Hydration of an α-Helical Peptide: Comparison of Theory and Molecular Dynamics Simulation

Angel E. García,[1]* Gerhard Hummer,[1] and Dikeos Mario Soumpasis[2]
[1]*Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545*
[2]*Biocomputation Group, Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany*

ABSTRACT We present a statistical mechanical description of biomolecular hydration that accurately describes the hydrophobic and hydrophilic hydration of a model α-helical peptide. The local density of water molecules around a biomolecule is obtained by means of a potential-of-mean-force (PMF) expansion in terms of pair- and triplet-correlation functions of bulk water and dilute solutions of nonpolar atoms. The accuracy of the method is verified by comparing PMF results with the local density and site-site correlation functions obtained by molecular dynamics simulations of a model α-helix in solution. The PMF approach quantitatively reproduces all features of the peptide hydration determined from the molecular dynamics simulation. Regions of hydrophobic hydration near the C_α and C_β atoms along the helix are well reproduced. The hydration of exposed polar groups at the N- and C-termini of the helix are also well described by the theory. A detailed comparison of the local hydration by means of site-site radial distribution functions evaluated with the PMF theory shows agreement with the molecular dynamics simulations. The formulation of this theory is general and can be applied to any biomolecular system. The accuracy, speed of computation, and local character of this theory make it especially suitable for studying large biomolecular systems. Proteins 27:471–480, 1997.
© 1997 Wiley-Liss, Inc.

INTRODUCTION

Understanding the role of water in determining the thermodynamic stability and function of proteins has been the subject of many theoretical and experimental studies over the last 30 years.[1] Hydrophobic interactions, for instance, are known to be of paramount importance in protein folding and allosteric effects. Much effort has been devoted toward studying the water structure at the surface and in the interior of proteins, nucleic acids, and their complexes. Binding of water molecules at specific sites plays a determinant role in the specificity and function of proteins as catalysts and regulatory proteins.[2] The subject of protein structural hydration has been extensively reviewed in the literature.[3–10] Experimentally, X-ray[8] and neutron diffraction,[11] and multidimensional nuclear magnetic resonance (NMR)[12,13] are the most reliable methods of studying biomolecular hydration structure. The geometry of water interactions with main- and side-chain structural elements in high-resolution crystal structures has been investigated by Thanki et al.[14–16] and Teeter.[17] General patterns of interactions have emerged from such analysis.[18] Theoretically, molecular simulations are the preferred method to study biomolecular hydration.[19–28] Molecular dynamics (MD) and Monte Carlo (MC) simulations of biomolecular systems in aqueous solution require long computations of large systems. In trying to overcome this limitation, alternative approaches, based both on single-water-molecule hydrogen bonding[29] and on methods[30] derived from the known hydration in high-resolution crystal structures, have been developed.

We present a statistical mechanical approach for calculating the local density of water around a biomolecule (or any inhomogeneity). This method is based on a potential of mean force (PMF) expansion of the N-point correlation function of water in an inhomogeneous water-solute system in terms of a hierarchy of approximations involving pair- and triplet-correlation functions of water molecules in bulk solution. This method has previously been applied to study the local water density at a water-ice interface[31] and to explore hydration patterns of ideal[32–35] and high-resolution single crystals of DNA, RNA, and DNA-drug complexes.[36] These studies of the hydration of nucleic acids proved that the method can reproduce the hydration patterns obtained by MD simulations[37] and high-resolution X-ray crystal and neutron fiber data.[38] To facilitate a meaningful comparison of theoretical predictions with molecular simulation results, extensive configurational averages are required to obtain adequate density statis-

---

tics even in a coarse (i.e., 1 Å³) grid. The poor statistics are a consequence of the open liquid water structure. This article describes the application of the PMF approach to determine the hydration structure of a model α-helix, Ala_{18}. The water density obtained by the PMF approach is found to agree with the density obtained by 1.125 ns MD simulations of the peptide in aqueous solution. Information on both the pair- and triplet-correlations is necessary in the PMF expansion to describe the tetrahedral features of the water structure.

THEORETICAL BACKGROUND
Local Density of Water in Terms of N-Point Correlation Functions

We are interested in studying a system consisting of a biological macromolecule and water molecules. The biomolecule will be treated as a source of inhomogeneity in an otherwise homogeneous (bulk) water bath. The local density of water near the biomolecule can be described by a position-dependent one-particle density of water $\rho(\vec{s})$. Our approach toward describing the local density of water around a biomolecule consists of deriving an approximate expression for $\rho(\vec{s})$ in terms of known structural properties of bulk water.

The equilibrium density of water at a point $\vec{s}$ near a biomolecule consisting of n atoms of various chemical identities $(\alpha_1, \ldots \alpha_n)$ at positions $(\vec{r}_1, \ldots, \vec{r}_n)$ is given by the conditional probability density

$$\rho_{\alpha_1, \ldots \alpha_n, w}(\vec{s}|\vec{r}_1, \ldots, \vec{r}_n)$$

$$= N_w \frac{\int d\vec{s}_2 \ldots d\vec{s}_{N_w} \exp[-\beta U(\vec{r}_1, \ldots, \vec{r}_n, \vec{s}_1, \ldots, \vec{s}_{N_w})]}{\int d\vec{r}_1 \ldots d\vec{s}_{N_w} \exp[-\beta U(\vec{r}_1, \ldots, \vec{s}_{N_w})]}. \quad (1)$$

Here U represents the total potential energy of the system comprising n solute atoms and N_w solvent molecules. In the MD simulations to be presented here, U is a pairwise-additive potential energy function modeled by the OPLS force field.[39] For a fixed biological molecule the interaction potential can be split into two parts such that the solute appears as an external field acting on the solvent molecules. The conditional solvent density at a point $\vec{s}$ is then expressed as an integral over the configurational space of the solvent molecules, where the solvated macromolecule appears as an external field. This conditional probability is, in turn, related to multiparticle correlation functions, g,

$$\rho_{\alpha_1, \ldots \alpha_n, w}(\vec{s}|\vec{r}_1, \ldots, \vec{r}_n)$$

$$= \rho_0 \frac{g_{\alpha_1, \ldots, \alpha_n, w}(\vec{s}, \vec{r}_1, \ldots, \vec{r}_n)}{g_{\alpha_1, \ldots, \alpha_n}(\vec{r}_1, \ldots, \vec{r}_n)} \quad (2)$$

where $\rho_0 = N_w/V$ is the solvent bulk density, that is, the local density of water at a point $\vec{s}$ is proportional to the $(n + 1)$-point correlation function between n fixed atoms that comprise the biomolecule and a water molecule in the water bath. This identification is only of formal importance since multiple-point correlation functions are usually calculated and measured only for pairs. However, the $(n + 1)$-point correlation functions can be accurately approximated by a combination of two- and three-particle correlation functions.

When only pair-correlation functions $g^{(2)}$ are known, the Kirkwood superposition approximation (KSA) is commonly used.[40] The KSA approximates the local water density by

$$\rho_{\alpha_1, \ldots \alpha_n, w}(\vec{s}|\vec{r}_1, \ldots, \vec{r}_n)$$

$$\approx \rho_0 \prod_{i=1}^{n} g^{(2)}_{\alpha_i, w}(|\vec{r}_i - \vec{s}|) = \rho^{KSA}. \quad (3)$$

The KSA approximation is accurate for most (r, s, t) triplets in bulk water and is a good approximation for monoatomic molecules and ions. A PMF approach based on the KSA has been used to calculate differences in the ionic free energy of DNA oligomers in monovalent ionic solution.[34,41–45] This simple PMF has been shown to describe well all known structural transitions driven by changes of monovalent salt concentrations, specially at high ionic concentrations ($c_s \geq 0.1$ M). The use of this approximation for describing the structure of water around polar groups cannot be expected to be as accurate, since two-point correlation functions are not sufficient to describe the tetrahedral character of liquid-water organization.

Recent calculations of triplet-correlation functions for water[31,46] and water-methane solutions have motivated the use of approximations of higher-order correlation functions in terms of pair- and triplet-correlation functions. In these cases pair- and triplet-correlation functions can be used in the Fisher-Kopeliovich superposition approximation (FKSA) to approximate the local water density by[47]

$$\rho(\vec{s}|\vec{r}_1, \ldots, \vec{r}_n) = \rho_0 \prod_{i=1}^{n} g^{(2)}(r_{i0})$$

$$\cdot \prod_{i,j>i} \frac{g^{(3)}(r_{i0}, r_{j0}, r_{ij})}{g^{(2)}(r_{i0}) g^{(2)}(r_{j0}) g^{(2)}(r_{ij})}$$

$$= \rho^{KSA} \prod_{i,j>i} \Gamma(r_{i0}, r_{j0}, r_{ij}), \quad (4)$$

where $r_{kl} = |\vec{r}_k - \vec{r}_l|$, and $r_{k0} = |\vec{r}_k - \vec{s}| \cdot \rho^{KSA}$ is defined in Eq. (3), and

$$\Gamma(r, s, t) = \frac{g^{(3)}(r, s, t)}{g^{(2)}(r) g^{(2)}(s) g^{(2)}(t)}. \quad (5)$$

The local density calculations that follow are based on Eq. (4). Triplet-correlation functions between each kind of atom pair (α_i, α_j) and water must be known to calculate $\rho_{\alpha_1,\ldots,\alpha_n,w}(\vec{s}|\vec{r}_1,\ldots,\vec{r}_n)$. These functions are calculated from MC or MD simulations of bulk water. The fact that the calculation of pair- and triplet-correlation functions is done only once makes the theory powerful and efficient. The correlation functions are tabulated and used to calculate the local density, as defined by Eq. (4), for any biomolecule in any fixed configuration. At this point Eq. (4) represents a general theory and does not rely on any specific model for the solute or the solvent.

Atom Classes—Polar and Nonpolar

The indices $\alpha_1, \ldots, \alpha_n$ account for different atom types in the biomolecule. A large number of correlation functions would be required to describe the hydration structure of a solvated macromolecule using Eq. (4) directly. The calculation of three particle correlation functions for all possible triplets of particle types (i.e., $g^{(3)}_{\alpha_i \alpha_j w}$) is not realistic. It is important to simplify the theoretical description with regard to the number of different atom types used to model the macromolecule. To accomplish this we will only label atoms as *polar* and *nonpolar*.

In previous work it has been shown that the PMF expansion predicts the inhomogeneous water structure around nonpolar molecular solutes.[48,49] For solutes consisting of bonded sites there is a combinatorial explosion of terms in the PMF expansion in the limit of complete site overlap, indicating that higher order ($n > 3$) correlation functions are necessary to describe the inhomogeneous hydration of nonpolar solutes correctly. However, it was found that proximity approximations can represent well the local structure of water near nonpolar solutes.[48] The proximity approximation assumes that water organization near nonpolar solutes is only locally sensitive to the structural details of the nonpolar molecular solute. In the simplest of the proximity approximations the local structure of water due to nonpolar solutes is determined by water correlations with only the nearest nonpolar solute site. For the sake of simplicity we assume that the hydration structure of nonpolar atoms in the biomolecule is represented by the hydration structure of a methane molecule in water.

Another simplification consists of assuming that the local water densities surrounding various electronegative atoms (e.g., N, O) are similar. Electronegative oxygen and nitrogen atoms exhibit a strong tendency to form hydrogen bonds with vicinal water molecules. The geometric features of these hydrogen bonds such as bond length and bond angle are similar to those formed between water molecules. Therefore, the correlation functions between electronegative atoms and water molecules are represented by the correlation functions between water molecules in bulk water.

To differentiate between hydrogen bond donor and acceptor atoms, we keep the triplet-correlation function for covalently bonded O-H (or N-H) groups and a hydrogen bond acceptor in Eq. (4). This triplet-correlation function is different from the others since it includes the correlation between atoms that are covalently bonded. The contribution of this term, which is not already included in the KSA for polar groups, is given by

$$\chi_{O-H\ldots O} = \frac{g^{(3)}_{O-H\ldots O}(r_{OO}, r_{HO}, r_{O-H})}{g^{(2)}_{OO}(r_{OO})}. \tag{6}$$

The term $\chi_{O-H\ldots H}$ represents the correction to the local density due to hydrogen orientation in polar groups. These approximations allow us to simplify Eq. (4) to

$$\rho(\vec{s}|\vec{r}_1, \ldots \vec{r}_n) = \rho_0 \prod_i^{polar} g^{(2)}_{OO}(r_{i0})$$

$$\cdot \prod_{i,j>i}^{polar} \frac{g^{(3)}_{OOO}(r_{i0}, r_{j0}, r_{ij})}{g^{(2)}_{OO}(r_{i0})g^{(2)}_{OO}(r_{j0})g^{(2)}_{OO}(r_{ij})}$$

$$\times g^{(2)}_{Me-O}(r^{min}_{lo})$$

$$\cdot \prod^{\substack{polar \\ hydrogens}} \chi_{O-H\ldots O}(\vec{r}_{donor}, \vec{r}_{H_j}, \vec{s}), \tag{7}$$

where $g^{(2)}_{Me-O}$ is the methane-water oxygen pair-correlation function and r^{min}_l is the distance between $\vec{s}$ and the closest nonpolar atom. $\chi_{O-H\ldots O}$ is the three-particle correction to the local density for OH . . . O atoms where an O-H (or N-H) pair is bonded (i.e., at a fixed distance), and the other O belongs to an oxygen atom in the water solvent. Explicit triplet correlations involving water, a polar site, and a nonpolar site have not been included. However, these terms are implicitly included within the proximity approximation for nonpolar atoms and the KSA for polar atoms. This simplified expression will be used to describe the local density of water around a model peptide. The assumptions introduced above can be easily relaxed by calculating triplet-correlation functions for various molecular groups in water solutions. At this point we have reduced the local density of water around a biomolecule to an expression that only depends on the pair- and triplet-correlation functions of bulk water and of methane in water. These correlation functions are obtained from MD simulations.

METHODS
Calculation of Water Pair- and Triplet-Correlation Functions

The triplet-correlation functions for water were calculated for a TIP3P water model.[50,51] We used the OPLS force field parameters for methane. Correlation functions were calculated by averaging over configurations sampled by 1.5 ns MD simulations of 216 water molecules immersed in a cubic box with sides $L = 18.78$ Å ($\rho_O = 1$ g/cm^3), at T = 300 K. The Coulomb interactions were modeled by a generalized reaction field model (GRF) (with cut-off distance, $R_c = 7.5$ Å, and a reaction field dielectric constant $\epsilon_{RF} = 65.0$).[31,52] Histograms of the occurrences of triangles with sides (r, s, t) formed by all possible triplets of particles were collected in a cubic grid for triangles with sides up to 6.2 Å, with a resolution of 0.2 Å. The proper normalization factors were used to obtain $g^{(3)}(r, s, t)$.[53] In addition, triplet-correlation values for triads including a pair of bonded polar atoms in the biomolecule that are closer than non-bonded pairs in the solvent are needed for the calculation of the local density of water. Constrained MD simulations were performed to determine the triplet-correlation functions for triangles with any side shorter than 2.6 Å. In these calculations, one pair of water molecules was forced to maintain a fixed distance between oxygen atoms (from 1.9 to 2.8 Å), while all other 214 water molecules were not constrained. A typical constrained simulation extended from 0.25 to 1.0 ns. Pair-correlation functions were calculated in a 0.02 Å grid for distances up to 8.0 Å.

Local Density of Water Around a Rigid Peptide

The local density of water around an ideal α-helix is calculated using Eq. (7), with two- and three-point correlation functions obtained for a TIP3P water model. A tri-linear interpolation function was used to interpolate the values of $g^{(3)}$ off the grid points. A linear interpolation was used for $g^{(2)}(r)$. A grid size of 0.25 Å was used to evaluate $\rho(x, y, z)$ on a cubic box containing the model peptide. To get a quantitative description of the local density of water we also calculated the radial distribution function $g(r)$ around each atom in the α-helix. $g(r)$ is defined as the average water oxygen density (in units of the bulk density ρ_0) on a sphere with radius r around a given atom in the peptide. This density was calculated by choosing 10^5 points randomly distributed on spheres with radii r at intervals of 0.1 Å.

We also performed MD simulations of the ideal, rigid α-helix in aqueous solution to calculate the configurational averages described in Eq. (1). The local density of water around the polypeptide was determined by averaging the occupancy numbers of water molecules in a uniformly spaced three-dimensional grid covering the simulation volume. Constant (N, V, T) simulations have been extended for 1.0 ns, after an equilibration period of 125 ps. The Verlet algorithm with an integration step of 0.002 ps was used to integrate the equations of motion of the system. All water molecules were simulated as rigid molecules during the simulation. An exact Lagrange multiplier scheme was used to maintain the molecules rigid.[54] The OPLS force field and a TIP3P water model were used for the simulations.

The polypeptide CH_3-Ala_{18}-$NHCH_3$ was constructed in an ideal α-helical conformation (indexed as Ala_2 to Ala_{19}). The (ϕ, ψ) angles were taken as ($-48°$, $-57°$). A united atom representation of the molecule was adopted, with uncharged N and C terminal residues. The peptide Ala_{18} contains 112 atoms (including united atoms and hydrogen atoms forming part of polar groups). This system was immersed in a previously equilibrated cubic box with sides $L = 21.7275$ Å containing 343 water molecules ($\rho_0 = 1.0$ g/cm^3). The peptide was held fixed in an ideal α-helical conformation with its helix axes aligned along the cube's principal diagonal. Atoms from different unit cells are at least 9 Å apart. All water molecules in this box that were within a distance of 2.7 Å to a non-hydrogen atom in the peptide were deleted from the system. The resulting system contains 266 water molecules. The Coulomb interactions were modeled by the GRF with a cut-off of 9.5 Å. The system was equilibrated for 125 ps at T = 300 K. Configurations were saved at a rate of 10 configurations/ps of simulation. The local water density was calculated in a cubic grid with ~1 Å (21.7275 Å/22) grid spacing. The local density of water was calculated for configurations obtained during 1.0 ns of simulations, and for eight subsets of 125 ps simulation segments.

RESULTS
Description of Triplet-Correlation Functions for Water

The triplet-correlation functions are easily interpreted in terms of the ratio of the probabilities of forming triangles with sides (r, s, t) in a liquid and the probability of forming the same triangle by a non-interacting gas. Figure 1 shows the triplet-correlation functions, $g^{(3)}_{OOO}$, $g^{(3-KSA)}_{OOO}$, and $\Gamma^{(3)}_{OOO} = g^{(3)}_{OOO}/g^{(3-KSA)}_{OOO}$, corresponding to isosceles triangles (r, r, t). The KSA differs from the calculated $g^{(3)}$ at significant places. The KSA approximation overestimates the probability of forming equilateral triangles at hydrogen-bond distance $r = s = t = 2.8$ Å, while $g^{(3)}_{OOO}(r, s, t)$ shows a maximum probability for $r = 2.8$ Å, $s = 2.8$ Å, $t = 4.5$ Å (and permutations of r, s, and t). The most probable configuration by the KSA ($r = s = t = 2.8$ Å) corresponds to three water molecules arranged in an equilateral triangle, with sides corresponding to the most probable distance
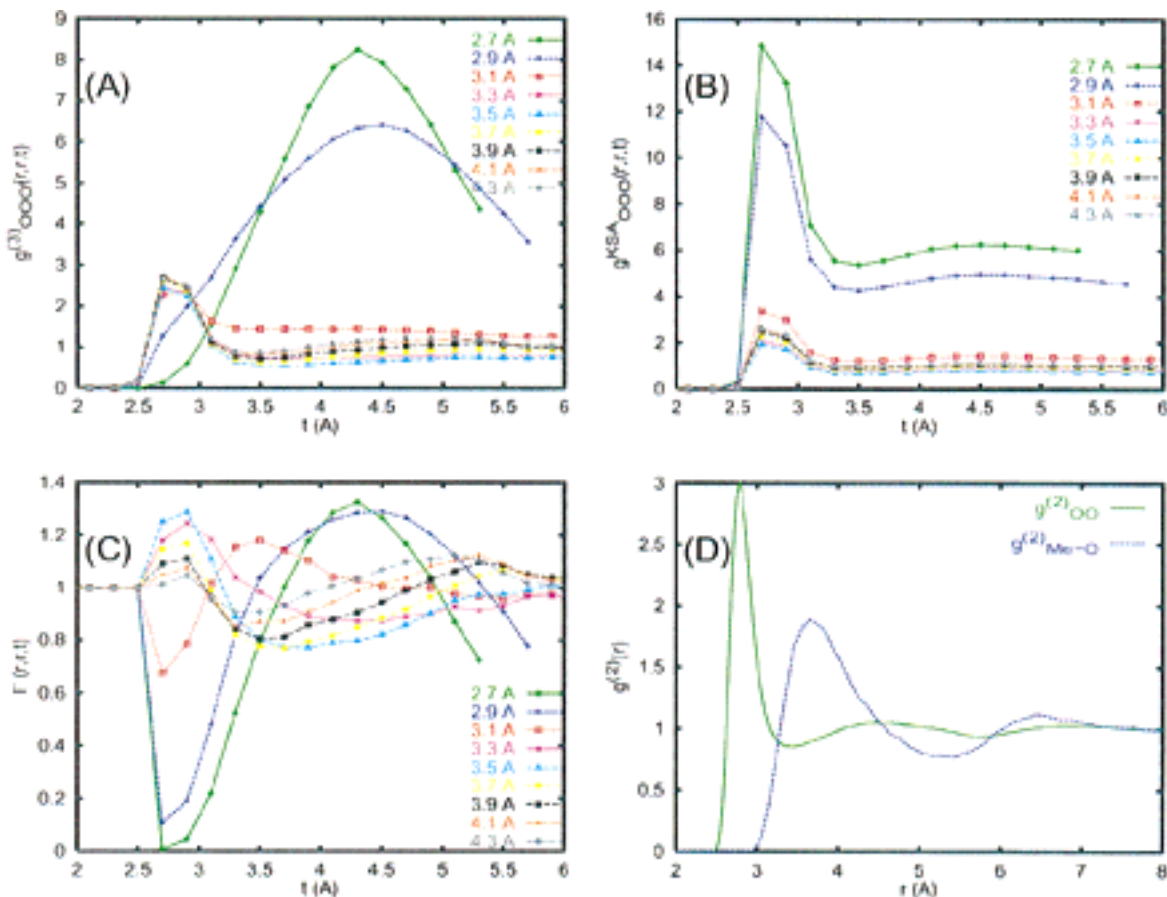
Fig. 1. **A:** Triplet-correlation function for water oxygen atoms, $g^{(3)}_{OOO}(r, r, t)$ forming isosceles triangles with sides r and base t. **B:** Triplet-correlation function obtained from the Kirkwood superposition approximations. **C:** Correction to the triplet-correlation function, $\Gamma_{OOO} = g^{(3)}_{OOO}/g^{(3-KSA)}_{OOO}$, introduced by the Fisher-Kopeliovich superposition approximation. $\Gamma \sim 1$ implies that the KSA is accurate. **D:** Radial distribution functions for water oxygen atoms around other oxygen atoms, $g^{(2)}_{OO}$, in bulk water and for oxygen atoms around a methane molecule, $g^{(2)}_{Me-O}$, in water, at infinite dilution. All correlation functions were calculated using the TIP3P water model, at 300 K and at a water density $\rho_0 = 1$ g/cm^3.

between a pair of water molecules forming hydrogen bonds. In $g^{(3)}_{OOO}$ the most probable configuration ($r = s = 2.8$ Å, $t = 4.5$ Å) corresponds to an isosceles triangle where the base is formed by a segment between two water molecules in van der Waals contact but without forming hydrogen bonds between them, with the other two sides corresponding to the distance between two water molecules forming hydrogen bonds. The large discrepancy between the KSA and the calculated triplet-correlation functions shown above is a definite proof that the inclusion of triplet-correlation functions is necessary to describe the tetrahedral character of the water structure. The correction to the local water density due to triplet-correlation functions is given by $\Gamma(r, s, t)$. A value of $\Gamma(r, s, t) = 1$ implies that the KSA approximation is satisfied. Large corrections are obtained for small values of (r, s) while $\Gamma(r, s, s) \to 1$ for larger values of (r, s) (Fig. 1). The largest correc-

tion to the KSA by the FKSA is found for isosceles triangles with sides near 2.7–2.9 Å, as discussed above. For isosceles triangles with equal sides near 2.8 Å and base between 3.1 and 3.5 Å, the KSA underestimates the actual triplet-correlation function. The largest Γ is 1.54 for $(r, s, t) = (2.9, 3.1, 3.3)$ Å.

Fluctuations in the Local Water Density

The local water density around the α-helix was calculated by means of the PMF approach and by MD simulations. Before proceeding to discuss the similarities between the PMF and MD simulation results we must set the basis for comparison. As a consequence of the open liquid water structure, simulation methods suffer from poor density statistics. That is, the number of counts in the local density calculation is extremely small and subject to large fluctuations. Averages over a large set of independent configurations are required to obtain ad-
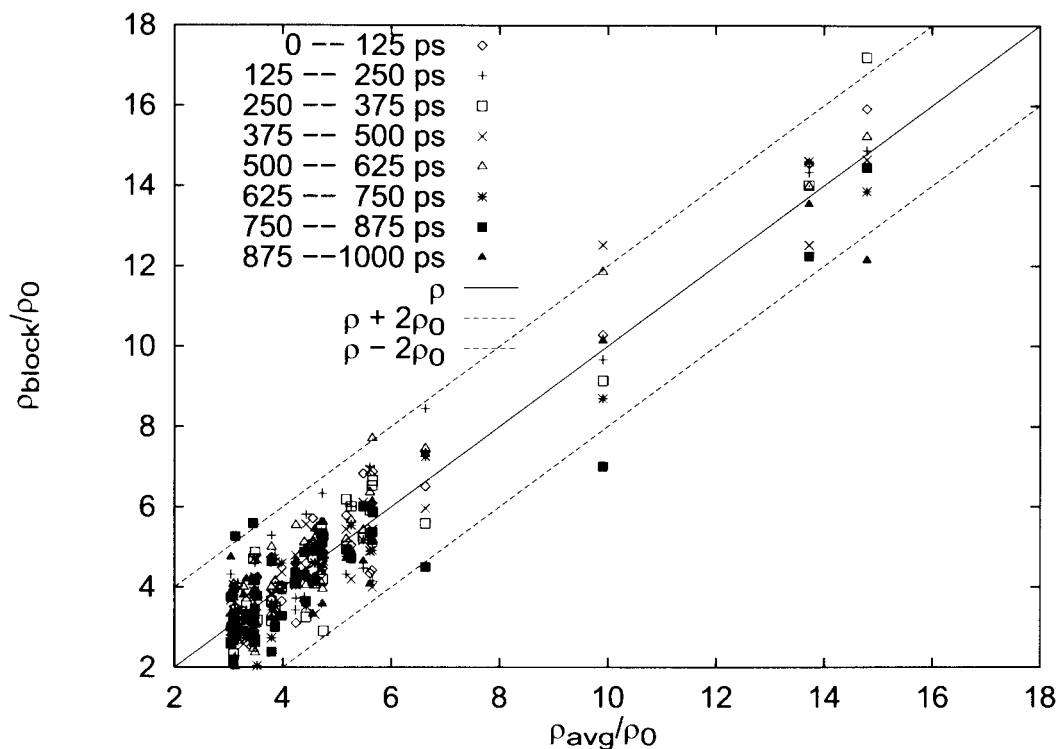
Fig. 2.   Scatter plot of the local density of water calculated at grid points around an ideal α-helix by molecular dynamics simulation. The abscissa represents the average water density in a ~1 Å³ grid over a 1 ns simulation. The ordinate represents the average local density obtained in the same grid over 125 ps block averages. The solid and dashed lines represent perfect correlation and deviations by ±2 ρ_0 density units.

equate density statistics in a coarse (i.e., 1 Å³) grid. To understand the origin of these fluctuations we consider the fluctuations $\sigma_N^2 = \langle N^2 \rangle - \langle N \rangle^2$ in the number of particles N in a volume ΔV. These quantities are related to the isothermal compressibility χ_T through $\sigma_N^2 = \langle N \rangle \rho_0 \kappa_B T \chi_T$ where $\langle N \rangle = \rho_0 \Delta V$, for large volumes. For water under normal conditions, $\chi_T \sim 4.5 \times 10^{-10}$ m² N⁻¹, and $\rho_0 \kappa_B T \chi_T \sim 0.06$. If we assume that the local water density varies considerably within 1.0 Å³, we obtain $\langle N \rangle \sim 0.0334$ and $\sigma_N / \langle N \rangle \sim 1$.

The size of the fluctuations in the local water density around the model peptide studied here was determined from 125 ps block averages over saved MD configurations. Figure 2 shows a plot of the calculated average water density at all points (in a 1 Å³ grid) during blocks of 125 ps of MD configurations as a function of the corresponding density averaged over the entire simulation. Density values below 3 ρ_0 are not shown. A straight line with slope 1 and intercept 0 indicates perfect correlation. We have also plotted lines with slope 1 but with intercepts at 2 and −2 to illustrate that variations of ±2 ρ_0 are common. MD water densities in a 1.0 Å³ grid give densities as high as 15 ρ_0. The local water density fluctuates by up to 6 bulk density units.

Comparison of PMF and Molecular Dynamics Densities

To compare the local water densities obtained from PMF and MD we label grid points of high density with spheres of various sizes (cyan for PMF results and magenta for MD results). We divide the local densities into three classes, $(3 \leq \rho/\rho_0 < 5)$; $(5 \leq \rho/\rho_0 < 7)$; and $(\rho/\rho_0 > 7)$. These grid points are illustrated by using spheres of sizes 0.3, 0.5, and 0.7 Å, respectively. Figure 3 shows a stereo plot of the local water density around the ideal α-helix. Water molecules are localized around the peptide bonds. The local water densities calculated by both methods are in good agreement. Quantitative agreement is better appreciated when we compare radial distribution functions, $g(r)$. Figure 4 shows the radial distribution functions of water around polar (N and O) and nonpolar atoms (C_α, C_β, and C) in amino acids located at the N-terminus, at the center of the helix, and at the C-terminus. Water localization is mainly determined by polar groups. The strong correlations between the methyl groups of the peptide and water seen in Figure 4 are similar to those of methane in water where water molecules are correlated in the radial direction but do not show any angular preference. On the other hand, O
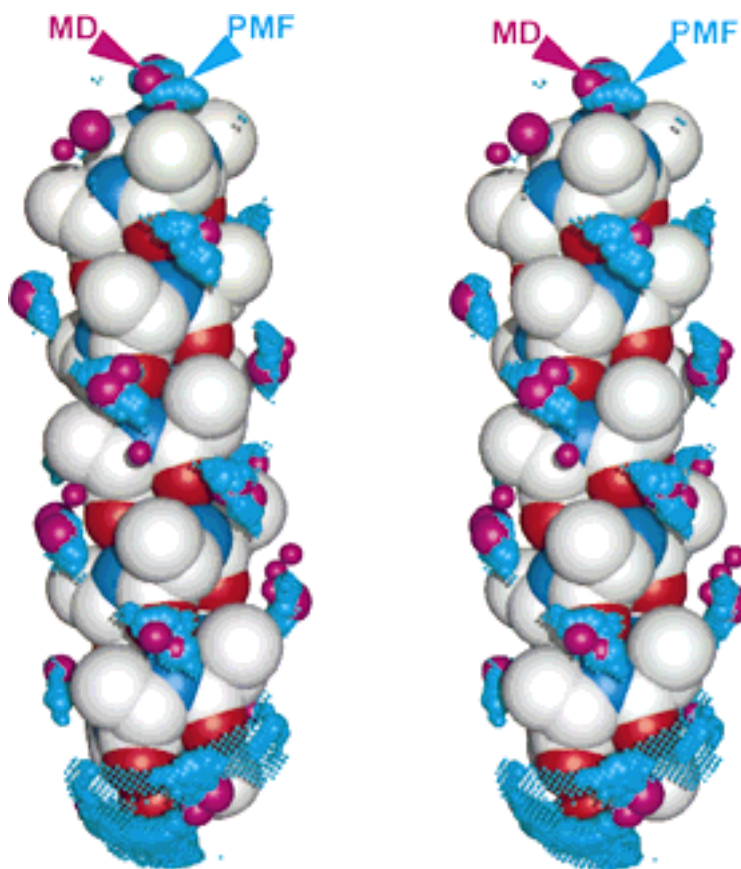
Fig. 3.   Stereo plot illustrating the sites around an ideal α-helix that show large localization of water. We labeled grid points of high density with spheres of various sizes (cyan for PMF results and magenta for MD results): $(3 \leq \rho/\rho_0 < 5)$; $(5 \leq \rho/\rho_0 < 7)$; and $(\rho/\rho_0 > 7)$. These grid points are illustrated by spheres of sizes 0.3, 0.5, and 0.7 Å for the PMF and 0.5, 0.7, and 0.9 Å for the MD results. The grid size is 0.25 Å for the PMF and 1.0 Å for the MD calculations. Larger radii are used for the MD results since there are 64 times more sampled grid points for the PMF than for the MD calculations. Atoms in the peptide are colored according to atom type: gray for C, blue for N, red for O, and white for H.

and N atoms show an angular preference of interaction with water molecules resulting in the water density pattern seen in Figure 3.

The PMF calculation reproduces well the correlation functions $g(r)$ over the distance range $0 < r \leq 6$ Å, including peak positions and heights. The agreement is best for the fully exposed C-terminal end and the central part of the helix. At the N-terminal side, MD gives somewhat higher densities (about 20%) than the PMF calculation. The correlation functions $g(r)$ of atoms in the central part of the helix (Ala$_9$ in Fig. 4) show little structure because the excluded volume dominates the spherical averages. The correlation functions of solvent-exposed atoms are more structured (atoms of Ala$_2$ and Ala$_{19}$ as well as C$_\beta$ atoms).

DISCUSSION

We have described a method of studying the structural hydration of biomolecules. This method is based on a statistical mechanical description of the local water density in terms of particle correlation functions. Our approach involves four main approximations: 1) in the expansion of the local water density in terms of lower order correlation functions it is sufficient to include two- and three-particle correlations; 2) the hydration near nonpolar solutes is local and is determined by the hydration of the nearest nonpolar atom and all nonpolar atoms are represented by methane; 3) inter-particle correlations involving one or two polar atoms in the solute and the solvent are well represented by bulk water oxygen and hydrogen correlation functions; and 4) the role of ions and dynamical fluctuations in the biomolecule's hydration structure is neglected.

The effects of each of these approximations in the local density can be assessed. Approximation 1 has been tested by comparing calculated water densities with extensive Monte Carlo simulations of a water-ice interface.[31] These calculations have shown that three-particle correlations cannot be excluded. Upon
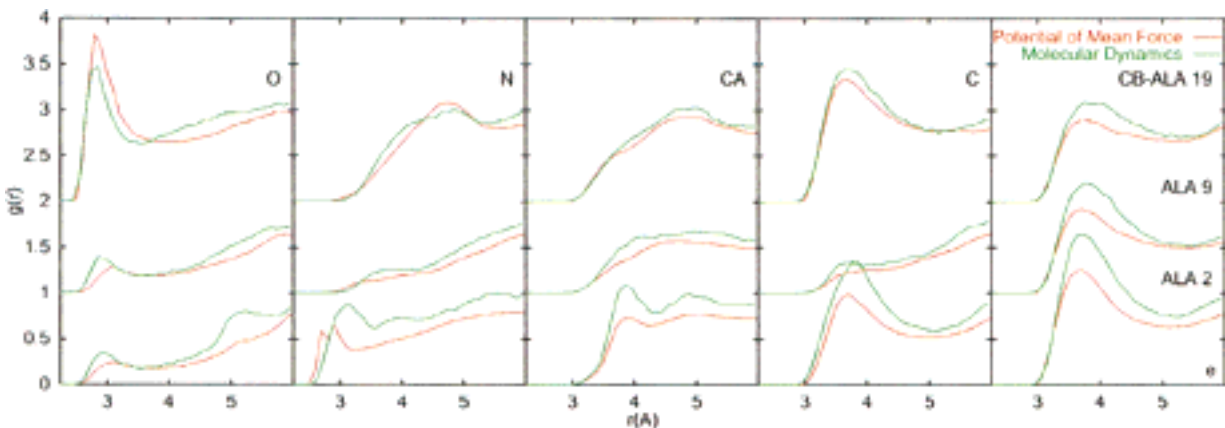
Fig. 4. Radial distribution functions of water around polar (O and N) and nonpolar atoms (C$_\alpha$, C, and C$_\beta$) in amino acids located at the N-terminus (Ala-2), at the center of the helix (Ala-9), and at the C-terminus (Ala-19). The red and blue curves show the PMF and molecular dynamics radial distribution functions, respectively. Curves within one plot are vertically displaced by one unit from the previous curve.

inclusion of three-particle correlation functions the local density of the region closest to the interface was well reproduced by the PMF approach. The local density at regions farther from the water-ice interface showed slight variations, indicating that higher order correlations might be important for larger distances. However, the contribution due to higher order correlations is significantly smaller than errors introduced by other approximations. Approximation 2 has been tested by comparing extensive Monte Carlo simulations of ideal nonpolar solutes composed of nonpolar (methane) atoms within bonding distance.[48] We found that the proximity approximation introduced by Ashbaugh and Paulatis[49] reproduced the radial and angular distribution of water around these solutes. Two proximity approximations, proximity-1 (where only the nearest nonpolar neighbor determines the contribution from all nonpolar atoms) and proximity-2 (where two nearest nonpolar neighbors determine the contribution to the local hydration from all nonpolar atoms) gave similar results. Proximity-1 is the simplest of these approximations and was used in our PMF calculations.

The PMF description of the structural hydration and radial distribution functions was accurate. The PMF overestimates the extent of local water density near the C-terminus of the α-helix. However, the radial distribution functions are qualitatively and quantitatively in agreement with the MD simulation results. In the MD simulations the carbonyl group (CO) localizes water over a much narrower solid angle around the CO bond that can be accounted for by the bulk water correlation function. This discrepancy in the local water densities can be attributed to approximations 2 and 3 concerning the number of atom-types considered in the calculations. Both approximations can, in principle, be removed by includ-

ing correlations between more nonpolar groups or by explicitly including the C-O multiple point correlation functions.

The effects of dynamical fluctuations and ions, approximation 4, have been avoided by studying rigid systems without ions. When comparing PMF results with crystallographic data, the effect of hydrogen delocalization due to bond rotations can be accounted for by neglecting the directional O-H and N-H contributions to the local density, contributed by the $\chi_{O-H\cdots O}$ correlation function in Eq. (7). Large-scale atomic fluctuations involving large groups of atoms, characteristic of proteins, cannot be easily included. However, the local structure (atoms within 5–7 Å) determines most of the hydration structure. On this scale, positional fluctuations of protein atoms is mostly limited to a fraction of an Ångstrom. Thus, the structural hydration is, in most cases, only slightly affected by protein dynamics.

CONCLUSIONS

We have presented a robust and accurate method of describing the local density of water around any biomolecule. This method is based on the statistical mechanics of liquids. The accuracy of the method has been tested directly by studying a model peptide. The PMF approach takes as input pair- and triplet-correlation functions of bulk water and a simple nonpolar solute in water. Although the calculation of these correlation functions requires extensive simulations of bulk water or dilute solute-water systems, these functions are calculated only once. Thereafter, they are tabulated and used for all subsequent calculations. Additional correlation functions can be added when necessary. The PMF approach truncated to triplet-correlation functions involves O(N) opera-

tions per density point, where N is the number of atoms in the biomolecule. The PMF method is also local, i.e., we can calculate the water density at a point of interest without calculating the density elsewhere since the correlations among various points on a biomolecule's surface are included in the correlation functions used to calculate the local density. Simulation methods involve thousands of O(N^2) iterations, and the correlations among water molecules are included by way of explicit configurational averaging. These features make the PMF computation dramatically more efficient. The PMF method is particularly useful for calculations of large biomolecular systems (i.e., t-RNA, photosynthetic reaction center, antigen-antibody complexes, protein-DNA complexes) where high-resolution structures are available but the extensive simulations required to obtain moderate accuracy in the local water density are not yet possible.[55] The efficiency of the PMF method also allows for testing the role of single point mutations in the localization of water molecules and for studying the hydration of the active sites of large proteins.[56]

ACKNOWLEDGMENTS

REFERENCES

1. Nemethy, G., Scheraga, H.A. The structure of water and hydrophobic bonding in proteins. III. The thermodynamic properties of hydrophobic bonds in proteins. J. Phys. Chem. 66:1773–1789, 1962.
2. Singer, P.T., Smalas, A., Carty, R.P., Mangel, W.F., Sweet, R.M. The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. Science 259:669–673, 1993.
3. Kuntz Jr., I.D., Kauzmann, W. Hydration of proteins and polypeptides. Adv. Protein Chem. 28:239–345, 1974.
4. Finney, J.L., Gellatly, B.J., Golton, I.C., Goodfellow, J. Solvent effects and polar interactions in the structural stability and dynamics of globular proteins. Biophys. J. 32:17–33, 1990.
5. Savage, H.F.J. Water-structure in crystalline solids: Ices to proteins. Water Sci. Rev. 2:67–148, 1986.
6. Saenger, W. Structure and dynamics of water surrounding biomolecules. Annu. Rev. Biophys. Biophys. Chem. 16:93–114, 1987.
7. Rupley, J.A., Careri, G. Protein hydration and function. Adv. Protein Chem. 41:37–172, 1991.
8. Teeter, M.M. Water-protein interactions—theory and experiment. Annu. Rev. Biophys. Biophys. Chem. 20:577–600, 1991.
9. Meyer, E. Internal water molecules and H-bonding in biological macromolecules: A review of structural features with functional implications. Protein Sci. 1:1543–1562, 1992.
10. Schoenborn, B.P., Garcia, A.E., Knott, R. Hydration in protein crystallography. Prog. Rep. Biophys. 64:105–119, 1996.
11. Cheng, X., Schoenborn, B.P. Hydration in protein crystals: A neutron diffraction analysis of carbonmonoxymyoglobin. Acta Crystallogr. B 46:195–208, 1990.
12. Otting, G., Liepinsh, E., Wüthrich, K. Protein hydration in aqueous solution. Science 254:974–980, 1991.
13. Kubinec, M.G., Wemmer, D.E. NMR evidence for DNA bound water in solution. J. Am. Chem. Soc. 114:8739–8740, 1992.
14. Thanki, N., Thornton, J.M., Goodfellow, J.M. Distributions of water around amino acid residues in proteins. J. Mol. Biol. 202:637–657, 1988.
15. Thanki, N., Thornton, J.M., Goodfellow, J.M. Influence of secondary structure on the hydration of serine, threonine and tyrosine residues in proteins. Protein Eng. 3:495–508, 1990.
16. Thanki, N., Umrania, Y., Thornton, J.M., Goodfellow, J.M. Analysis of protein main-chain solvation as a function of secondary structure. J. Mol. Biol. 221:669–691, 1991.
17. Roe, S.M., Teeter, M.M. Patterns for prediction of hydration around polar residues in proteins. J. Mol. Biol. 229:419–427, 1993.
18. Savage, H.F.J., Finney, J.L. Repulsive regularities of water structure in ices and crystalline hydrates. Nature 322:717–720, 1986.
19. Clementi, E., Corongiu, G. Interactions of water with DNA single-helix in the A conformation. Biopolymers 18:2431–2450, 1979.
20. Parak, F., Hartmann, H., Schmidt, M., Corongiu, G., Clementi, E. The hydration shell of myoglobin. Eur. Biophys. J. 21:313–320, 1992.
21. Steinbach, P.J., Brooks, B.R. Protein hydration elucidated by molecular dynamics simulation. Proc. Natl. Acad. Sci. USA 90:9135–9139, 1993.
22. Loncharich, R.J., Brooks, B.R. Temperature dependence of dynamics of hydrated myoglobin. J. Mol. Biol. 215:435–455, 1990.
23. Levitt, M., Sharon, R. Accurate simulations of protein dynamics in solution. Proc. Natl. Acad. Sci. USA 85:7557–7561, 1988.
24. Levitt, M. Molecular dynamics of macromolecules in water. Chem. Scripta, 29a:197–203, 1989.
25. García, A.E., Stiller, L. Computation of the mean residence time of water in the hydration shells of biomolecules. J. Comp. Chem. 14:1396–1406, 1993.
26. Brunne, R.M., Liepinsh, E., Otting, G., Wüthrich, K., van Gunsteren, W.F. Hydration of proteins: A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. J. Mol. Biol. 231:1040–1048, 1993.
27. Gerstein, M., Lynden-Bell, R.M. Simulation of water around a model protein helix. 1. Two-dimensional projections of solvent structure. J. Phys. Chem. 97:2982–2990, 1993.
28. Gerstein, M., Lynden-Bell, R.M. Simulation of water around a model protein helix. 2. The relative contributions of packing, hydrophobicity, and hydrogen bonding. J. Phys. Chem. 97:2991–2999, 1993.
29. Vedani, A., Huhta, D.W. An algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. J. Am. Chem. Soc. 113:5860–5862, 1991.
30. Pitt, W.R., Murray-Rust, J., Goodfellow, J. Aquarius 2: Knowledge-based modeling of solvent sites around proteins. J. Comp. Chem. 14:1007–1018, 1993.
31. Hummer, G., Soumpasis, D.M. Computation of the water density distribution at the ice-water interface using the potentials-of-mean-force expansion. Phys. Rev. E 49:591–596, 1994.
32. Hummer, G., Soumpasis, D.M. A new approach to calculate the hydration of DNA molecules. In: "Structural Biology: The State of the Art; Proceedings of the Eighth Conversations in the Discipline of Biomolecular Stereodynamics. Vol. 2. Sarma, R.H., Sarma, M.H. (eds.). Schenectady, NY: Adenine Press, 1994:273–278.
33. Hummer, G., Soumpasis, D.M. Statistical mechanical treatment of the structural hydration of biological macromolecules: Results for B-DNA. Phys. Rev. E 50:5085–5095, 1994.
34. Hummer, G., Soumpasis, D.M., Garcia, A.E. Potential-of-mean-force description of ionic interactions and structural hydration in biomolecular systems. In: Nonlinear Excitations in Biomolecules. Peyrard, M. (ed.). Les Editions de Physique. Berlin: Springer, 1995:83–99.
35. Garcia, A.E., Hummer, G., Soumpasis, D.M. Theoretical

description of biomolecular hydration—application to A-DNA. In: Neutrons in Biology. Schoenborn, B.P., and Knott, R. (eds) Plenum Press, New York, 1997.

36. Hummer, G., García, A.E., Soumpasis, D.M. Hydration of nucleic-acid fragments: Comparison of theory and experiment for high-resolution crystal structures of RNA, DNA, and DNA-drug complexes. Biophys. J 68:1639–1652, 1995.

37. Forester, T.R., McDonald, I.R. Molecular dynamics studies of the behaviour of water molecules and small ions in concentrated solutions of polymeric B-DNA. Mol. Phys. 72:643–660, 1991.

38. Forsyth, V.T., Mahendrasingam, A., Pigram, W.J., Greenall, R.J., Bellamy, K., Fuller, W., Mason, S.A. Neutron fibre diffraction study of DNA hydration. Int. J. Biol. Macromol. 11:236–240, 1989.

39. Jorgensen, W.L., Tirado-Rives, J. The OPLS potential energy functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110:1657–1666, 1988.

40. Kirkwood, J.G., Statistical mechanics of fluid mixtures. J. Chem. Phys. 3:300–313, 1935.

41. Soumpasis, D.M. Statistical mechanics of the B-Z transition of DNA: Contribution of diffuse ionic interactions. Proc. Natl. Acad. Sci. USA 81:5116–5120, 1984.

42. García, A.E., Soumpasis, D.M. Harmonic vibrations and thermodynamic stability of a DNA oligomer in monovalent salt solution. Proc. Natl. Acad. Sci. USA 86:3160–3164, 1989.

43. García, A.E., Gupta, G., Soumpasis, D.M., Tung, C.S. Energetics of the hairpin to mismatched duplex transition of d(GCCGCAGC) on NaCl solution. J. Biomol. Struct. Dyn. 8:173–186, 1990.

44. Klement, R., Soumpasis, D.M., Jovin, T.M. Computation of ionic distributions around charged biomolecular structures: Results for right-handed and left-handed DNA. Proc. Natl. Acad. Sci. USA 88:4631–4635, 1991.

45. Chen, Y.Z., Prohofsky, E.W. Salt dependent premelting base pair opening probabilities of B and Z DNA poly[d(G-C)] and significance for the B-Z transition. Biophys. J. 64:1394–1397, 1993.

46. Soumpasis, D.M., Procacci, P., Corongiu, G. Part A: On the computation of N-particle correlations in classical fluids via computer simulations. Part B: Triplet correlations in NCC water. IBM DSD Report. IBM, October 1991.

47. Fisher, I.Z., Kopeliovich, B.L. On a refinement of the superposition approximation in the theory of fluids. Dokl. Akad. Nauk. SSSR, 133:81–83, 1960. [1960. Sov. Phys.-Dokl. 5:761–763].

48. Garde, S., Hummer, G., García, A.E., Pratt, L.R., Paulaitis, M.E. Hydrophobic hydration: Inhomogeneous water structure near nonpolar molecular solutes. Phys. Rev. E. 53: R4310–R4313, 1996.

49. Ashbaugh, H.S., Paulaitis, M.E. Entropy of hydrophobic hydration—extension to hydrophobic chains. J. Phys. Chem. 100:1900–1913, 1996.

50. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79:926–935, 1983.

51. Jorgensen, W.L., Madura, J.D., Swenson, C.J. Optimized intermolecular potential functions for liquid hydrocarbons. J. Am. Chem. Soc. 106:6638–6646, 1984.

52. Pratt, L.R., Hummer, G., García, A.E. Ion pair potentials-of-mean-force in water. Biophys. Chem. 51:147–165, 1994.

53. Hummer, G., Soumpasis, D.M. Correlations and free energies in restricted primitive model descriptions of electrolytes. J. Chem. Phys. 98:581–591, 1993.

54. Ryckaert, J.P., Cicotti, G., Berendsen, H.J.C. Numerical integration of the cartesian equations of motions of a system with constraints: Molecular dynamics of n-alkanes. J. Comp. Phys. 23:327–341, 1977.

55. Hummer, G., García, A.E., Soumpasis, D.M. A statistical mechanical description of biomolecular hydration. Faraday Discuss. Chem. Soc. 113, 1997 in press.

56. Oprea, T.I., Hummer, G., García, A.E. Identification of a functional water channel in cytochrome P450 enzymes. Proc. Natl. Acad. Sci. USA, 1997 in press.