sigma_{HH}}^2 - 1) \\
+ n_{HM}g_{HM}(\sigma_{HM})^2(\langle v_{RO} \rangle_{\sigma_{HM}} - 1)) \\
+ \frac{4\pi}{3} \rho_{\text{bulk}} g_{SO}(\sigma_{SO})^2,
\]

(13)

where \( \sigma_{OO} \equiv \sigma_O \), \( \sigma_{HM} = \lambda \), \( \sigma_{HH} = \sigma \), and \( n_{ij} \) is the number of interacting \( i-j \) pairs between two molecules (for EPM4 \( n_{MM} = 0 \) since there is no interaction between these sites).

All simulations were carried out with \( N = 1024 \), and preferential sampling was used to sample the surrounding of the solute efficiently. The ratio of accepted moves per single molecule was adjusted to about one third. The total number of accepted moves in the final runs was about \( 3 \times 10^9 \), which followed after at least 30 million equilibration steps. In the case of water between the walls, the true bulk phase was well developed in the central portion of the slab, separating two symmetric regions perturbed by the walls. Although it was rather time consuming to reach symmetric distribution at the walls, no pathological behaviour [18, 22] was observed. The results reported for the hard wall systems represent an average over both regions at each wall.

3. Results

In [6] we studied the hydrophobic effects for different solutes for a series of temperatures \( T \). The latter dependence does not seem to be so important, and in this paper we confine our considerations to only one typical temperature for each model: \( \beta = 6 \) for the EPM4 solvent, and \( \beta = 5 \) for the EPM5 solvent.

We considered hard sphere solutes of diameters \( d = 4 \) and \( 5 \), and carefully resimulated also the solutions with \( d = 1, 2, \) and \( 3 \). The limit of an infinitely large solute, \( d = \infty \), was represented by a pure water system between two parallel walls. All calculations reported in this paper were performed in an NVT ensemble with the appropriately adjusted partial packing fraction of the solvent, \( \eta_W = (\pi/6)\rho_W \), so as to yield, within the estimated errors, the same pressure for all systems (table 1). For the inhomogeneous system, the total volume, i.e. \( V = L_z^2 L_x^2 \), was used to define \( \eta_W \); \( L_z = 12.53 \) for the EPM4 solvent, and \( L_z = 13.14 \) for the EPM5 solvent.

3.1. Structure of the solvation shell

The solute–oxygen correlation functions for the EPM4 and EPM5 solvents and a number of solute diameters are shown in figure 2. In both cases the trend of \( g_{SO} \) at contact to go to higher values with increasing \( d \) is observed. Also, the locations of extrema of the correlation functions are practically independent of the solute diameter, \( r_{min} - r_c \approx 0.7 \) (we recall that the site–site separation is measured from the contact distance). All these observations are in full agreement with those found for small and medium diameters in [6]. Nonetheless, a qualitative difference in the behaviour of the two solutions can also be noted. Whereas the EPM4 solvent exhibits quite a flat profile and therefore no layering either around the solute or at the wall, the EPM5 solvent seems somewhat structured: the second peak of \( g_{SO} \) is well pronounced and also is shifted to slightly larger separations.

The solute–oxygen correlation function determines the coordination number of water molecules around the solute, and hence the occupancy number, equation (10). These numbers are given in table 1. As is seen, the occupancy number increases and seems to converge, approximately, to the value corresponding to normal fluids (e.g. for a dense fluid of hard spheres the occupancy number around an arbitrarily chosen hard sphere is about 1.0). It means that the relatively open original structure of water around the solute becomes denser with increasing solute size. One might thus intuitively expect that this trend should be accompanied by an increase in the energy, but the opposite holds true (see table 1 and figure 3). Since in the unperturbed water structure the molecules exhibit a certain saturation with respect to hydrogen bonding, the molecules pene-
Table 1. Simulated systems and their thermodynamic properties: the partial packing fraction of the solvent $\eta_w$, pressure $P$, coordination number $N_c$, occupancy number $N_o$, the average energy of the water molecules in the first solvation shell $E^{(1)}$, the total average energy per water molecule $E$, the excess chemical potential of the solute $\mu_S^{\mathrm{ex}}$, and the excess chemical potential of water $\mu_W^{\mathrm{ex}}$. $\beta = 6$ for the EPM4 model, and $\beta = 5$ for the EPM5 model.

<table>
<thead>
<tr>
<th>$d$</th>
<th>$\eta_w$</th>
<th>$\beta P$</th>
<th>$N_c$</th>
<th>$N_o$</th>
<th>$-2E^{(1)}$</th>
<th>$-2E$</th>
<th>$\beta\mu_S^{\mathrm{ex,(der)}}$</th>
<th>$\beta\mu_S^{\mathrm{ex,(ext)}}$</th>
<th>$\beta\mu_W^{\mathrm{ex}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure water</td>
<td>0.35</td>
<td>3.65</td>
<td>—</td>
<td>—</td>
<td>3.11</td>
<td>5.9</td>
<td>5.5</td>
<td>7.49</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.35</td>
<td>3.71</td>
<td>13</td>
<td>0.82</td>
<td>3.04</td>
<td>3.12</td>
<td>27.6</td>
<td>25.7</td>
<td>7.46</td>
</tr>
<tr>
<td>2</td>
<td>0.3497</td>
<td>3.64</td>
<td>24</td>
<td>0.82</td>
<td>2.94</td>
<td>3.14</td>
<td>164</td>
<td>157</td>
<td>7.25</td>
</tr>
<tr>
<td>3</td>
<td>0.346</td>
<td>3.57</td>
<td>44</td>
<td>0.88</td>
<td>2.88</td>
<td>3.12</td>
<td>77</td>
<td>72</td>
<td>7.26</td>
</tr>
<tr>
<td>4</td>
<td>0.3417</td>
<td>3.62</td>
<td>62</td>
<td>0.97</td>
<td>2.82</td>
<td>3.10</td>
<td>164</td>
<td>157</td>
<td>7.25</td>
</tr>
<tr>
<td>5</td>
<td>0.335</td>
<td>3.64</td>
<td>90</td>
<td>0.93</td>
<td>2.80</td>
<td>3.08</td>
<td>302</td>
<td>290</td>
<td>7.39</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.3418</td>
<td>3.63</td>
<td>86</td>
<td>1.07</td>
<td>2.55</td>
<td>3.29</td>
<td>—</td>
<td>—</td>
<td>6.66</td>
</tr>
</tbody>
</table>

EPM5 model

<table>
<thead>
<tr>
<th>$d$</th>
<th>$\eta_w$</th>
<th>$\beta P$</th>
<th>$N_c$</th>
<th>$N_o$</th>
<th>$-2E^{(1)}$</th>
<th>$-2E$</th>
<th>$\beta\mu_S^{\mathrm{ex,(der)}}$</th>
<th>$\beta\mu_S^{\mathrm{ex,(ext)}}$</th>
<th>$\beta\mu_W^{\mathrm{ex}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure water</td>
<td>0.3</td>
<td>3.58</td>
<td>—</td>
<td>—</td>
<td>3.68</td>
<td>5.7</td>
<td>6.1</td>
<td>5.73</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>3.55</td>
<td>13</td>
<td>0.71</td>
<td>3.66</td>
<td>3.69</td>
<td>26.1</td>
<td>27.4</td>
<td>5.70</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
<td>3.59</td>
<td>23</td>
<td>0.80</td>
<td>3.63</td>
<td>3.70</td>
<td>73</td>
<td>76</td>
<td>6.48</td>
</tr>
<tr>
<td>3</td>
<td>0.2978</td>
<td>3.59</td>
<td>39</td>
<td>0.85</td>
<td>3.62</td>
<td>3.68</td>
<td>158</td>
<td>163</td>
<td>6.29</td>
</tr>
<tr>
<td>4</td>
<td>0.295</td>
<td>3.60</td>
<td>58</td>
<td>0.85</td>
<td>3.60</td>
<td>3.68</td>
<td>302</td>
<td>300</td>
<td>6.03</td>
</tr>
<tr>
<td>5</td>
<td>0.289</td>
<td>3.59</td>
<td>81</td>
<td>0.87</td>
<td>3.58</td>
<td>3.67</td>
<td>292</td>
<td>300</td>
<td>6.03</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.2986</td>
<td>3.59</td>
<td>90</td>
<td>1.00</td>
<td>3.47</td>
<td>3.66</td>
<td>—</td>
<td>—</td>
<td>6.04</td>
</tr>
</tbody>
</table>

$\alpha\beta\mu_S^{\mathrm{ex\,(der)}}$ denotes the results obtained using equation (15).

$\beta\mu_S^{\mathrm{ex\,(ext)}}$ denotes the results obtained using [6], equation (10).

Figure 2. Solute-oxygen pair correlation functions for different solute diameters. The contact distance $r_c$ equals $(1 + d)/2$ for the spherical solute of diameter $d$. 

![Figure 2](image)

Hydrophilic hydration II

Trating the first solvation shell do not find easily a partner to establish a hydrogen bond. To make this happen, the original hydrogen bond network must gradually be disrupted, leading to an increase in number of molecules with fewer hydrogen bonds and hence to the drop of energy per molecule, which equals minus half of the number of bonds per molecule. To better understand the difference between the two models and the large decrease in the number of bonds per molecule in the hydration shell of the EPM4 model, figure 4 shows the distribution of molecules in the first solvation shell with respect to the number of established hydrogen bonds. The distribution for pure fluids without any solute is included also. A number of differences between the two pictures are discernible. The most significant difference found appears to be the most probable number of bonds: whereas for the EPM4 model it is 3, followed by 4 and 2, the EPM5 model features a distinct maximum for 4 bonds, with about 25% of molecules having 3 bonds, about 3% molecules having 2 bonds, and less than 0.2% molecules having 1 bond; the number of non-bonded molecules is negligible (0.004%). This is not the case with the EPM4 model, with about 0.3% of non-bonded molecules and 4% molecules creating only 1 bond. We also note that, unlike the EPM5 model, in which for steric reasons the number of hydrogen bonds cannot exceed 4, with probability less than 2% the EPM4 model creates 5 bonds and, in extremely rare cases, even 6 bonds. This finding is in agree-
ment with the behaviour of realistic models of water, which also allow for the creation of five or even more bonds with a small probability [23]. In any case, the capability of the EMP4 model to create more than 4 bonds has no effect on our discussion.

The bonding of the EPM4 molecules in the hydration shell also exhibits a greater sensitivity to the presence of a solute. As the solute size increases, the number of molecules with 4 bonds is nearly halved for \( d = 5 \) and drops to only 13% for the hard wall, resulting in an increased percentage of molecules with 2 bonds or 1. The fraction of non-bonded molecules for \( d = 5 \), 0.6% is twice that in pure water, and for the hard wall system it doubles further to 1.2%. The number of molecules having 3 bonds remains approximately the same for all solutes. The decrease in bonding in the solvation shell of the EPM5 solvent with increasing solute size, seen in table 1 and figure 3, is also clear from figure 4, although it is much smaller than that for the EPM4 model. Even for the hard wall, the population of states with 0 or 1 bond remains negligible, 0.013% and 0.5%, respectively. Therefore, the only observed trend is the reduction in the number of molecules with 4 bonds from 72% for pure solvent to 64% for \( d = 5 \) and 55% for the hard wall, which is transformed predominantly into an increase in the percentage of molecules with 3 bonds and partially also of those with 2 bonds.

The hydrogen bonded structure in the hydration shell at the wall is shown in figure 5 for both solvents. In the view along the normal to the wall, the broken structure in the EPM4 model is seen clearly, whereas the EPM5 solvent still shows the network structure, though with vacancies. In the lateral view (i.e. along the wall) we see a distortion of the tetrahedral bonding of molecules. For the EPM5 solvent, the three bonds of a molecule closest to the wall are nearly planar, resulting in an increase in the angle between the bonds from the tetrahedral value 109° to close to 120°, while the fourth bond pointing to the bulk forms nearly a right angle with the other bonds. Thus the EPM5 solvent allows the bond angles to vary to some extent, which enables the molecules to accommodate the solute and to preserve the hydrogen bonding at the same time. The EPM4 model does not exhibit such plasticity of bonds, which results in a loss of bonding between the molecules closest to the solute. As the solute size increases and its surface curvature...
decreases, it becomes more and more difficult for molecules of the EPM4 solvent to form hydrogen bonds in its vicinity.

### 3.2. Orientational ordering

The orientational ordering of molecules in the vicinity of the solute or interface is the key quantity for understanding the aspects of molecular ordering that determine the macroscopic properties of these systems. In [6] we found that for small solute diameters and low temperatures the dependence of angles $\alpha$ and $\gamma$ on diameter $d$ was qualitatively the same for both primitive water solvents. However, this is not the case for large $d$. Before we proceed to consider separately the behaviour of the two models for large solutes, a brief summary of the orientational alignment found in [6] for small solutes is in order. In general, we found that the preferred orientations are those preserving hydrogen bonding. Specifically, for both models the preferred alignments were those with (i) one hydrogen site pointing towards the bulk and three bonds straddling the surface of the solute, and (ii) the dipole pointing towards the solute and the hydrogens pointing towards the solute at angles $\alpha \approx 55^\circ$ large enough not to prevent bonding of these sites to other molecules.

#### 3.2.1. EPM4 model

The distribution of angles $\alpha$ and $\gamma$ is shown in figure 6. Based on this information and supported by snapshots of the structure of selected molecular configurations, we can infer the alignment of water molecules in the vicinity of the solute. Both angles $\alpha$ and $\gamma$ must be analysed simultaneously to identify orientations consistent with the observed distribution of angles.

For small solutes, two distinct orientations are preferred: the first orientation, denoted by (1) in [6], figure...
8, with the dipole pointing towards the solute, $\gamma = 0$, and both hydrogens at $\alpha = 55^\circ$, and the second, denoted by (2), with one hydrogen pointing towards the bulk ($\alpha = 180^\circ$) and the other therefore around $\alpha = 70^\circ$ and $\gamma \approx 120^\circ$. Of course, configurations which can be tilted from these idealized orientations and still keep the same bonding pattern are observed which give rise to a broad maximum for angles $45^\circ < \alpha < 80^\circ$ with a peak at $\alpha \approx 60^\circ$ instead of the expected two maxima at $55^\circ$ and $70^\circ$.

Orientation (2) becomes less favourable with increasing solute size, as confirmed by a decrease in the distribution of $\alpha$ around $180^\circ$ and of $\gamma$ for $\gamma > 90^\circ$, while alignment (1) with dipole pointing towards the solute predominates for all solute sizes up to $d = 5$. Also, we observe that the maximum at $\alpha \approx 60^\circ$, which increases with solute size up to $d = 2$, decreases again and shifts slightly towards larger $\alpha$ when the solute size is increased further. We note that it was the decrease in this maximum for $d = 3$ with respect to $d = 2$, found in [6], that drew our attention to a possibility of some restructuring for large solutes in this solvent and motivated this study.

The most important trend, however, is that the configurations with one hydrogen pointing towards the solute, $\alpha = 0$, which are unfavourable for small solutes because of the lack of hydrogen bond, become favourable for large solutes. Clearly the two maxima of distribution of $\alpha$ at $\alpha = 0$ and $\alpha \approx 60^\circ$ for $d = 5$ cannot originate from the positions of two hydrogens of a single molecule. On the contrary, they indicate that the resulting distribution is a sum of contributions from two kinds of molecule: (i) those with proper hydrogen bonds (orientation (1)) contributing to angles $\alpha \approx 55^\circ$ and $\gamma \approx 0$, and (ii) those sacrificing one hydrogen bond by pushing one hydrogen towards the solute, $\alpha \approx 0$, with the other hydrogen around $\alpha \approx 100^\circ$ and dipole at $\gamma \approx 50^\circ$. For $d = 5$, the orientation with the proper hydrogen bonding still prevails. However, the increasing ratio of weakly bonded molecules is already responsible for the shoulder of $\alpha$ at small angles, the shift of the maxima at $\alpha \approx 60^\circ$ towards large values due to the hydrogen pointing parallel to the surface or slightly away from it, and the shift of the maxima of $\gamma$ from 0 to $\gamma \approx 40^\circ$. The weakly bonded configurations with the hydrogen pointing towards the solute predominate for the infinitely large solute. The contribution from the weakly bonded configurations dictates the shape of distributions of the two angles $\alpha$ and $\gamma$, although the shift of the maxima of $\gamma \approx 40^\circ$ and $\alpha \approx 80^\circ$, as opposed to the values originating exclusively from the orientation with one hydrogen pointing directly to the solute, $\gamma = 55^\circ$ and $\alpha = 109^\circ$, still indicates the presence of strongly bonded molecules.

The question is, what causes the restructuring of the hydration shell of the EPM4 model in the vicinity of large solutes. Beyond any doubts, it is the interplay between hydrogen bonding (energy contribution) and steric, exclusion effects (entropy contribution). We analysed the role of these two factors in [6], comparing the behaviour of the extended primitive models with that of these models by switching off completely the hydrogen bonding and leaving just the hard sphere site–site repulsions (the so-called pseudo-hard bodies, PHB) [24, 25]. We established that switching off the hydrogen bonds completely inverted the preferred orientations found for the primitive model solvent around the solute. The angular distribution of the PHB fluid originating in the EPM4 fluid (i.e. the fluid with any hydrogen bonding) around the solute of diameter $d = 1$, shown in [6], figure 7, is in fact strikingly similar to the angular distribution found for the EPM4 fluid at a hard wall.

The fact that the angular distribution in the hydration shell of EPM4 water at a hard wall results from a sum of contributions from strongly and weakly bonded molecules is confirmed in figure 7, where we show the angular distribution of angles $\alpha$ and $\gamma$ for three subshells.
of the hydration shell based on their distance from the solute. The molecules in the innermost layer (i.e. closest to the wall) push one hydrogen to the wall, i.e. to the region inaccessible for oxygens but penetrable for hydrogens. The exclusion effects overrule the hydrogen bonding in this layer. In the outermost subshell of the hydration shell, hydrogen bonding is favoured. The intermediate layer shows mixed behaviour. Though the inner layer is very thin ($\approx 0.3$), it is highly populated (figure 2). As a result, the preferential alignment with one hydrogen pointing towards the wall predominates when taking the average over the entire hydration shell. This finding explains the sharp drop in the average number of hydrogen bonds per molecule seen in figure 3 for the layer closest to the solute. A picture similar to figure 7 can be obtained for all solute diameters $d > 2$, i.e. in all these cases the surface layer reveals the orientation corresponding to poor hydrogen bonding. As the solute is enlarged, this contribution becomes more pronounced and significant.

3.2.2. EPM5 model

The situation for the EPM5 solvent seems much simpler. As figure 8 shows, with increasing solute size the favourable values of $\alpha$ become more pronounced. The locations of extrema are nearly the same for all $d$, with a slight trend towards larger values of $\alpha$ as $d$ increases. The positions of the maxima at $\alpha \approx 60^\circ$ and $\alpha \approx 180^\circ$ suggest that in the most preferred configurations one site (H and M sites are fully symmetric in this model) points towards the bulk and the other three straddle the surface of the solute. This was our conclusion drawn in [6]. However, it cannot explain by itself the maximum of the distribution of angle $\gamma$ at 0. The evolution of the angular distribution with solute size, and particularly the change in the shape of distribution of $\gamma$ found for the largest solute and the hard wall, help us to identify two preferred configurations that are responsible for all observed features. Before proceeding it is worth recalling that, because of the symmetry of H and M sites in the EPM5 model, the distribution of $\gamma$ should be symmetric (the bisectors of HOH and MOM are antiparallel). In simulations, sites H and M are evidently distinguishable and therefore the observed distribution of $\gamma$ is not strictly symmetric. Its relatively high degree of symmetry indicates that the structure of water around the solute is not frozen but allows for reorientation of the molecules and swappings of the H and M sites. This is confirmed by the fact that the distributions of angles $\alpha \equiv \angle \text{SOH}$ and $\angle \text{SOM}$ are indistinguishable (they were collected independently in the course of simulations).

The orientations with one site pointing towards the bulk (which we denote as (2) in accordance with the equivalent orientation in the EPM4 solvent and [6], figure 8) is responsible for the distribution of $\alpha$ with maxima at $\alpha = 70^\circ$ and $\alpha = 180^\circ$, and $\gamma = 55^\circ$ and $\gamma = 125^\circ$. The other preferred orientation is that corresponding to the configuration denoted as (1) for the EPM4 model, i.e. with two sites pointing towards the solute at angle $\alpha = 55^\circ$ and the other two sites pointing away from the solute at angle $\alpha = 125^\circ$. If the two sites pointing towards the solute are of the same kind, the corresponding $\gamma$ is 0 for the sites H and 180$^\circ$ for the sites M. However, if the two sites are of a different kind, the resulting dipole is parallel to the wall, $\gamma = 90^\circ$. Though these two cases are distinguishable, it seems that their ability to create bonds in the vicinity of the solute is comparable, since they share the same geometry.

Small differences in the distributions of $\alpha$ and $\gamma$ for different solute sizes seem to be understood, having noted that with increasing $d$ the distribution of $\alpha$ becomes only more pronounced. All three findings, namely (i) the maximum of $\alpha$ around 60–65$^\circ$ (i.e. between the angles $\alpha = 55^\circ$ and $\alpha = 70^\circ$ of the two preferred orientations), (ii) the preference for all values $\alpha > 150^\circ$, and (iii) the flat distribution of $\gamma$ for all sizes of the solute, indicate that both the preferred orientations we have identified are always present. In fact, the
term ‘preferred orientation’ is an idealization from which the actual orientation differs and it is not difficult for a molecule to switch between orientations (1) and (2) by a simple tilting. For \( d = 1 \), the distribution of \( \alpha \) closer to those angles corresponding to orientation (1) and the maxima of \( \gamma \) at 0, 90° and 180° indicate that the orientation with two sites straddling the sphere is the most frequent. On the other hand, near the wall we observe an opposite shift in the distribution of \( \alpha \), and the two maxima of \( \gamma \) also indicate that the orientation with one site pointing towards the bulk and three sites straddling the surface is favoured. For both these orientations, all sites can participate in hydrogen bonding, and therefore we observe only a small decrease in the number of hydrogen bonds per molecule in the hydration shell of the EPM5 solvent. From a figure analogous to figure 7 (not shown) one would see that there is a unique angular distribution of \( \alpha \) and \( \gamma \) in the entire first hydration shell for any size of the solute, i.e. the orientation of molecules in the hydration shell does not depend on the distance from the solute, and even the molecules closest to the solute participate in strong hydrogen bonding, as shown in figure 5.

3.3. Thermodynamic properties

Differences in the structure of different solutions should be reflected also in their thermodynamic behaviour. Thermodynamic functions of interest are calculated both for the entire system and in the regions of inhomogeneity around the solute.

The total internal energy of dilute solutions does not seem to be affected by different solutes, as is seen from table 1, and regardless of the solute, the overall internal energy (which equals the energy of the solvent) per water molecule \( E \) is (within estimated simulation errors) the same for the given solvent. However, systematic changes are observed for \( E^{(1)} \), i.e. for the energy per water molecule in the first hydration shell. These changes were discussed in § 3.1.

A very important thermodynamic property is the chemical potential. In [6] we computed the excess chemical potential for small solutes (up to \( d = 1.7 \)) by Widom’s virtual insertion, equation (12). For large solutes the probability of insertion is extremely small, and therefore the virtual insertion method is inapplicable. However, if the solute–oxygen correlation function is known for a number of solute diameters, one can make use of the exact relationship between the contact value of the solute–oxygen pair correlation function and the derivative of the excess chemical potential [20],

\[
\frac{\partial \mu_{ex}^{S}(d')}{\partial d'} \bigg|_{d'=d} = 3(1 + d^2) \rho_W(r_c),
\]

(14)
to determine the chemical potential. To make use of this equation in numerical calculations, we assume, as in [6], that the chemical potential of the solute can be approximated by a cubic polynomial,

\[
\beta \mu_{ex}^{S}(d) = a_0 + a_1 d + a_2 d^2 + a_3 d^3.
\]

(15)

From the four unknown coefficients \( a_i \), only one, \( a_2 \), is in fact adjustable, with the remaining ones being subject to certain physical constraints (for details see [6]). \( a_0 \) and \( a_1 \) are given by the exact zero diameter limit,

\[
a_0 = -\ln (1 - \eta); \quad a_1 = 3\eta/(1 - \eta),
\]

(16)

and \( a_3 \) is related to the large solute limit,

\[
\frac{\partial \beta \mu_{ex}^{S}(d)}{\partial d} \bigg|_{d=\infty} = \frac{\pi}{2} d^2 \beta P,
\]

(17)

which yields

\[
a_3 = \frac{\pi}{6} \beta P.
\]

(18)

Using the data on \( g_{SO} \) for the values of \( d \) given in table 1 and the above constraints we determined \( a_2 \) by a least-squares fit, with the result:

EPM4: \( a_0 = 0.431, a_1 = 1.615, a_2^{(der)} = 1.788, a_3 = 2.051 \);

EPM5: \( a_0 = 0.357, a_1 = 1.286, a_2^{(der)} = 2.426, a_3 = 1.889 \).

We mention in passing that if the parameter \( a_3 \) is not constrained and is considered as an adjustable parameter, the fitting procedure nonetheless yields, numerically, the same value, confirming accuracy of equation (17). The resulting excess chemical potential \( \beta \mu_{ex}^{S(ex)} \) is given in table 1 along with the result of the extrapolation \( \beta \mu_{ex}^{S(ex)} \) of the fit of our previous data [6] on the chemical potentials for small solutes (up to diameter \( d = 1.7 \)) obtained by virtual insertions. As is seen, agreement between the two predictions is very good, and confirms that these simple ways provide reliable methods for predicting the chemical potential of the solute. For small solutes, \( \beta \mu_{ex}^{S(ex)} \) gives better agreement with the measured chemical potential, while \( \beta \mu_{ex}^{S(ex)} \) is believed to be more accurate for larger solutes.

To obtain the complete thermodynamic picture, we computed also the excess chemical potential of the solvent using the usual Widom’s method. In the models of water considered here, the Boltzmann factor in expression (12) can take only six values (0 for insertion resulting in overlap, 1 for insertion without any bond, exp \( \beta \) for insertion with creation of a single bond, etc.) which may differ by several orders of magnitude. As a
result, the average of the Boltzmann factor is very sensitive to the number of insertions resulting in the creation of two or more hydrogen bonds, since each of these instances has the same weight as thousands of insertions without creating a bond. This severe problem of a reliable determination of $\beta \mu_W^{ex}$ is aggravated by the low insertion probability (i.e. the probability of successful insertion without an overlap), which is about $1 \times 10^{-5}$ for EPM4, and $4 \times 10^{-7}$ for EPM5. As a result, we had to perform an enormous number $(O(10^6))$ of trial insertions to obtain at least a reasonable estimate of $\beta \mu_W^{ex}$.

Due to inhomogeneity in the surrounding of the solute, the excess chemical potential of water also varies with distance from the solute. Therefore we recorded the histograms for $\mu_W^{ex}$ in shells around the solute and evaluated $\rho_W^{ex}$ as a function of $r$. Since the total chemical potential $\mu_W^{ex}$,

$$\beta \mu_W = \ln \rho(r) + \beta \mu_W^{ex}(r) = \ln \rho_W + \beta \mu_W^{ex},$$

must be constant throughout the entire system, this condition served also as a test of consistency of the calculations, and an example is shown in figure 9. Allowing for relatively large simulation errors we see that this condition is satisfied. The results for the excess chemical potential $\mu_W^{ex}$ are given in table 1. Besides an apparent decrease for the hard wall systems of EPM4, the excess chemical potential does not seem to exhibit any trend with increasing solute size for either solvent model and seems to be, within simulation uncertainty, constant.

4. Conclusion

We have studied hydrophobic hydration of large hard sphere solutes of diameters up to five times the size of the water molecule, and the limiting case of water in a wide slit using two qualitatively different extended primitive models of water. Unlike the case of small solute discussed in [6], where both models give qualitatively the same results in close agreement with those for realistic potentials, we found qualitatively different behaviour of the two primitive water solvents for large solutes and near the hard wall.

In the EPM5 solvent, the hydrogen bonding exhibits a plasticity that makes it possible for the perfect hydrogen bonded network to extend to the ultimate vicinity of the wall. While one hydrogen points to the bulk, the remaining arms, straddling the wall, are nevertheless able to create bonds with other molecules of the hydration shell. In contrast to the EPM4 model, in the case of the EPM5 model the hydration shell can be divided into two layers. The outer layer exhibits features similar to the hydration shell of the EPM5 model; hydrogen bonding is relatively strong in this layer. In the inner layer closest to the wall, competition between hydrogen bonding and volume exclusion effects takes place which leads to an orientation re-ordering. In the most probable configuration, one hydrogen of a molecule is stuck to the wall, sacrificing a bond for better packing of molecules.

The results obtained point to some sensitivity of the structure of water around the solute to details of the model used. What we do not know is the extent to which the results for the model solutions considered in this paper correspond to reality. Such a comparison, however, does not seem to be simple and unambiguous. First, in realistic models both the solvent and the solute are represented by realistic models whose effects are hard to separate. The solute is in most cases a Lennard-Jones particle whose counterpart in the limit of an infinite diameter is a wall with Steele’s average potential [37]; this potential can give rise to a profile quite different from a simple structureless and purely repulsive wall [9]. In the models considered here, the density around the solute increases with solute diameter, whereas the opposite trend seems to be the case for realistic solutes [28]. Similarly, different models of solvent may invoke different structures (e.g. ST2 water at a wall exhibits quite a flat density profile [30] where MCY water exhibits well pronounced layering [31]). Second, different treatment of the long range Coulombic interactions (simple spherical cutoff, modified reaction field method, or different variants of the Ewald summation) may also contribute to differences in the case of a hard wall. Nonetheless, there seems to be a consensus on the orientational ordering of water molecules at a structureless wall: a water molecule at a hydrophobic surface tends to sacrifice one hydrogen bond by orienting one potentially hydrogen bonding group towards the surface. The result for the EPM4 model is fully consistent with this observation.
All the known facts show that hydrophobicity is a subtle equilibrium between entropic and energetic (enthalpic) effects, where the effect of the solute becomes more pronounced. This seems to be the case also for the EPM4 and EPM5 models studied. Both models seemed equivalent as regards the model of pure water, and only solutions with an extreme (very large) solute have revealed the difference. However, the degree to which they mimic real water remains an open question, just as it is for the much more realistic models.

The advantage of the water and solute models that form the basis for the work presented in this paper is that the computation time is considerably shorter when compared with calculations using realistic models, and their Hamiltonian model is simple enough to be amenable to theoretical study, for at least some thermodynamic properties. Simplicity of the primitive models makes it possible also to understand their different solvent properties and to trace them back to the difference in geometric mechanism of hydrogen bonding in both models. However, this may not be so for complex realistic models. In any case, the fact that the results are model dependent should always be borne in mind, and any conclusions on the hydration of apolar solutes should be drawn with caution.

This research was supported by the Grant Agency of the Academy of Sciences of the Czech Republic under Grant No. A4072908, and by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences, US Department of Energy.

References