UCRL-JRNL-234592

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

First Principals and Classical Molecular Dynamics Simulations of Solvated Benzene

M. Allesch, F. Lightstone, E. Schwegler, G. Galli

September 13, 2007

Journal of Chemical Physics

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Updated October 14, 2003

First principles and classical molecular dynamics simulations of solvated benzene

Markus Allesch,[1,2] Felice C. Lightstone,[2] Eric Schwegler,[2] and Giulia Galli[2,3]

[1]*Department of Theoretical and Computational Physics, Graz University of Technology, Austria*

[2]*Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550*

[3]*University of California, Davis, CA 95616*

(Dated: October 15, 2007)

Abstract

We have performed extensive *ab initio* and classical MD simulations of benzene in water in order to examine the unique solvation structures that are formed. Qualitative differences between classical and *ab initio* MD simulations are found and the importance of various technical simulation parameters is examined. Our comparison indicates that non-polarizable classical models are not capable of describing the solute-water interface correctly if local interactions become energetically comparable to water hydrogen bonds. In addition, a comparison is made between a rigid water model and fully flexible water within *ab initio* MD simulations which shows that both models agree qualitatively for this challenging system.

I. INTRODUCTION

The accurate description of the interaction between hydrophobic solutes and interfaces with water play a key role in understanding many different processes, such as protein folding and stability[1]. However, the description of these complex interfacial systems has proven to be particularly difficult since properties cannot be simply extrapolated from the bulk and the behavior at the interface strongly depends on subtle details of the solute-solvent interaction and external conditions[2].

For the different types of MD simulation techniques developed over the last decades the hydrophobic hydration of benzene (B) represents a challenging test case. It can serve as a model for many biological systems containing delocalized π-electron systems, it is a small enough system for extensive *ab initio* simulations and it contains two distinct regions that interact differently with water[3]. Solvated benzene also has significantly different properties than comparable solvated aliphatic hydrocarbons (e.g. a much higher solubility) and has been the subject of several computational studies using classical MD techniques[4–14]. Most recently two studies of solvated benzene based on empirical potentials[12,13] showed qualitative differences for computed radial distribution functions, one of the basic properties that is usually employed to compare empirical potentials with experimental results. These two sets of classical simulations used different empirical models for water and for benzene leading to a qualitatively different description of the weak π-hydrogen bond which is characteristic for this system[15]. Since this important feature contributes significantly to the special solvation properties of benzene[13] and also occurs in many other biological systems, it deems necessary to resolve these existing discrepancies between MD simulations and to obtain an accurate description of this system by systematically investigating the different methods and approximations used.

In a previous study dealing with the hydrophobic hydration of benzene, the structural and electronic properties of solvated benzene were examined with a rigid water model in *ab initio* MD simulations[3]. Figure 1 gives an overview of the hydrophobic solvation of benzene by showing spatial distribution functions (SDF) of oxygen for two different isosurfaces (1 and 2.6). Clearly visible are the first and parts of the second solvation shell in the left graph. At a higher level of the isosuface, cage-like solvation structures can be seen around the equatorial region, which closely resembles other small hydrophobic solutes[16]. In contrast,

due to the formation of weak π-hydrogen bonds, distinct disk-like distributions are found near the axial regions of the solute.

These results for benzene have been compared with analogous data for solvated hexafluorobenzene (HFB) to highlight the different nature of the solvent interaction with the π-electron systems. Due to a reversed quadrupole moment, it was found that solvated HFB prefers a π-lone pair interaction with water in the axial region, whereas benzene displays a weak π-hydrogen bond. In both cases, the equatorial regions around the aromatic molecules behaves like an ordinary hydrophobic solute leading to the typical water tilt angle of about 120°.

In this study we have continued to examine the solvation properties of benzene in water by carrying out an *ab initio* MD simulation using fully flexible water molecules and extensive classical MD simulations. The comparison and evaluation of these different types of simulation methods and the approximations contained within are the main goal of the present study.

The rest of the paper is structured as follows: First we compare the structural properties obtained from two sets of *ab initio* simulations using rigid and flexible water models. In section III.B we contrast these results with analogous data obtained from extensive classical MD simulations. Finally, we investigate the role of several technical simulation parameters which could be responsible, in part, for some of the observed differences between results from different MD methods.

II. METHOD

A. *Ab initio* MD technique

Our *ab initio* MD simulations[17,18] of solvated benzene are based on electronic structure calculations within density functional theory (DFT)[19,20] utilizing the PBE generalized gradient approximation (GGA)[21]. Plane wave basis sets were used to represent the valence wavefunctions and charge density and were truncated in reciprocal space at 85 and 340 Ry, respectively. Valence-core interactions were described by Hamann type norm-conserving pseudopotentials[22,23]. We note that binding energy calculations for the benzene-water dimer at this level of theory yield good agreement compared with quantum chemical studies[3].

All of our *ab initio* simulations consisted of 73 water molecules surrounding a benzene molecule in a cubic cell with a lateral dimension of 13.25 Å and periodic boundary conditions. The system included the first and most of the second solvation cell at a density of 0.994 g/cc and ambient pressure and temperature conditions. A weakly coupled velocity scaling thermostat was initially used as previously described[3] before statistics were collected within the NVE ensemble. The total simulation times and parameters of all MD simulations are reported in Table I.

The propagation of electrons and ionic cores in the simulations was performed following the Car-Parrinello (CP) technique, which is based on the use of a Lagrangian that couples together the system's electronic and ionic degrees of freedom[24,25]. A central parameter in the CP scheme is the fictitious electron mass, μ, which has to be chosen very carefully to avoid overlapping of ionic and electronic degrees of freedom and ensure adiabaticity[26,27]. If water molecules in CP simulations are fully flexible it has been shown that μ can be safely set to 340 a.u.[26], whereas in the case of a fully constrained (rigid) water model larger values for μ become feasible, allowing the use of a larger MD timesteps dt[28].

Despite the rapidly increasing capability of modern supercomputers, it is still very difficult for normal *ab initio* MD (which includes all intramolecular modes) to reach statistically significant timescales for hydrophobic systems. In addition, we note that at a temperature of 300 K, $k_BT \sim 200$ cm^{-1}, whereas the high frequency intramolecular modes in water range from ∼1000 to 3500 cm^{-1}. In other words, the amount of thermal energy available to excite vibrational modes is much smaller than the lowest possible intramolecular vibrational excitations,

$$\hbar\omega \gg k_BT. \tag{1}$$

As such, a real quantum system will essentially be restricted to its vibrational ground state at 300 K. Therefore, one could argue that a rigid water model is a closer representation of the quantum system than a classical flexible model. The use of a fully constrained (rigid) water model[28] also offers the advantage of using timesteps up to 10 a.u. This extension of simulation timescales by a factor of three opens up the possibility to carry out *ab initio* trajectories beyond 100 ps at reasonable computational costs.

Since the geometry of water molecules is not expected to be significantly different in the vicinity of the solute at ambient conditions[29] compared to its bulk values, we were able to utilize the geometry of our recently developed rigid water model[28] for some of our calculations

presented in this study (simulation B in Table I). This model has been originally developed for simulations of pure water[28] and has been successfully applied to solvated Ca^{2+} [30]. Despite the success of the model and similar encouraging results from other authors[31], we have verified the results against data obtained from simulations using a fully flexible model for water. It is important to note that despite the rigid geometry that is enforced, the electronic wavefunctions are still computed from first principles and electronic charges can distribute freely - in contrast to rigid classical water models like TIP4P.

Spectroscopic studies on benzene - water clusters have shown carbon-hydrogen stretch fundamentals of benzene to be on the order of $3100 cm^{-1}$ for a $benzene - water_8$ cluster[32]. To utilize the advantages of the rigid approximation, i.e. its possibility of longer timesteps, we have also constrained the C-H bonds of benzene to their optimized gas phase values.

B. Classical MD technique

For our classical MD simulations we have used the GROMACS 3.2.1 program[33] with the OPLS-AA (optimized potential for liquid simulation - all atom) force field[34,35] and the (rigid) TIP4P model for water[36,37]. Non-bonded interactions in classical simulations are typically composed of Lennard-Jones (LJ) and Coulomb potential contributions:

$$V_{LJ}(\boldsymbol{r}_{ij}) \;=\; 4\varepsilon_{ij}\Big((\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}\Big). \tag{2}$$

$$V_C(r_{ij}) \;=\; \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r r_{ij}}. \tag{3}$$

The parameters for non-bonded LJ interactions have been constructed according to the following combination rule,

$$C_{ij}^{(6)} \;=\; \Big(C_i^{(6)}\, C_j^{(6)}\Big)^{\frac{1}{2}} \tag{4}$$

$$C_{ij}^{(12)} \;=\; \Big(C_i^{(12)}\, C_j^{(12)}\Big)^{\frac{1}{2}}, \tag{5}$$

representing the geometric mean of the atom type specific σ_i- and ε_i-parameters. The empirical parameters for benzene in the classical simulations, taken from the OPLS-AA force field, are listed in Table II. The non-bonded parameters for benzene were originally developed by W.L. Jorgensen and D.L. Severance in Ref. 7. Among the five adjustable

parameters for benzene (four Lennard-Jones and one Coulomb parameter) the two σ and ϵ for O and H had to be chosen consistent with earlier force field parameters[38]. The remaining charge parameter, $q^C = -q^H$, was fitted to results of Monte Carlo simulations for liquid benzene primarily to the observed heat of vaporization and density. For comparison with the existing literature dealing with solvated benzene[11–13] we have also included one other empirical force fields for benzene[39,40] in Table II. Bonded interactions did not play a role for the dynamics since all molecules were constrained to rigid geometries, i.e. TIP4P geometry for water and experimental bond lengths for benzene as specified in Ref. 7.

The simulation box contained one benzene molecule surrounded by either 73 or 512 water molecules using periodic boundary conditions. Both NPT and NVT ensembles were simulated (T=300K; P=1atm) for up to 100 ns in the smaller simulation cell and 10 ns in the larger cell at ambient conditions. An overview of all classical and *ab initio* simulations is given in Table I.

For all systems an initial energy minimization was performed by running 500 steps of the steepest descent (SD) energy minimization method followed by 100 ps of initial MD equilibration, which were discarded and not used to collect statistics. The timestep used for the numerical integration of the equations of motion was 2 fs. Coordinates were recorded every 100 steps (0.2 ps). The list of neighboring non-bonded atoms, which is the basis for evaluating non-bonded interactions, was updated after every step using a simple search during the energy minimization and every two steps using a grid search during the MD runs. Electrostatic interactions were treated using the Particle Mesh Ewald (PME) summation[41] with cubic interpolation and van der Waals interactions were handled using twin range cut-offs[33]. All cut-offs (Coulomb, van der Waals, short range neighbor list) were chosen to be as large as possible, i.e. half of the cell size. Long range dispersion corrections were used for energy and pressure.

The rigid intramolecular geometry of all molecules in the simulation box was enforced by a SHAKE algorithm[42] with a relative geometrical tolerance of 0.001. Different initial temperature distributions were realized for classical MD simulations by randomly assigning statistical velocity distributions to all particles in the box. Temperature control was achieved with a Berendsen thermostat[43] set to a target temperature of 300K and with a time constant of 1 ps. Barostats used for isotropic pressure coupling for NPT ensembles were also of the Berendsen type using the same time constant of 1 ps.

III. RESULTS AND DISCUSSION

A. Comparison of rigid and flexible water models in *ab initio* MD simulations of solvated benzene

The most commonly calculated quantities to investigate solvation structures are radial distribution functions (RDF, or $g(r)$)[44]. In the case of benzene the mirror plane symmetry of the solute molecule can be exploited to improve statistics for the reported RDFs. To get a clearer picture of the solvation structure we have divided the space around the solute into an axial region, which is defined as the space within 20° of the axis perpendicular to the plane of the solute (z-axis), and an equatorial region, which consists of the space within 20° on both sides of the symmetry plane.

In Figures 2 and 3 we compare the RDFs for benzene-oxygen (B-O) and benzene-hydrogen (B-H) calculated for the equatorial and axial region around benzene using flexible and rigid water models (simulations A and B). In the equatorial region (Figure 2) both models agree very well with differences in peak heights around 0.2 units and in peak positions of about 0.1 Å. For both g(r)s in this region the flexible models shows slightly more structure than obtained from the rigid approximation - an observation which is consistent with results for pure water[28].

In the axial region (see Figure 3) this trend is reversed and differences between both models are far more pronounced. For $g_{B-H}(r)$ first peaks differ by 0.6 units, second peaks by only 0.05 units but with a 0.2 Å outward shift. For the sharpest peak in the system, $g_{B-O}(r)$, the rigid model gives a maximum of 3.58 whereas the flexible model yields a much softer structure with a peak height of only 2.25. However, the overall agreement is still satisfactory in the axial region since the same qualitative relation between $g_{B-H}(r)$ peaks has been established and quantitative differences are within the estimated error bars considering the relatively short simulation time (simulation A: 81 ps). A more detailed discussion of simulation times and statistical errors can be found in Section III.C. It should be noted that these differences in the axial region appear much smaller when a slightly larger definition of the axial region is chosen.

In Ref. 11 a similar comparison between structural properties of solvated benzene obtained from classical models also includes results from a flexible empirical water model (F3C/ENCAD)[45]. While this flexible model yields the softest structure among all tested classical models in both regions, our *ab initio* results for flexible water show a different trend, i.e. comparable structures in the equatorial but far less structure in the axial region. Despite the uncertainty due to the large statistical error bars for data obtained for the axial region (see Section III.C), this difference between the axial and equatorial g(r)s obtained from rigid and flexible models is significant and, again, points at the very different type of interactions in the two regions of this hydrophobic solute[3].

Since geometric constraints can alter the dynamics of some molecular systems (e.g. angular constraints in the case of macromolecules[46]), we have also calculated preferred orientations of water molecules with and without intramolecular flexibility. Measurements of such angular distribution functions are commonly performed by calculating tilt angles of water molecules which are defined as the smaller angles between the bisecting vector of the H_2O molecule and the radial vector between oxygen and the center of the solute. To resolve possible orientational effects in the equatorial and the axial region separately, we have chosen to plot the cosine of the tilt angle against the cosine of the polar angle (Θ) which is measured from the z-axis defined by the normal of the molecular plane. As a consequence the upper region close to $\cos\Theta = 1$ in Figure 4 corresponds to the axial region, while the lower end of the y-axis shows the equatorial region. Figure 4 shows the tilt angle distributions within the first solvation shell computed for the two *ab initio* simulations A (left graph) and B (center graph). Both distributions display two separate maxima, one in the axial region at a tilt angle of $\sim 46^\circ$ and broader distribution in the equatorial region around 120°. The dominant axial peak is a result of the π-H bond, whereas the orientational preference in the equatorial region is typical for small hydrophobic solutes which do not interact strongly with water. Most notable is the different relation between the two peaks in both graphs. While both tilt angle distributions agree qualitatively, the flexible model yields orientational effects of similar magnitude in both the equatorial and the axial region - an effect which can also be inferred from g(r)s in Figures 2 and 3. In contrast to that the rigid model shows a significantly stronger and sharper orientational preference for the π-H bond in the axial region while the broad peak in its equatorial region around 120° is less pronounced than for the flexible model.

B. Comparison of classical and *ab initio* MD results

Figure 5 shows the characteristic solvation structure of benzene from classical MD. Clearly visible are the first solvation shell and the difference between oxygen and hydrogen SDFs in the axial region coming from the weak π-H bond. If these classical SDFs are plotted using the same representation as for the *ab initio* data in Figure 1, both distributions become almost identical on the chosen scale. Despite these similarities in terms of SDFs, significant differences occur when RDFs of classical and *ab initio* simulations are compared. In Figure 6 we have added the g(r) from a comparable classical simulation (simulation C) to the two *ab initio* results. The large overstructuring in classical results is immediately obvious with peak heights more than twice as high. Peak positions agree somewhat better but maxima obtained from classical MD are shifted inward on average by about 0.2 Å. Even more notable is the different relation between the two axial $g_{B-H}(r)$ peaks. Both *ab initio* simulations shows a first $g_{B-H}(r)$ peak, coming from one hydrogen atom which is pointed towards the π-system, which is about half of the second peak. In contrast, classical data yields the opposite relation, where the first peak is twice as pronounced as the second one and at the same time several times stronger in magnitude. In terms of solvent orientation this corresponds to a more favorable benzene-water interaction (through a weak π-H bond) in classical MD. As mentioned earlier, it should be noted that the differences between classical and *ab initio* MD results are far less obvious if the axial region is defined as a 45° cone around the z-axis.

In the equatorial region differences between classical and *ab initio* g(r) are relatively small with classical peak maxima about 0.15 units smaller than the corresponding values from the rigid *ab initio* model and peak positions shifted outward by about 0.1 Å.

Analogous to the two *ab initio* simulations we have also computed tilt angle distributions for a classical NVT simulation (Simulation C). The right graph in Figure 4 displays the two characteristic peaks for the two regions of the system which have been found before and shows good qualitative agreement with the *ab initio* data, in particular with the rigid model. However, a close comparison reveals that classical MD gives a significantly stronger axial peak (tilt angle $\sim 46^\circ$) than the *ab initio* results but a slightly softer distribution in the equatorial region around tilt angles of about 120° (as stated before from RDFs). The

smoother distribution for the classical MD data shown in Figure 4 is due to better statistics from a longer trajectory.

The similarities between the classical models and the *ab initio* simulations for the equatorial regions, and the significant differences near the axial regions is perhaps not too surprising. Typically, empirical parameters of classical water force fields are chosen to reproduce bulk liquid properties. In the case of small hydrophobic solutes it has been shown that the hydrogen bond network mostly persists - although in a rearranged form[47,48]. Because the local hydrogen bonds are not broken, the water molecules near the solute see a environment that looks similar to the bulk and the classical models work well. However, in cases such as the axial regions around benzene where solute-water interactions become significant, the local environment changes when the water-water hydrogen bonding is no longer preserved. We note that a similar breakdown of simple classical water models should occur around large hydrophobic solutes and surface as it is not possible to preserve the water-water hydrogen bonding around an extended interface[47,48].

In the case of solvated benzene, the interfacial water molecules in the axial region are weakly bound to the π-electron system of benzene instead of forming water-water hydrogen bonds, which leads to smaller water molecule dipole moments by about 0.15 Debye when compared to the bulk as shown in Figure 8. A similar decrease in the dipole moment of the interfacial waters near the equatorial region of the solute does not occur because water-water hydrogen bonding persists in this region.

We therefore speculate that the classical water models do not properly describe the interface when solute-water interactions become large enough to disrupt the local hydrogen bond network around the solute. Similar conclusions were also inferred from studies of solvated cations where classical water models were shown to coordinate stronger to the ion and form a more rigid solvation cage[49].

C. Convergence of simulation results

In this section we use a series of long time-scale classical MD simulations to investigate the sensitivity of the computed structural properties to the choice of certain parameters and approximations that are often used. This is done in an effort to understand the differences between published results for the hydrophobic solvation of benzene, to increase the repro-

ducibility of results obtained from MD simulations and to estimate realistic error bars for our relatively short *ab initio* MD simulations.

First, we have compared the influence of different bin sizes, Δr, used for collecting statistics on the obtained RDFs. By comparing the obtained peak heights for the axial g(r) for B-O (which is the "sharpest" peak requiring the best sampling) from simulation C using bin sizes of 0.03, 0.05, 0.07, 0.09, 0.11, 0.13 and 0.2 Å it can be shown that a bin size of 0.07 Å or smaller is required to resolve correct peak heights for this most sensitive peak of our system.

Another factor which is often not reported but which quantitatively influences peak heights is the normalization of the obtained "bined" data for the g(r). In many cases the raw data is simply normalized with respect to the average particle density in the simulation box. If one is interested in the g(r) of solvent atoms around a solute in a rather small simulation cell then this "relative" normalization leads to a cell size dependency which can be misleading if comparisons are made with data from larger cells where the volume occupied by the solute becomes less relevant. Instead, normalization with respect to the ambient particle density of the pure solvent gives RDFs independent of the simulation size.

Despite ever more powerful supercomputers *ab initio* MD simulations are still limited to relatively small system sizes containing 500 - 1000 atoms and total simulation times around 100 ps[50]. To obtain an estimate for possible finite size effects we have compared classical MD simulations of solvated benzene containing 73 and 512 solvent molecules. Figure 9 shows the most sensitive RDFs, $g_{B-H}(r)$ and $g_{B-O}(r)$ in the axial region of the solute, for the NVT ensemble with 73 and 512 solvent molecules (simulations C, D). Peak heights for these NVT simulations are in almost perfect agreement with differences below 0.1, demonstrating no significant size effects on obtained RDFs despite the relatively small number of solvent molecules. Size effects on RDFs obtained from the equatorial region around benzene are not visible on the reported scale.

Another test examining finite size effects was performed for tilt angle distributions of water around the hydrophobic solute. Both distributions obtained from simulations C and D show the two preferred orientations (as in the right plot of Fig. 4) and agree quantitatively. The only notable difference is a coarser resolution of the distribution obtained from the shorter trajectory (simulation D).

Limited total simulation times constitute another serious limitation of *ab initio* MD.

While a recent study[26] has investigated the time scales required to obtain well converged structural properties in simulations of pure water, we have obtained a similar estimate for the error bars (of RDFs) in simulations of hydrophobic solutes. Since a rigorous determination of the error bars for the *ab initio* MD simulations is out of reach we obtained an estimate from our classical MD simulations. Again, we have used the g(r)s for benzene-oxygen and benzene-hydrogen in the axial region since they are the most sensitive and characteristic peaks stemming from the π-hydrogen bond in this system. Figure 10 shows ten uncorrelated g(r)s obtained from 100 ps trajectories (NVT ensemble, 73 solvent molecules, 300 K), which illustrates the magnitude of the expected variance for this structural property on a 100 ps timescale. The maxima of the first peaks of $g_{B-O}(r)$ and $g_{B-H}(r)$ both vary by about 1.2 units, whereas the second peaks of $g_{B-H}(r)$ differ by 0.6 units. The same type of analysis for ten 10 ns trajectories shows differences of 0.23 units for the first B-H peak and about 0.1 units for the second B-H peak.

A comparison of NVT vs. NPT ensembles with consistent average pressure and volume (simulations D and F) is shown in Figure 11. The small difference of the $g_{B-O}(r)$ RDFs is within the estimated statistical error bar for simulations of 10 ns.

D. Sensitivity with respect to empirical force field parameters

Recently two studies have investigated different aspects of the hydrophobic solvation of benzene[11,13]. In Ref. 11 (which was reanalyzed in Ref. 12) the authors have used different force fields and water models to investigate the hydration of benzene and compare it to cyclohexane. While shape and peak positions of comparable benzene RDFs are the same, peak heights of the B-O and B-H peaks in the axial region differ quantitatively when compared to our results. Results from simulation F, which are shown in Figure 11, display peak heights of 6.10 and 4.12 Å whereas Ref. 11 reports 5.44 and 3.60 Å for axial $g_{B-O}(r)$ and $g_{B-H}(r)$, as shown in Figure 12.

Schravendijk and van der Vegt (Ref. 13) have studied the effect of benzene-water hydrogen bonding on hydrophobic hydration compared to a "van der Waals model" of benzene which is not capable of forming such weak H-bonds. When comparing RDFs from different studies, it should be noted that the obtained axial distributions are highly sensitive with respect to the definition of this axial region. The definition of the axial region in Ref. 13 includes

a 90° cone (45° around the z-axis) which leads to far less structured RDFs for both B-O and B-H because it includes a significant amount of water molecules which do not bind to the π-electron system through a weak H-bond. Results from simulations C-F are almost indistinguishable if compared for this extended region (45° instead of 20°) and on this scale. After recalculating RDFs for this differently defined axial region we find differences in benzene-hydrogen first peak heights of about 0.2 and a somewhat different shape of the benzene-oxygen double peak (Figure 13). The first $g_{B-O}(r)$ peak consists of two maxima, one from H_2O molecules which take part in the weak hydrogen bond and one from ordinary solvent molecules of the first solvation shell. Analogous to the results for benzene-hydrogen our classical data shows a stronger first benzene-oxygen maximum corresponding to a more favorable π-hydrogen bond compared to the results from Ref. 13.

The most likely cause for the differences in Figure 13 is the use of different force fields (GROMOS43A1 vs. OPLS-AA, see Table II) and water models (SPC vs. TIP4P). In the GROMOS43A1 parameter set atomic charges on carbon and hydrogen of benzene are slightly smaller which could lead to the observed softer solvation structure around benzene.

From our extensive analysis of different classical simulation parameters for this system, only part of which is shown in this paper, we are able to draw two main conclusions: Firstly, the results from classical MD simulations vary quantitatively depending on the chosen force field, water model and on subtle technical details of the simulation and statistical analysis of the results. Secondly, despite those quantitative differences between classical results all of the empirical-model based simulations examined here as well as in other work[11] show systematic qualitative differences when compared to *ab initio* results for the interfacial region where solute-solvent interactions are significant, i.e. they yield a much stronger π-hydrogen bond between water and benzene.

IV. CONCLUSION

Solvated benzene constitutes a unique test case for molecular simulations because it presents two different interfaces to the solvent. The equatorial region resembles normal hydrophobic solvation of aliphatic hydrocarbons, whereas the axial region displays a weak π-hydrogen bond between solute and solvent molecules which alters the electronic structure at the interface. Both flexible and rigid *ab initio* water models represent the hydration

structure at both interfaces well and show good agreement. To make statistically meaningful comparisons between simulations we have carefully investigated the role of several simulation parameters such as finite size and time scale effects.

Commonly used (nonpolarizable) classical water models give a robust description of the equatorial region where strong water hydrogen-bonds persist and no significant solute-solvent interactions are present. In the axial region these simple water models cannot adapt to the altered environment and there is a tendency to "overbind" the π-hydrogen bond, which leads to qualitatively different solvation structures. We conclude that this limitation of simple (yet commonly used) classical models occurs when solute-solvent interactions become comparable in strength to regular water hydrogen-bonds, which in turn leads to a disruption of the water-water hydrogen bond network near the solute. Although the exact magnitude of the differences between *ab initio* and classical MD simulations for the axial region may be influenced by subtle effects, such as possible inaccuracies in the description of dispersion interactions, the overall trends due to the disruption of the hydrogen bond network should persist. Based on these findings we expect that simple classical models should provide a qualitatively correct description of small solvated aliphatic hydrocarbons, but may not adequately represent small solutes that involve significant solute-solvent interactions, such as the aromatic groups that are abundant in biochemistry.

The authors would like to thank Giancarlo Cicero and Edmond Y. Lau for helpful discussions. This work was performed under the auspices of the U.S. Dept. of Energy at the University of California/Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48 and was supported in part by DOE grant DE-FG02-04ER25625.
UCRL-JRNL-234592

---

[1] H. A. Scheraga, M. Khalili, and A. Liwo, Annual Review of Physical Chemistry **58**, 57 (2007).

[2] A. Wallqvist, E. Gallicchio, and R. Levy, J. Phys. Chem. B **105**, 6745 (2001).

[3] M. Allesch, E. Schwegler, and G. Galli, J. Phys. Chem. B **111**, 1081 (2007).

[4] G. Ravishanker, P. K. Mehrotra, M. Mezei, and D. L. Beveridge, J. Am. Chem. Soc. **106**, 4102 (1984).

[5] P. Linse, G. Karlström, and B. Jönsson, J. Am. Chem. Soc. **106**, 4096 (1984).

[6] P. Linse, J. Am. Chem. Soc. **112**, 1744 (1990).

[7] W. L. Jorgensen and D. L. Severance, J. Am. Chem. Soc. **112**, 4768 (1990).

[8] S. Urahata, K. Coutinho, and S. Canuto, Chem. Phys. Lett. **274**, 269 (1997).

[9] S. Urahata and S. Canuto, Chem. Phys. Lett. **313**, 235 (1999).

[10] A. Laaksonen, P. Stilbs, and R. E. Wasylishen, J. Chem. Phys. **108**, 455 (1998).

[11] T. M. Raschke and M. Levitt, J. Phys. Chem. B **108**, 13492 (2004).

[12] T. M. Raschke and M. Levitt, PNAS **102**, 6777 (2005).

[13] P. Schravendijk and N. F. A. van der Vegt, J. Chem. Theory Comput. **1**, 643 (2005).

[14] S. Ikawa, J. Chem. Phys. **123**, 244507 (2005).

[15] S. Suzuki *et al.*, Science **257**, 942 (1992).

[16] J. C. Grossman, E. Schwegler, and G. Galli, J. Phys. Chem. B **108**, 15865 (2004).

[17] F. Gygi, *GP 1.24.0: A general ab initio molecular dynamics program*, Lawrence Livermore National Laboratory, 2003.

[18] F. Gygi, *The Qbox code*, Lawrence Livermore National Laboratory, 2005.

[19] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).

[20] W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).

[21] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **80**, 891 (1998).

[22] D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).

[23] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[24] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[25] G. Galli and M. Parrinello, in *Computer Simulation in Materials Science*, edited by M. Meyer and V. Pontikis (Kluwer Academic Publishers, Netherlands, 1991), pp. 283–304.

[26] J. C. Grossman *et al.*, J. Chem. Phys. **120**, 300 (2004).

[27] E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli, J. Chem. Phys. **121**, 5400 (2004).

[28] M. Allesch, E. Schwegler, F. Gygi, and G. Galli, J. Chem. Phys. **120**, 5192 (2004).

[29] Y. Danten, T. Tassaing, and M. Besnard, J. Phys. Chem. A **103**, 3530 (1999).

[30] F. C. Lightstone *et al.*, ChemPhysChem **6**, 1745 (2005).

[31] K. Leung and S. B. Rempe, Phys. Chem. Chem. Phys. **8**, 2153 (2006).

[32] C. J. Gruenloh *et al.*, Science **276**, 1678 (1997).

[33] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Mod **7**, 306 (2001).

[34] G. Kaminski, F. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, J. Phys. Chem. B **105**, 6474 (2001).

[35] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[36] W. L. Jorgensen *et al.*, J. Chem. Phys. **79**, 926 (1983).

[37] W. L. Jorgensen and J. D. Madura, Mol. Phys. **56**, 1381 (1985).

[38] W. L. Jorgensen, J. Am. Chem. Soc. **110**, 1657 (1988).

[39] W. R. P. Scott *et al.*, J. Phys. Chem. A **103**, 3596 (1999).

[40] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. V. Gunsteren, J. Comput. Chem. **25**, 1656 (2004).

[41] U. Essmann *et al.*, J. Chem. Phys. **103**, 8577 (1995).

[42] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[43] H. J. C. Berendsen *et al.*, J. Chem. Phys. **81**, 3684 (1984).

[44] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).

[45] M. Levitt *et al.*, J. Phys. Chem. B **101**, 5051 (1997).

[46] W. F. van Gunsteren and M. Karplus, Macromolecules **15**, 1528 (1982).

[47] D. Chandler, Nature **417**, 491 (2002).

[48] D. Chandler, Nature **437**, 640 (2005).

[49] J.White, E. Schwegler, G. Galli, and F. Gygi, J. Chem. Phys. **113**, 4668 (2000).

[50] F. Gygi *et al.*, in *SC '05: Proceedings of the 2005 ACM/IEEE conference on Supercomputing* (IEEE Computer Society, Washington, DC, USA, 2005), p. 24.

List of Figures

1 The hydrophobic solvation structure of benzene in the axial and equatorial region of simulation B, illustrating the spheroidal shape of the solvation shell (isosurface 1, on the left). The right graph shows the cage-like structure in the equatorial region of the first shell and the high concentration on the faces (isosurface 2.6). The SDF of oxygen is colored in red, carbon atoms in grey. Hydrogen atoms belonging to the benzene molecule are not shown. Adapted from Ref. 3. 20

2 RDFs for the equatorial region obtained from *ab initio* simulations using flexible and rigid water models (simulations A and B). B-O denotes the RDF between the benzene (center) and oxygen, B-H refers to the benzene (center)-hydrogen RDF. 21

3 RDFs for the axial region obtained from *ab initio* simulations using rigid and flexible water models (simulations A and B). Due to the narrow definition of the axial region this RDF mostly represents the weak π-H bond between one water molecule and benzene. 22

4 Tilt angle distribution of water molecules up to a distance of 6.625 Å from the benzene center (simulations A, B and C). Results from the flexible *ab initio* model are shown on the left, rigid *ab initio* model in the center and classical results on the right. 23

5 The hydrophobic solvation structure of benzene in the axial and equatorial region (90° sections; isosurface 1) from simulation B. The SDF of oxygen is colored in red, hydrogen (if different) in green, carbon atoms in grey. Hydrogen atoms belonging to the benzene molecule are not shown. 24

6 RDFs for the axial region obtained from classical MD and *ab initio* simulations using rigid and flexible water models (simulations A, F and G). 25

7 RDFs for the equatorial region obtained from classical MD and *ab initio* simulations using rigid and flexible water models (simulations A, B and C). . 26

8 Calculated dipole moments of the binding water in the axial region around benzene[3], in the first shell of the equatorial region of simulation B and for bulk water[28]. 27

9 Comparison of g(r) peak heights for Benzene-Oxygen and Benzene-Hydrogen RDFs obtained from two different system sizes, 73 and 512 solvent molecules, in a NVT ensemble (simulations C, D). The dominant peaks come from water molecules forming a weak H-bond with the delocalized electron system of benzene. The distance between the first B-O and B-H peaks (0.94 Å) almost exactly corresponds to the O-H distance of the (rigid) TIP4P water model (0.9572 Å). 28

10 Estimation of the statistical error bar for B-O and B-H g(r)s in the axial region from ten 100 ps classical MD simulations with 73 solvent molecules. . 29

11 Comparison between RDFs from the two different ensembles, NVT and NPT (at the same average thermodynamic conditions), with 512 solvent molecules (simulations D, F). 30
12 Axial RDFs obtained from simulation F compared with those from Ref. 11 using the same force fields (TIP4P, OPLS-AA). 31
13 Axial RDFs obtained from simulation D compared with those from Ref. 13. For this graph the axial region was defined as a 90° cone perpendicular to the plane of the molecule. 32

Figures

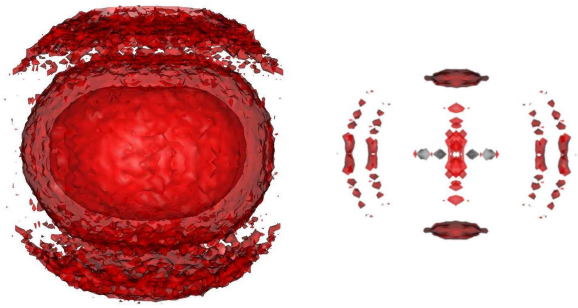

FIG. 1:

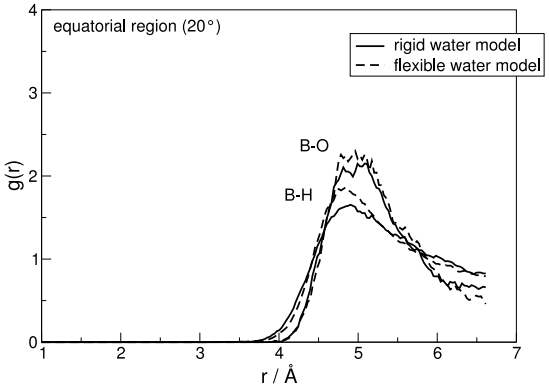
equatorial region (20°)
rigid water model
flexible water model
B-O
B-H
g(r)
r / Å
0
1
2
3
4
1
2
3
4
5
6
7

FIG. 2:

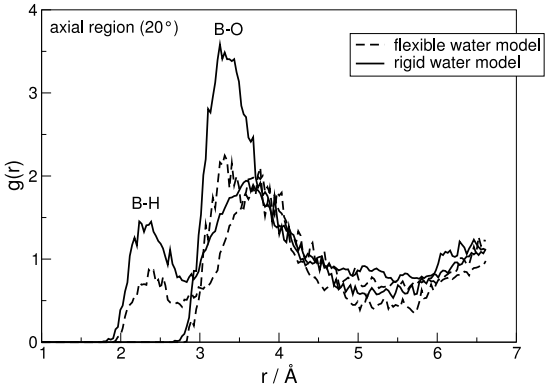
axial region (20°)
B-O
B-H
flexible water model
rigid water model
g(r)
r / Å
0
1
2
3
4
1
2
3
4
5
6
7

FIG. 3:

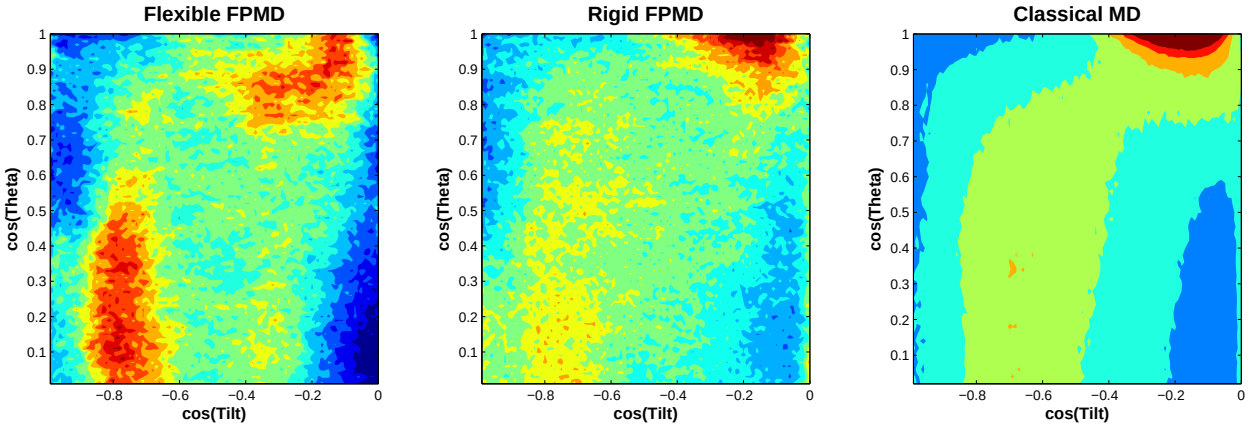
Flexible FPMD
Rigid FPMD
Classical MD
cos(Theta)
cos(Tilt)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
−0.8
−0.6
−0.4
−0.2
0

FIG. 4:

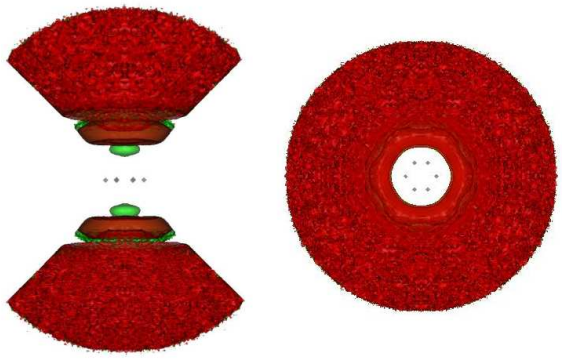

FIG. 5:

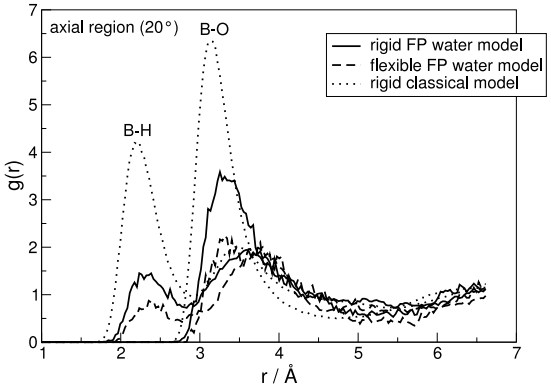
axial region (20°)
B-O
B-H
rigid FP water model
flexible FP water model
rigid classical model
g(r)
r / Å
0
1
2
3
4
5
6
7
1
2
3
4
5
6
7

FIG. 6:

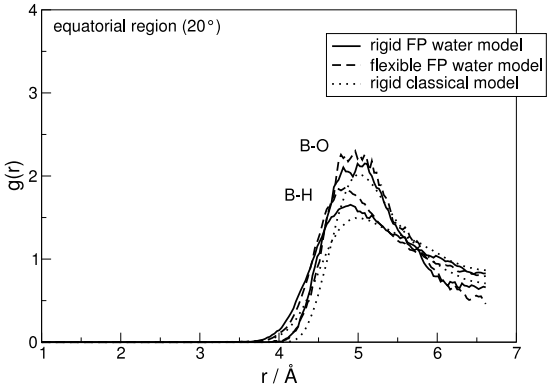
equatorial region (20°)
rigid FP water model
flexible FP water model
rigid classical model
B-O
B-H
g(r)
r / Å
0
1
2
3
4
1
2
3
4
5
6
7

FIG. 7:

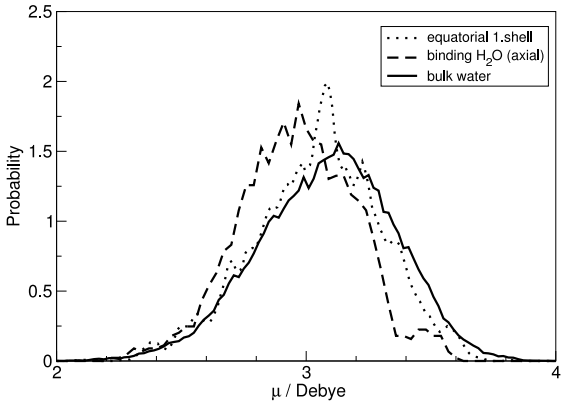
equatorial 1.shell
binding H_2O (axial)
bulk water
Probability
μ / Debye
2.5
2
1.5
1
0.5
0
2
3
4

FIG. 8:

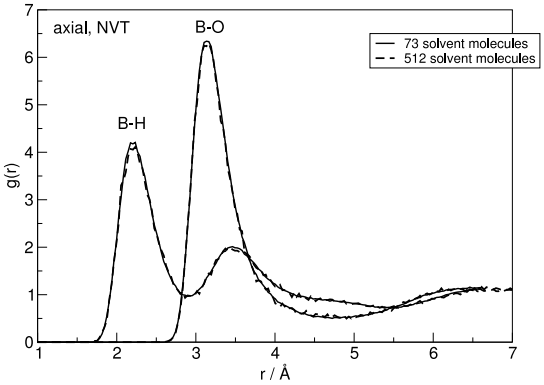
axial, NVT
B-O
B-H
73 solvent molecules
512 solvent molecules
g(r)
r / Å
0
1
2
3
4
5
6
7
1
2
3
4
5
6
7

FIG. 9:

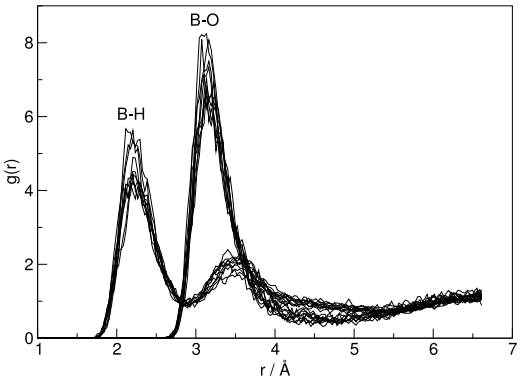

FIG. 10:

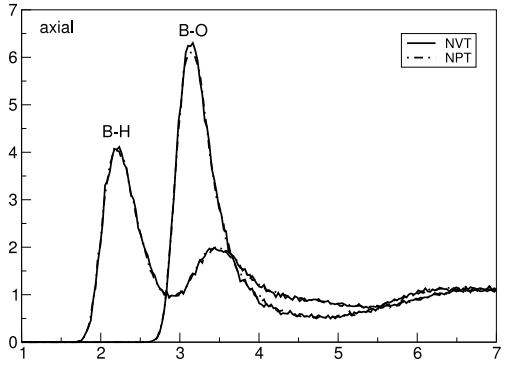

FIG. 11:

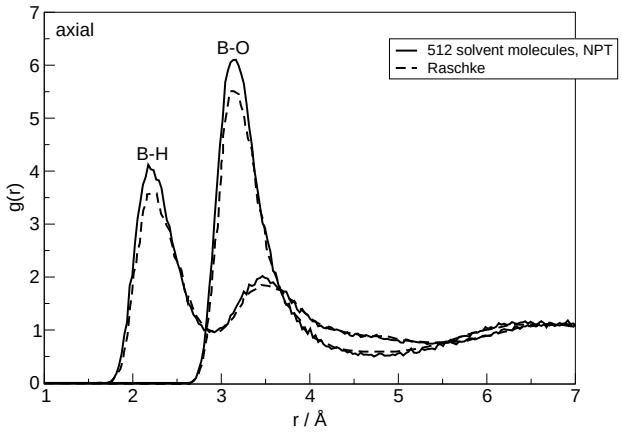
axial
B-O
B-H
512 solvent molecules, NPT
Raschke
g(r)
r / Å
0
1
2
3
4
5
6
7
1
2
3
4
5
6
7

FIG. 12:

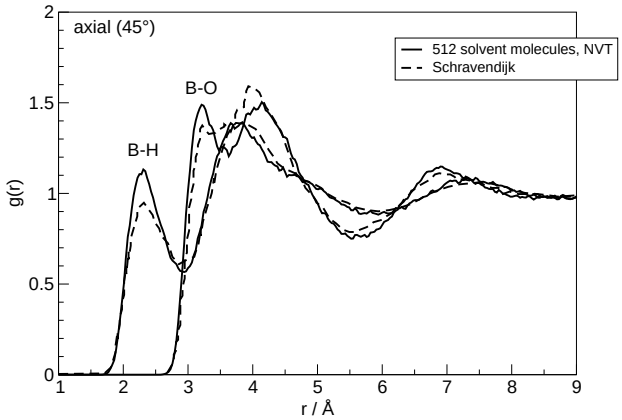
axial (45°)
512 solvent molecules, NVT
Schravendijk
B-O
B-H
g(r)
r / Å
0
0.5
1
1.5
2
1
2
3
4
5
6
7
8
9

FIG. 13:

TABLE I: Overview of the MD simulations: Size refers to the number of water molecules included in the simulation cell. In our rigid FP simulation a rigid water geometry was enforced and the carbon-hydrogen bonds of benzene were constrained to eliminate fast vibrational modes in the system.

Name	Method	Geometry	size	T (K)	dt (fs)	μ (a.u.)	time (ns)
A	FP	Flexible	73	297.5 (NVE)	0.07	340	0.081
B	FP	Rigid	73	303.3 (NVE)	0.24	1100	0.152
C	Classical	Rigid	73	(NVT)	2	-	100
D	Classical	Rigid	512	(NVT)	2	-	10
E	Classical	Rigid	73	(NPT)	2	-	30
F	Classical	Rigid	512	(NPT)	2	-	10

TABLE II: Empirical nonbonded interaction parameters for the benzene-water interactions used in classical MD simulations.

Force Field	$q^C(e)$	$q^H(e)$	σ^C (Å)	ϵ^C (kcal/mol)	$\sigma^H(\mathring{A})$	ϵ^H (kcal/mol)
OPLS-AA [a]	-0.115	0.115	3.5500	0.07	2.4200	0.03
GROMOS 43A1	-0.100	0.100	3.3611	0.09697	2.3734	0.02828

[a]OPLS-AA parameters for benzene were originally taken from Ref.[7]. The GROMOS43A1 force fields was used in Ref.[13], OPLS-AA was also used in Ref.[11].