Published in final edited form as:
J Phys Chem B. 2011 January 20; 115(2): 319–328. doi:10.1021/jp102587q.

Calculation of Local Water Densities in Biological Systems — A Comparison of Molecular Dynamics Simulations and the 3D-RISM-KH Molecular Theory of Solvation

Martin C. Stumpe[*], **Nikolay Blinov**[†], **David Wishart**[‡], **Andriy Kovalenko**[§], and **Vijay S. Pande**[¶]

[*]Departments of Bioengineering and Chemistry, Stanford University, 318 Campus Drive West, Stanford, California 94305. martin.stumpe@stanford.edu [†]Department of Mechanical Engineering, University of Alberta, and National Institute for Nanotechnology, National Research Council of Canada, 11421 Saskatchewan Drive, Edmonton, Alberta T6G 2M9, Canada. nblinov@ualberta.ca [‡]Departments of Computing Science and Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E8, and National Institute for Nanotechnology, National Research Council of Canada, 11421 Saskatchewan Drive, Edmonton, Alberta T6G 2M9, Canada. david.wishart@ualberta.ca [§]National Institute for Nanotechnology, National Research Council of Canada, and Department of Mechanical Engineering, University of Alberta, 11421 Saskatchewan Drive, Edmonton, Alberta T6G 2M9, Canada. andriy.kovalenko@nrc-cnrc.gc.ca [¶]Departments of Chemistry, Computer Science, and Structural Biology, Stanford University, 318 Campus Drive West, Stanford, California 94305. pande@stanford.edu

Abstract

Water plays a unique role in all living organisms. Not only is it nature's ubiquitous solvent, but it also actively takes part in many cellular processes. In particular, the structure and properties of interfacial water near biomolecules like proteins are often related to the function of the respective molecule. It can therefore be highly instructive to study the local water density around solutes in cellular systems, particularly when solvent-mediated forces like the hydrophobic effect are relevant. Computational methods like molecular dynamics (MD) simulations seem well suited to study these systems at the atomic level. However, due to sampling requirements, it is not clear that MD simulations are indeed the method of choice to obtain converged densities at a given level of precision. We here compare the calculation of local water densities with two different methods, MD simulations and the three-dimensional reference interaction site model with the Kovalenko-Hirata closure (3D-RISM-KH). In particular, we investigate the convergence of the local water density to assess the required simulation times for different levels of resolution. Moreover, we provide a quantitative comparison of the densities calculated with MD and with 3D-RISM-KH, and investigate the effect of the choice of the water model for both methods. Our results show that 3D-RISM-KH yields density distributions that are very similar to those from MD up to a 0.5 Å resolution, but for significantly reduced computational cost. The combined use of MD and 3D-RISM-KH emerges as an auspicious perspective for efficient solvent sampling in dynamical systems.

Keywords

interfacial water; solvation; confined water; hydrophobic effect; MD; box size

1 Introduction

Water plays a central role in virtually all cellular systems. This role is not only limited to providing a passive solvent environment for proteins, DNA, and other biomolecules, but water also takes an active part in many processes. One of the most important properties of water is that it exerts the hydrophobic effect, which stabilizes cell membranes and micelles [1-3], and constitutes the main driving force for protein folding [4]. This ubiquitous role of water, as well as its peculiar physicochemical properties [5,6] make it a topic of ongoing research since decades, with experimental [7] as well as computational and theoretical [8-10] approaches.

Due to the remarkable complexity of aqueous systems, one still falls short of fully understanding or even characterizing water in these systems. This is also reflected in the variety of different water models commonly used in molecular dynamics (MD) simulations, none of which are yet able to fully describe all physicochemical properties of water correctly, although progress can be observed in the development of new models [11,12].

A significant portion of the current research on water addresses its function and properties when in contact with macromolecules and other surfaces, and how this interfacial water differs from bulk water [9,10,13,14]. Recent studies have suggested that the specific properties of interfacial water in systems like the ribosomal exit tunnel [15] or the GroEL/ES protein folding machinery [16,17] might be the determinant factor which allow these systems to fulfill their specific physiological functions.

To advance our understanding of these and other systems where interfacial water plays a significant role, it is crucial to have an accurate description of the local water properties. In particular, due to different effects like solvent-solute hydrogen bonding on the one hand, or dewetting [18,19] on the other hand, the local water density around solutes can differ significantly [9,10], and it hence can be used as a signature for the effect a solute exerts on its solvent environment.

Molecular dynamics simulations are very well suited to provide insight into structure and dynamics with atomistic detail, and have successfully demonstrated their capability to elucidate biological processes in many prominent examples. However, MD simulations of systems of the typical size of protein complexes are computationally quite costly, and in many cases it is not obvious that the results do not suffer from limited sampling.

Another computational approach to calculate the distribution of solvent molecules like water around a solute is the reference interaction site model (RISM) [20]. While its application was originally limited to simple systems like molecular liquids [21] and polar liquids [22] (with extensions to di- or quadrupolar liquids [23]), recent enhancements to the 3D-RISM theory [24-27] have rendered it capable to successfully describe even complex biological systems like solvation water around a protein [28]. Other recent successful applications include the calculation of free energies upon protein hydration [29,30] and folding [31], the experimentally verified prediction of ion binding sites at a protein [32,33], and an investigation of urea-water mixtures [34] in very good agreement with previous results from MD simulations [35].

MD and RISM are conceptually very different. While MD is based on the time-integration of Newton's equation of motion, RISM is an integral equation theory that is based on the Ornstein-Zernicke equation [36]

$$h(r_{12})=c(r_{12})+\int dr_3 c(r_{13})\rho h(r_{32}), \quad (1)$$

which presents the total correlation function $h(r_{12})$ between two particles as contributions from a direct correlation $c(r_{12})$ and from indirect contributions of $c(r_{12})$ propagating through all possible chains of mediating particles which sum up to $\int dr_3 c(r_{13})\rho h(r_{32})$ (see Ref. [37] for a more detailed description of the RISM concept). Consequently, and in contrast to MD, RISM does not provide any information about the solvent dynamics of a system, but yields the equilibrium statistical mechanical distribution of solvent particles around a given solute. It should be noted that there has been a non-equilibrium statistical mechanical theory, the generalized Langevin equation in the interaction site model formalism (using the solutions from the equilibrium RISM theory as initial conditions in the time evolution), that does provide full statistical microscopic information about solvent dynamics [38-40]. The conformational degrees of freedom of a solute molecule can be accounted for in the hybrid approaches coupling RISM with Monte Carlo (MC) [41], or 3D-RISM with MD [42,43]. RISM was combined with the multicanonical and simulated annealing MC methods [44,45] for predicting conformations of biomolecules stabilizing by solvent [41]. Recently, the three-dimensional version of RISM approach (3D-RISM) has been combined with MD in the Amber molecular dynamics package to simulate dynamics of a macromolecule with the adiabatic description of solvent degrees of freedom [43].

For the calculation of an equilibrated solvent distribution around a given biomolecular solute, 3D-RISM is orders of magnitude faster than MD. The solvent distribution can be found by using the 3D-RISM method with a precision defined by the grid spacing and the time allocated for the self-consitent iterative solution of the 3D-RISM equations. Thus, in some cases, 3D-RISM can be efficiently used to address questions currently not tractable with MD.

However, it is unclear how the results from MD and 3D-RISM match each other. Can one method be used instead of, or in conjunction with the other? Additionally, it is not clear how the local water densities are affected by the choice of the water model that is used with these methods. We here address these questions by comparing the calculation of local water densities with these two methods. Since 3D-RISM and MD can use the same Hamiltonian (force field), this comparison directly addresses differences between the two methods. Further, we investigate the effect of the choice of the water model for both methods by calculating the water density using two different force fields with each method. In particular, we address the following questions:

1. How long does the convergence of local water densities in MD take on different levels of resolution?
2. How strongly do the local water densities depend on the choice of the water model?
3. How similar are local water densities calculated by MD and by 3D-RISM?

To that end, we performed MD simulations and 3D-RISM-KH calculations of precisely the same system setup and compared the results. As solute, we chose the GroEL/GroES chaperonin complex, which is a representative biologically relevant molecule where interfacial water plays a crucial role for its function [17].

GroEL/ES is a chaperonin system, which facilitates the correct folding of substrate proteins *in vivo* [46]. GroEL consists of two identical rings (*cis* and *trans*) which are stacked back-to-back to each other. Each of these rings is a heptamer of seven identical monomers, which are each 547 amino acids long. Folding of substrate proteins takes place within the hydrophobic cavity of each ring, while the smaller heptameric GroES functions as lid for the cavity. Recent studies have suggested that interfacial water inside the cavity plays a central role for substrate folding by exerting a stronger hydrophobic effect on the substrate than

bulk water [16]. For more information on the GroEL/ES system and theories on its mechanism, see Refs. 16,47,48.

The water models we compare are the relatively new TIP4Pew as a very accurate water model [11], and TIP3P [49] and SPC/E [50] as more established and computationally less expensive models. In addition, we investigate whether the results depend on the box size, and thus whether the methods compared here have different levels of agreement for interfacial water as for bulk water. Moreover, we address the question how well local features in the local density distributions of the two methods agree, in particular the solvation shell densities around specific residue types and the locations of high density water locations which may correspond to trapped or bound water molecules.

2 Methods

System setup

The GroEL/ES wildtype crystal structure 1PCQ [46] of the *cis* cavity was solvated in a (20 nm)^3 simulation box. K^+ and Cl^- ions were added to neutralize the system at a physiological ion concentration. The total number of atoms in the simulation box was ≈ 1, 000, 000. After initial energy minimization was performed using steepest descent, an MD equilibration phase of 850 ps was done with harmonic restraints on the Cα atoms with pressure coupling (NpT ensemble) to allow for side chain relaxation and pressure equilibration. The last frame of this equilibration phase was used for the actual MD and 3D-RISM calculations.

Figure 1A shows a snapshot of the simulation box together with an illustration of local water densities in slices through the box. The different levels of resolution (4 Å, 2 Å, 1 Å, 0.5 Å) investigated here are displayed in panel B.

Molecular dynamics simulations

All MD simulations were performed using Gromacs 3.3.1 [51,52] with the Amber 2003 port for Gromacs [53]. TIP4Pew [11] and TIP3P [49] were used as water models. Particle Mesh Ewald summation (PME) was used to calculate the long-range electrostatic interactions with a fourier grid size of 160 in each direction, and an interpolation order of 4. A cutoff of 1.0 nm was used for the short-range Coulomb and the Lennard-Jones interactions. The simulation temperature was fixed at 298 K using Berendsen-type temperature-coupling [54] with a coupling coefficient of τ_T = 0.5 ps. After the initial volume equilibration phase in NpT, production runs were performed with constant volume condition (NVT ensemble) in order to have precisely the same system setup as for the 3D-RISM calculations. The integration timestep was set to 2 fs. The positions of all protein atoms and ions were held completely fixed using freeze groups. The total simulation times were 20 ns with TIP4Pew, 20 ns with TIP3P, and an additional 10 ns for additional tests addressing ion mobility and side chain flexibility (see below).

3D-RISM calculations

The solvent structure around a solute molecule can be characterized in terms of the three-dimensional (3D) distribution functions defined for each site of a molecular solvent. These functions describe an excess or depletion of the solvent density with respect to its bulk value. The distribution functions can be calculated in explicit solvent molecular dynamics simulations. Alternatively, they can be obtained based on the statistical mechanical molecular theory of solvation also known as the 3D reference interaction site model (3D-RISM). Additionally to the solvent distribution functions $g_\gamma(\mathbf{r})$ for every solvent site γ, closely related total correlation functions $h_\gamma(\mathbf{r}) = g_\gamma(\mathbf{r}) - 1$, the theory introduces the 3D

direct correlation functions, $c_\gamma(\mathbf{r})$. The correlation functions can be found from the 3D-RISM integral equation

$$h_\gamma(\mathbf{r}) = \sum_\alpha \int d\mathbf{r}' c_\alpha(\mathbf{r} - \mathbf{r}') \chi_{\alpha\gamma}(r'), \quad (2)$$

complemented by a so-called closure relation. The exact closure is a diagrammatic series in terms of multiple integrals of combinations of the total correlation functions. It is extremely cumbersome, and in practice is replaced with amenable approximations.

The 3D-RISM integral equation can be derived from the six-dimensional molecular Ornstein-Zernike integral equation [55] by averaging out the orientation degrees of freedom of solvent molecules while keeping the full 3D description of the orientation of the solute molecule [56,57]. In this equation, the site-site susceptibility of pure solvent, $\chi_{a\gamma}(r)$, describes the correlations between solvent sites a and γ at the site-site separation r. The susceptibility is given by the equation $\chi_{a\gamma}(r) = \omega_{a\gamma}(r) + \rho_a h_{a\gamma}(r)$, where the intramolecular matrix $\omega_{\alpha\gamma}(r) = \delta\left(r - l_{\alpha\gamma}\right) / \left(4\pi l_{\alpha\gamma}^2\right)$ specifies the intramolecular correlations of solvent molecules with the geometry given by the matrix of site separations $l_{a\gamma}$, and the intermolecular part is given by the site-site radial correlation functions of the bulk solvent, $h_{a\gamma}(r)$, times the bulk solvent site number density ρ_a.

The 3D-RISM theory with the hypernetted chain (HNC) closure was conceptually sketched by Chandler and co-workers in their derivation of the density functional theory for classical site distributions of molecular liquids [58,59] and then introduced in that way by Beglov and Roux for polar liquids [25]. The 3D-HNC closure to the 3D-RISM equation (2) is given by $g_\gamma(\mathbf{r}) = \exp(-\beta u_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r}))$. Here $u_\gamma(\mathbf{r})$ is the interaction potential between solvent site γ and the solute, $\beta = 1/k_B T$ is the inverse temperature with the Boltzmann constant k_B. Because of the drawback of the HNC approximation overestimating the effect of attractive interaction potentials, for a relatively strong attractive potential between a solute molecule with multiple partial charges and water solvent (in particular, with ions), the 3D-RISM-HNC equations may have no solutions [56, 57, 60, 61]. Kovalenko and Hirata proposed the closure approximation (KH closure)

$$g_\gamma(\mathbf{r}) = \begin{cases} \exp\left(\eta_\gamma(\mathbf{r})\right) & \text{for} \quad \eta_\gamma(\mathbf{r}) \le 0 \\ 1 + \eta_\gamma(\mathbf{r}) & \text{for} \quad \eta_\gamma(\mathbf{r}) > 0 \end{cases}, \quad (3)$$
$$\eta_\gamma(\mathbf{r}) = -\beta u_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})$$

that automatically applies the 3D-HNC treatment to repulsive cores and other regions of density depletion ($g_\gamma(\mathbf{r}) < 1$) due to repulsive interactions and steric constraints, and the 3D mean spherical approximation (MSA), $c_\gamma(\mathbf{r}) = -\beta u_\gamma(\mathbf{r})$, to the distribution peaks due to associative forces and other density enhancements with $g_\gamma(\mathbf{r}) > 1$, including long-range distribution tales for structural and phase transitions in fluids and mixtures. The 3D-KH approximation (3) enforces the proper long-range asymptotics of the direct correlation function $c_\gamma(\mathbf{r})$ in the same way as the original HNC and MSA closures do. The MSA-type linearization prevents the artifact of the distribution function diverging due to the regions of a large interaction potential. This partial linearization somewhat reduces and widens high peaks of the distribution functions, whereas it much less affects the coordination numbers of the solvation shells.

In the current study, we used the 3D-RISM integral equation (2) with the KH closure (3D-RISM-KH theory) to calculate the three-dimensional water distributions (for oxygen and hydrogen sites) around the GroEL protein. Before 3D-RISM calculations, the bulk solvent susceptibility $\chi_{a\gamma}(r)$ was obtained from the dielectrically consistent RISM theory (DRISM)

developed by Perkyns and Pettitt [62]. This approach provides a dielectrically consistent description of the dielectric properties of ions in polar solvent. The combination of 3D-RISM-KH and DRISM approaches (referred below for brevity as 3D-RISM) ensures proper account for the short-range solvation structure effects, such as chemical specificities of solute and solvent molecules, hydrogen bonding, hydrophobicity, steric effects, nanoconfinement, etc. [56, 63-71].

The 3D-RISM-KH equations were solved on a 512 × 400 × 400 grid in a 512.0 × 400.0 × 400.0 $Å^3$ rectangular box, of which the central 400.0 × 400.0 × 400.0 $Å^3$ were used for all density comparisons. The Amber 2003 force field parameters [53] were used for the solute-solvent interaction potential. For water, the SPC/E [50] and TIP3P [49] models were employed with the van der Waals radius of the water hydrogen set to 0.4 Å to properly describe the thermodynamic and structural properties of bulk water. All calculations were performed at the temperature T = 298.15 K with the water number density of ρ = 0.997 g/cm^3 and the dielectric constant ε = 78.4. The convolution in Eq. (2) was handled by using the 3D fast Fourier transform (3D-FFT) technique. The solute-solvent electrostatic potential was calculated with the Particle-Mesh Ewald method. To ensure proper treatment of electrostatics (otherwise causing large errors in the solvation structure and thermodynamics), the long-range electrostatic parts of the direct and total correlation functions (including the limit of $\mathbf{k} \rightarrow 0$) are separated out and handled analytically in the calculation of the Fourier transforms and convolutions, for both the radial solvent-solvent and 3D solute-solvent correlations [56,60,61].

Correlation and convergence analysis

Water densities $\rho = \rho(x, y, z)$ are calculated on a three-dimensional grid and compared to each other by calculating a correlation coefficient

$$R(\rho_i, \rho_j) = \frac{C(\rho_i, \rho_j)}{\sqrt{C(\rho_i, \rho_i) C(\rho_j, \rho_j)}}, \tag{4}$$

where $C(\rho_i, \rho_j)$ is the covariance between density sets ρ_i and ρ_j:

$$C(\rho_i, \rho_j) = \sum_{x,y,z} (\langle \rho_i \rangle - \rho_i(x, y, z)) \cdot (\langle \rho_j \rangle - \rho_j(x, y, z)). \tag{5}$$

Using this definition, R ranges between 0 (totally uncorrelated data) and 1 (completely correlated data). All analysis were done using the coordinates of the oxygen atom.

This correlation analysis was also utilized to assess the convergence of the local water density in the MD simulations. To that aim, we employed a calculation similar to mutual information. A water density distribution $\rho_1(T, x, y, z)$ was calculated for a given simulation length T. A second water density distribution $\rho_2(T, x, y, z)$ was then calculated from another, independent, simulation with the same length T, and the correlation coefficient $R = R(T)$ calculated between these two density distributions. This procedure was done for multiple simulation lengths T, yielding a time dependent correlation coefficient (see Fig. 2 for illustration). Longer sampling times T will lead to better convergence, in particular:

$$R(T_1) > R(T_2) \quad for \quad T_1 > T_2, \tag{6}$$

and

$$\lim_{T\to\infty} R(T)=1. \tag{7}$$

Hence, $R(T)$ can be used as a quantitative measure how converged the local water density distributions are.

The convergence time is calculated by treating density convergence (i.e., phase space sampling) as a kinetic process with a relaxation constant τ. A single-exponential fit

$$R(T)=1 - a\exp(T/\tau) \tag{8}$$

seemed to underestimate the convergence time for larger T (see Fig. 2). We therefore used the slower relaxation constant τ_2 from a bi-exponential fit

$$R(T)=1 - a_1\exp(T/\tau_1) - a_2\exp(T/\tau_2) \tag{9}$$

to estimate the convergence time.

3 Results and Discussion

Distribution of water densities

We calculated the local water density in the GroEL simulation box using four different grid spacings of 4 Å, 2 Å, 1 Å, 0.5 Å. Figure 3 shows a histogram of local water densities calculated on the 0.5 Å grid. The distribution can be fit with a gaussian $h(\rho)=h_0 \cdot e^{-\left(\frac{\rho-\rho_0}{2\sigma}\right)^2}$ around a midpoint of $\rho_0 = 0.9779 g/cm^3$ and a width of $\sigma = 0.1632\ g/cm^3$ (goodness of fit $r^2 = 0.9989$).

Some water molecules were found to persist at the same position all the time (i.e., bound and crystallographic water). For very fine grid spacings (e.g., 0.5 Å), these water molecules can give rise to very high peaks in the local density. Similarly, the volume occupied by the chaperonin or by the fixed ions always has a water density of zero (see Fig. 3). Although these systematic effects of the grid-based approach of calculating densities would not change the correlation analysis qualitatively, they would decrease the contrast and partially occlude the correlation of the actually relevant water densities around $\approx 1.0 g/cm^3$ (see Fig. 3). Therefore, we excluded these very low or very high densities from the correlation analysis and limited the range to [$0.3\ g/cm^3 \le \rho \le 2\ g/cm^3$]. Note that the choice of these thresholds is, within certain limits, arbitrary and affects the correlation coefficients only marginally.

Convergence times of water densities

Figure 4 shows the convergence of local water densities in dependence on sampling time, for four different levels of resolution (grid spacings). Using the bi-exponential fit (Eq. 9), the convergence times τ of local water densities are: 0.86±0.24 ns (4 Å grid), 5.3±0.6 ns (2 Å grid), 14±1 ns (1 Å grid), and 59±3 ns (0.5 Å grid).

Effect of the water model in MD

To investigate the effect of the water model on the local densities and the convergence times, we have performed a similar analysis for the simulations with the TIP3P water model. The convergence times for TIP3P on the different levels of resolution were similar to those found for TIP4Pew (data not shown).

In order to obtain a quantitative measure for the similarity, we calculated the time-averaged water density distribution $\rho(x, y, z)$ from 20 ns simulation time with each of the two water models, TIP4Pew and TIP3P, and calculated the correlation coefficient between both density distributions (see Eq. 4). To compare the water models independent of potentially incomplete sampling, we normalize the correlation coefficient of the density distributions from TIP4Pew and TIP3P by that of the density distributions between two different sets of TIP4Pew simulations:

$$rel.corr. = \frac{R(\rho_{20\ ns}^{TIP4Pew}, \rho_{20\ ns}^{TIP3P})}{R(\rho_{20\ ns}^{TIP4Pew(1)}, \rho_{20\ ns}^{TIP4Pew(2)})} \quad (10)$$

This normalization is permitted because of the similar convergence times for both water models. When comparing density distributions between different force fields or calculations methods (e.g., MD versus 3D-RISM), the solvent excluded volume due to protein and ion atoms can be slightly different between two density distributions. Hence, there is no unambiguous definition which volume to exclude to reduce the trivial contributions resulting from (near) zero density. An additional advantage of this normalization done here is that these trivial contributions cancel out, effectively enhancing the contrast of the correlation analysis.

Figure 5 shows these relative correlations for the different grid spacings. In contrast to several properties which have been found to differ significantly between the two water models (e.g., surface tension, vapour-liquid equilibria, temperature dependence of the density, self-diffusion, [72]), the local water density distributions are remarkably similar. Even for the very fine grid spacing of 0.5 Å, a relative correlation of $R = 0.96$ is found. This observation is supported by the visual representation of the density distributions from the two water models in slices through the simulation box shown in Fig. 7. The high correlation coefficients would still allow for an offset or linear scaling factor between the density distributions for TIP4Pew and TIP3P. A gaussian-fit to the TIP3P density histogram shows that the distribution is centered around approximately the same midpoint of $\rho_0 = 0.9789\ g/cm^3$ (TIP4Pew: $\rho_0 = 0.9779\ g/cm^3$) However, it is marginally sharper with $\sigma = 0.1593 g/cm^3$ (TIP4Pew: $\sigma = 0.1632 g/cm^3$). The density histogram is shown in Fig. 6.

Comparison between MD and 3D-RISM

The comparison between water density distributions $\rho(x, y, z)$ obtained from MD and from 3D-RISM was performed similarly as the comparison between the TIP3P and TIP4Pew densities from MD. Since a four-point water model like TIP4Pew is not yet supported by the 3D-RISM package employed here, we used TIP3P water in both, MD and 3D-RISM, for a direct comparison of the two methods. The time-averaged MD water density $\rho^{MD}(x, y, z)$ was correlated with the water density $\rho^{3D-RISM}(x, y, z)$ obtained from the 3D-RISM calculation with a completely identical system setup. A normalization with the correlation between two different TIP3P water simulations was performed to account for incomplete sampling in the MD data as described above.

The relative correlation coefficients are shown in Fig. 8, and a visual display of representative slices through the simulation box is shown in Fig. 7 for the 0.5 Å grid spacing. As can be seen in Fig. 7, qualitative features like the spatial size, shape, and magnitude of most of the solvation shells around the solutes (protein and ions) are very similar between MD and 3D-RISM. Most markedly, however, the local density distributions obtained from 3D-RISM are much smoother than those from MD. This was to be expected, since the fluctuations in the MD give rise to noise in the average densities, whereas 3D-

RISM directly yields only the equilibrium distributions. This is also reflected in the water density histograms for 3D-RISM, which are substantially sharper than those for MD (see Fig. 6). Furthermore, minor differences in some of the details are also discernible (like the exact size of some solvation shells) at the fine grid spacing of 0.5 Å.

Effect of the water model in 3D-RISM

We also investigated the effect of the water model in the 3D-RISM calculations. To this end, the same calculations were performed with SPC/E instead of TIP3P, and the resulting density distributions compared. The density histograms for the two water models are very similar to each other (see Fig. 6). A small difference can be seen at densities around $\rho = 0.90 - 0.95$ g/cm^3, which are found slightly more frequent with TIP3P (green line) than with SPC/E (blue line). The correlation between the density distributions $\rho(x, y, z)$ obtained with the two water models is particularly high (see Fig. 9).

While our main aim here was to investigate the convergence times and similarity for one given solute conformation, we note that we have also performed two additional sets of shorter MD simulations to assess the effect of ion mobility and side chain flexibility. The first set of these simulations was performed with mobile ions. The results suggested that the influence of the ions on the average local water density and on the convergence times is not very significant, except in regions where ions have longer residence times, e.g., due to salt-bridges. A second set of simulations was performed with mobile ions and flexible side chains of the protein. The local water density in these simulations seemed to have a much longer convergence time. This result is not surprising, regarding that introducing side chain flexibility effectively convolutes the process of water density sampling for one given solute conformation (as investigated here) with the side chain conformational sampling process. A more detailed investigation of these issues might be addressed in future work.

Influence of the box size

The correlations and convergence times investigated so far were all calculated for the whole simulation box with a box size of 20 nm^3. However, a large portion of the box is occupied by bulk water and the question arises whether the correlation coefficients presented here are actually different for bulk water than for interfacial water. It is therefore essential to verify that the correlation coefficients presented here are not dominated by the bulk contributions, but instead are representative of the interfacial water. To that end, we restricted the volume of the analyses by shrinking the box length from 20 nm to 12 nm in each dimension (effectively removing most of the bulk water from the analysis, as shown in Fig. 10A+B), and investigated the dependence of the correlation coefficients and the convergence times on the box size.

When going from the full box size (20 nm^3) to the minimal box size (12 nm^3), an effect of the box size is seen on the convergence times (see Fig. 10C). In particular, the local water densities converge faster when reducing the box size and removing bulk water — which shows that the local densities converge slower for bulk water than for interfacial water.

However, virtually no effect is seen on the correlations coefficients describing the similarities between different method and force field combinations discussed above (see Fig. 10D). These results shows that the correlation coefficients between the two different methods (MD, 3D-RISM) and the three different force fields (TIP4Pew, TIP3P, SPC/E) depend only marginally on the box size, and that the agreement between the different methods and force fields is similar for bulk and interfacial water.

Solvation shell densities around particular residue types

So far, overall water densities calculated from MD and from 3D-RISM have been found to agree very well with each other. The question arises whether the two methods perform differently around different types of residues, e.g. in the solvation shells of polar or hydrophobic side chains. To address this question, we have calculated the water densities in the solvation shell around each individual amino acid type in GroEL and around the ions (K^+ and Cl^-). The density of all grid points (using the 1.0 Å grid with 200^3 points) within a given cutoff distance was calculated, where the distance was defined as the distance of each grid point to the closest heavy atom of all residues of the respective type. Note that this definition includes grid points which are in the excluded volume of the protein, and therefore the absolute values of the densities are lower than one would expect.

Figure 11 shows the densities for each residue type within a cutoff distance of 3 Å. As can be seen, the average densities in the solvation shell of each residue type are very similar for the four calculation methods MD:TIP4Pew (black bars), MD:TIP3P (red bars), 3D-RISM:TIP3P (green bars), and 3D-RISM:SPC/E (blue bars). In particular, an effect of the force field on the average densities is hardly visible for either MD or 3D-RISM. This observation suggests that, in contrast to other properties (like the diffusion coefficient or density temperature dependence), local water densities are relatively robust to changes in the force field – probably because the density at room temperature is one of the observables which is commonly and easily used for the parametrization of water force fields.

While the differences between the two force fields in either MD or 3D-RISM are very small, systematic differences between the densities calculated with MD and those calculated with 3D-RISM are visible. In particular, 3D-RISM seems to systematically underestimate the local water densities as compared to MD in the solvation shell with 3 Å radius, but no significant dependence on the residue types can be seen.

The average solvation shell densities for cutoff distances other than 3 Å (shown in Fig. 11) look very similar. This is also illustrated in Fig. 12A, which shows the dependence of the density correlations on the cutoff distance around the protein: the correlation is almost independent of the distance to the protein, which was also found when investigating the effect of the box size (see Fig. 10). However, although the correlation of the densities does not depend on the solvation shell size, we found that the degree to which 3D-RISM underestimates the local water densities as compared to MD does show a systematic dependence on the cutoff radius. As shown in Fig. 12B, the ratio of the average densities from MD and 3D-RISM is higher for smaller solvation shells. Or in other words, 3D-RISM underestimates the local water densities closer to the solute, whereas it levels out for densities above 10 Å. This small but systematic underestimation of water densities in the solvation shell is due to a partial linearization in the 3D-KH closure relation, which causes a small attenuation in the density distribution in areas with high density [56].

Location of density maxima

Aside from the water density in the solvation shells around particular residue types, another interesting question is how well the location of maxima in the water density distribution agrees between the different methods. Density maxima can correspond to trapped or bound water molecules, which may have a functional role in protein function [73], and have previously been compared between 3D-RISM calculations and X-ray scattering data [74]. Here, we quantitatively investigate to what degree the locations of density maxima from MD and from 3D-RISM agree with each other.

To that end, we have identified the top n % of grid points which have the highest water density for each method, and calculated which fraction of these top n % high density grid

points agree between the different methods or force fields. Figure 13 shows the percentage of high density grid points which agree between two data sets, calculated on the 1 Å grid. More than 80 % of the grid points which correspond to the top 1 % highest water densities agree in MD with using either TIP4Pew or TIP3P (Fig. 13A). When not only comparing the exact bin positions, but allowing a tolerance of ±1 or ±2 grid points, the overlap is almost complete (dashed and dash-dotted line, respectively). Between MD and 3D-RISM, both using TIP3P, about half of the top 1 % density bins share the same location (Fig. 13B), and about 25 % for the top 0.1 % density. Again, when allowing a tolerance of ±1 or ±2 grid points, the overlap between the two data sets is significantly higher and over 90 %. Finally, the agreement of the locations of top density bins for 3D-RISM using two different force fields (Fig. 13C) is very high (≈ 90 %), and almost complete when including a tolerance.

In summary, we find that the locations of high water density grid points, as potential locations for bound or trapped water molecules, are highly conserved between the force-fields when using either MD or 3D-RISM. When comparing MD and 3D-RISM, we find that the overlap of the exact bin positions is lower than when using the same method with different force-fields, and about half of the high density grid positions agree when considering the top 1 % density percentile. However, the agreement between the two methods increases significantly to about 90 % when allowing a tolerance of one or two grid points (i.e. Å) distance – which appears to be sufficiently accurate for practical applications to identify bound or trapped water molecules.

4 Summary and Conclusion

We have investigated and compared two methods to calculate local water densities: molecular dynamics simulations (MD), and the 3D reference interaction site model with the Kovalenko-Hirata closure relation (3D-RISM-KH). The main aim was to assess the applicability of 3D-RISM to be used instead of MD to study the solvent structure around solutes, with a particular focus on biological systems.

Using a correlation analysis similar to mutual information, the convergence of local water densities in MD simulations was found to be describable as a (bi-)exponential kinetic process with a time-constant τ. Depending on the desired level of resolution, the required simulation times for converged local water densities ranged from approximately 1 ns for a resolution of 4 Å to approximately 60 ns for a 0.5 Å resolution (for box sizes which have a typical separation between the solute and box boundaries). While the required sampling time on the coarse 4 Å resolution is easily obtainable even for large systems like the GroEL/ES protein complex investigated here, the computational effort to sufficiently sample the local water density on the finer resolutions quickly surmounts to magnitudes which are out of reach for most applications. For instance, the calculations presented here were performed on modern 3 GHz Quadro-core CPUs with about 100 ps per day. To sufficiently sample the local water densities on the 0.5 Å resolution, it would take 2000 cpu days – and this is just for one single fixed solute conformation. While a coarser resolution than 0.5 Å might be sufficient for many biophysical questions, the computational effort is still substantial even for more coarse resolutions, and can easily become out of reach again when trying to investigate whole ensembles of structures.

A comparison of TIP4Pew and TIP3P in the MD simulations, and of TIP3P and SPC/E in the 3D-RISM calculations, has shown that the choice of the water model does not significantly affect the local water densities, suggesting that results obtained with different models should be robust and comparable to each other.

The water densities calculated with 3D-RISM-KH show a remarkable degree of similarity to those from MD — in particular considering the completely different approach of the two methods. Qualitative features like the size, shape, and magnitude of the solvation shells around the protein and the ions were found to be very similar, with some differences in the details. The correlations between the density distributions from MD and from 3D-RISM-KH ranged from 0.98 for the 4 Å resolution to 0.53 for the 0.5 Å resolution. The computational effort for the calculations was orders of magnitude smaller than for MD — it took about 3 cpu-days to obtain convergence of the 3D-RISM-KH equations on the 0.5 Å grid resolution.

The degrees of similarity between MD and 3D-RISM-KH, as well as between the different water force fields, was found to be virtually constant when bulk water was removed from the analyses, showing that the agreement between them is similar for bulk water as for interfacial water. We also found that local densities of bulk water converge slower than those of interfacial water.

While the correlation between densities from MD and 3D-RISM-KH was also found to be extremely high in a residue type based solvation shell analysis, we found that 3D-RISM-KH has a tendency to slightly underestimate higher densities in the first few Angströms of the solvation shell while overestimating peak widths, due to a partial linearization in the 3D-KH closure relation used in the current study [56].

We found no significant dependence of the correlation between MD and 3D-RISM-KH water densities on the solute residue type. However, due to the heterogeneity of the chemical environment of each amino acid in a large system like this, GroEL might not be the best test system for this particular question. Future studies might address this issue by choosing different test systems (e.g. isolated amino acids or GXG tripeptides [75]).

The locations of high water density grid points, as potential locations for bound or trapped water molecules, are highly conserved between the force-fields when using either MD or 3D-RISM-KH. When comparing MD and 3D-RISM-KH, we found that the overlap of the exact bin positions is lower than the overlap observed when using the same method with different force-fields. As discussed above, due to the partial linearization in the 3D-KH closure the localization peaks obtained from 3D-RISM-KH are somewhat lower but wider than those from MD, resulting in a better agreement for the coordination number in a density peak (in a solvation shell). When considering the top 1 % of the highest density bins, about half of the corresponding density grid positions agree. However, the overlap between 3D-RISM-KH and MD increases significantly to about 90 % when allowing a tolerance of one or two grid points (i.e. Å) distance – which appears to be sufficiently accurate for practical applications to identify bound or trapped water molecules.

In summary, 3D-RISM-KH seems to be very capable to of calculating local densities with sufficient accuracy as compared to MD — particularly since a resolution of 1 Å or 2 Å might completely suffice to investigate solvent effects in most biophysical processes. These results motivate that MD and 3D-RISM-KH can be used in a complementary manner to very effectively investigate the solvent structure in biological systems: MD can be used to generate Boltzmann-weighted ensembles, (e.g., of different side chain or protein conformations), and 3D-RISM-KH can then be used to efficiently sample the solvent structure for these static solute conformations with high accuracy, similarly to the recently implemented MD/3D-RISM-KH coupled method [43] and MM/3D-RISM-KH post-processing of MD [76]. In addition, 3D-RISM-KH could also be used to already provide equilibrated solvent structures to be used as input in MD simulations. We expect that the combined use of MD and 3D-RISM-KH calculations in such manner will substantially contribute to our understanding of water in biological systems.

Acknowledgments

MCS and VSP acknowledge support from the NIH Nanomedicine Center (PN1 EY016525-02), the NIH Simbios Center for Biomedical Computation (NIH U54 Roadmap GM072970), and NSF award CNS-0619926 for computer resources. NB, DW, and AK acknowledge financial support by the Alberta Prion Research Institute and the National Research Council of Canada.

References

1. Tanford, Charles. The Hydrophobic Effect: Formation of Micelles and Biological Membranes. John Wiley & Sons Inc; 1973.
2. Pratt LR, Chandler D. J Chem Phys. 1977; 67(8):3683–3704.
3. Tanford, Charles. Science. 1978
4. Dill, Ken A. Biochemistry. 1990; 29(31):7133–7155. [PubMed: 2207096]
5. Franks, F., editor. Water: A comprehensive Treatise. Plenum Press; 1981.
6. Mallamace, Francesco. Proc Natl Acad Sci USA. 2009; 106:15097–15098. [PubMed: 19805244]
7. Huang C, Wikfeldt KT, Tokushima T, Nordlund D, Harada Y, Bergmann U, Niebuhr M, Weiss TM, Horikawa Y, Leetmaa M, Ljungberg MP, Takahashi O, Lenz A, Ojamäe L, Lyubartsev AP, Shin S, Pettersson LGM, Nilsson A. Proc Natl Acad Sci USA. 2009; 106:15214–15218. [PubMed: 19706484]
8. Errington, Jeffrey R.; Debenedetti, Pablo G. Nature. 2001; 409(6818):318–321. [PubMed: 11201735]
9. Mittal, Jeetain; Hummer, Gerhard. Proc Natl Acad Sci USA. 2008; 105:20130–20135. [PubMed: 19074279]
10. Patel, Amish J.; Varill, Patrick; Chandler, David. J Phys Chem B. 2009; 114:1632–1637. [PubMed: 20058869]
11. Horn, Hans W.; Swope, William C.; Pitera, Jed W.; Madura, Jeffry D.; Dick, Thomas J.; Hura, Greg L.; Head-Gordon, Teresa. J Chem Phys. 2004; 120:9665–9678. [PubMed: 15267980]
12. Abascala JLF, Vega C. J Chem Phys. 2005; 123:234505. [PubMed: 16392929]
13. Rasaiah, Jayendran C.; Garde, Shekhar; Hummer, Gerhard. Annual Review of Physical Chemistry. 2008; 59:713–740.
14. Godawat, Rahul; Jamadagni, Sumanth N.; Garde, Shekhar. Proc Natl Acad Sci USA. 2009; 106
15. Lucent, Del; Snow, Christopher D.; Aitken, Colin Echeverría; Pande, Vijay S. PLoS Comp Biol. 2010; 6:e1000963.
16. England, Jeremy; Lucent, Del; Pande, Vijay. Curr Opin Struct Biol. 2008; 18:163–169. [PubMed: 18291636]
17. England, Jeremy L.; Lucent, Del; Pande, Vijay S. J Am Chem Soc. 2008; 130(36):11838–11839. [PubMed: 18710231]
18. Wallqvist A, Berne BJ. J Phys Chem. 1995; 99:2893–2899.
19. England, Jeremy L.; Pande, Vijay S.; Haran, Gilad. J Am Chem Soc. 2008; 130:11854–11855. [PubMed: 18707183]
20. Chandler, David; Andersen, Hans C. J Chem Phys. 1972; 5:1930–1937.
21. Chandler, David. Annu Rev Phys Chem. 1978; 29:441–471.
22. Hirata, Fumio; Rossky, Peter J. Chem Phys Lett. 1981; 83:329–334.
23. Hirata, Fumio; Montgomery Pettitt, B.; Rossky, Peter J. J Chem Phys. 1982; 77:509–520.
24. Beglov, Dmitrii; Roux, Benoît. J Chem Phys. 1995; 103:469602.
25. Beglov, Dmitrii; Roux, Benoît. J Phys Chem B. 1997; 101:7821–7826.
26. Kovalenko, Andriy; Hirata, Fumio. Chem Phys Lett. 1998; 290:237–244.
27. Du, Qishi; Beglov, Dmitrii; Roux, Benoît. J Phys Chem B. 2000; 104:796–805.
28. Imai T, Kovalenko A, Hirata F. Molecular Simulation. 2006; 32:817–824.
29. Imai, Takashi; Kovalenko, Andriy; Hirata, Fumio. Chem Phys Lett. 2004; 395:1–6.

30. Imai, Takashi; Harano, Yuichi; Kinoshita, Masahiro; Kovalenko, Andriy; Hirata, Fumio. J Chem Phys. 2006; 125:024911.
31. Imai, Takashi; Harano, Yuichi; Kinoshita, Masahiro; Kovalenko, Andriy; Hirata, Fumio. J Chem Phys. 2007; 126:225102. [PubMed: 17581082]
32. Yoshida, Norio; Phongphanphanee, Saree; Maruyama, Yutaka; Imai, Takashi; Hirata, Fumio. J Am Chem Soc. 2006; 128:12042–12043. [PubMed: 16967934]
33. Yoshida, Norio; Phongphanphanee, Saree; Hirata, Fumio. J Phys Chem B. 2007; 111:4588–4595. [PubMed: 17428086]
34. Yamazaki, Takeshi; Kovalenko, Andriy; Murashov, Vladimir V.; Patey, Grenfell N. J Phys Chem B. 2010; 114:613–619. [PubMed: 19947642]
35. Stumpe, Martin C.; Grubmüller, Helmut. J Phys Chem B. 2007; 111(22):6220–6228. [PubMed: 17497766]
36. Hansen, J-P.; McDonald, IR. Theory of Simple Liquids. 2. Academic Press; London: 1986.
37. Kast, Stefan M. Comp Phys Comm. 2004; 5:449–455.
38. Chong, S-H. Dynamical Processes in Solution. Vol. 24. Kluwer Academic Publishers; Dordrecht, The Netherlands: 2003.
39. Kobryn AE, Yamaguchi T, Hirata F. J Chem Phys. 2005; 122(18):184511. [PubMed: 15918733]
40. Kobryn AE, Hirata F. J Chem Phys. 2007; 126(4):044504. [PubMed: 17286484]
41. Kinoshita, Masahiro. Conformational stability of biomolecules in solution, volume 24 of Understanding Chemical Reactivity. Kluwer Acedemic Publishers; Dordrecht, The Netherlands: 2003.
42. Miyata T, Hirata F. J Comput Chem. 2008; 29(6):871–882. [PubMed: 17963231]
43. Luchko T, Gusarov S, Roe DR, Simmerling C, Case DA, Tuszynski J, Kovalenko A. J Chem Theory Comput. 2010; 6:607–624. [PubMed: 20440377]
44. Kinoshita M, Okamoto Y, Hirata F. J Comput Chem. 1997; 18:1320.
45. Kinoshita M, Okamoto Y, Hirata F. J Am Chem Soc. 1998; 120:1855–1863.
46. Xu, Zhaohui; Horwich, Arthur L.; Sigler, Paul B. Nature. 1997; 388:741–750. [PubMed: 9285585]
47. Tang, Yun-Chi; Chang, Hung-Chun; Roeben, Annette; Wischnewski, Dirk; Wischnewski, Nadine; Kerner, Michael J.; Hartl, F Ulrich; Hayer-Hartl, Manajit. Cell. 2006; 125:903–914. [PubMed: 16751100]
48. Farr, George W.; Fenton, Wayne A.; Horwich, Arthur L. Proc Natl Acad Sci USA. 2007; 104(13): 5342–5347. [PubMed: 17372195]
49. Jorgensen, William L.; Chandrasekhar, Jayaraman; Madura, Jeffry D. J Chem Phys. 1983; 79(2): 926–935.
50. Berendsen HJC, Grigera JR, Straatsma TP. J Phys Chem. 1987; 91:6269–6271.
51. Berendsen HJC, van der Spoel D, van Drunen R. Comp Phys Comm. 1995; 91:43–56.
52. Lindahl, Erik; Hess, Berk; van der Spoel, David. J Mol Model. 2001; 7:306–317.
53. Sorin, Eric J.; Pande, Vijay S. Biophys J. 2005; 88:2472–2493. [PubMed: 15665128]
54. Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR. J Chem Phys. 1984; 81(8): 3684–3690.
55. Hansen, J-P.; McDonald, I. Theory of Simple Liquids. 2. Academic; London: 1986.
56. Kovalenko, A. Three-dimensional RISM theory for molecular liquids and solid-liquid interfaces volume 24 of Understanding Chemical Reactivity. Kluwer Acedemic Publishers; Dordrecht, The Netherlands: 2003.
57. Kovalenko A, Hirata F. J Chem Phys. 1999; 110:10095–10112.
58. Chandler D, McCoy J, Singer S. J Chem Phys. 1986; 85(10):5971–5976.
59. Chandler D, McCoy J, Singer S. J Chem Phys. 1986; 85(10):5977–5982.
60. Kovalenko A, Hirata F. J Chem Phys. 2000; 112:10391–10402.
61. Kovalenko A, Hirata F. J Chem Phys. 2000; 112:10403–10417.
62. Perkyns JS, Pettitt BM. J Chem Phys. 1992; 97(10):7656–7666.
63. Kovalenko A, Hirata F. Chem Phys Lett. 2001; 349:496–502.

64. Kovalenko A, Hirata F. J Theor Comput Chem. 2002; 1:381–406.
65. Yoshida K, Yamaguchi T, Kovalenko A, Hirata F. J Phys Chem B. 2002; 106:5042–5049.
66. Omelyan I, Kovalenko A, Hirata F. J Theor Comput Chem. 2003; 2:193–203.
67. Yamazaki T, Imai T, Hirata F, Kovalenko A. J Phys Chem B. 2007; 111:1206–1212. [PubMed: 17266276]
68. Yoshida N, Imai T, Phongphanphanee S, Kovalenko A, Hirata F. J Phys Chem B Feature Article. 2009; 113:873–886.
69. Imai T, Kovalenko A, Hirata F, Kidera A. Journal of Interdisciplinary Sciences: Computational Life Sciences. 2009; 1:156–160.
70. Imai T, Oda K, Kovalenko A, Hirata F, Kidera A. J Am Chem Soc. 2009; 131:12430–12440. [PubMed: 19655800]
71. Yamazaki T, Kovalenko A, Murashov VV, Patey GN. J Phys Chem B. 2010; 114:613–619. [PubMed: 19947642]
72. Vega C, Abascal JLF, Conde MM, Aragones JL. Faraday Discuss. 2009; 141:251–276. [PubMed: 19227361]
73. Rhodes, Maria M.; Réblová, Kamila; Sponer, Jirí; Walter, Nils G. Proc Natl Acad Sci USA. 2006; 103:13380–13385. [PubMed: 16938834]
74. Imai, Takashi; Hiraoka, Ryusuke; Kovalenko, Andriy; Hirata, Fumio. J Am Chem Soc. 2005; 127:15334–15335. [PubMed: 16262373]
75. Stumpe, Martin C.; Grubmüller, Helmut. PLoS Comp Biol. 2008; 4(11):e1000221.
76. Blinov, Nikolay; Dorosh, Lyudmyla; Wishart, David; Kovalenko, Andriy. Biophys J. 2010; 98:282–296. [PubMed: 20338850]

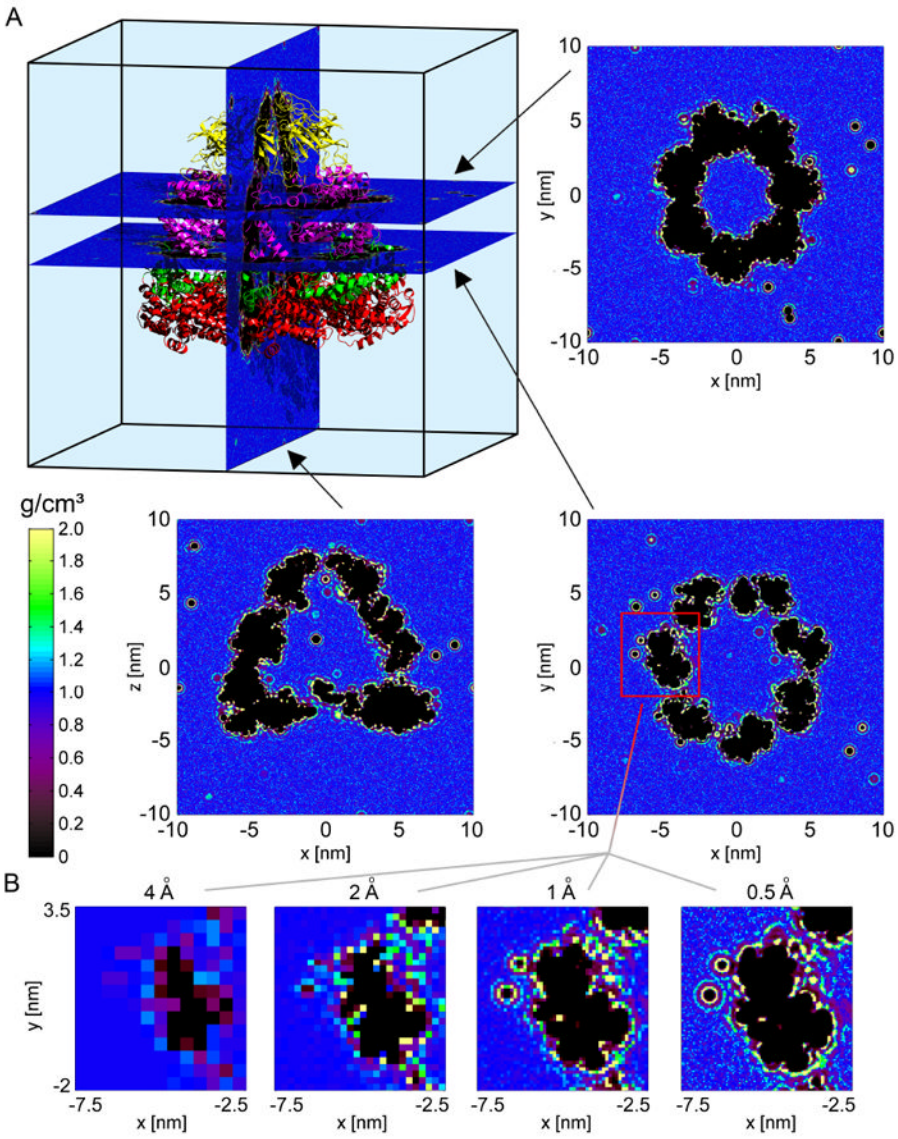

Figure 1.

A) Illustration of the GroEL simulation system. Coloring for the chaperonin complex is according to its domains (red: equatorial, green: intermediate, magenta: apical, and yellow: GroES). The density maps show the water densities in slices through the simulation box, as indicated by the respective arrow. B) Illustration of the four levels of resolution (grid spacings) we have investigated here. From left to right: 4 Å, 2 Å, 1 Å, 0.5 Å.

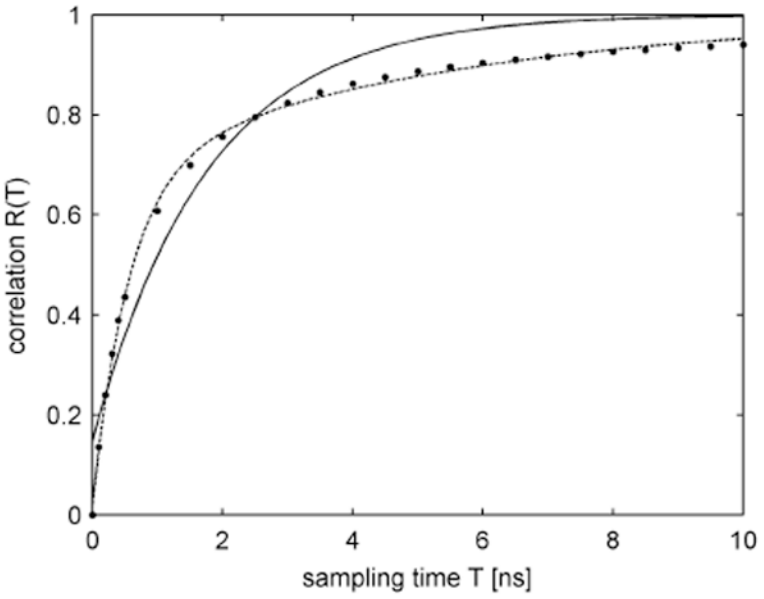

Figure 2.
Time-dependent correlation coefficient $R(T)$ between the water densities $\rho(x, y, z)$ from two different simulations. The raw data (dots) are much better described by a bi-exponential fit (dashed line) than by a mono-exponential fit (solid line).

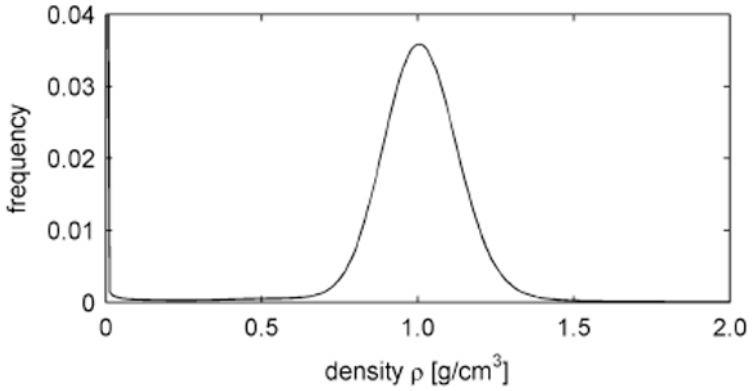

Figure 3.
Water density distribution histogram from the MD with TIP4Pew water.

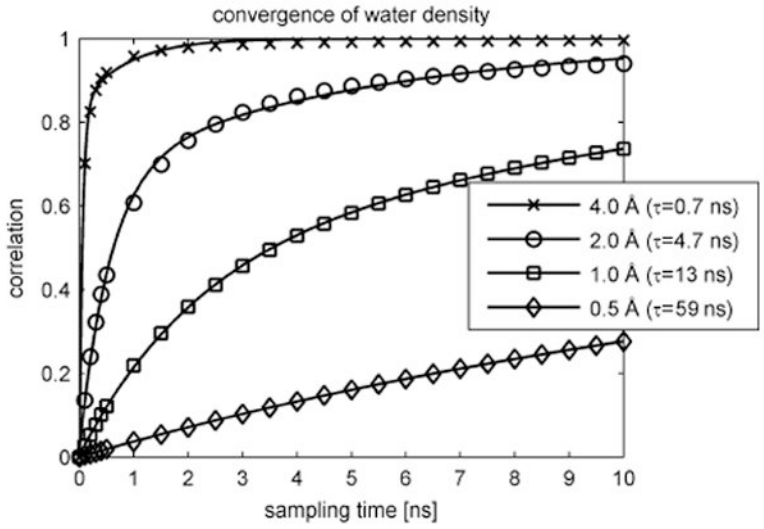

Figure 4.

Convergence times of water densities for different levels of detail (red: 4 Å, green: 2 Å, blue: 1 Å, black: 0.5 Å). The lines represent bi-exponential fits to the raw data, which were used to calculate the convergence times.

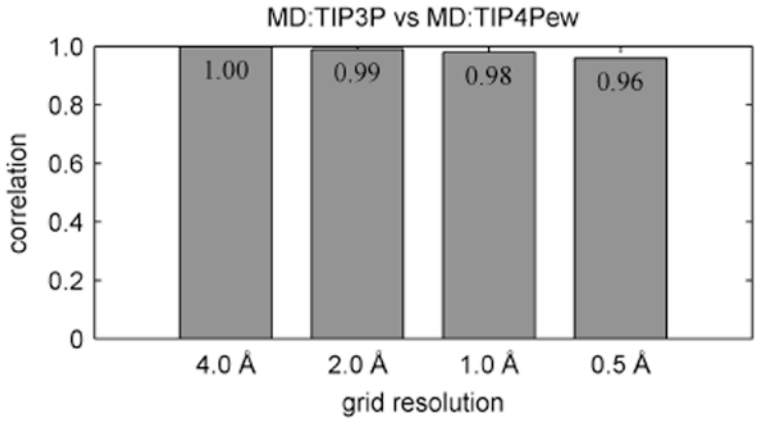

Figure 5.
Relative correlation between the water densities calculated with TIP3P compared to TIP4Pew, both using MD.

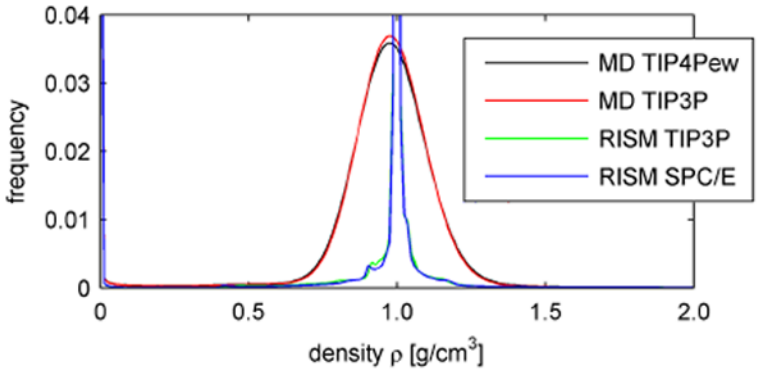

Figure 6.
Water density distribution histograms for MD with TIP4Pew (black) and TIP3P (red) water, and for 3D-RISM-KH with TIP3P (green) and SPC/E (blue) water.

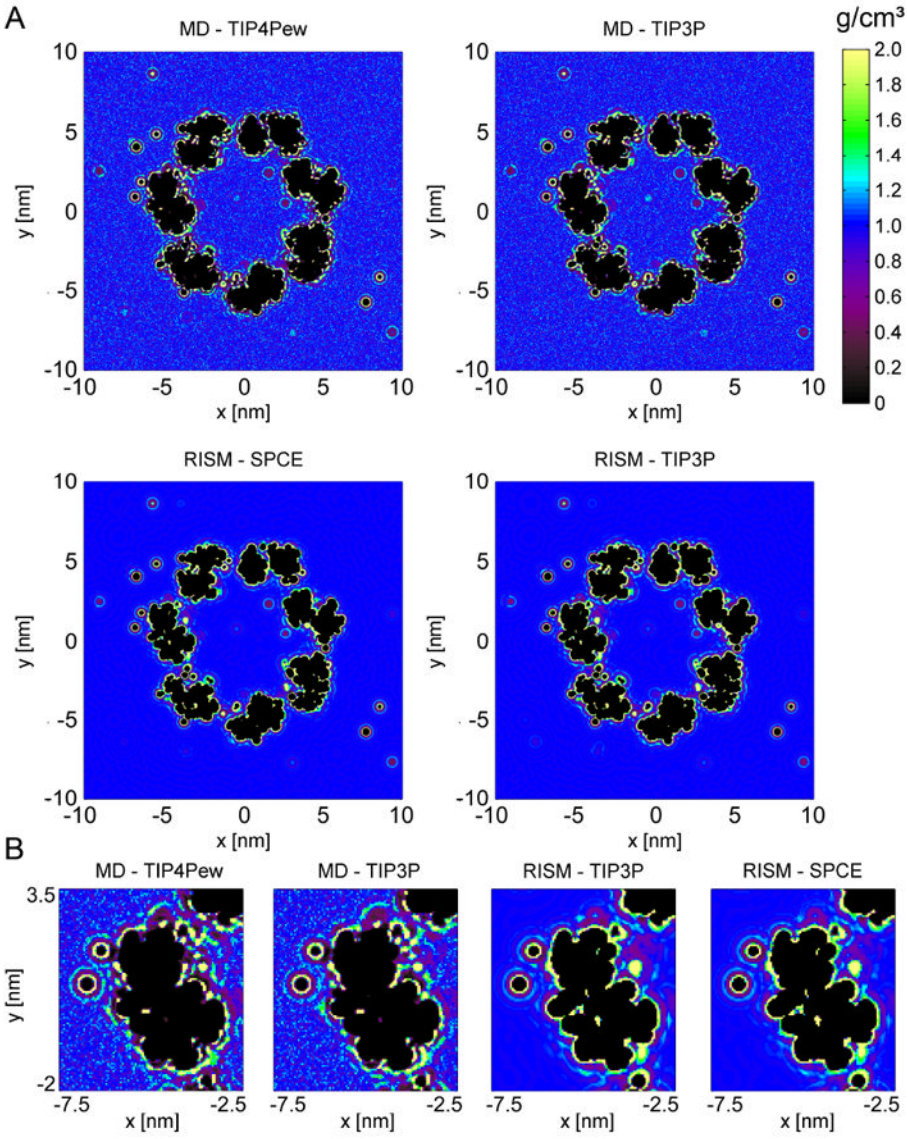

Figure 7.

Water density, in one slice through GroEL, obtained from MD and 3D-RISM-KH with two different force fields each. (A) shows a slice through the whole system, (B) a magnification around one GroEL chain to illustrate details.

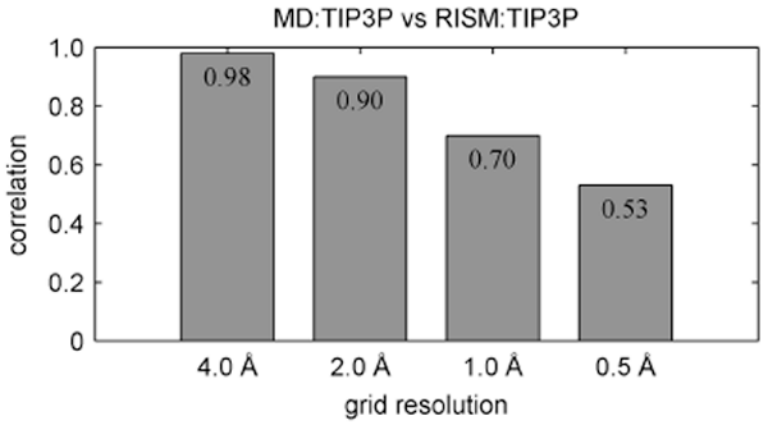

Figure 8.
Relative correlation between the water densities calculated with 3D-RISM-KH compared to those from MD, both with TIP3P water.

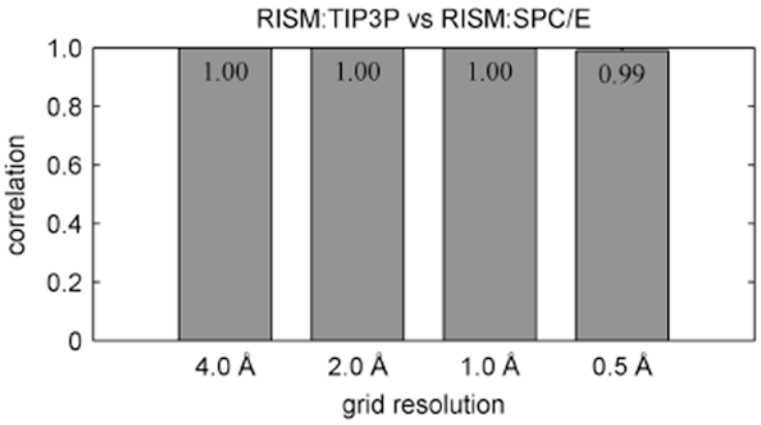

Figure 9.
Correlation between the water densities calculated with TIP3P compared to SPC/E, both using 3D-RISM-KH.

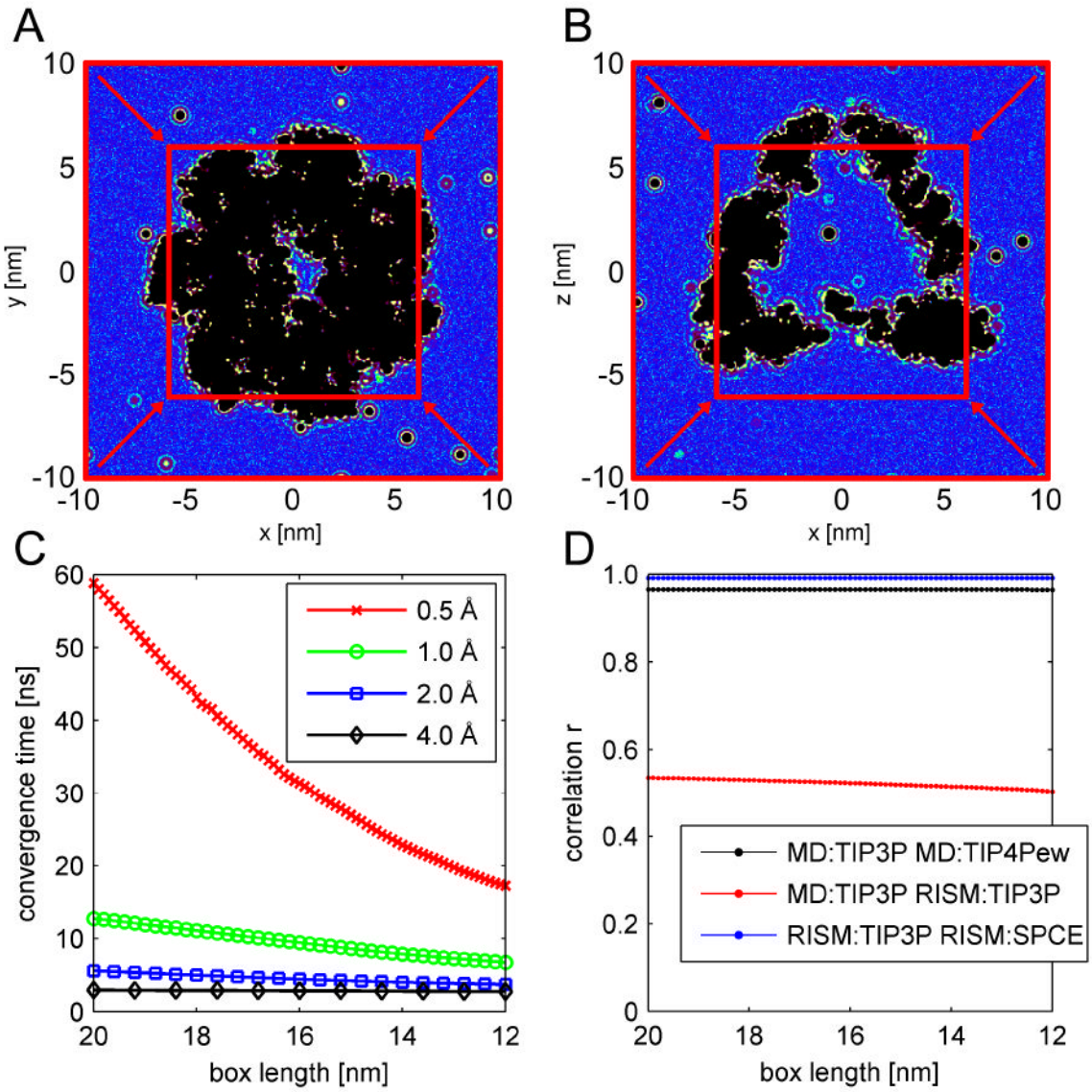

Figure 10.

(A + B) Illustration of the box-cutout. A: top-view as intersection at the z-position where GroEL has the largest diameter, B: side-view through the box-center. The outer red box shows the maximum box size (the complete 20 nm^3 box) used for the analysis, and the inner box depicts the minimum box size (12 nm^3, 22% of the original box volume) to which the analysis volume is shrunken to investigate the influence of the box size on the correlations. (C) Dependence of the convergence times on the box length. The values for the box length of 20nm are the same as derived above in Fig. 4. (D) Dependence of the correlation coefficients on the box length. Shown are the correlations for the grid spacing of 0.5 Å, the correlations for the larger grid spacings are even less dependent on the box length (not shown). The values for the box length of 20 nm are the same as in Figs. 5 (black curve), 8 (blue curve), and 9 (red curve).

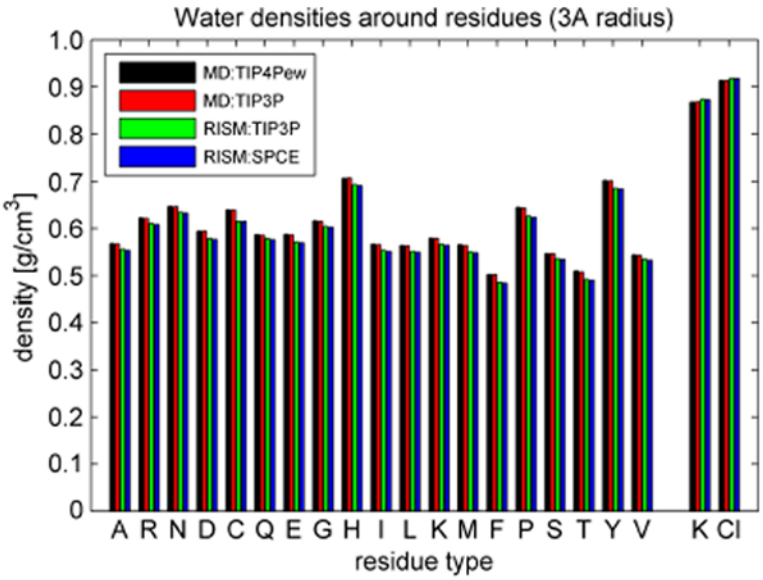

Figure 11.

Average local water densities in a 3 Å solvation shell around residue types in GroEL. Note that due to excluded volume, the absolute numbers of the averages are lower than expected without excluded volume.

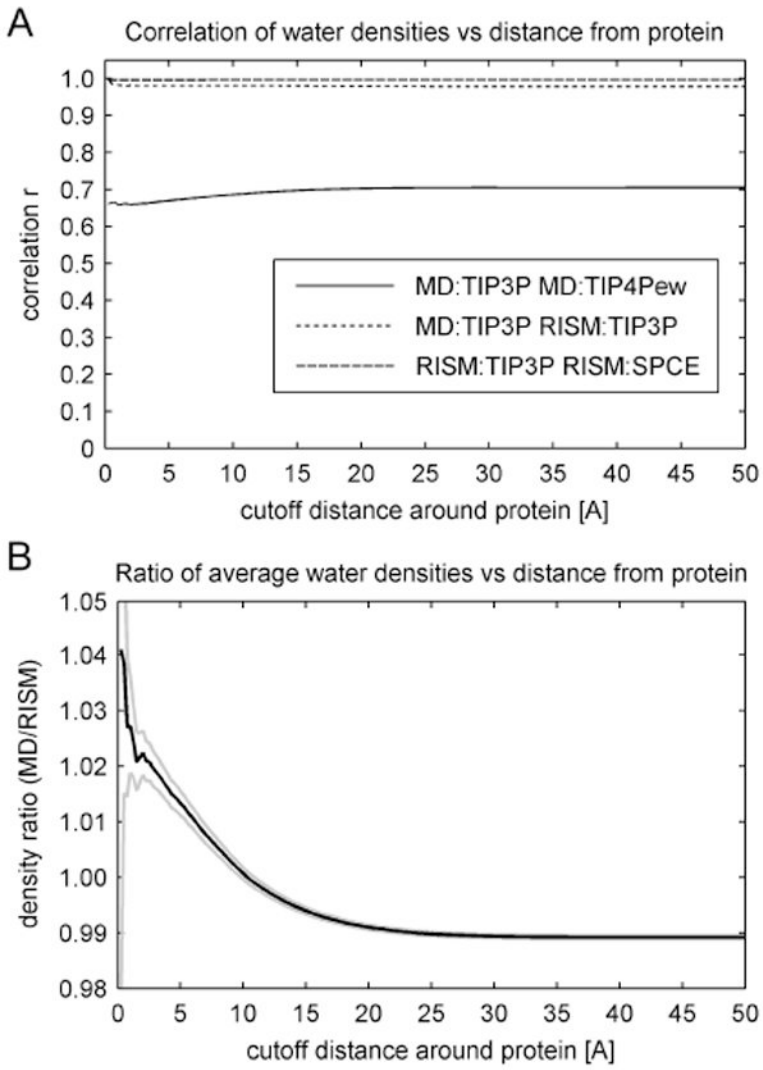

Figure 12.

A) The correlation of the local water density distribution for the different methods in dependence of the solvation shell radius. B) The ratio of the local water density distribution for the different methods in dependence of the solvation shell radius.

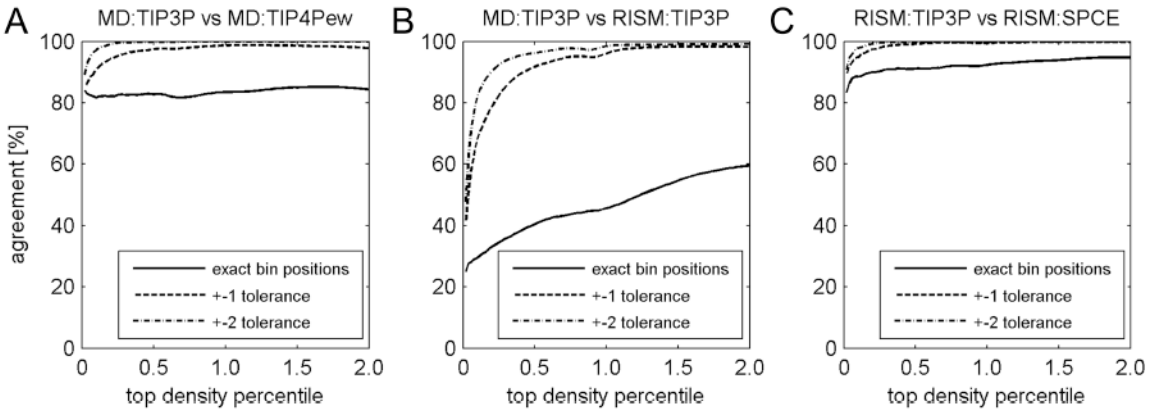

Figure 13.

Agreement of the grid positions of the top n % water densities, for the different method and force field combinations on the 1 Å grid. A) comparing TIP3P and TIP4Pew in MD. B) comparing MD and 3D-RISM, both using TIP3P. C) comparing 3D-RISM using TIP3P and SPC/E. Solid lines: exact bin positions, dashed lines: ±1 grid point tolerance, dash-dotted lines: ±2 grid points tolerance.