UNIVERSITY OF CALIFORNIA
RIVERSIDE

Applications of Molecular Theory in Solvation of Pharmaceutical Solutes, Ions and Amine-Grafted Silica Gel

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemical and Environmental Engineering

by

Shijie Sheng

December 2017

Dissertation Committee:
Dr. Jianzhong Wu, Chairperson
Dr. Bryan M. Wong
Dr. Chia-En Chang

The Dissertation of Shijie Sheng is approved:

Committee Chairperson

University of California, Riverside

Acknowledgement

I feel grateful to many people for their great support over the years. First and foremost, I would like to thank my supervisor Prof. Jianzhong Wu for his superb and insightful guidance. Learned from him, I start to know how to think as a theoretician, to solve problems in an elegant way, and to enjoy the beauty of statistical mechanics. His strong enthusiasm motivates me and shows an excellent example on how to be a successful chemical engineer and professor.

I sincerely appreciate the members of my PhD committee, Prof. Bryan M. Wong and Prof. Chia-en A. Chang, for their great advices and help in my qualifying exam and final defense. I also wish to express my special thanks to Prof. Shuangliang Zhao, Dr. Jia Fu, Dr. Yun Tian, Dr. Yu Liu, Dr. Justin Neal and Dr. Jehoon Kim for their valuable support during my PhD studying. Also acknowledged are all the good friends that I have made in Riverside for their encouragement throughout this special period of my life.

What's more, I greatly appreciate the help of my parents and my girlfriend during this long journey, who support and encourage me more than anyone else to pursue an academic career.

This dissertation contains certain materials that have already been published in the Molecular Physics (Volume 114, Issue 16-17, Page 2351-2363, February 2016), Journal of Physical Chemistry B (Volume 128, Issue 28, Page 6898-6908, June 2017), Journal of Physical Chemistry C (Volume 121, Issue 40, Page 22156-22163, September 2017).

ABSTRACT OF THE DISSERTATION

Applications of Molecular Theory in Solvation of Pharmaceutical Solutes, Ions and Amine-Grafted Silica Gel

by

Shijie Sheng

Doctor of Philosophy, Graduate Program in Chemical and Environmental Engineering
University of California, Riverside, December 2017
Dr. Jianzhong Wu, Chairperson

Solvation and solvent effects play an important role in diverse chemical processes ranging from reaction kinetics to molecular recognition, solubility, solvato- chromism and phase separations. Despite enormous activities in this field, quantitative solvation calculations remain an enormous intellectual challenge.

My thesis is focused on development and application of molecular density functional theory (MDFT) and molecular dynamics (MD) simulation to predicting solvation properties. Accomplishments include 1) improved the average unsigned error of MDFT predictions for the room-temperature solvation free energies (SFE) of 504 pharmaceutical molecules in water from 1.04 kcal/mol to 0.66 kcal/mol; 2) established a more reliable numerical procedure to calculate the direct correlation functions (DCF) of solvent from MD simulations; 3) extended MDFT prediction of SFE to different temperatures and calibrated the theoretical results with experimental data for the hydration free energies of 5 nitrotolunenes and a library of 197 solutes at 277 K, 298 K

and 313 K. In addition, I investigated the 3-dimensional (3D) solvation structure of amine-grafted silica gel in liquid water by applying a spherical harmonics expansion method to the MD trajectories. The simulation results provide evidence on the strong influence of the silica surface on hydration structure, which is often ignored in the theoretical analysis of surface reactions. Furthermore, I developed a hybrid method for predicting the SFE of spherical ions by combining MDFT with MD simulations. The numerical analysis justifies the universality of the bridge functional that can be reasonably approximated by the modified fundamental measure theory (MFMT) for hard-sphere systems.

Results from this thesis demonstrate that the DCFs are important in application of MDFT to SFE predictions. Based DCF from on integral-equation methods, MDFT can also capture the temperature effect on SFE in good agreement with experiment. In addition, the hybrid MDFT-MD method provides accurate predictions of hydration free energies for charged solutes and the numerical analysis sheds light on future theoretical development. The efficient sampling method for generating 3D density profiles from MD may open up opportunities for application of MDFT to more complex systems, for example, protein solvation and enzyme kinetics. By studying the solvation structure of amine-grafted silica shell, I found that the silica surface affects not only the distribution of surrounding water but also the hydrogen-bonding network. This surface effect is long-ranged and can be reduced with longer grafted amine chains.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Molecular Theories of Solvation

Solvent effects are fundamental to understanding chemical processes in many ways. The presence of a solvent can alter the color of a solute, which is known as solvato-chromism. Marini et al.[1] explained this phenomenon with time-dependent density functional theory (TDDFT) calculations. The formation of hydrogen bonds with solvent molecules changes the configurations of proteins, which is important, for example, for understanding protein folding. Piana et al[2] tested 4 common force fields on protein folding and demonstrated that molecular dynamics (MD) simulation could capture the solvent behavior. The polarity of solvent molecules could not only affect solvent-solute interactions but also change the rates of chemical reactions due to its strong influence on electronic structure of the solute. Polarizable continuum model (PCM) was first introduced into ab initio calculations by Foresman et al[3]. Recent theoretical interest has been mostly devoted to understanding the microscopic details of solute-solvent interactions and the effects of local solvent structure on the chemical and biochemical affinities of dissolved species. Such information is key to studying the chemical and biochemical processes in organic or aqueous systems including chemical reactions and relation dynamics, stability of bio-macromolecules, and 'lock-key' interactions for rational drug design.

Despite enormous activities in this field, it remains a theoretical challenge in some quantitative calculations, for example, in high-throughput solvent screening and

predictions of solvation free energies. Approximately, molecular theories to investigate SFE may be classified into three categories: continuous approaches, molecular simulations, and liquid-state theories. With solvent molecules depicted as a dielectric medium, a continuous model describes the solvation free energy in terms of the geometric measures of the solute-solvent boundary, such as the solute size, solvent-accessible surface area, surface curvature and various energetic contributions due to solute-solvent electrostatic and van der Waals interactions. Neglecting the microscopic details of solvent molecules makes the phenomenological approach computationally extremely efficient and thus convenient for practical applications.

Simulation methods have been a popular choice to study solvent effects over the past few decades. Nowadays many molecular simulation packages (LAMMPS, GROMACS, AMBER, NAMD) are readily available. The simulation results can be analyzed with many open-source tools in this large community. The development of general force fields (e.g., GROMOS, OPLS and GAFF) further widens the applications of simulation methods. However, in spite of the growing of computation power, simulations of SFE with the standard thermo-integration method are computationally demanding. As a result, simulation results for the SFE of a large library of chemicals are rarely reported.

Liquid-state theories represent a compromise between continuum models and molecular simulations. Next section will provide a brief introduction of the liquid-state method used in this thesis.

1.2. Molecular Density Functional Theory (MDFT)

MDFT is established on the Hohenberg-Kohn-Mermin (HKM) theorem. The basic idea is that the free energy can be expressed as a functional of the atomic/site density profile. In the grand canonical ensemble, the grand potential is given by

$$\Omega_v = \Omega_v\left[\rho_i(\mathbf{r})\right] \tag{1}$$

where ρ_i is the one body density of site i at position $\mathbf{r}$. Conventionally, the grand potential is split into three terms

$$\Omega_v = F^{id}\left[\rho_i(\mathbf{r})\right] + \sum_i \int d\mathbf{r}\left[V_i^{ext}(\mathbf{r}) - \mu_i\right]\rho_i(\mathbf{r}) + F^{ex}\left[\rho_i(\mathbf{r})\right] \tag{2}$$

where F^{id} is the intrinsic Helmholtz free energy for an ideal-gas system of the same atomic density profiles; V_i^{ext} is the external potential; μ_i is the chemical potential; and F^{ex} is the excess intrinsic Helmholtz free energy. The ideal-gas term can be derived as

$$F^{id} = k_B T \int d\mathbf{R}\left(\ln \rho_M \Lambda_M^3 - 1\right)\rho_M + \int W \rho_M \, d\mathbf{R} \tag{3}$$

Similar to ideal-gas term in atomic system, $\rho_M(\mathbf{R})$ is the molecular density at configuration $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots)$; W is the intra-molecular potential; and Λ_M is the molecular thermal wavelength. Unlike atomic system, there is an extra term accounting for bonding energy.

An exact formula for the excess free energy F^{ex} of an inhomogeneous molecular system is not available. In practical applications, we seek to approximate this functional

with perturbation methods or by expanding the functional with respect to site density ρ_i at the homogeneous condition,

$$F^{ex} = F_b^{ex} + \sum_i \int d\mathbf{r} \mu_{i,b}^{ex} \Delta\rho_i(\mathbf{r}) - \frac{1}{2}\sum_{i,j} \int d\mathbf{r} \int d\mathbf{r}' c_{ij}(\mathbf{r}-\mathbf{r}') \Delta\rho_i(\mathbf{r}) \Delta\rho_j(\mathbf{r}') + F_B \quad (4)$$

where subscript b stands for bulk properties, $\Delta\rho_i(\mathbf{r}) = \rho_i(\mathbf{r}) - \rho_{b,i}$, c_{ij} is the direct correlation functions of site i and j, and F_B is the bridge functional. Effectively, the bridge functional represents all higher-order terms. The detail discussion of DCF is given in the next two Chapters. According to the universality ansatz of bridge functional, F_B can be approximated by that of a reference hard-sphere system with same density profiles. In this work, the bridge functional is formulated with the modified fundamental measure theory (MFMT)[4]:

$$F_B \simeq F_B^{MFMT} \quad (5)$$

1.3. Dissertation Organization

This dissertation is focused on the application of molecular density functional theory (MDFT) and molecular dynamics (MD) simulations to solvation calculations. Special attention is given to hydration of small pharmaceutical solutes, ions and amine-grafted silica gels. The rest of dissertation is organized as follows.

Chapter 2 describes the numerical procedure for sampling the site-site direct correlation functions (DCF) for 3-site and 4-site water models. It is demonstrated that accurate evaluation of DCF improves the theoretical performance in application of

MDFT to predicting the solvation free energies (SFE). In the Appendix for Chapter 2, I provide a generalized method for sampling DCFs of solvents with arbitrary molecular configurations.

Chapter 3 incorporates DCFs into MDFT and applies it to solvation systems at different temperatures. The theoretical method is calibrated with experimental data for 5 nitro-toluenes and 197 solutes at three temperatures

Chapter 4 examines the surface effect on the solvation structure of amine-grafted silica amines. Both the pH of solution and chain length of grafted amines are investigated. In sampling 3-dimentional density profiles, I used an efficient sampling method based on the spherical-harmonics expansion.

Chapter 5 provides a hybrid method for SFE calculation by combing MDFT and MD simulation. This method is firstly tested with water self-solvation to reproduce the experiment value and further calibrated with the hydration free energies of 32 ions of different valences. With the 3D density profiles as input, I demonstrate that the bridge functional is essential for quantitative prediction of hydration free energies. The numerical results may shed light on future liquid-state theory development.

Finally, in Chapter 6 I summarize the main conclusions of the thesis and provide perspective for future work.

Chapter 2. Direct Correlation Functions for Three-site and Four-site Water Models

Direct correlation functions (DCFs) play a pivotal role in modern liquid-state theories and are often indispensable for non-mean-field implementation of the classical density functional theory (cDFT). Whereas analytical expressions have been derived for the DCFs of simple fluids and electrolytes, DCFs for molecular systems are attainable only through numerical solution of the integral-equation theories in combination with molecular simulation or an approximate closure. Unlike radial distribution functions (RDFs), DCFs reflect the variational of the local chemical potential of individual atoms/interaction sites with the density-profile fluctuations, which are difficult be sampled from the simulation trajectory. This article presents an improved numerical procedure to calculate the DCFs of 3-site (SPC/E, TIP3P) and 4-site (TIP4P-Ew) water models based on the Reference Interaction Site Model (RISM) and molecular dynamics (MD) simulations. In combination with the modified fundamental measure theory for the bridge functionals, the DCFs have been utilized to predict the hydration free energies of 504 small organic molecules and yield an average unsigned error of 0.66 kcal/mol in comparison with the simulation data, better than that (~1 kcal/mol) reported in previous cDFT calculations.

2.1. Direct Correlation Functions (DCF)

In classical density functional theory (cDFT) for molecular fluids[5], the grand potential Ω of a one-component open system at temperature T, volume V and chemical potential μ is defined as a functional of the molecular density profile, $\rho(\mathbf{x})$:

$$\Omega[\rho(\mathbf{x});\mu,V,T] = k_B T \int d\mathbf{x}\rho(\mathbf{x})\{\ln[\rho(\mathbf{x})\Lambda^3]-1\} + \int d\mathbf{x}[\Psi^{ext}(\mathbf{x})-\mu]\rho(\mathbf{x}) + F^{ex}[\rho(\mathbf{x})] \quad (6)$$

where $\mathbf{x}$ is a composite vector specifying the molecular configuration (*i.e.,* the atomic positions of individual molecules), k_B stands for the Boltzmann constant, and Λ is an *effective* thermal wavelength. In Eq.(6), the first term on the right side corresponds to the *intrinsic* Helmholtz energy of an ideal system with the molecular density profile the same as that of the real system under consideration; the second term accounts for a one-body energy for each molecule due to external potential $\Psi^{ext}(\mathbf{x})$; and the last term, $F^{ex}[\rho(\mathbf{x})]$, defines the excess intrinsic Helmholtz energy, *i.e*., the deviation from that of the ideal system due to intermolecular interactions. At equilibrium, the grand potential is minimized with respect to the molecular density profile, leading to the Euler-Lagrange equation for the molecular density profile

$$\rho(\mathbf{x}) \sim \exp[-\beta\Psi^{ext}(\mathbf{x}) - \delta\beta F^{\mathrm{ex}} / \delta\rho(\mathbf{x})] \quad (7)$$

where $\beta = 1/(k_{\mathrm{B}}T)$, and the proportionality constant is fixed by the average molecular density of a bulk system at the same temperature and chemical potential. Eq.(7) is formally exact. With analytical expressions for the external potential and the excess intrinsic Helmholtz energy, I can calculate the molecular density profile and, subsequently, all thermodynamic properties of the molecular system under consideration[6-7].

To a good approximation, the molecular configuration is fixed for a system of polyatomic molecules with low molecular weight so that the molecular density profile

can be expressed as a function of the molecular position and orientation[8]. For flexible polyatomic molecules in general, the high-dimensional function is most conveniently expressed in terms of the atomic (or site) density profiles[9-10]

$$\rho_i(\mathbf{r}) = \rho_i^0 \exp[-\lambda_i(\mathbf{r})] < \exp\{-\sum_{j\neq i} \lambda_j(\mathbf{r}_j)\} >_{\mathbf{x}'} \tag{8}$$

where ρ_i^0 is the number density of atom i in the bulk, $<\cdots>_{\mathbf{x}'}$ represents the Boltzmann average over the molecular configurations with the position of atom i fixed at $\mathbf{r}$, and $\lambda_i(\mathbf{r})$ corresponds to a reduced one-body potential for atom i. This one-body potential arises from the external field, here designated as $\varphi_i^{\text{ext}}(\mathbf{r})$, and the deviation of the local excess chemical potential from its bulk value:

$$\lambda_i(\mathbf{r}) = \beta\varphi_i^{\text{ext}}(\mathbf{r}) + \delta\beta F^{\text{ex}} / \delta\rho_i(\mathbf{r}) - \beta\mu_i^{\text{ex}} \tag{9}$$

where μ_i^{ex} stands for the atomic excess chemical potential in the bulk. For a uniform polyatomic system at given temperature and chemical potential, both the atomic density and the atomic excess chemical potential can be fixed from the stoichiometric relations[10]. In writing Eq.(9), I assume that the external molecular potential can be formally decomposed into contributions from the individual atoms

$$\Psi^{ext}(\mathbf{x}) = \sum_i \varphi_i(\mathbf{r}_i)\,. \tag{10}$$

Eq.(10) is fully consistent with the pairwise additivity assumption as typically adopted in a conventional non-polarizable molecular force field.

The essential task in applications of the cDFT is to formulate an analytical expression for the excess *intrinsic* Helmholtz energy as a functional of atomic or molecular density profiles. Without loss of generality, I may express the unknown excess *intrinsic* Helmholtz energy relative to that of a bulk system of the same temperature, volume, and chemical potential

$$F^{ex}[\rho_i(\mathbf{r})] = F^{ex}[\rho_i^{\,0}] + \sum_i \mu_i^{ex} \int d\mathbf{r} \Delta\rho_i(\mathbf{r}) - \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r} \int d\mathbf{r}' \Delta\rho_i(\mathbf{r}) \Delta\rho_j(\mathbf{r}') c_{ij}(|\mathbf{r}-\mathbf{r}'|) + F_B[\rho_i(\mathbf{r})] \tag{11}$$

In Eq.(11), $\Delta\rho_i(\mathbf{r}) \equiv \rho_i(\mathbf{r}) - \rho_i^0$ represents the deviation of the local atomic density from the bulk value; $c_{ij}(r)$ denotes the direct correlation function (DCF) between atoms *i* and *j* for the bulk system, and $F_B[\rho_i(\mathbf{r})]$ stands for the bridge functional, *i.e.,* all higher-order terms beyond the quadratic functional expansion.

Even though the site-site DCFs appeared in Eq.(11) are independent to molecular orientations, the cDFT procedure outlined above is formally exact, immaterial to the specific forms of the intermolecular potential. In cDFT, the orientational effects are implicitly accounted for in computation of the atomic density profiles as shown in Eq.(8). In other words, cDFT is able to describe angular correlations that are intrinsically neglected in the reference interaction site model (RISM). Mathematically, the first three terms on the right side of Eq.(11) correspond to a functional quadratic expansion of the excess *intrinsic* Helmholtz energy, and the bridge functional accounts for all high-order terms in the functional Taylor expansion. Without the bridge functional, Eq. (11) is often

referred to as the hypernetted chain approximation (HNC)[11], the homogeneous reference fluid (HRF) approximation[12], or the Chandler-McCoy-Singer (CMS) theory[13]. It has been shown that the quadratic expansion conforms to the exact results at both the "ideal gas" and "ideal liquid" limits for the long-range components of the intermolecular potentials[14]. Besides, it has been well documented that HNC (and similar approximations) performs rather well for systems in which thermodynamic non-ideality is dominated by long-range interactions. The good performance for HNC to account for long-range interactions implies that the bridge functional is mostly affiliated with short-range interactions, insensitive to the mathematical details of the long-range components of intermolecular forces[15]. The insensitivity of the bridge functional to the precise form of intermolecular potentials was recognized first by Rosenfeld for simple fluids[16-18]. Using the fundamental measure theory (FMT) as an input, Rosenfeld demonstrated that *the universality of bridge functional* performs well for a wide variety of thermodynamic systems, including charged Yukawa fluids and plasma mixtures (see ref[18] for an overview).

The universality *ansatz* allows us to represent the bridge functional with that of a hard-sphere (HS) reference system

$$\begin{aligned} F_B[\rho_i(\mathbf{r})] \approx F^{HS}[\rho_i(\mathbf{r})] - F^{HS}(\rho_i^0) - \sum_i \mu_i^{HS} \int d\mathbf{r} \Delta\rho_i(\mathbf{r}) \\ + \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r}_1 \int d\mathbf{r}_2 \Delta\rho_i(\mathbf{r}_1) \Delta\rho_i(\mathbf{r}_2) c_{i,j}^{HS}(|\mathbf{r}_1 - \mathbf{r}_2|) \end{aligned} \tag{12}$$

where $F^{HS}[\rho_i(\mathbf{r})]$ is the excess Helmholtz energy functional of the HS system; $\mu_{0,i}^{HS}$ and $F_{HS}^{ex}(\rho_i^0)$ are, respectively, the excess chemical potential and the excess Helmholtz energy

of the HS system at bulk density ρ_i^0; and $c_{ij}^{HS}(|\mathbf{r}_1-\mathbf{r}_2|)$ is the HS direct correlation functions in the bulk. In the previous work[4, 19-20], I have developed accurate expressions for the excess Helmholtz energy functional and the direct correlation functions of inhomogeneous hard-sphere systems.

With a good expression for the bridge functional at hand, an important task in application of the cDFT is to obtain the direct correlation functions (DCFs) of uniform bulk systems. Historically, DCFs were introduced within the framework of the integral-equation theories following the diagrammatic expansion of the grand partition function[15]. Within the framework of cDFT, DCFs can be succinctly expressed as the second-order functional derivatives of the reduced excess intrinsic Helmholtz energy[13]

$$c_{ij}(\mathbf{r},\mathbf{r}') = -\frac{\delta^2 \beta F^{ex}}{\delta\rho_i(\mathbf{r})\delta\rho_j(\mathbf{r}')} \quad . \tag{13}$$

For an isotropic system, the density profiles are uniform and the functional derivative takes the uniform limit. In that case, the DCF depends only on the center-to-center distance between atoms, *i.e.*, $c_{ij}(\mathbf{r},\mathbf{r}') = c_{ij}(|\mathbf{r}-\mathbf{r}'|;\rho_i^0)$. As the local excess chemical potential corresponds to the first-order functional derivative of the excess intrinsic Helmholtz energy, $c_{ij}(\mathbf{r},\mathbf{r}')$ reflects the change in the one-body potential for atom *i* at position $\mathbf{r}$ in response to a variation of the local density of atom *j* at position $\mathbf{r}'$ or vice versa.

For a uniform system, the responses in the one-body atomic potential are formally connected with the two-body density-density correlation functions through the Reference Interaction Site Model (RISM)[21]

$$\hat{c}(k) = \hat{\omega}^{-1}(k)\hat{h}(k)\left[\hat{\omega}(k) + \rho_b \hat{h}(k)\right]^{-1} \tag{14}$$

where ρ_b stands for the molecular number density of the bulk system; $\hat{h}, \hat{\omega}, \hat{c}$ represent, respectively, the matrix forms of the site-site total correlation functions (TCFs), intramolecular correlation functions (ICFs), and DCFs; and k is the wavenumber (*i.e.*, the magnitude of the wave vector) introduced from the 3-dimensional Fourier transforms of the two-body correlation functions. For a polyatomic system containing identical rigid molecules, ICFs are known exactly, independent of the thermodynamic conditions of the system,

$$\omega_{ij}(k) = \delta_{ij} + \left(1 - \delta_{ij}\right)\frac{\sin kl_{ij}}{kl_{ij}} \tag{15}$$

where δ_{ij} is the Kronecker delta function, l_{ij} is the distance between atoms *i* and *j*. TCFs are related to the site-site radial distribution functions (RDFs), $g_{ij}(r)$,

$$h_{ij}(r) = g_{ij}(r) - 1 \tag{16}$$

If each molecule is represented by a spherical particle, there is only one element in each correlation function matrix and $\hat{\omega}(k) = 1$. In that case, the RISM equation reduces to the Ornstein-Zernike equation (OZ) for simple fluids.

I may follow two complementary procedures to calculate the DCFs of a uniform system. One is the so-called "closure" approach, which introduces an additional relation among the total and direct correlation functions. Like those for the OZ equation for simple fluids, the RISM "closures" can be partially justified with diagrammatic expansions of the total and direct correlation functions [22-24]. Alternatively, I can evaluate the DCFs by combining the RISM equation with molecular simulations[25-26]. While the closure approach has the advantage of numerical efficiency, most RISM closures fail to account for the effects of intramolecular correlations and may lead to erroneous predictions of the dielectric constant[27]. Despite the caveat, the closure approach provides a reasonable description of the TCFs of polyatomic systems and has been a popular choice in the literature[28-29]. Conversely, molecular simulation provides an accurate description of TCFs but has numerical issues in calculation of the DCFs from the RISM equation. In previous publications[25, 30], I proposed a hybrid procedure to calculate the DCFs by combing an asymptotic analysis of the two-body correlation functions with molecular dynamics (MD) simulations. The main purpose of this work is to further improve the hybrid procedure for generating the DCFs of different water models and to calibrate their utility in cDFT predictions of hydration free energies for a large library of small organic molecules. The DCFs for both 3-site and 4 site models are compared with those predicted by a modified HNC (MHNC) approximation.

The rest of this article is organized as follows. Section II describes the asymptotic behavior for the long-range components of DCFs. In Section III, I introduce an improved procedure to calculate the DCFs by a combination of MD simulation, asymptotic analysis,

and the RISM equation. In Section IV, the DCFs from the hybrid method are compared with those from the RISM/MHNC integral-equation theory. In Section V, the numerical performance of the new DCFs for cDFT calculations was tested with MD simulation data for the hydration free energies of 504 small molecules. Finally, Section VI summarizes the main conclusions from this work and offers some perspectives for future developments.

2.2. Long-range Behavior of DCFs

At large distance, the DCFs of simple fluids are known to be proportional to the reduced pair potential [31]

$$c_{ij}(r) = -\beta u_{ij}(r), \quad \text{when } r \to \infty \tag{17}$$

where $u_{ij}(r)$ is the pair potential between particles i and j. Because of intramolecular correlations, a similar asymptotic relation for molecular systems would lead to a bulk dielectric constant the same as that of an ideal gas [32-33]. Such a trivial result for molecular systems can be avoided by a semi-empirical modification of the long-range component of the DCFs [25, 34-35]

$$c_{ij}(r) = -A_{ij} / r, \quad \text{when } r \to \infty \tag{18}$$

where A_{ij} are coefficients chosen to be consistent with the bulk dielectric constant. Strictly speaking, the simple scaling of the electrostatic interactions is not valid for systems containing polyatomic molecules with highly symmetric structures (e.g., linear triatomic molecules such as CO_2). Neither is Eq.(18) applicable to solutions with a solute

of finite concentration in a molecular solvent. In the latter case, the asymptotic behavior of the DCFs can be captured with the dielectrically consistent reference interaction site model (DRISM) theory[36]. In this work, I are concerned primarily with a pure liquid (*viz*. water) and solvation of individual solute molecules. For a 3-site water model, it has been shown that the long-range component of the DCFs is proportional to the reduced Coulomb potential[25]

$$A_{ij} = \zeta \frac{\beta q_i q_j}{4\pi\epsilon_o} \tag{19}$$

where q stands for the atomic charge, ϵ_o is the permittivity in vacuum, and the pre-factor ζ is related to static dielectric constant ϵ and molecular dipole moment d

$$\zeta = 1 + \frac{1}{\epsilon - 1} - \frac{1}{3y} \tag{20}$$

and $y = 4\pi\rho_b d^2 / (9k_B T)$. Eqs.(19) and (20) are not applicable to 4-site water models like TIP4P-Ew. Due to the increase in the dimensionality of the two-body correlation function matrices, there is no generic relationship between coefficients A_{ij} and the bulk dielectric constant ϵ. In that case, A_{ij} may be calculated through numerical extrapolations as detailed in Section III.

2.3. Evaluation of DCFs through MD Simulations

The TCFs for liquid water can be calculated from MD simulations. While MD simulation is presumed to be reliable for TCFs in the real space, direct Fourier transforms

of TCFs are numerically problematic owing to the long-range nature of integration and the limited box size used in MD simulations. In particular, molecular simulation provides little information on TCFs in the Fourier space at the long-wavelength limit, i.e., as wavenumber k→0.

The near zero *k* behavior of TCFs is affiliated with the compressibility and the dielectric constant of the bulk system[15]. Two complementary procedures can be used to calculate the long-wavelength limit. One procedure, first proposed by Salacuse[37], is to remove the finite size effects by adding a correction term to TCFs in the Fourier space

$$h_{ij}(k,R) = h_{ij}^{MD}(k,R) + \frac{4\pi R^3\left[1+\rho_b h_{ij}^{MD}(0,R)\right]}{3N - 4\pi\rho_b R^3}\frac{3j_1(kR)}{kR} \tag{21}$$

where $j_1(kR)$ is the spherical Bessel function of the first kind, N is the total number of water molecules within the cutoff, and $h_{ij}^{MD}(k,R)$ is the Fourier integration of TCF within the cutoff distance R

$$h_{ij}^{MD}(k,R) \equiv 4\pi\int_0^R dr\, r^2\left[\frac{\sin(kr)}{kr}\right]h_{ij}(r) \tag{22}$$

Because the size correction depends on the cutoff distance R, a large simulation box is typically required to get converged results for the TCFs. Another way to calculate the long-wavelength limit of TCFs is through MD sampling of the partial structure factors (PSFs) [38]

$$\overline{S}_{ij}(k) \equiv \frac{1}{N}\left\langle \sum_{M\neq M'} \exp\left[-i\mathbf{k}\cdot\left(\mathbf{r}_i^M - \mathbf{r}_j^{M'}\right)\right]\right\rangle \tag{23}$$

The PSFs and TCFs related through

$$\overline{S}_{ij}(k) = \rho_b h_{ij}(k) \tag{24}$$

In Eqs.(23) and (24), $\overline{S}_{ij}(k)$ represents the partial structure factor for a pair of atoms *i* and *j* (e.g., O and H atoms for liquid water); superscripts *M* and *M'* denote different molecules; $\mathbf{r}_i^M$ and $\mathbf{r}_j^{M'}$ are the local coordinates of the interaction sites. To simplify notation, I refer to the Fourier transform of TCFs directly from real space sampling as TCF-F, and those obtained from the partial structure factor as TCF-D. In this work, $h_{ij}(k)$ at small k values are obtained by sampling the partial structure factor because it is relatively easy to implement and has less numerical uncertainties.

The PSFs and TCFs at $k = 0$ cannot be calculated directly from MD simulation. In the limit of small *k* values ($k \to 0$), I may express the TCFs as a low-order polynomial[30]

$$h_{ij}(k) = h_{ij}(0) + h_{ij}^{(2)} k^2 + h_{ij}^{(4)} k^4 + \ldots \tag{25}$$

where $h_{ij}(0)$ is related to the isothermal compressibility χ_T of the bulk system[15]

$$h_{ij}(0) = k_B T \chi_T - 1/\rho_b \ . \tag{26}$$

As shown in our previous work[25, 30], the second-order coefficients in Eq. (25) can be related to the static dielectric constant ε and dipole moment d :

$$\Delta h^{(2)} \equiv \sum_{i,j} q_i q_j h_{ij}^{(2)} = \frac{1-1/\epsilon}{4\pi\beta\rho_b^2} - \frac{d^2}{3\rho_b} \tag{27}$$

Eq.(27) is applicable to both 3-site and 4-site water models. While the molecular dipole moment is model specific, the dielectric constant and density are bulk properties accessible via experiments. In this work, I use $\rho_b =$ 0.997 g/cm^3 and $\varepsilon = 78.4$ for bulk water at the ambient condition.

With the simulation results for $h_{ij}(k)$ at small k as the input, I can estimate the coefficients in the polynomial expansion. It is worth noting that the numerical values for these coefficients, and DCFs in subsequent inverse Fourier transform, are highly sensitive to the range of k values and the order of the polynomial used in the numerical fitting. To minimize numerical uncertainties, I fix $h_{ij}(0)$ based on the simulation results for the isothermal compressibility and the bulk density of the liquid water (see Table 2-1). The second-order coefficients are obtained by best linear fitting subject to the constraint given by Eq.(27). Finally, the fourth-order coefficients are obtained by another linear fitting of $h_{ij}(k) - h_{ij}(0) - h_{ij}^{(2)}k^2$ in terms of k^4 for $k <$ 0.8 Å^{-1}. Different from direct polynomial fitting, the coefficients from the two-step linear correlations are relatively insensitive to the range of k .

For convenience, I refer to the TCFs obtained from the 4th-order polynomial equation as TCF-fit. In combination with the Fourier transform of $h_{ij}(r)$ directly from MD simulations, I now have a full spectrum of TCFs in the Fourier space, *i.e.,* TCF-fit

for k from 0 to 0.8 Å^{-1}, TCF-D for k from 0.8 Å^{-1} to 1.4 Å^{-1}, and TCF-F for $k >$ 1.4 Å^{-1}. In stark contrast to direct fitting of $h_{ij}(k)$ from Fourier transform in the full range as proposed by Chuev et al [26, 35], the additional results from PSFs and the asymptotic analysis lead to numerical values for $h_{ij}(k)$ more reliable at small k. For each *k* value, I then calculate the DCFs from the TCFs using the RISM equation. Similar to $h_{ij}(k)$, $c_{ij}(k)$ can also be expressed as a polynomial (starting from the $1/k^2$ term due to the long-range behavior Coulomb interactions).

$$c_{ij}(k) = c_{ij}^{(-2)} / k^2 + c_{ij}^{(0)} + c_{ij}^{(2)} k^2 + \ldots \tag{28}$$

Once I have DCF data for small *k* values, the long-range asymptotic coefficients $c_{ij}^{(-2)}$ could be calculated from zero-point extrapolation. For a 3-site water model, $c_{ij}^{(-2)}$ could be alternatively calculated from Eq.(19).

I calculated the RDFs and PSFs using MD simulation in the NPT ensemble. The simulations were based on cubic simulation boxes of 28.8 Å, 32.7 Å and 56.0 Å for SPC/E, TIP3P and TIP4P-EW water models, respectively. For the purpose of conformity, all simulations are carried out at bulk water density 0.997 g/cm^3. The temperature is controlled with a Nose-Hoover thermostat at 298 K. The particle mesh Ewald summation is used for solving the Coulomb potential. All MD simulations are carried out with 1 fs step length and the trajectories are generated every 100 time steps for over 2 ns after the system reaches equilibrium. The trajectory files are analyzed through a C++ parallel computing toolkit developed in this group. In order to examine the effects of bulk density

and the box size, I checked the DCF results corresponding to two box sizes (box length 28.8 Å and 33.0 Å for SPC/E water) and two densities (0.986g/cm^3 and 0.997g/cm^3) for TIP/3P water.

2.4. Correlation Functions in Real and Fourier Spaces

I first present the radial distribution functions (RDF) for each water model and compare the results with the experimental data and pervious simulations from the literature. As shown in Figure 2-1, the RDF between oxygen atoms predicted by the TIP4P-Ew and SPC/E models agree well with the experimental results except that both slightly underestimate the first-peak position. As expected, the TIP3P model misses slightly the height and the position of the first peak and flattens at the second peak for the RDF between oxygen atoms.

In order to evaluate the TCFs in the small k region, I calculate the partial structural factors (PSFs) directly through MD simulations. As shown in Figure 2-2A and 2B, the TCFs calculated from PSFs (TCF-D) compare well with those from the Fourier transform (TCF-F) for both TIP3P and SPC/E water models when k is not too small. However, for the reasons discussed above, the TCF-F curves do not satisfy the compressibility condition given in Eq.(26), and exhibit unphysical undulations at very small values of k. Conversely, the TCF-D curves show no oscillations at small k values. By a linear fitting of the TCF-D data from 0 to 0.8 Å^{-1} according to Eq.(25), I obtain a polynomial representation of the TCFs (shown as dashed lines in Figs. 2C and 2D and Table 2 presents the coefficients).

Figure 2-3 shows the DCFs calculated from Eq.(14), based on the TCFs at different regions of *k* values. While the three regions of DCFs link well at the boundaries, I find significant uncertainties in the intermediate region. To attain a smooth function for each pair of DCFs, I interpolate the numerical results at low and high values of k:

$$k^2 c_{ij}^{SM}(k) = \sum_{n=-1}^{1} c_{ij}^{(2n)} k^{2n+2} \exp\left[-(k/\tau)^4\right] + k^2 c_{ij}(k)\left(1 - \exp\left[-(k/\tau)^4\right]\right) \quad (29)$$

where $c_{ij}^{(2n)}$ are the first three asymptotic coefficients of DCFs, $c_{ij}(k)$ is calculated from TCFs at high k values, and τ is a tuning parameter. Because the interpolation does not change the first 3 coefficients in the asymptotic expansion of DCFs (Eq.(28)), such a procedure is numerically more reliable than the DCFs directly obtained from fitting TCFs in the whole range of k.

Figure 2-4 shows the DCFs in real space obtained from the reversed Fourier transformation and results from previous calculations. In general, the DCF curves correspond well with each other, in particular for HH and OH pairs. The long-range components of the DCFs are virtually the same because the same density and dielectric constant are used in the asymptotic analysis. The main difference lies in the short-range component of the DCF between oxygen atoms. The discrepancy arises from different numerical procedures for fitting the simulation data.

Unlike the 3-site water models, TIP4P-Ew includes a pseudo-site M at the bisector of HOH angle. The M site bears a negative partial charge to account for the decentralized distribution of electrons in the oxygen atom. In the TIP4P-Ew water model, the oxygen site is neutral, interacting through the Lennard-Jones potential.

The 4 interaction sites affiliated with the TIP4P-Ew model result in 6 pairs of inter-related TCFs. As shown in Figure 2-7, the RDFs for MM, MO and OO pairs virtually identical due to the closeness of M and O sites. Similarly, the RDFs for OH and MH pairs are close to each other and differ only in the height of the first peak, i.e., the first peak of $g_{\mathrm{OH}}(r)$ is slightly higher than that of $g_{\mathrm{MH}}(r)$. Figure 2-8 shows the TCFs in the Fourier space. Following the same procedure for the 3-site water models, I calculate $h_{ij}(k)$ by a combination of the Fourier transform for large k and the direct sampling of the partial structural factors (PSFs) in the small k region. Figure 2-9 shows the asymptotic limit of $h_{ij}(k)$ near $k=0$ obtained from a linear fitting with Eq.(25), subject to the constraint on the bulk compressibility.

With the TCFs at different regions of k, I can readily calculate the corresponding DCFs using the RISM equation. Figure 2-10 shows the DCFs in the Fourier space for the TIP4P-Ew model. To connect the numerical values at different regions of k, I again use Eq.(29) to obtain a smooth fitting of the simulation data before the DCFs in the Fourier space are transformed into real space. The specific values of $C_{ij}^{(-2)}$ are also tabulated in Table 2-2. As aforementioned, Eq.(19) is not valid for a 4-site water model; otherwise, zero values would be expected for $C_{OO}^{(-2)}$, $C_{OH}^{(-2)}$, $C_{MO}^{(-2)}$ because the TIP4P-Ew model gives no partial charge on the oxygen site.

Figure 2-11 shows the DCFs in real space according to the TIP4P-Ew model. Unlike the RDF curves, the DCF curves for OO, MO and MM pairs are clearly separated.

The distinction of DCFs among these sites can be attributed to the intramolecular correlation functions $\omega_{ij}(r)$ as given by Eq.(15).

2.5. DCFs from Modified HNC Closure

The modified HNC closure (MHNC) accounts for the long-range limit of the DCFs using an effective potential [39]

$$h_{ij}(r)+1=\exp\left[-\beta u_{ij}^{eff}(r)+h_{ij}(r)-c_{ij}(r)\right] \tag{30}$$

where

$$u_{ij}^{eff}(r)=\zeta u_{ij}^{Coul}(r)+u_{ij}^{LJ}(r) \tag{31}$$

As shown in Eq.(19), the long-range limits of the DCFs for 3-site water models are proportional to the negative of the Coulomb potential times a pre-factor ζ. According to Eq.(20), the ζ values for SPC/E and TIP3P are 0.9604 and 0.9594, respectively. Because of the long-range correction, I expect that the DCFs from MHNC be more accurate than those from the original RISM/HNC or RISM/KH theory.

I use the Picard iteration to solve Eq.(30) together with the RISM equation for the TCFs and DCFs. Figures 12 and 13 compare the DCFs for two 3-sites water models with the results from MD simulations. The difference between the DCFs calculated from the (semi-) analytical and numerical methods mainly lies in the short range. Overall, the DCFs from RISM/MHNC correspond well with those from MD simulations, suggesting that the DCFs obtained from the numerical procedure may provide a valuable alternative to time-consuming MD simulation.

2.6. DCFs for Solvation Free Energy Predictions

DCFs are not directly measureable through experiments or simulations. To validate the reliability of DCFs obtained from this work, I test their applicability to solvation free energy calculations using cDFT. As in our previous publications[40-42], I define the solvation free energy as the reversible work to transfer a solute molecule from a vacuum into the pure solvent at fixed temperature and the chemical potential of solvent molecules. For a given conformation of the solute molecule $\mathbf{x}$, the solvation free energy can be related to the change in the grand potential of the pure solvent upon inserting the solute molecule [43]

$$F_s[\rho(\mathbf{x});\mu,V,T] = \Omega[\rho(\mathbf{x});\mu,V,T] - \Omega_0[\mu,V,T] \tag{32}$$

where Ω_0 is the grand potential of the pure solvent. If the solute molecule takes multiple conformations, the overall solvation free energy is given by

$$\beta F_s = -\ln\left\langle \exp[-\beta F_s(\mathbf{x})]\right\rangle_0 \tag{33}$$

where $<\cdots>_0$ denotes an ensemble average of the solute conformations in vacuum. In this work, I confine our interest in the solvation free energies of small molecules so that the solvation free energy can be estimated based on a single solute conformation. With the solute-solvent interactions treated as an external potential, I solve the site distributions of the solvent molecules from Eq.(8) and, subsequently, calculate the solvation free energy from

$$\beta F_s[\rho_i(\mathbf{r})] = -\frac{1}{M_s}\sum_i \int d\mathbf{r}\Delta\rho_i(\mathbf{r}) + \frac{1}{2}\sum_{i,j}\iint \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' c_{ij}^{(2)}(|\mathbf{r}-\mathbf{r}'|)[\rho_i(\mathbf{r})\rho_j(\mathbf{r}') - \rho_i^{bulk}\rho_j^{bulk}] + \beta F^{\mathrm{B}}[\rho_i(\mathbf{r})] - \sum_i \int \mathrm{d}\mathbf{r}\rho_i(\mathbf{r})B_i(\mathbf{r}) \quad (33)$$

where M_s stands for the number of interaction sites, and $B_i(\mathbf{r})$ is the bridge function of site *i*

$$B_i(\mathbf{r}) \equiv \delta F_B / \delta\rho_i(\mathbf{r}) . \quad (34)$$

The exact expressions for F_B and the selection of an effective hard-sphere diameter for the reference system can be found in our previous publications[40-42].

As a benchmark to test the DCFs obtained from in this work, I compare the hydration free energies of 504 small organic molecules predicted from cDFT with MD simulations and experimental data [44]. The solute force fields used in cDFT and MD simulations are identical[44]. Figure 2-14 shows that the new DCFs improve the theoretical predictions from an average unsigned error of ~1 kcal/mol as reported in a previous publication[42] using the same data set to 0.66 kcal/mol. For comparison with the experimental data, however, the new DCF yields a similar level of cDFT performance, indicating that the main limitation in hydration free-energy calculations is related to the accuracy of the semi-empirical force field for solute-solvent interactions rather than approximations introduced in cDFT. Because the cDFT calculations are far more efficient than MD simulations, I conjecture that cDFT provides a valuable alternative to

simulation methods for high-throughput prediction of solvation free energies as required in many practical applications[28].

2.7. Conclusions

In this work, I propose a numerical procedure for accurate evaluation of the direct correlation functions (DCFs) for liquid water using MD simulation and the reference interaction site model (RISM). Along with the DCF calculations, I present the site-site radial distribution functions (RDFs), the partial structural factors (PSFs), and the total correlation functions (TCFs) for two 3-site (SPC/E, TIP3P) and one 4-site (TIP4P-Ew) water models. The DCFs for the 3-site models can be reproduced near quantitatively with the modified hypernetted chain (MHNC) closure for the RISM equation.

One of the foremost important applications of the DCFs is to use them as an input for formulation of the excess intrinsic Helmholtz energy functional in the classical density functional theory (cDFT). When the new DCFs obtained from this work are employed in the cDFT predictions of the hydration free energies of 504 small molecules, the average unsigned error is 0.66 kcal/mol in comparison with the simulation data, better than ~1 kcal/mol reported in our earlier publication[42]. The improved performance suggests that the newly derived DCFs are more reliable. Because the DCFs generated from MD simulation compare fairly well with the RISM/MHNC predictions, I expect that the integral-equation theory will play a significant role in broader applications of the cDFT to molecular systems.

2.8. Appendix: Generalized Method of Direct Correlation Functions

This appendix will discuss the details of how to sample DCFs of any arbitrary molecules without specific symmetry.

Reference Interaction Site Model (RISM)

RISM was first proposed by Chandler and Anderson[45] and was based on site-site Ornstein-Zernike (OZ) equation as pointed out later by Cummings and Stell[46]. This equation relates the direct correlation functions (DCF) to total correlation functions (TCF) and thus provide a method to sample DCF from MD simulations. To understand this, first consider an inhomogeneous grand canonical one-component molecular system. The grand partition function Ξ can be written as

$$\Xi = Tr\left\{\exp\left[-\sum_{i,m}\left[\beta V_i^{ext}\left(\mathbf{r}_i^m\right)-\beta\mu_i\right]-\frac{1}{2}\sum_{i,m}\sum_{j,m'}\beta\Gamma\left(\mathbf{r}_i^m,\mathbf{r}_j^{m'}\right)-\sum_m V_B\left(\mathbf{r}_0^m,\mathbf{r}_2^m,\mathbf{r}_3^m,...\right)\right]\right\} \qquad (35)$$

In the equation, Tr is the trace on the Hilbert space of Mayer function in the grand-canonical ensemble and $Tr=\sum_{N=0}^{+\infty}\frac{1}{\left(v\Lambda\right)^N}\int\prod_{m,i}d\mathbf{r}_i^m$, where N is the number of molecules in the grand canonical ensemble, v is the symmetry number of molecule, Λ is the molecular thermo wavelength, subscript m stands for molecule id, i is the molecular site, β is $1/k_BT$, V_i^{ext} is the external potential on site i, Γ is the two-body interaction potential, and V_B is the intra-molecular bonding potential. In statistic mechanics, density can be expressed as fluctuations of summation of Dirac delta functions

$$\rho_i(\mathbf{r}) = \left\langle \sum_m \delta\left(\mathbf{r} - \mathbf{r}_i^m\right) \right\rangle \tag{36}$$

where $\delta\left(\mathbf{r} - \mathbf{r}_i^m\right)$ is the Dirac delta function of site *i* at position $\mathbf{r}_i^m$. Combing Eq.(35) and (36), one find relation between grand potential $\Omega = -k_B T \ln \Xi$ and density as

$$\rho_i(\mathbf{r}) = \frac{\delta \Omega}{\delta V_i^{ext}(\mathbf{r})} \tag{37}$$

Performing functional derivative on both sides of Eq.(37) and define χ_{ij} as

$$\begin{aligned} \chi_{ij}(\mathbf{r},\mathbf{r}') &\equiv -\frac{\delta \rho_i(\mathbf{r})}{\delta \beta V_j^{ext}(\mathbf{r}')} \\ &= \left\langle \sum_m \sum_{m'} \delta\left(\mathbf{r} - \mathbf{r}_i^m\right) \delta\left(\mathbf{r}' - \mathbf{r}_j^{m'}\right) \right\rangle - \left\langle \sum_m \delta\left(\mathbf{r} - \mathbf{r}_i^m\right) \right\rangle \left\langle \sum_m \delta\left(\mathbf{r}' - \mathbf{r}_j^m\right) \right\rangle \\ &= \rho_{ij}^{(2)}(\mathbf{r},\mathbf{r}') + \rho_b \delta(\mathbf{r} - \mathbf{r}') - \rho_i(\mathbf{r}) \rho_j(\mathbf{r}') \end{aligned} \tag{38}$$

where $\rho_{ij}^{(2)}$ is the two-body distribution function and ρ_b is the bulk density. This equation shows that the functional derivative of density profile with external potential is related to two-body distribution function. It will later be demonstrated that the inverse functional is actually related to DCF.

Breaking Ω into three terms with respect to a reference ideal atomic system, which has the same site density but has no intra or inter interactions.

$$\Omega = F^{ia} + \left(F - F^{ia}\right) + \sum_i \int d\mathbf{r} V_i^{ext}(\mathbf{r}) \rho_i(\mathbf{r}) \tag{39}$$

where F is the intrinsic Helmholtz free energy, superscript ia stands for the reference ideal atomic system. For ideal atomic system, F^{ia} is analytical.

$$\beta F^{ia} = \sum_i \int d\mathbf{r}\left(\ln \rho_i \Lambda_i^3 - 1\right)\rho_i \tag{40}$$

At equilibrium,

$$\frac{\delta\Omega}{\delta\rho_i(r)} = 0 = k_B T \ln \rho_i \Lambda_i^3 + \frac{\delta\left(F - F^{ia}\right)}{\delta\rho_i(r)} + V_i^{ext}(\mathbf{r}) - \mu_i \tag{41}$$

Performing functional derivative on both sides and define $D_{ij}(\mathbf{r})$

$$D_{ij}(\mathbf{r},\mathbf{r}') \equiv -\frac{\delta^2\left(\beta F - \beta F^{ia}\right)}{\delta\rho_i(\mathbf{r})\delta\rho_j(\mathbf{r}')} = \frac{\delta_{ij}\delta(\mathbf{r}-\mathbf{r}')}{\rho_i(\mathbf{r})} + \frac{\delta\beta V_i^{ext}(\mathbf{r})}{\delta\rho_j(\mathbf{r}')} \tag{42}$$

Inserting the definition of χ_{ij}, equation above reduces to

$$D_{ij}(\mathbf{r},\mathbf{r}') = \frac{\delta_{ij}\delta(\mathbf{r}-\mathbf{r}')}{\rho_i(\mathbf{r})} - \chi_{ij}^{-1}(\mathbf{r},\mathbf{r}') \tag{43}$$

where χ_{ij}^{-1} is the inversion of functional derivative and has the formula

$$\int d\mathbf{r}'' \sum_k \chi_{ik}^{-1}(\mathbf{r},\mathbf{r}'')\chi_{kj}(\mathbf{r}'',\mathbf{r}') = \delta_{ij}\delta(\mathbf{r}-\mathbf{r}') \tag{44}$$

In homogeneous limit $\rho_i(\mathbf{r}) \to \rho_b$, Eq.(43) becomes

$$D_{ij}(\mathbf{r}-\mathbf{r}') = \frac{\delta_{ij}\delta(\mathbf{r}-\mathbf{r}')}{\rho_b} - \chi_{ij}^{-1}(\mathbf{r}-\mathbf{r}') \tag{45}$$

Similarly, we consider an ideal chain in homogeneous limit,

$$D_{ij}^{id}\left(\mathbf{r}-\mathbf{r}'\right)=\frac{\delta_{ij}\delta\left(\mathbf{r}-\mathbf{r}'\right)}{\rho_b}-\chi_{ij}^{id,-1}\left(\mathbf{r}-\mathbf{r}'\right) \tag{46}$$

Substract Eq.(46) from (45),

$$D_{ij}-D_{ij}^{id}=\left.\frac{\delta\left(F-F^{ex}\right)}{\delta\rho_i\left(\mathbf{r}\right)\delta\rho_j\left(\mathbf{r}'\right)}\right|_{\rho_b}=\chi_{ij}^{id,-1}\left(\mathbf{r}-\mathbf{r}'\right)-\chi_{ij}^{-1}\left(\mathbf{r}-\mathbf{r}'\right) \tag{47}$$

According to the definition of DCF, intra-molecular correlation (ICF) and TCF, and rewrite Eq.(47) in matrix format in Fourier space

$$\rho_b c=\omega^{-1}-\left(\omega+\rho_b h\right)^{-1} \tag{48}$$

where c is DCF, ω is ICF in ideal chain condition (or single chain), h is TCF. With simple algebra, Eq.(48) leads to the well know RISM equation.

Asymptotic Behavior of DCF

As pointed out by the previous chapter, the asymptotic behavior of DCF in molecular system is different from that in atomic system. And result from MD simulation of TIP4P-Ew water model proves that the 'scaling method' doesn't work any more.

$$c_{ij}\not\gtrsim -\zeta\frac{\beta q_i q_j}{r} \tag{49}$$

In order to get the correct asymptotic relation of DCF, DCF is spited into three terms

$$c_{ij}=c_{ij}^{S}+c_{ij}^{L}+b_{ij} \tag{50}$$

where c_{ij}^{S}, c_{ij}^{L}, b_{ij} are short range, long range and correction parts, respectively.

$$c_{ij}^{L} = -\frac{\beta q_i q_j}{r} \tag{51}$$

And RISM is rewritten in a more general way

$$h - \zeta_d = \left(\omega + \rho_b \zeta_d\right)\left(c - b\right)\left(\omega + \rho_b \zeta_d\right) + \left(\omega + \rho_b \zeta_d\right)\left(c - b\right)\rho_b \left(h - \zeta_d\right) \tag{52}$$

where ζ_d is the screened dielectric correction function[47] which renormalizes TCF h and ICF ω. By equating the original RISM and generalized RISM equations,

$$b = \omega^{-1} - \left(\omega + \rho_b \zeta_d\right)^{-1} \tag{53}$$

As demonstrated by Perkyns and Pettitt[36], the correct dielectric constant of electrolyte solution can be achieved if correction term ζ_d is written in Fourier space as

$$\begin{aligned} \zeta_{ij}^{d}\left(k\right) = j_0\left(kx_i\right) j_0\left(ky_i\right) j_1\left(kz_i\right)\left(\frac{\epsilon - 1}{y} - 3\right) \\ \exp\left(-s^2 k^2 / 4\right) j_0\left(kx_j\right) j_0\left(ky_j\right) j_1\left(kz_j\right) \end{aligned} \tag{54}$$

where $\left(x_i, y_i, z_i\right)$ are the coordinates of solvent molecule site i, and j_0, j_1 are spherical Bessel functions, ϵ is the dielectric constant of the solution, $y = 4\pi\rho_b d^2 / (9k_B T)$, s is the separation parameter and does not affect the final result. To be noted is that this relation works only for rigid molecules.

$$\omega_{ij}(k) = \delta_{ij} + \left(1 - \delta_{ij}\right)\frac{\sin kl_{ij}}{kl_{ij}} \tag{55}$$

where δ_{ij} is the Kronecker delta function, l_{ij} is the distance between interaction sites *i* and *j*.

With molecular structure as an input into Eq.(54) and Eq.(53), the asymptotic limit of DCF can be derived. And this method can be generally applied to all rigid molecular solutions.

General Method to DCF

This section discusses a practical way to sample DCF from MD simulations of arbitrary molecular solution, including flexible solvent like octanols. As explained above, the asymptotic behavior of DCF for rigid molecules can be analytically derived. For non-rigid molecules, the task is much simpler.

According to the definition, ICF of flexible molecules can be directly sampled in Fourier space from a single chain simulation, which is trivial to implement.

TCF is sampled from a standard homogeneous bulk solution simulation. And by using the technique described in the previous Chapter, small k values of TCF can be directly sampled.

Insert the values of ICF and TCF into Eq.(53) , one may get raw values of DCF. The asymptotic behavior of DCF can be numerically derived by a linear fitting of a small k range in $k^2 c_{ij}$.

$$k^2 c_{ij} \approx c_{ij}^{(-2)} + c_{ij}^{(0)} k^2 \tag{56}$$

Finally, after subtracting the long-range part of DCF from the raw DCF data, a smooth function is used to link the remaining short-range part. And DCF can be sampled.

The whole procedure is similar to sample DCF in TIP4P-Ew except for that ICF is sampled from a single chain simulation instead of using Eq.(55).

This method is general and one can sample DCF even without pre-knowledge of the asymptotic behavior of DCF. The long-range part of DCF can be sampled from simulations.

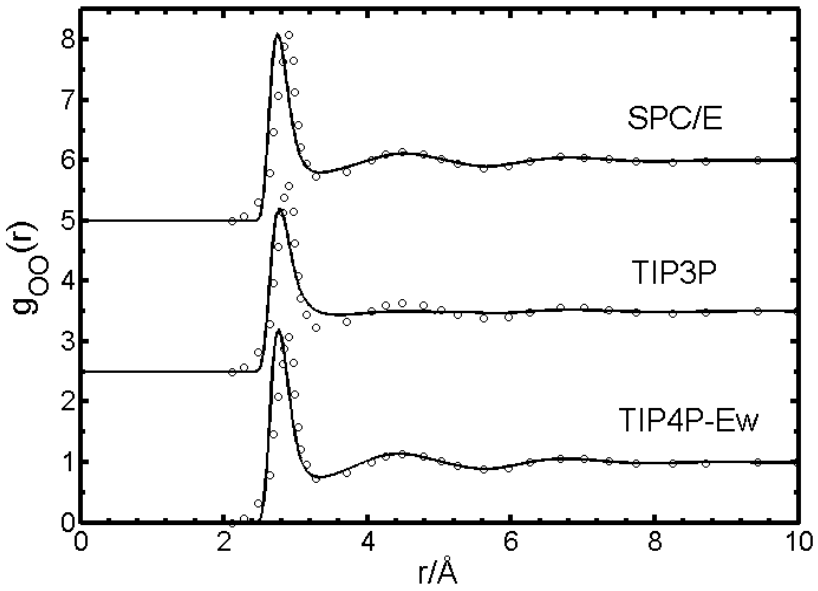
Table 2-1. Properties of different water models

	$\chi(10^{-6}\,bar^{-1})$	$\rho_b(g/cm^{-3})$	Dipole (D)	q_H (e)	ζ
SPC/E	44.97	0.997	2.350688	0.4238	0.9604
TIP3P	50.57	0.997	2.346846	0.417	0.9594
TIP4P-EW	45.78	0.997	2.320908	0.52422	\

Table 2-2. Coefficients for the polynomial expansions of TCFs and DCFs at small k

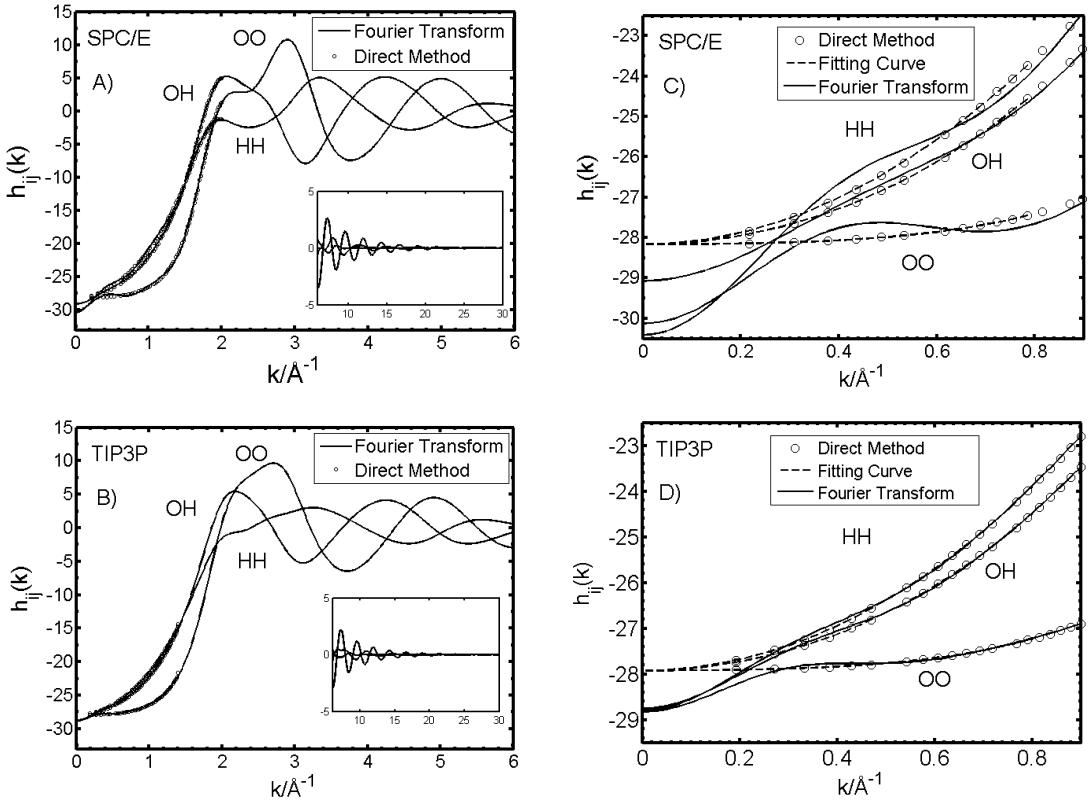
		$h_{ij}^{(0)}$, Å^3	$h_{ij}^{(2)}$, Å^5	$h_{ij}^{(4)}$, Å^7	$c_{ij}^{(-2)}$, Å
SPC/E	OO	-28.155	0.3879	1.2172	-4862.0
	OH	-28.155	5.3472	0.8002	2431.0
	HH	-28.155	7.148	0.0498	-1215.5
TIP3P	OO	-27.9238	0.3493	1.1299	-4702.14
	OH	-27.9238	4.8359	0.8102	2351.07
	HH	-27.9238	6.0707	0.3196	-1175.53
TIP4P-Ew	OO	-28.168	0.1070	1.3664	-12598.80
	OH	-28.168	4.6868	1.0554	4769.80
	HH	-28.168	6.0129	0.5649	-1978.98
	MM	-28.168	0.2176	1.3544	-5042.25
	MH	-28.168	4.1216	1.1091	-2790.04
	MO	-28.168	0.2366	1.3556	7830.72

Figure 2-1



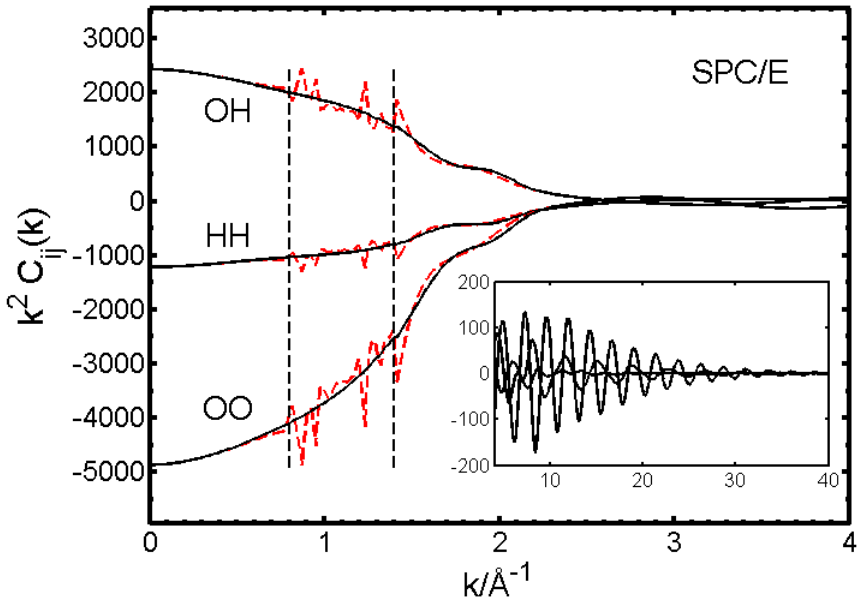
The radial distribution function between oxygen atoms according to SPC/E, TIP3P and TIP4P-Ew water models. For comparison, the experimental results [48] are shown in open circles.

Figure 2-2



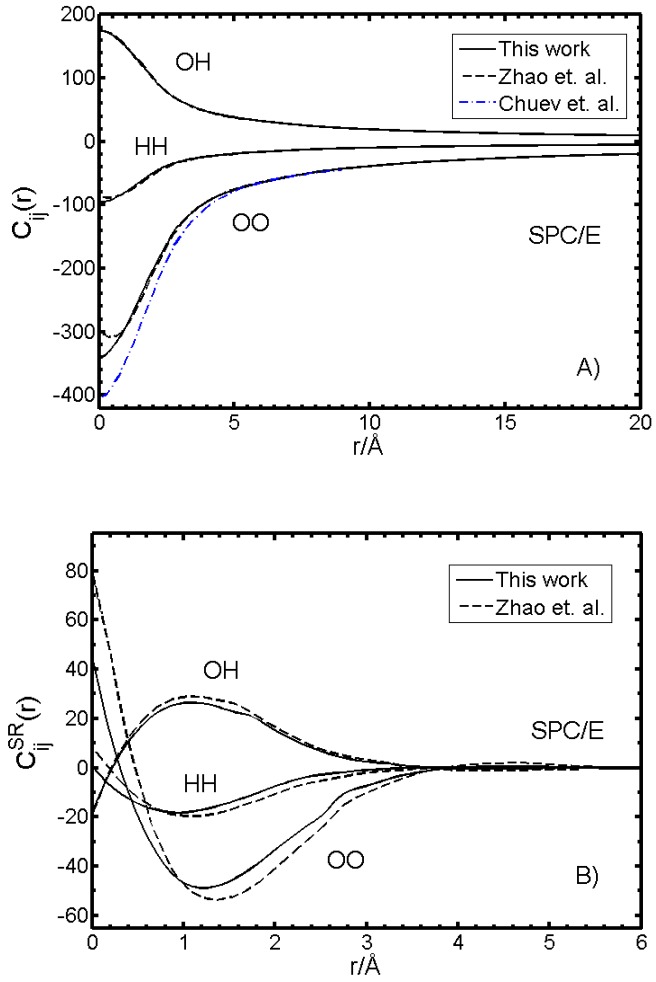
Site-site total correlation functions in the Fourier space for SPC/E and TIP3P models. Here the solid lines are from the Fourier transforms of the radial distribution functions (TCF-F), circles are from the direct sampling of the partial structure factors (TCF-D), and the dashed lines are polynomial fitting curves (TCF-fit).

Figure 2-3



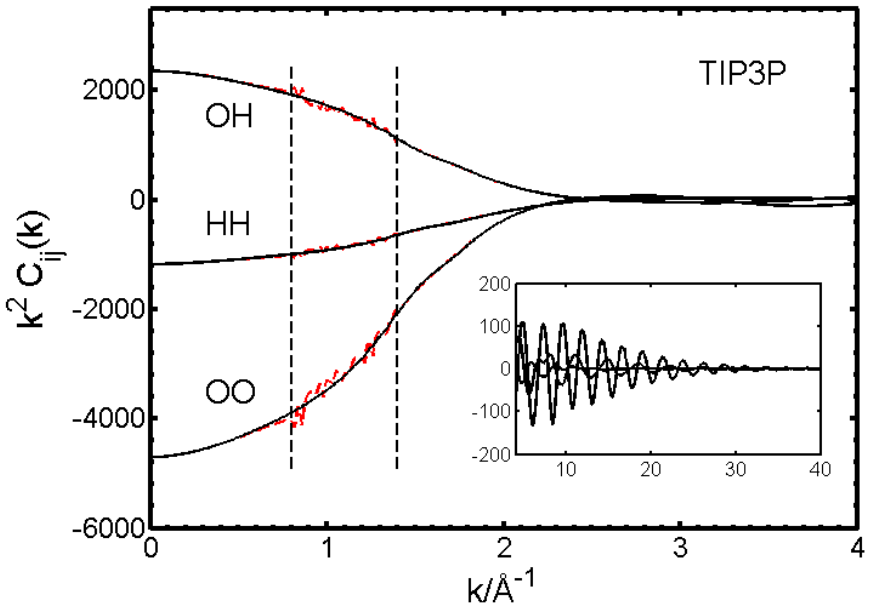
The Fourier transforms of the direct correlation functions for SPC/E water calculated from the total correlation functions according to the RISM equation (dashed line). The smoothed lines are obtained by an interpolation function, Eq.(29). Depending on the k values, the inputs for the RISM equation are from TCF-fit ($k<0.8$ Å^{-1}), TCF-D ($0.8 \le k \le 1.4$ Å^{-1}), and TCF-F ($k > 1.4$ Å^{-1}).

Figure 2-4



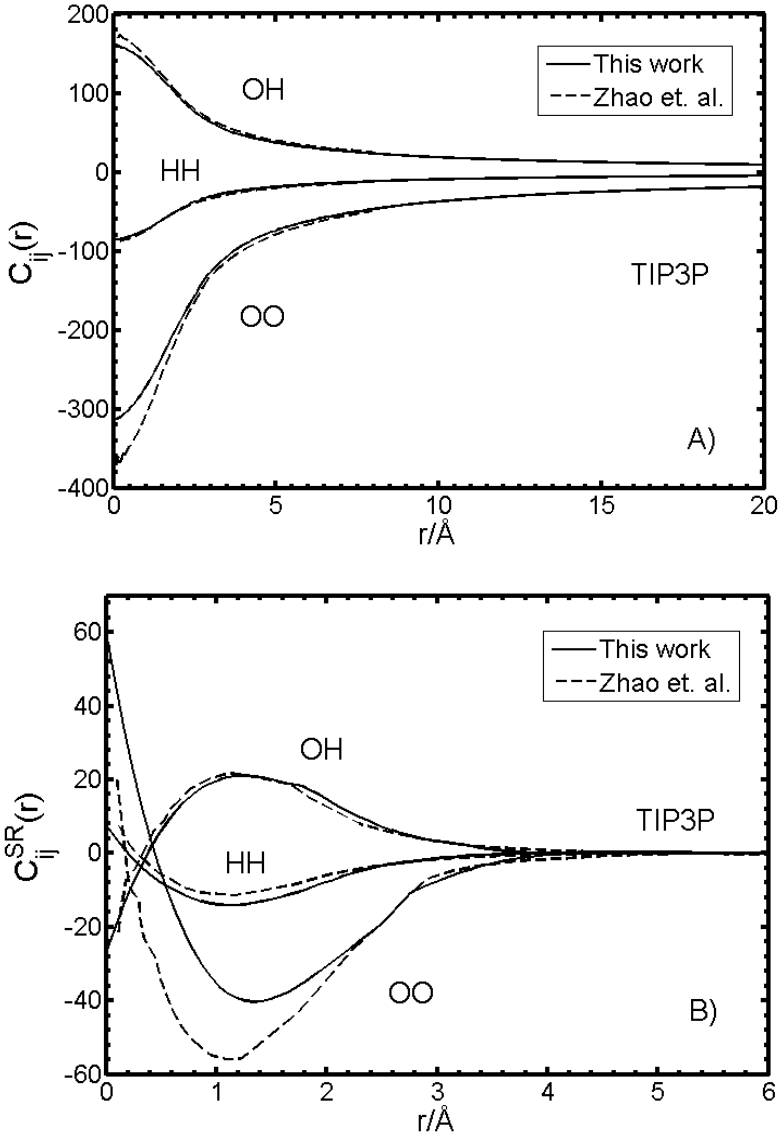
Direct correlation functions for SPC/E water from this work (solid lines) in comparison with those from Zhao *et al* (dashed lines)[25] and Chuev et al (dash-dotted lines)[35]. Panel A shows the DCFs up to 2 nm and panel B is the short-range component of the DCFs.

Figure 2-5



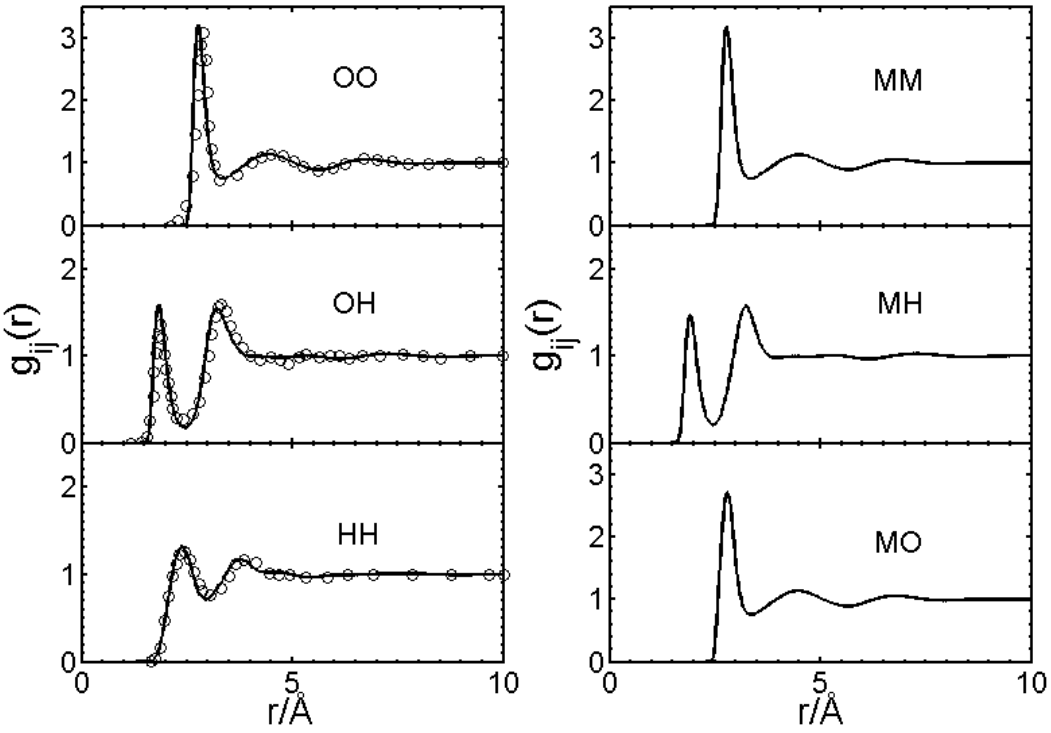
The same as Fig. 2-3 but for the TIP3P model.

Figure 2-6



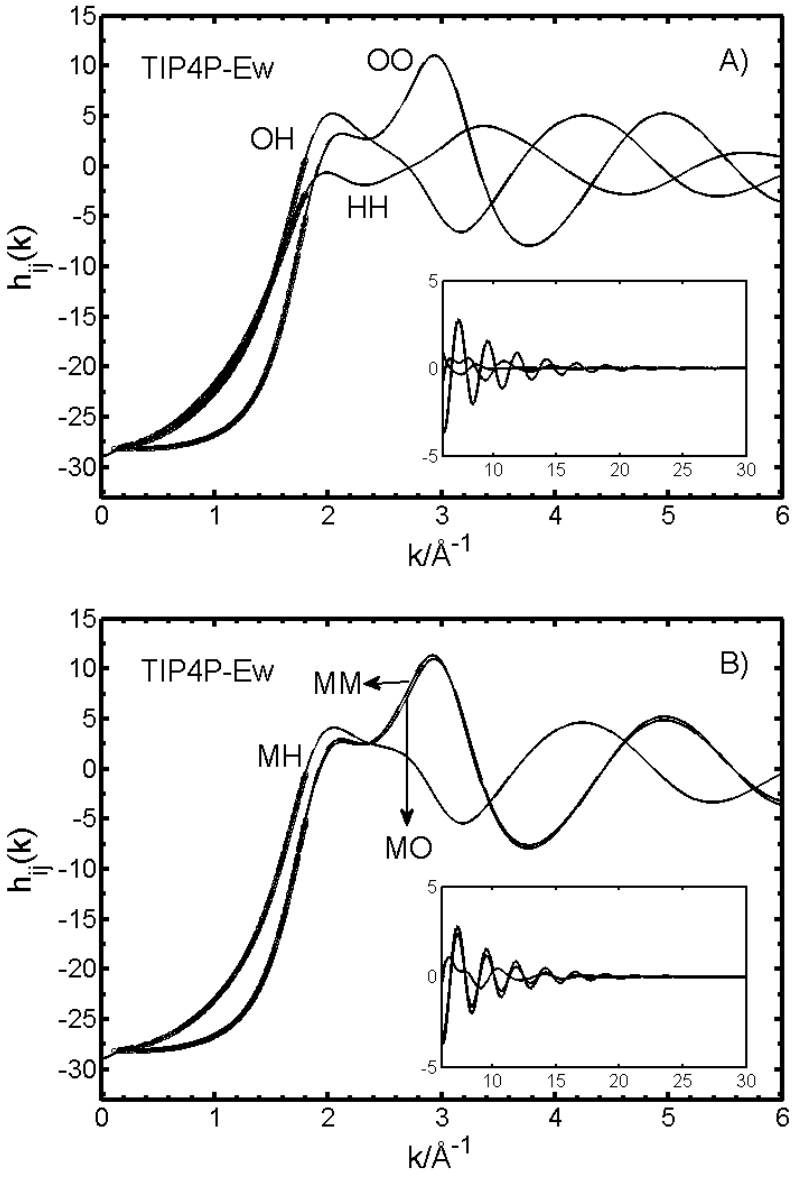
The same as Fig. 2-4 but for the TIP3P model.

Figure 2-7



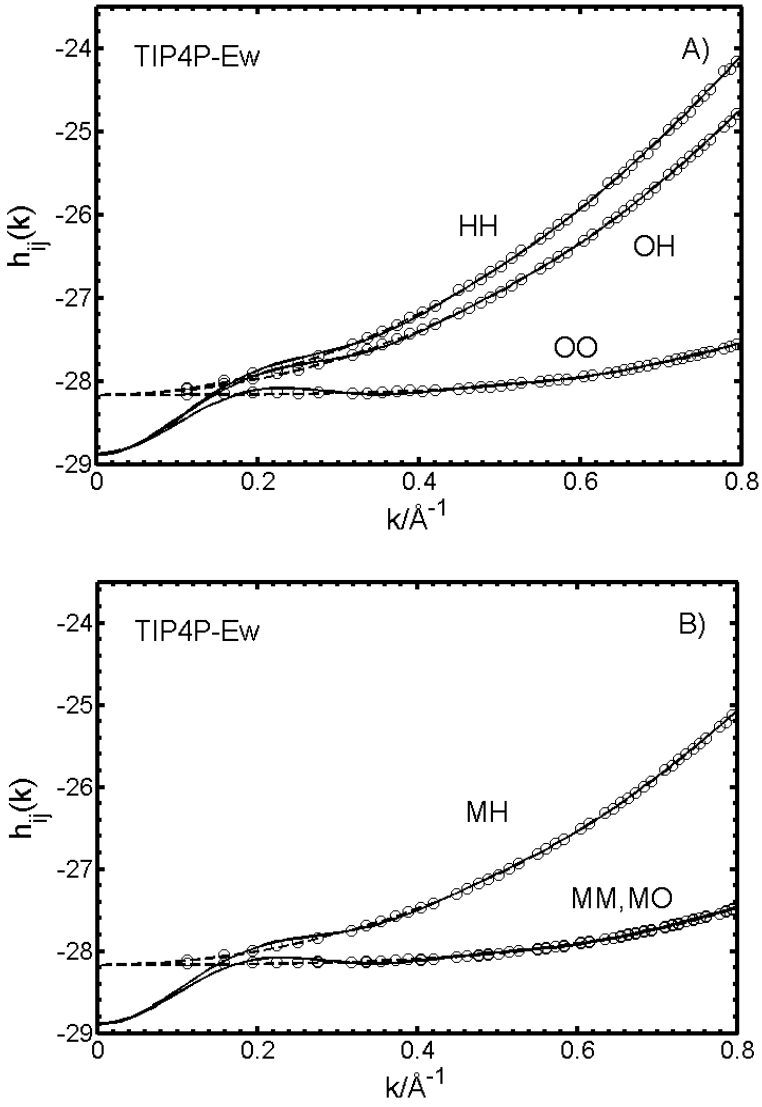
The site-site radial distribution functions according to the TIP4P-Ew model. For comparison, the experimental data are shown as circles [48].

Figure 2-8



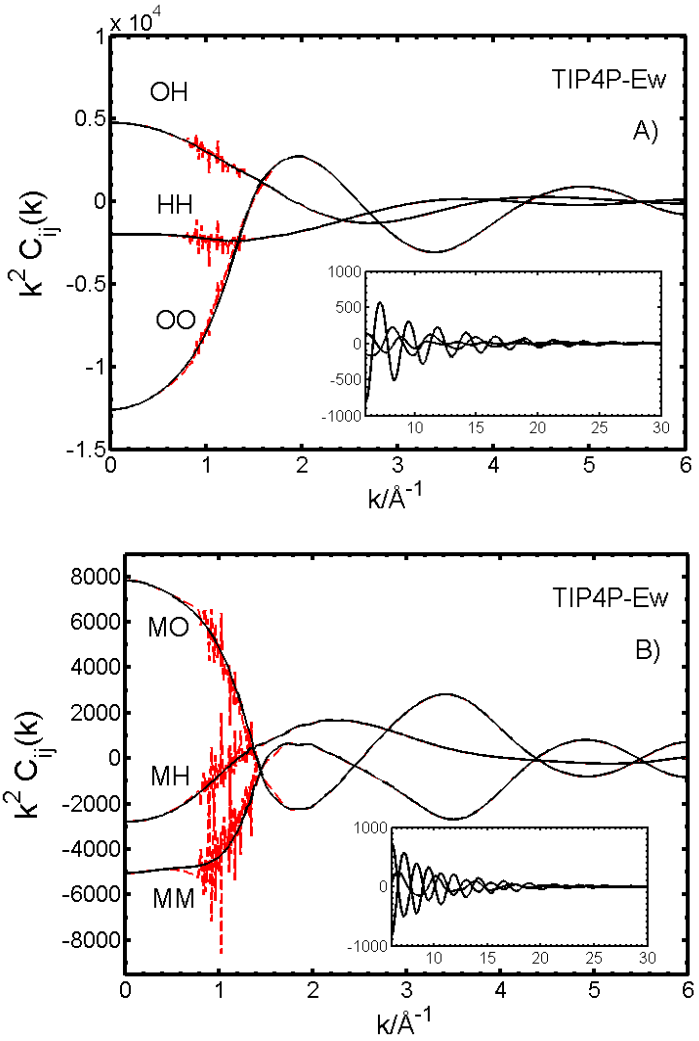
The site-site total correlation functions for the TIP4P-Ew model calculated from the RDFs (solid lines) and from the partial structure factors (circles).

Figure 2-9



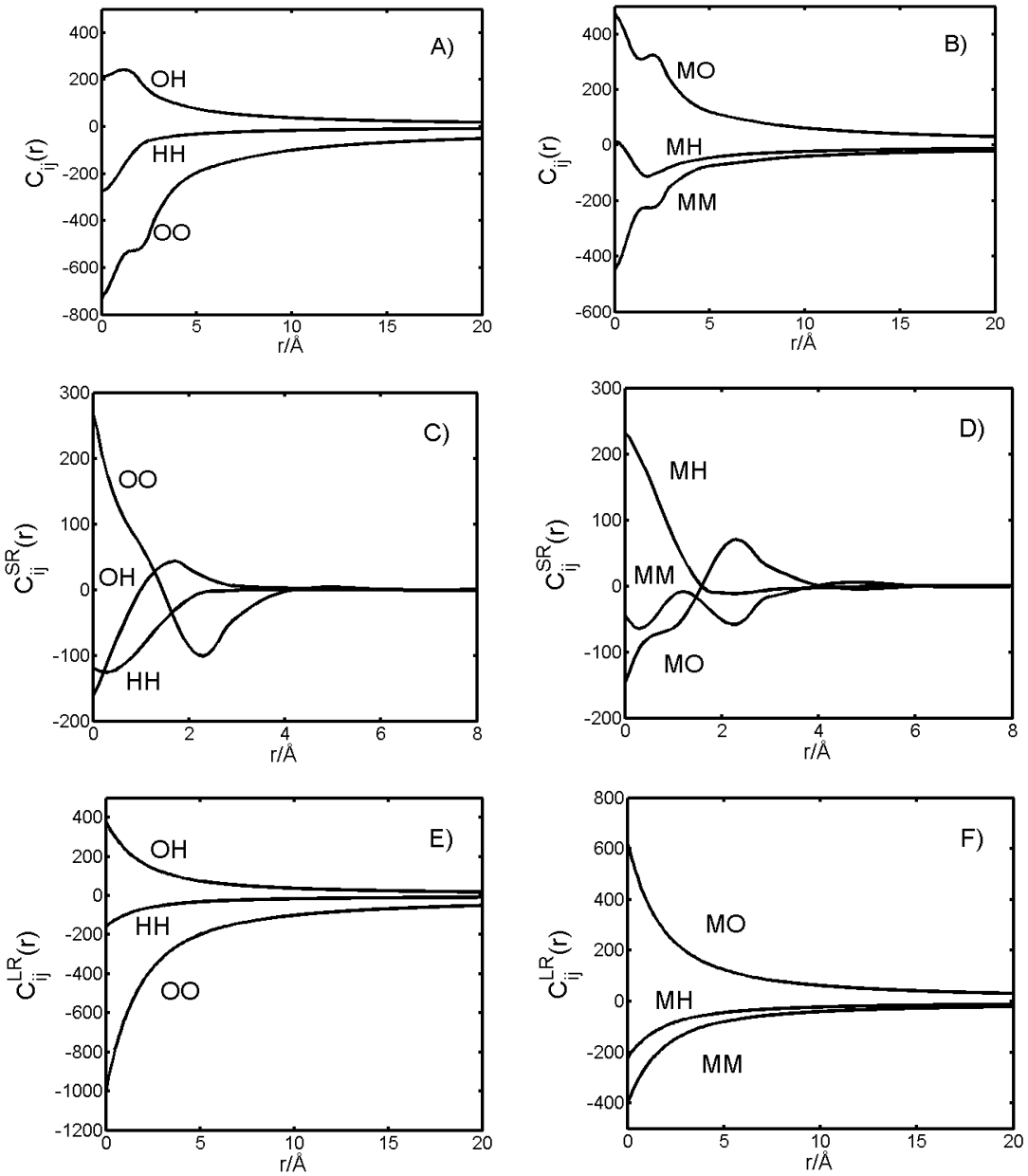
The asymptotic behavior of the TCFs for TIP4P-Ew from MD sampling (symbols), the Fourier transform (solid lines) and polynomial fitting (dashed lines).

Figure 2-10



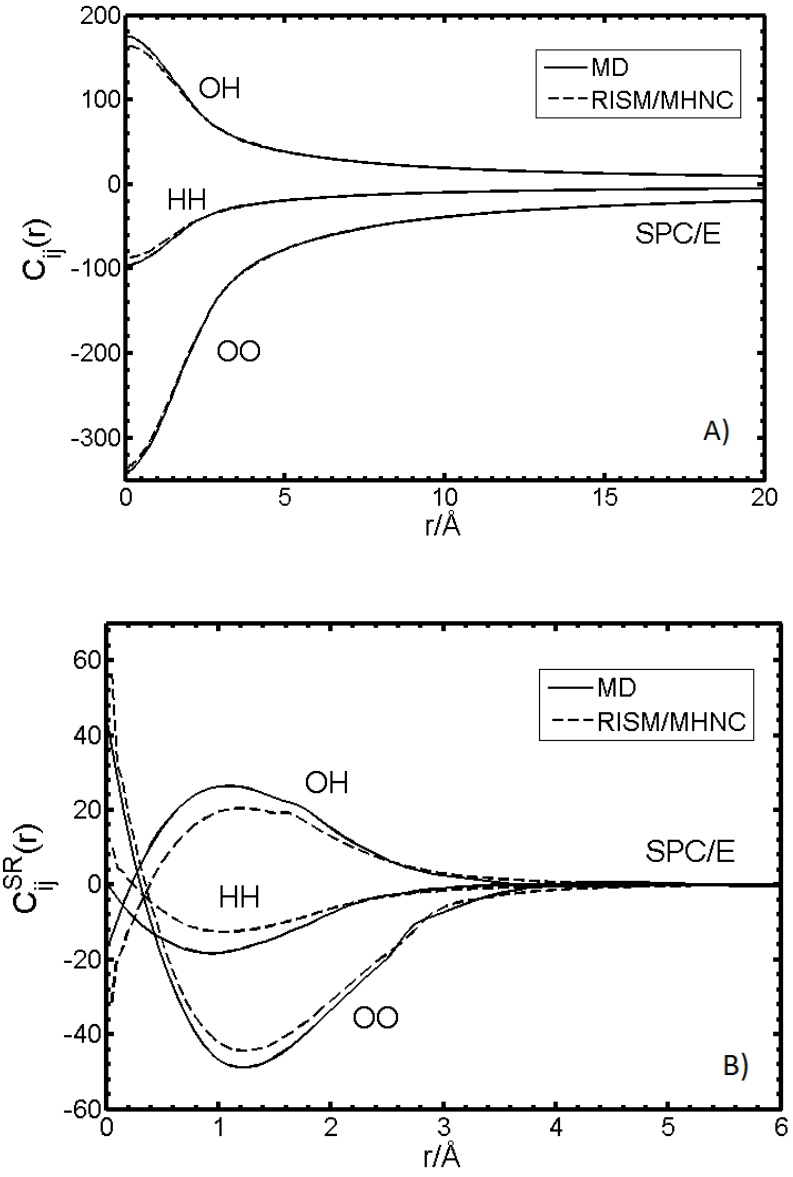
The Fourier transforms of the direct correlation functions for TIP4P-Ew calculated from the total correlation functions according to the RISM equation (dashed line). The smoothed lines are obtained by an interpolation function, Eq.(29).

Figure 2-11



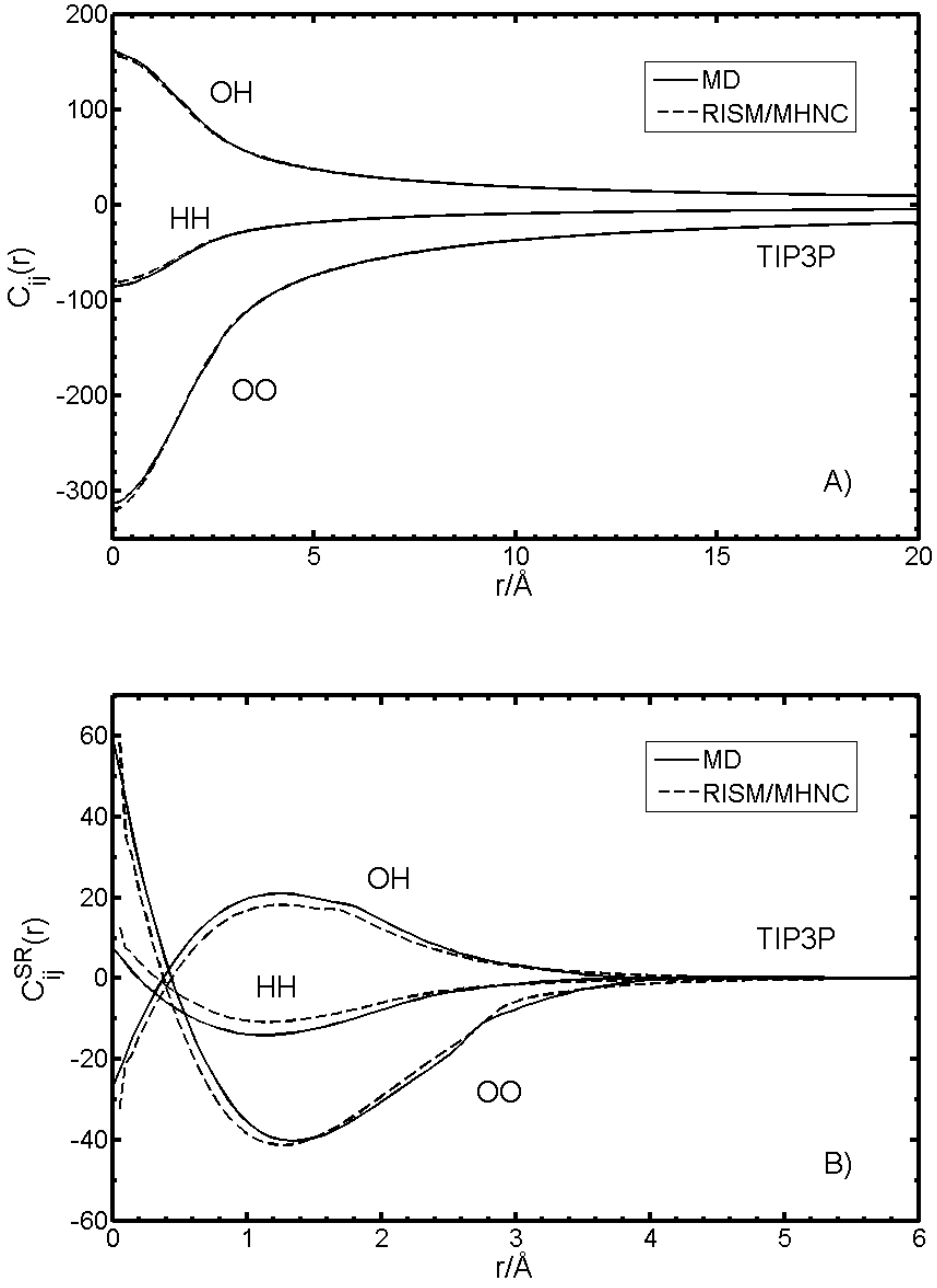
The direct correlation functions of TIP4P-Ew water in the real space. The upper panels are the DCFs for the entire range, the middle panels are the short-range component of the DCFs, and the bottom panels correspond to the long-range component of DCFs.

Figure 2-12



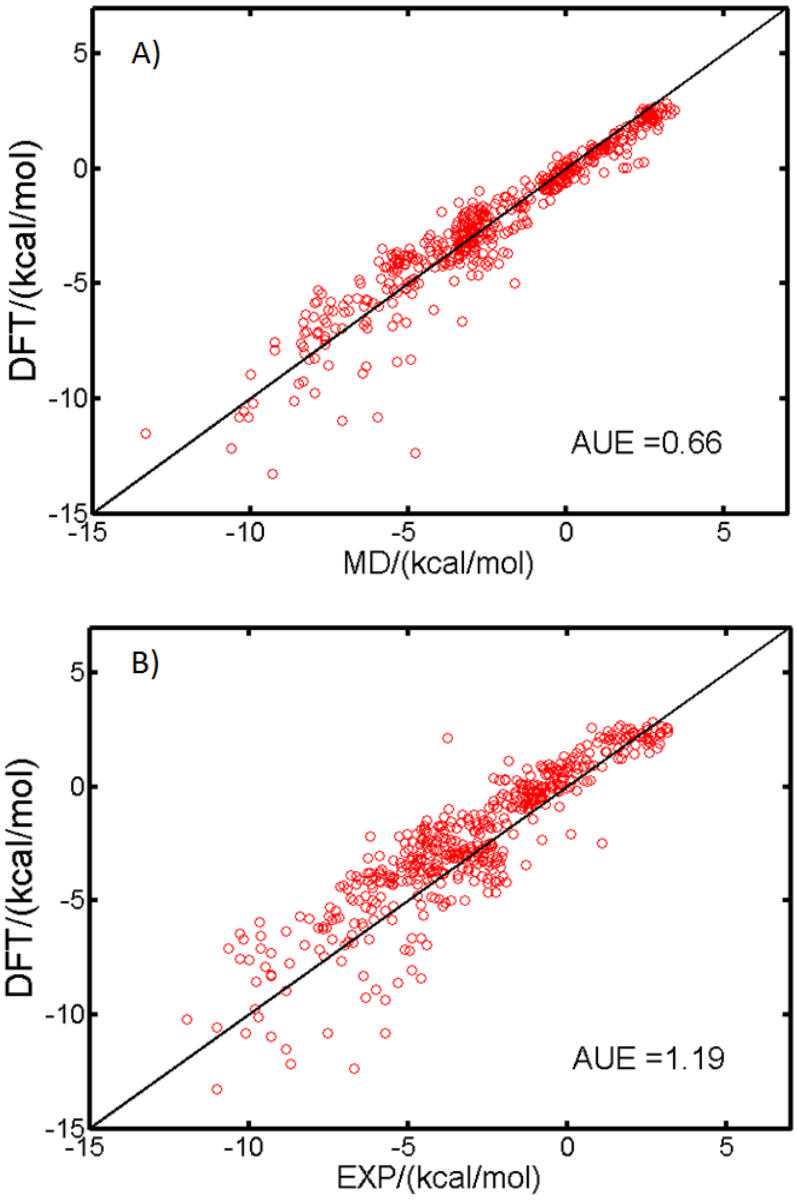
A comparison of DCFs from RISM/MHNC(dashed lines) and from MD simulation (solid lines) for SPC/E water. Panel A shows the DCFs over the entire range of separation, and panel B magnifies the short-range part of the DCFs.

Figure 2-13



The same as Fig. 2-12 but for the TIP3P model.

Figure 2-14



Theoretical predictions of the hydration free energies of 504 small organic molecules in comparison with the MD simulation (A) and experimental data (B) [44].

Chapter 3. A Molecular Theory of Hydration at Different Temperatures

Solvation plays an important role in diverse chemical processes ranging from reaction kinetics to molecular recognition, solubility and phase separations. Despite a long-history of theoretical exploration, quantitative prediction of solvation remains a theoretical challenge without relying on the macroscopic properties of the solvent as an input. Here I present a molecular density functional theory that provides a self-consistent description of the solvation structure and thermodynamic properties of small organic molecules in liquid water at different temperatures. Based on the solute configuration and force-field parameters generated from first-principles calculations, the theoretical predictions are found in good agreement with experimental data for the hydration free energies of 197 organic molecules in a temperature range from 0 to 40 °C. In addition to calibration with experimental results, the theoretical predictions are compared with recent molecular dynamics simulations for the hydration of five highly explosive nitrotoluenes. This work demonstrates the potential of the liquid-state theory for high-throughput prediction of solvation properties over a broad range of temperatures.

3.1. Solvation at Different Temperatures

Solvation is a fundamental chemical process involving the interaction of a solute with a large number of solvent molecules. The multi-body effect is pertinent not only to the solubility and the phase behavior of diverse chemical species in liquid solutions but also to the self-organization of biomacromolecules including protein folding[49-50], the chemical kinetics of liquid-phase reactions[51], protein-ligand associations[52], and solvatochromism[1, 53], to mention but a few applications. Among various thermodynamic

quantities commonly used to assess the solvent effects, the solvation free energy has been a focal point of theoretical and experimental studies[54-56]. Not only is such knowledge essential in solution chemistry but it is critically needed from a practical perspective for solvent design and selection important for many engineering applications[57].

The solvation free energy is commonly defined as the reversible work to transfer a solute molecule from the vacuum into a liquid at given temperature and pressure. To predict the solvation free energy and its dependence on temperature and other thermodynamic conditions, I may use a wide variety of implicit and explicit solvent models[58]. In general, different solvation methods are complementary and their selection often reflects a compromise of the computational cost and the degree of microscopic details. An implicit solvent model neglects the microscopic structure of solvent molecules, which makes the theoretical analysis of the solvated molecules convenient and computationally efficient. Although the local solvent properties and structure are distinctively different from those in the bulk, implicit solvent models are able to provide accurate prediction of the solvation properties using few semi-empirical parameters[59-61]. For example, Chamberlin et al.[62] demonstrated that SM6T (solvation model 6 with temperature dependence) predicts the hydration free energies of 181 organic compounds over the temperature ranging from the freezing to the boiling point of liquid water (273 K to 373 K) with a remarkably small mean unsigned error (~0.08 kcal/mol). While the superior computational efficiency makes implicit solvent models a popular choice, one caveat is that they require the macroscopic properties of the solvent as an input, which may not be readily available under diverse thermodynamic conditions. Besides, an

implicit solvent model provides no information on the solvation structure important for understanding the chemical activity of the solvated molecules. Furthermore, the usefulness of macroscopic solvent properties and the validity of semi-empirical parameters can be questionable beyond the training set. To address these short-comings, recent continuum models that combines electron density functional theory requires only two or three parameters. For example, Andreussi et al.[63] applied their self-consistent continuum model predicts mean error of 1.3 kcal/mol over 240 neutral solutes with only 2 parameters. Furthermore, Sundararaman et al.'s empirical-parameter-free SaLSA[64] model reaches a RMS error of 1.3 kcal/mol in water and 0.8 kcal/mol in chloroform and carbon tetrachloride.

Explicit-solvent modeling of solvation is mostly based on molecular dynamics (MD) simulations[54]. While the procedure is in principle exact, MD simulation is computationally demanding even with modern computers, in particular if one is interested in high-throughput evaluation of solvation free energies. Unlike properties directly accessible from molecular configurations, the free-energy calculation relies on alchemical methods to sample a large number of microstates along various thermodynamic pathways to connect the vacuum and liquid states[65-66]. Although advanced simulation techniques are available to reduce the computational cost[67-73], simulation data for the solvation free energies of a large library of chemical systems are scarce[74]. One notable exception is the Monte Carlo (MC) simulation for the solvation properties of over 200 organic solutes in aqueous and organic solutions reported by Duff and Jorgensen[75]. More recently, Mobley *et al.*[44] simulated the hydration free energies of

504 small organic molecules at room temperature. The MD simulation yields a root-mean-square error (RMSE) of 1.24 kcal/mol in comparison to experimental data. The temperature effect on solvation free energies is less known. Paschek[76] investigated the hydration free energies of 4 noble gases and methane using MD and found that the results were in good agreement with experimental data. To our knowledge, no MD studies have been previously reported to investigate the temperature dependence of the solvation free energy for a large library of chemical compounds.

Liquid-state theories provide a valuable alternative to simulation methods for solvation calculations. Like MD simulations, liquid-state theories are able to predict both solvation structures and free energies with an explicit consideration of the solvent molecules. Over the past decades, there have been tremendous progresses in both integral-equation theories and classical density functional methods[5, 22, 28, 77-81]. The theoretical approaches have been demonstrated to yield accurate predictions of the hydration free energies for a large library of small organic molecules at room temperature. For example, Palmer et al.[82] reported that, with the partial molar volume (PMV) correction, the 3D-RISM (3-dimensional reference interaction site model) predictions yield a RMSE of 0.99 kcal/mol for 120 organic molecules at 298 K. Sergiievskyi et al.[83] suggested that the universal PMV correction may be justified in terms of different thermodynamic constraints underlying theoretical and experimental considerations. Although concerns were raised on the theoretical derivations[84], 3DRISM with the PMV correction has been successfully implemented by Palmer et al.[85] to predict the hydration free energies for 181 solutes from 278 K to 368 K with good accuracy in

comparison to the experimental data (RMSE of ~1.5 kcal/mol). In addition to the integral equations, the molecular density functional theory (MDFT)[43, 80, 86-91] is also able to provide fast predictions of the hydration free energies while capturing the microscopic details of the solvation shell. For example, MDFT predicts the hydration free energies of over 500 small organic molecules at 298 K with an average unsigned error of ~1 kcal/mol [83, 86]. It has also been used to investigate the performance of different force fields for predicting the solvation free energies of over ~2500 organic solutes in liquid water[87]. In contrast to the 3D-RISM model, MDFT takes proper account of the intramolecular correlations that are essential to describe the long-range nature of angular correlations between water molecules[92].

In this work, I propose a molecular density functional theory (MDFT) suitable for predicting solvation structure and thermodynamic properties of small organic molecules from first principles. I formulate the free energy functional following the universality of the bridge functional and the direct correlation functions (DCFs) of the bulk solvent. In combination with quantum-mechanical (QM) calculations for the solute structure and force-field parameters, the theoretical performance has been tested with the thermodynamic properties of solvation for 5 nitrotoluenes recently studied by MD simulations[93] and with the experimental data for the hydration free energies of 197 chemical species over a broad range of temperatures[94]. I demonstrate that the multi-scale procedure, i.e., a combination of QM calculations for the solute structure and partial charges of the solute atoms and the MDFT calculations of the thermodynamic properties of the solvent, is both computationally efficient and theoretically reliable for hydration

free energy predictions. For predicting the hydration free energies of the large library of chemical species, it yields an average unsigned error of ~1.3 kcal/mol over the entire range of temperatures (277-313 K).

The remainder of this article is organized as follows. Section II explains elements of the molecular density functional theory and theoretical justifications for formulation of the free energy functional employed in this work. In Section III, I give the computational details, in particular, the numerical methods to calculate the direct correlation functions at different temperatures, a key input for our MDFT predictions. In Section IV, I discuss the DCFs for pure water at different temperatures and the theoretical predictions of solvation in comparison with the experimental values. Finally, Section V summarizes the main results.

3.2. Molecular Density Functional Theory of Solvation

In this work, I confine our interest to solvation of small organic molecules in liquid water. As demonstrated in our previous work[86], the flexibility of the solute molecules can be taken into account by considering an ensemble average of different solute configurations and is relatively unimportant for small molecules.

With the assumption of a rigid configuration for each solute, the solvation free energy can be conveniently calculated in the context of the molecular density functional theory (MDFT). The solvent molecules surrounding the solute may be considered as an inhomogeneous system where the solute-solvent interaction is represented by an external potential. At given temperature T, system volume V, and chemical potential μ in the

bulk, the grand potential is a functional of the molecular density profile $\rho(\mathbf{x})$ of the solvent molecules[43]:

$$\begin{aligned}\Omega[\rho(\mathbf{x});\mu,V,T] = k_B T \int d\mathbf{x}\rho(\mathbf{x})\{\ln[\rho(\mathbf{x})\Lambda^3]-1\} \\ + \int d\mathbf{x}[V_B(\mathbf{x})+\varphi^{ext}(\mathbf{x})-\mu]\rho(\mathbf{x}) + F^{ex}[\rho(\mathbf{x})]\end{aligned} \tag{1}$$

In Eq.(1), k_B is the Boltzmann constant; $\mathbf{x}$ is a composite vector specifying the atomic positions of a solvent molecule; Λ represents an *effective* thermal wavelength which is irrelevant to this work; $V_B(\mathbf{x})$ and $\varphi^{ext}(\mathbf{x})$ are, respectively, the intramolecular bond potential and the external potential on a solvent molecule due to the presence of the solute. In MDFT calculations, the intra- and inter- molecular potentials are specified by a semi-empirical force field as in typical MD simulations.

The first term on the right side of Eq.(1) corresponds to the *intrinsic* Helmholtz energy of a non-interacting reference system, i.e., an ideal-gas system of the solvent molecules with the density profile the same as that of the real system under consideration (but without intermolecular interactions); the second term accounts for the one-body potential for each solvent molecule; and the last term, excess *intrinsic* Helmholtz energy, $F^{ex}[\rho(\mathbf{x})]$, represents the thermodynamic non-ideality arising from intermolecular interactions. At equilibrium, the grand potential is minimized with respect to the molecular density profile, leading to the Euler-Lagrange equation

$$\rho(\mathbf{x}) = \frac{1}{\Lambda^3}\exp[\mu - \beta V_B(\mathbf{x}) - \beta\varphi^{ext}(\mathbf{x}) - \delta\beta F^{\mathrm{ex}} / \delta\rho(\mathbf{x})] \tag{2}$$

where $\beta = 1/(k_{\mathrm{B}}T)$. Eq.(2) is formally exact in the context of the classical density functional theory[5]. With an explicit expression for the excess intrinsic Helmholtz energy, it allows us to calculate molecular density profiles and, subsequently, thermodynamic properties of solvation[6-7].

The dimensionality of the molecular density profile $\rho(\mathbf{x})$ grows rapidly with the number of atoms. If the solvent molecule has a rigid configuration, Eq.(2) can be numerically implemented by expressing the molecular position in terms of the center of mass and the orientation[8]. Alternatively, I can solve for the atomic density profiles of the solvent molecules[10, 95]

$$\rho_i(\mathbf{r}) = \rho_i^0 \left\langle \exp\left[\beta\mu_i^{\mathrm{ex}} - \beta\varphi_i(\mathbf{r}) - \delta\beta F^{\mathrm{ex}} / \delta\rho_i(\mathbf{r}) \right] \right\rangle_\omega \tag{3}$$

where ρ_i^0 is the number density of atom i in the bulk solvent, $<\cdots>_\omega$ stands for the Boltzmann average over all possible intramolecular configurations with atom i fixed at position $\mathbf{r}$, μ_i^{ex} stands for the atomic excess chemical potential in the bulk, and φ_i accounts for the external potential on atom i. In writing Eq.(3), I assume that the pair potential between solute-solvent molecules can be decomposed into contributions from individual atoms:

$$\varphi^{ext}(\mathbf{x}) = \sum_i \varphi_i(\mathbf{r}_i) \tag{4}$$

The pairwise additive assumption is commonly used in a non-polarizable force field.

The essential task in application of MDFT is to establish an explicit expression for the excess intrinsic Helmholtz energy as a functional of the molecular density profile,

$F^{ex}[\rho(\mathbf{x})]$. Without loss of generality, an analytical expression for the excess *intrinsic* Helmholtz energy can be expressed relative to that of a uniform reference system

$$F^{ex}[\rho_i(\mathbf{r})] = F^{ex}[\rho_i^0] + \sum_i \mu_i^{ex} \int d\mathbf{r} \Delta\rho_i(\mathbf{r}) - \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r} \int d\mathbf{r}' \Delta\rho_i(\mathbf{r}) \Delta\rho_j(\mathbf{r}') c_{ij}(|\mathbf{r}-\mathbf{r}'|) + F_B[\rho_i(\mathbf{r})] \quad (5)$$

where $\Delta\rho_i(\mathbf{r}) \equiv \rho_i(\mathbf{r}) - \rho_i^0$ stands for the deviation of the local density of atom *i* from the bulk value; $c_{ij}(r)$ is the direct correlation function (DCF) between atoms *i* and *j* of the uniform reference system; and $F_B[\rho_i(\mathbf{r})]$ represents the bridge functional, i.e., all higher-orders terms beyond the quadratic functional expansion of $F^{ex}[\rho(\mathbf{x})]$ relative to that of a uniform system. Eq.(5) remains formally exact, independent of the specific form of intramolecular interactions. It is worth noting that, without the bridge functional $F_B[\rho_i(\mathbf{r})]$, Eq.(5) reduces to the excess free-energy used in the Chandler-McCoy-Singer (CMS) theory[22]. The approximation is also equivalent to the hypernetted-chain approximation (HNC) for atomic systems or the homogeneous reference fluid (HRF) approximation proposed by Borgis and coworkers[90].

In our previous publications[43, 92], I demonstrated that the bridge functional, $F_B[\rho_i(\mathbf{r})]$, can be estimated from the *universality ansatz* for simple fluids[16-18]. Specifically, the bridge functional is approximated with that of a hard-sphere reference system $F_B^{HS}[\rho_i(\mathbf{r})]$ of the same atomic density profiles:

$$F_B[\rho_i(\mathbf{r})] \approx F_B^{HS}[\rho_i(\mathbf{r})] \equiv F_{HS}^{ex}[\rho_i(\mathbf{r})] - F_{HS}^{ex}(\rho_i^0) - \sum_i \mu_{0,i}^{HS} \int d\mathbf{r} \Delta\rho_i(\mathbf{r}) + \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r}_1 \int d\mathbf{r}_2 \Delta\rho_i(\mathbf{r}_1) \Delta\rho_i(\mathbf{r}_2) c_{ij}^{HS}(|\mathbf{r}_1 - \mathbf{r}_2|) \quad (6)$$

where $F_{HS}^{ex}(\rho_i^0)$ and $F_{HS}^{ex}[\rho_i(\mathbf{r})]$ are, respectively, the HS excess intrinsic Helmholtz energy at bulk density ρ_i^0 and at density $\rho_i(\mathbf{r})$; $\mu_{0,i}^{HS}$ is the HS excess chemical potential; and $c_{ij}^{HS}(|\mathbf{r}_1 - \mathbf{r}_2|)$ is the HS direct correlation functions. Accurate expressions for the Helmholtz energy and the DCFs of hard-sphere systems have been developed in the previous work[4, 19-20].

As mentioned above, the solvation free energy is defined as the reversible work to transfer a solute molecule from the vacuum to a bulk solvent at constant temperature T and solvent chemical potential μ. The reversible work corresponds to the difference between the grand potentials of the solvent system before and after the solute transfer

$$\Delta G_{sol} = \Omega[\rho(\mathbf{x}); \mu, V, T] - \Omega_0[\mu, V, T] \quad (7)$$

where Ω_0 stands for the grand potential of a bulk solvent of the same volume. By combing Eqs.(1) and (5)-(7), I derive an analytical expression for the solvation free energy

$$\beta\Delta G_{sol} = -\frac{1}{M_s} \sum_i \int d\mathbf{r} \Delta\rho_i(\mathbf{r}) + \frac{1}{2} \sum_{i,j} \iint \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' c_{ij}(|\mathbf{r} - \mathbf{r}'|)[\rho_i(\mathbf{r})\rho_j(\mathbf{r}') - \rho_i^0 \rho_j^0] + \beta F_B[\rho_i(\mathbf{r})] - \sum_i \int \mathrm{d}\mathbf{r} \rho_i(\mathbf{r}) B_i(\mathbf{r}) \quad (8)$$

where M_s represents the number of atoms in each solvent molecule, and $B_i(\mathbf{r})$ is defined as

$$B_i(\mathbf{r}) \equiv \delta F_B / \delta \rho_i(\mathbf{r}) \tag{9}$$

To calculate solvation free energy from Eq.(8), I need in advance the atomic density profiles of solvent molecules and the HS diameter for each atom that is used to estimate the bridge functional. The atomic density profiles can be obtained from Eq.(3), which is used for all solvation free energy calculations discussed in this work, or alternatively from molecular simulations[96]. For solvation in liquid water, the HS diameter of hydrogen atoms is chosen to be zero, which is consistent with many water models[97]. The HS diameter for oxygen atoms can be fixed by using either the self-solvation free energy or the hydration free energy of a single solute. In this work, I use σ_O = 2.91 Å for the HS diameter of oxygen atoms, which is obtained by reproducing the hydration free energy of benzene at 298 K. This value is close to the Barker-Henderson diameter[98] (2.97 Å) of the oxygen atom according to the SPC/E model.

3.3. Direct Correlation Functions at different temperatures

Our MDFT calculations require an explicit expression for the direct correlation functions (DCFs) of the pure solvent as an input. For interaction sites *i* and *j* from different solvent molecules, the DCF is defined as a second-order functional derivative of the reduced excess Helmholtz energy with respect to their density profiles:

$$c_{ij}(|\mathbf{r}-\mathbf{r}'|) \equiv -\frac{\delta^2 \beta F^{ex}}{\delta \rho_i(\mathbf{r}) \delta \rho_j(\mathbf{r}')} \tag{10}$$

Note that for a uniform system, each pair of DCF is a function of the center-to-center distance between sites *i* and *j* as well as of the thermodynamic variables defining the bulk

system. Physically, $c_{ij}(|\mathbf{r}-\mathbf{r}'|)$ represents the response of the local excess chemical potential of site i at position $\mathbf{r}$ to the variation of the local density of site j at position $\mathbf{r}'$ or vice versa.

As reported in a previous publication[99], I can evaluate the DCFs of a bulk molecular fluid using MD simulation in combination with the Reference Interaction Site Model (RISM)[32]:

$$\hat{c}(k) = \hat{\omega}^{-1}(k)\hat{h}(k)\left[\hat{\omega}(k) + \rho_b \hat{h}(k)\right]^{-1} \tag{11}$$

where ρ_0 represents the molecular number density of the bulk solvent; $\hat{h}, \hat{\omega}, \hat{c}$ are, respectively, the matrix forms of the site-site total correlation functions (TCFs), intramolecular correlation functions (ICFs), and DCFs in the Fourier space (k). For the SPC/E model considered in this work, the configuration of the solvent molecules is fixed. In that case, the intramolecular correlation functions are exactly known

$$\omega_{ij}(k) = \delta_{ij} + \left(1-\delta_{ij}\right)\frac{\sin kl_{ij}}{kl_{ij}} \tag{12}$$

where δ_{ij} stands for the Kronecker delta function, l_{ij} is the distance (or bond length) between interaction sites i and j from the same molecule. With the site-site total correlation functions $h_{ij}(r)$ calculated from MD simulation, I can obtain DCFs from the RISM equation, Eq.(11).

MD simulation is in general time-consuming even for calculating the bulk correlation functions. To minimize the computational cost, I can alternatively solve the TCFs and

DCFs together by combining the RISM equation with the modified HNC (MHNC) closure[39]

$$h_{ij}(r)+1=\exp\left[-\beta u_{ij}^{eff}(r)+h_{ij}(r)-c_{ij}(r)\right] \tag{13}$$

In Eq.(13), $u_{ij}^{eff}(r)$ represents the effective pair potential between sites *i* and *j*, which consists of a direct Coulomb energy $u_{ij}^{Coul}(r)$ and the Lennard-Jones (LJ) potential $u_{ij}^{LJ}(r)$:

$$u_{ij}^{eff}(r)=\zeta u_{ij}^{Coul}(r)+u_{ij}^{LJ}(r) \quad . \tag{14}$$

The scaling factor ζ is introduced such that the asymptotic limit of the DCF conforms to the exact result[95].

Because this work is concerned with solvation for a large library of chemical species at different temperatures, I use the DCFs calculated from the MHNC instead of from MD simulations. The RISM/MHNC calculations were performed with a homemade program. At each temperature, ζ is obtained in advance using the static dielectric constant of the bulk solvent, ϵ, and the molecular dipole moment, d,

$$\zeta=1+\frac{1}{\epsilon-1}-\frac{1}{3y} \tag{15}$$

where $y=4\pi\rho_b d^2/(9k_BT)$, and ρ_b is the number density of the bulk solvent.

3.4. Results and Discussions

The systems considered in this work include 5 nitrotoluenes as a test case for comparison with simulation results and a larger library of experimental data that consists

of 197 small organic molecules of different genres. Similar to MD simulations, application of MDFT to solvation requires as the input the solute structure as well as the force-field parameters to represent solute-solvent and solvent-solvent interactions. For all hydration free energy calculations, I use the SPC/E model for liquid water and assume that the solute configurations are assumed rigid. The non-bonded intermolecular interactions are described in terms of the Lennard-Jones (LJ) plus Coulomb potential. The solute structure and atomic partial charges were generated from various quantum-mechanical (QM) calculations as discussed in our earlier publications[87, 100].

DCFs of water at different temperatures

I calculated the DCFs for SPC/E water at seven temperatures (273K, 277K, 290K, 298K, 310K, 313K, 330K) using the RISM/MHNC equations. For numerical efficiency, the DCFs are divided into a long-range (LR) and a short-range (SR) component as discussed in our previous work[95]. The LR component conforms to the asymptotic limit

$$c_{ij}^{LR}(r) = -\zeta \frac{\beta q_i q_j}{4\pi\epsilon_0 r}\left[1-\exp\left(-\kappa r\right)\right] \tag{16}$$

where a dampening factor $\left[1-\exp(-\kappa r)\right]$, κ =1 Å^{-1}, was introduced to avoid divergence at $r=0$. The SR component of the DCFs, defined as

$$c_{ij}^{SR}(r) = c_{ij}(r) - c_{ij}^{LR}(r) \tag{17}$$

is solved by the Picard iteration.

Figure 3-1 shows the site-site DCF between oxygen atoms from different water molecules. When the site-site distance is larger than the Lennard-Jones diameter for the SPC/E water (σ =3.166 Å), the DCF is essentially the same as the effective pair

potential in the reduced form, $-\beta u_{ij}^{eff}(r)$. The asymptotic limit is consistent with the prediction of Eq.(13). The long-range behavior of the DCF conforms to the exact sum rules and has been thoroughly discussed in our previous work[95, 101]. As shown in Table 3-1, both the dielectric constant and the liquid density of the bulk fluid are approximately fixed in the temperature range considered in this work. To a good approximation, the long-range component of the DCF is inversely proportional to temperature. At short distance (r < 3.166 Å), the DCF between oxygen atoms is not much different from that for a hard-sphere fluid[4]. In both cases, the DCF falls monotonically to a maximum negative value as the site-site distance approaches zero. Notwithstanding the similarity, the absolute value of the maximum DCF between oxygen atoms is about one order of magnitude larger than that for hard spheres. While the hard-sphere system is athermal, i.e., the DCF and other thermodynamic quantities in dimensionless units are independent of temperature, the DCF between oxygen atoms is noticeably reduced when the temperature increases.

Figure 3-1(B) presents the short-range component of the DCF as defined by Eq.(17). It shows a non-monotonic dependence on r at short distance and vanishes beyond the Lennard-Jones diameter. The non-monatomic behavior implies that the short-range component of the DCF does not disappear simply by adjusting the screening parameter κ in Eq.(16). Interestingly, the short-range DCF is relatively insensitive to temperature except near its minimum at an intermediate distance. Besides, the minimum value (~-50) is not much different from that for a hard-sphere system of similar packing density. These similarities suggest that the short- and long-range components of DCF

may be represented by that for an effective hard-sphere system and that by hard-sphere-electrostatic correlations as predicted by the analytical solutions of mean spherical approximation (MSA) for simply electrolytes[102].

Figure 3-2 shows the three pairs of the DCFs predicted by the RISM/MHNC equations in comparison with MD simulation[101]. Here I present only the short-range component of the DCFs because the long-range component is represented by the same analytical equation, *viz*., Eq. (16). While there are discrepancies between theory and simulation near zero separation, I consider the RISM/MHNC predictions satisfactory because the maximum error, typically on the order of 10, is relatively small in comparison to the absolute values of the DCFs in the same range (>300).

Hydration of nitrotoluenes

Nitrotoluenes are important industrial agents used as pigments, photographic chemicals, pesticides, and explosives. I test our theoretical procedure against the hydration free energies of nitrotoluenes not only because of the practical relevance of such properties for environmental regulations but also for the readily availability of such results from previous experiment and simulation[93]. Specifically, I consider solvation of five nitrotoluenes in water from 273 K to 330 K. Figure 3-3 illustrates schematically the chemical structure of these nitrotoluenes: 2-nitrotoluene (2-NT), 4-nitrotoluene (4-NT), 2,4-dinitrotoluene (2,4-DNT), 2,6-dinitrotoluene (2,6-DNT), and 2,4,6-trinitrotoluene (2,4,6-TNT). In our MDFT calculations, the molecular parameters for each nitrotoluene molecular were generated according to a combination of quantum/statistical procedure optimized for nitro-compounds[100]. The atomic partial charges are generated through the

AM1-BCC model[103], and the Lennard-Jones parameters are taken from GAFF[104] using the convenient Antechamber tool kit[105].

Figure 3-4 compares the theoretical predictions of the hydration free energies for the five nitrotulenes with MD simulation and experimental values at four different temperatures (273 K, 290 K, 310 K and 330 K). Using the experimental data as a benchmark, I find that the average unsigned error (AUE) of the theoretical predictions is 0.74 kcal/mol, which is comparable to the AUE of 0.68 kcal/mol for MD simulation[93]. The magnitude of the solvation free energy increases with the number of nitro groups due to more extensive hydrophilic hydration. Both theory and experiment indicate that the hydration free energy is relatively insensitive to the positions of the nitro groups. While the hydration free energies for 2-NT and 4-NT are virtually identical, MDFT predicts that the hydration free energy for 2,4-DNT is slightly more negative than that for 2,6-DNT, which is in good agreement with both experimental and simulation data. The small difference may be attributed to somewhat enhanced van der Waals interaction between 2,4-DNT and water molecules due to its slightly more extended molecular structure.

Based on the hydration free energy data at different temperatures, I can estimate the hydration enthalpy using the van't Hoff equation

$$\frac{\partial \beta \Delta G_{sol}}{\partial \beta} = \Delta H_{sol} \quad . \tag{18}$$

The hydration entropy for each solute is simply related to the difference between the free energy and the enthalpy, $\Delta S_{sol} = (\Delta H_{sol} - \Delta G_{sol}) / T$. Figure 3-5 presents the temperature dependence of the hydration free energies of the five nitrotoluenes from experimental

data, MD simulation, and the theoretical predictions. The linear correlation between the hydration free energy and the reverse temperature suggests that the enthalpy of solvation is approximately constant within the range of temperatures considered in this work. Unlike hydrophobic solutes such as alkanes, the hydration enthalpies of nitrotoluenes are relatively insensitive to temperature because of the presence of both hydrophobic and hydrophilic groups.

Figure 3-6 presents the enthalpy and the entropy of hydration obtained from the van't Hoff equation. The hydration enthalpy becomes noticeably more negative as the number of nitro groups increases due to their favorable interactions with water molecules. However, the hydration energy does not linearly depend on the number of nitro groups due to the interference of neighboring groups. The spatial interference of hydration for different functional groups is evident in the distribution of water molecules in the solvation shell. As shown in Figure 3-7, the local density of oxygen atoms is strongly correlated with the position of nitro groups. Although the solvation energies of 2NT and 4NT are similar, their solvation structures are noticeably different, in particular for oxygen density near the nitro group. The oxygen density in the hydration shell becomes more localized as the number of nitro groups increase. Both MD and MDFT predict the enthalpy of solvation with about 1 kcal/mol deviation in comparison to the experimental data. Except for 2,4,6-TNT, both theory and simulation underestimate the enthalpy of solvation in absolute values. The discrepancy may be attributed to strong conjugation between the nitro group with the π-electrons of the benzene ring that is not adequately captured by conventional force fields[93].

Figure 3-6(B) and (C) show the entropic contributions to the hydration free energies of the five nitrobenzenes at two representative temperatures. In both cases, the entropic contributions rise only slightly as the number of nitro groups increases, implying that the hydration entropies are dominated by hydrophobic effects. Although the MDFT predicts different solvation structures for different isomers, the variations in hydration entropies are relatively insignificant.

High-throughput solvation calculation

A major objective of liquid-state methods, MDFT included, is to provide fast prediction of solvation free energies that otherwise are too time-consuming for MD simulations. Computational efficiency is important not only for high-throughput calculations but also for validation of molecular models/force fields with extensive experimental data. In addition to the five nitrobenzenes discussed above, I have also tested the MDFT performance with the hydration free energies for a large library of chemicals that consists of 197 small neutral organic compounds[94]. Since the solutes are small, and relatively rigid temperature-triggered conformational change is neglected. To generate the solute structure and the molecular parameters, I carried out quantum-mechanical (QM) calculations using the Hartree-Fock (HF) method with the SVP basis set at the vacuum condition (Vac). The partial charges are parameterized with the ChelpG scheme and the Lennard-Jones parameters are taken from the OPLS-AA force field[106]. This combination HF/SVP/Vac/OPLS-AA/ChelpG was found to give the best overall performance. ORCA 3.0.1[107] software package was used to perform all QM calculations.

Once the force field is assigned, the hydration free energies are predicted for each solute in the temperature range from 0°C to 40°C.

Figure 3-8 compares the MDFT predictions with experimental data for the hydration free energies of 197 organic compounds at three representative temperatures (277 K, 298 K and 313 K). Here the experimental results were obtained from Henry's constants at different temperatures. In general, the AUE values, from 1.08 kcal/mol at 313 K to 1.22 kcal/mol at 277 K, are comparable with 1.19 kcal/mol reported in our previous work for the hydration free energies of 512 compounds at room temperature[99]. The good correlation indicates that MDFT is able to provide reasonable prediction of the hydration free energies of small organic molecules at different temperatures.

Like MD simulation, the MDFT predictions are sensitive to the molecular model/force field parameters and its performance varies significantly with different groups of chemical species. On average, non-polar molecules have better prediction than polar molecules. Figure 3-9 presents the distributions of the AUE values based on different functional groups in the test set. I see that the MDFT predictions are poorest for peroxides with AUE around 3 kcal/mol, while the predictions for hydrophilic compounds such as phenols, alcohols, acids and nitrates are less than 2 kcal/mol. The MDFT performs best for small hydrophobic chemicals like alkanes, alkenes, ketones, with AUE all around 0.5 kcal/mol. The performance for different chemicals is partially related to the effective hard-sphere diameter used in the bridge functional, which was calibrated with the hydration free energy for methane. As demonstrated in our earlier work, the

theoretical performance can be further improved by calibrating the force field parameters for chemical species with a similar type of functional groups.

3.5. Conclusions

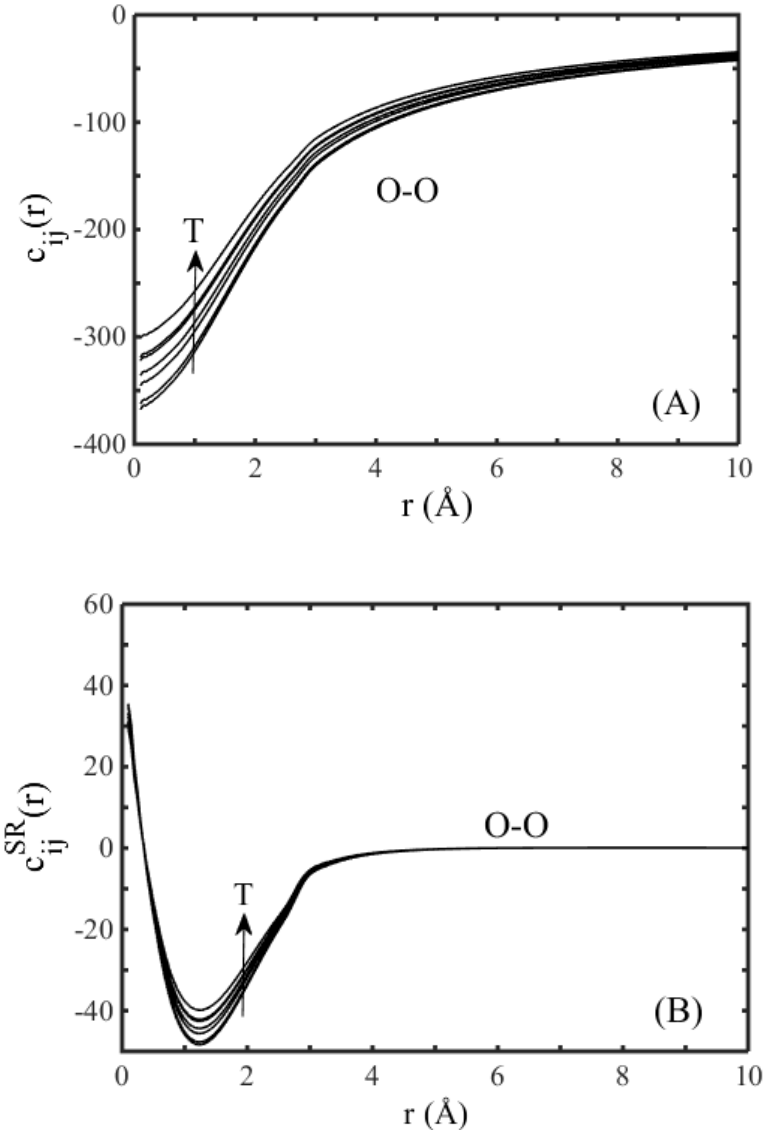
In this work, I have established a molecular density functional theory (MDFT) that is able to predict the solvation structure and corresponding thermodynamic properties using the site-site direct correlation functions (DCFs) of the pure solvent and the bridge functional of a reference hard-sphere fluid as the input. The DCFs can be calculated from molecular dynamics (MD) simulations in combination with the reference interaction site model (RISM) or from the numerical solution of the integral-equation theory (RISM/MHNC). I show that, to a good approximation, the long-range component of the DCFs for liquid water are can be represented by an effective pair potential similar to the mean-spherical approximation (MSA) for ionic systems. The short-range components of the DCFs are significant only when the site-site distance is smaller than the solvent diameter, and they are relatively insensitive to temperature changes. With the DCFs as the input, the molecular density functional theory (MDFT) is able to predict the hydration free energies of small organic compounds in good agreement with molecular dynamics simulations and experimental data over a broad range of temperatures.

The reduced hydration free energies of nitrotoluenes show a linear dependence on the inverse temperature, suggesting that their hydration enthalpies are relatively insensitive to temperature changes. Different from typical hydrophobic hydration such as those for alkanes that are entropy dominant, hydration of nitrotoluenes is dominated by hydrophilic interactions between the nitro groups and water molecules. While the

hydration entropies remain little changed as the number of nitro groups increases, the hydration enthalpies of nitrotoluenes become significantly more negative as the solute becomes more hydrophilic. The hydration energies and entropies of are similar for different isomers of nitrotoluenes, even though their local solvation structures are noticeably different.

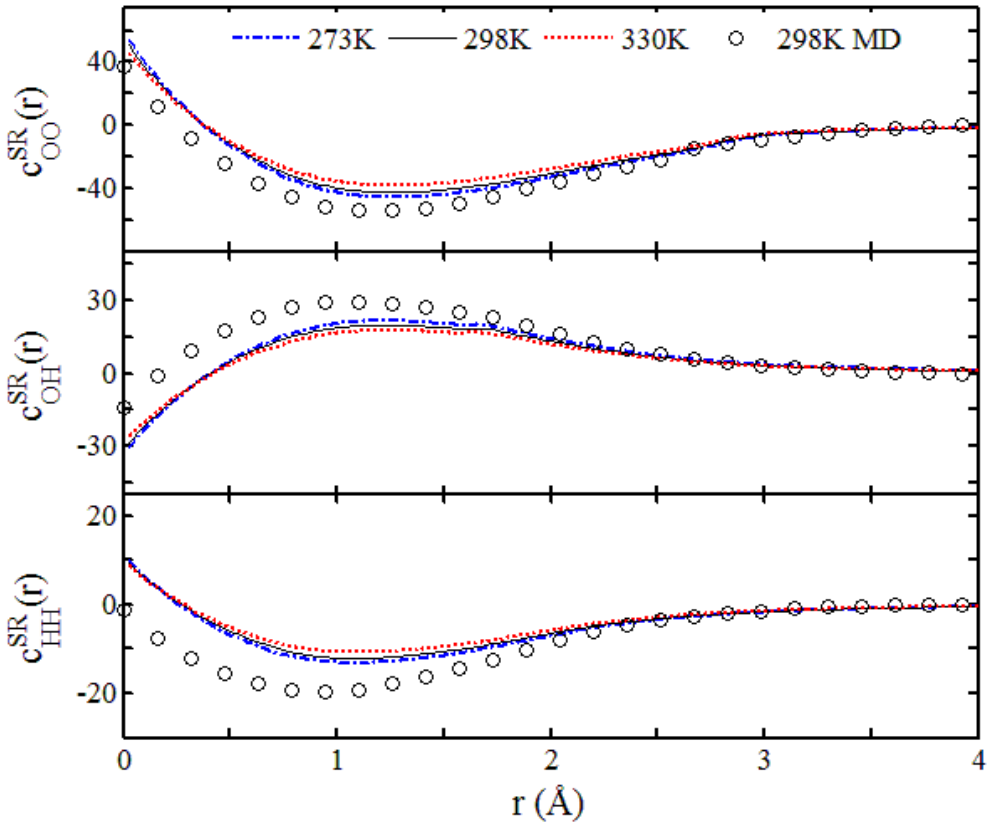
I demonstrated that MDFT is able to predict hydration free energies of 197 small organic molecules from 0 to 40 °C with the absolute unsigned error (AUE) slightly higher than 1 kcal/mol. Through categorizing the error according to the functional groups of different chemical species, I find that peroxides have the worst performance with an AUE at ~3 kcal/mol. This analysis provides the possibility for customized optimization of the molecular models/force field parameters for chemicals with similar functional groups. In combination with the customized model, the MDFT provides an alternative to predict solvation properties of chemical species for which experimental data are scarce or unavailable[108].

Figure 3-1



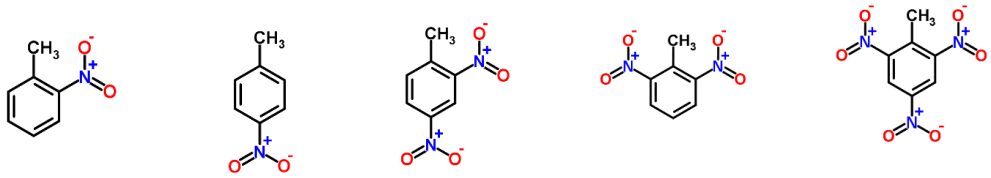
(A) The site-site direct correlation function (DCF) between the oxygen atoms of water molecules predicted from the RISM/MHNC equations. From bottom to top, the temperature varies from 277 K to 330 K. (B) The short-range (SR) component of the DCF at different temperatures.

Figure 3-2



Comparison of the RISM/MHNC predictions with MD data for the short-range (SR) components of the site-site direction correlation functions of the SPC/E water, $c_{OO}^{SR}(r)$, $c_{OH}^{SR}(r)$ and $c_{HH}^{SR}(r)$. The RISM/MHNC predictions indicate that $c_{ij}^{SR}(r)$ is relatively intensive to temperature changes.

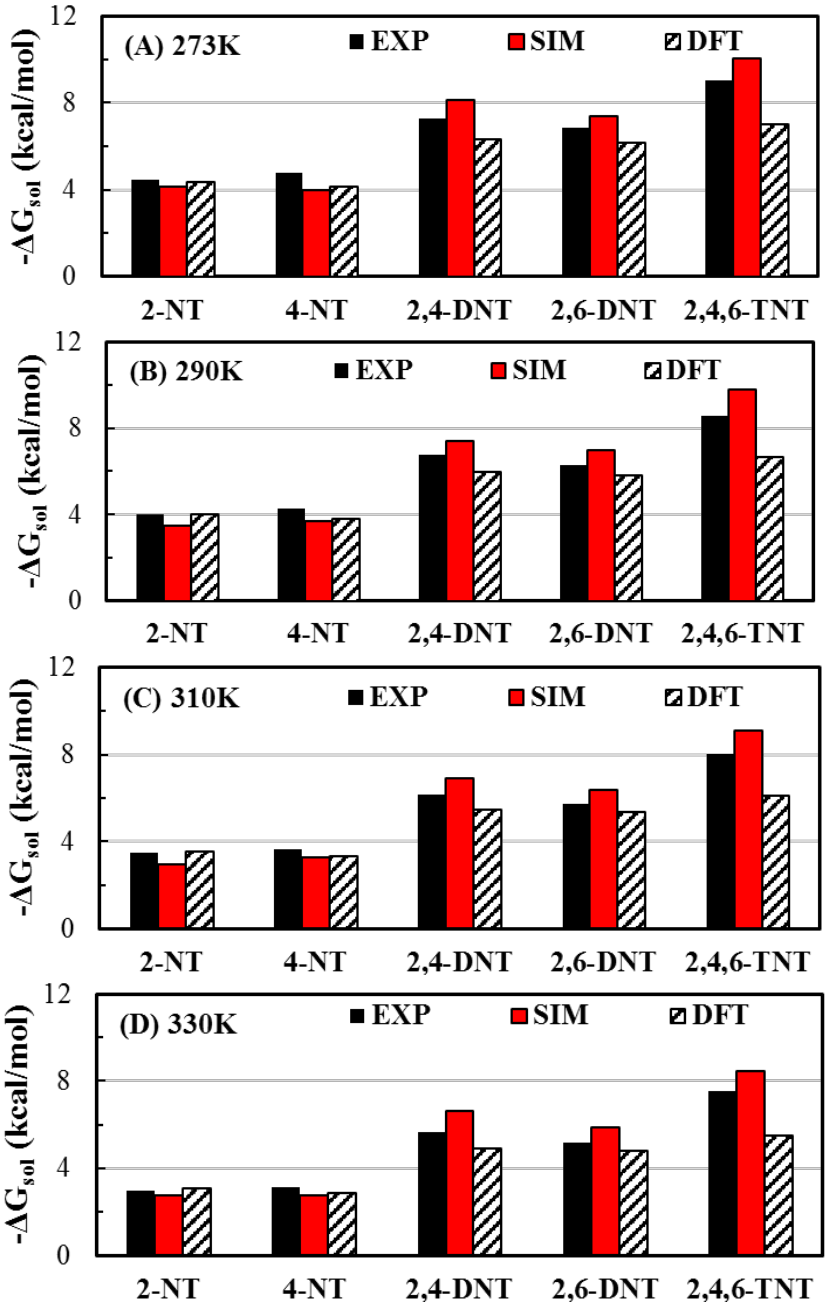
Figure 3-3



2-NT 4-NT 2,4-DNT 2,6-DNT 2,4,6-TNT

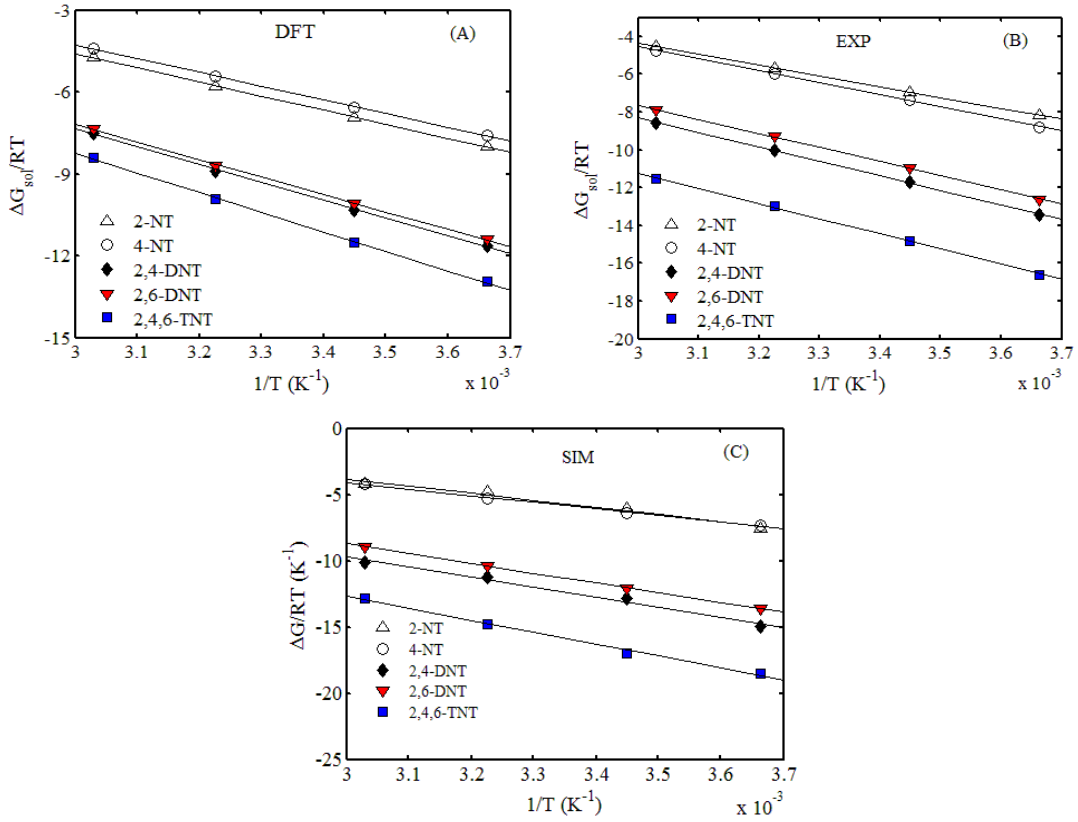
Chemical structures of five nitrotoluenes: 2-nitrotoluene (2-NT), 4-nitrotoluene (4-NT), 2,4-dinitrotoluene (2,4-DNT), 2,6-dinitrotoluene (2,6-DNT) and 2,4,6-trinitrotoluene (2,4,6-TNT).

Figure 3-4



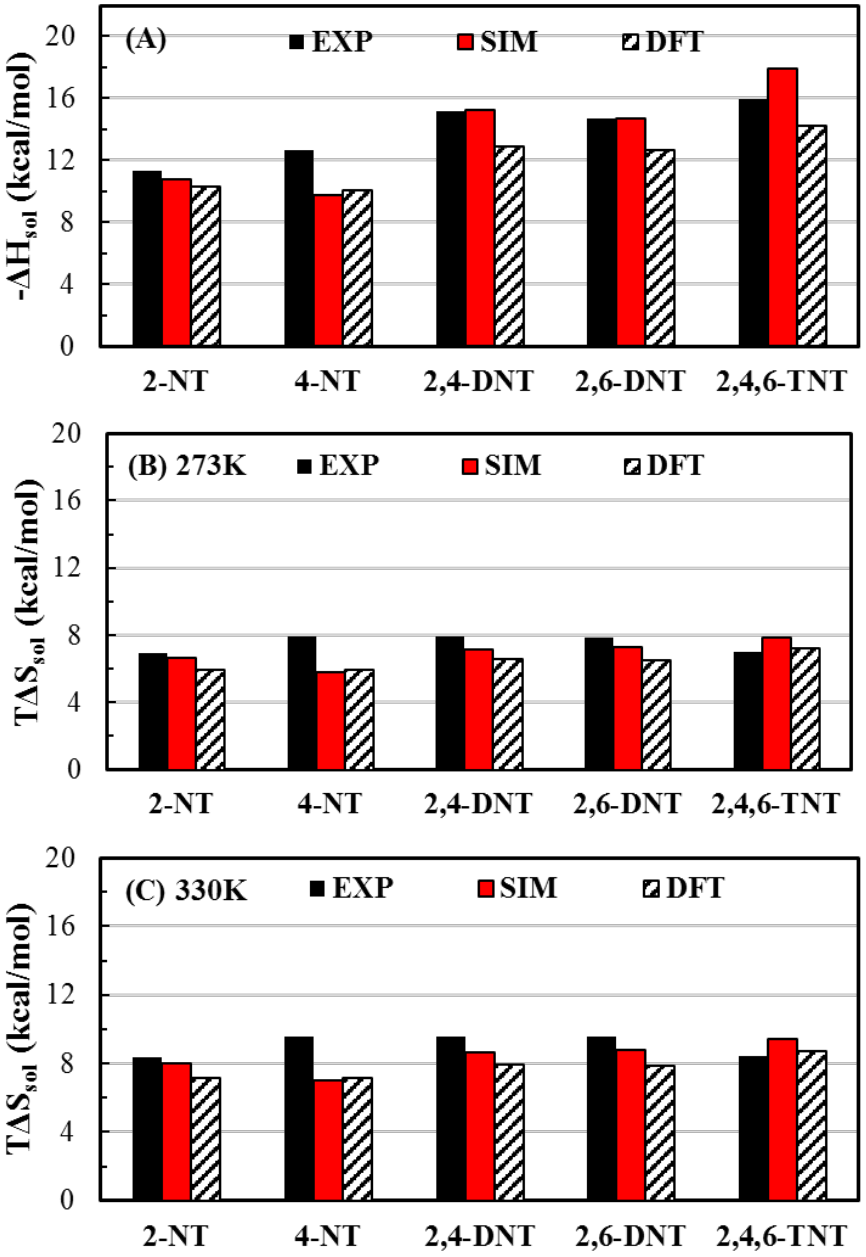
Comparison of theory (DFT), MD simulation (SIM) and experiment (EXP) for the solvation free energies of five nitrotoluenes at 273 K, 290 K, 310 K and 330 K.

Figure 3-5



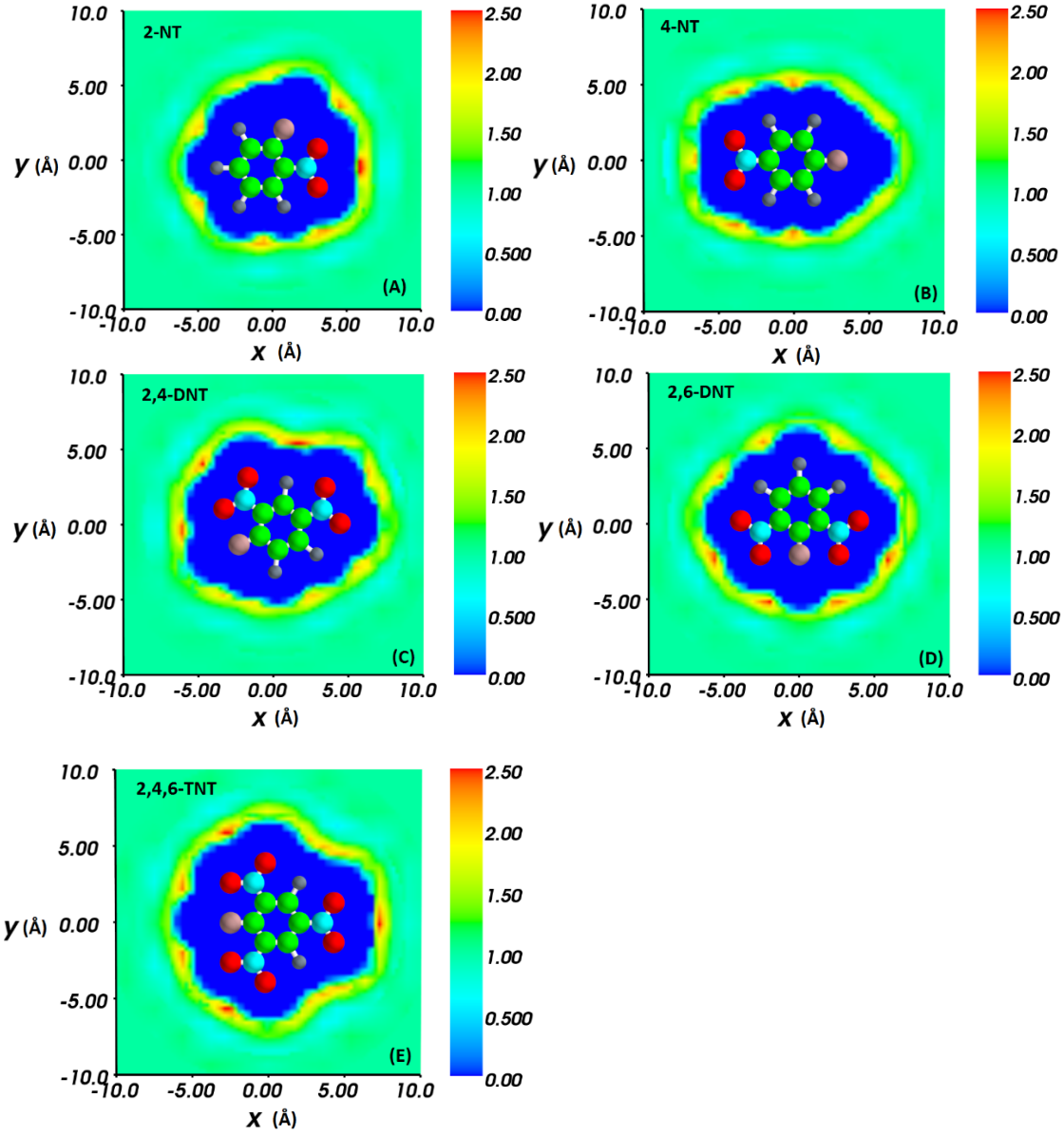
Linear fitting of the reduced solvation free energy $\Delta G_{sol} / RT$ versus $1/T$. The points are from A) DFT predictions, (B) experimental data and (C) MD data[93]

Figure 3-6



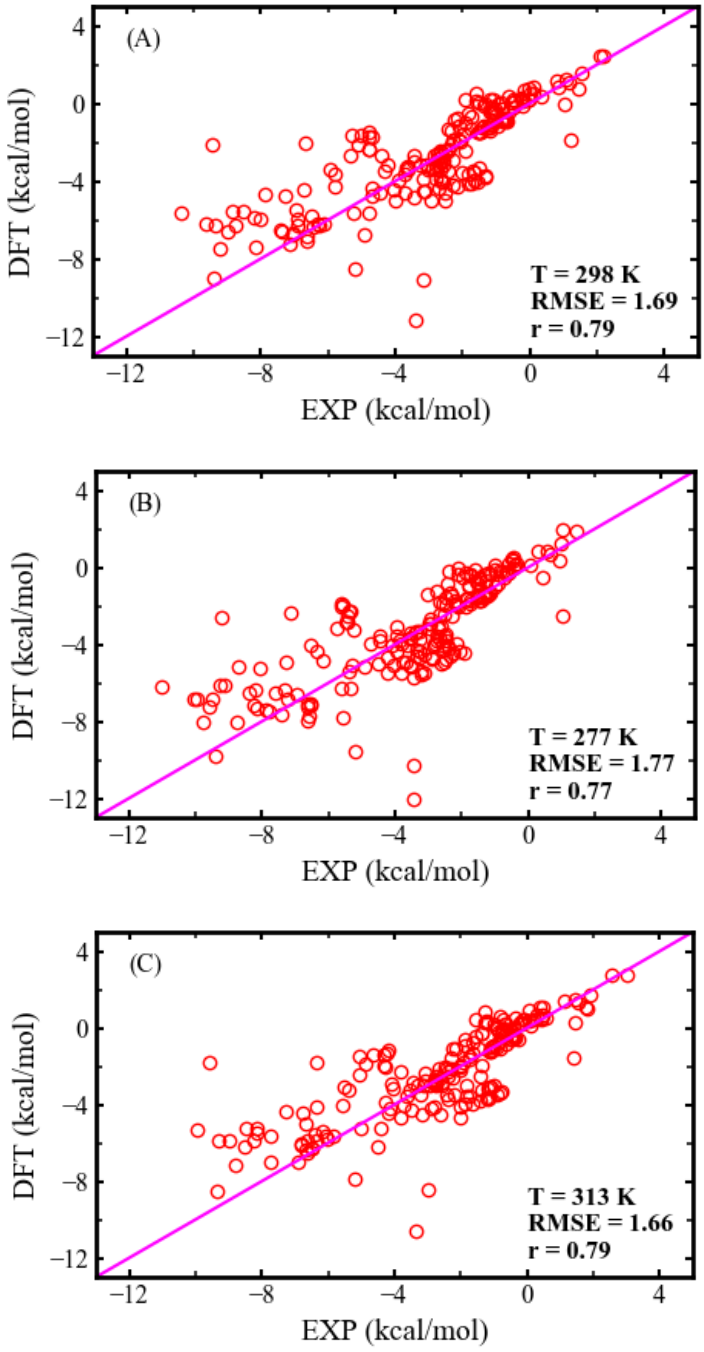
Comparison of theory (DFT), MD simulation (SIM) and experiment (EXP) for the hydration enthalpies of five nitrotoluenes (A) and the corresponding hydration entropies at 273 K (B) and 330 K (C).

Figure 3-7



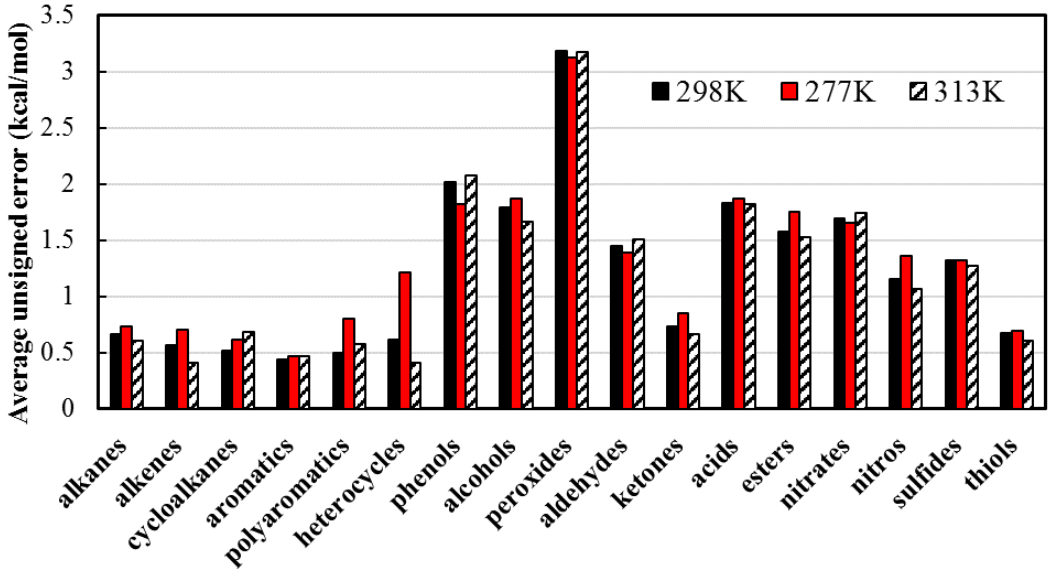
The density profiles of water molecules (at the position of the oxygen atom) near 5 solvated nitrotoluenes at room temperature. The sidebars show the ratio of the local oxygen density to the bulk density, $\rho_O(r)/\rho_O^b$.

Figure 3-8



Comparison between theory (DFT) and experiment (EXP) for the hydration free energies of 197 organic solutes. The average unsigned errors (AUE) are 1.22, 1.11 and 1.03 at 277 K (A), 298 K (B) and 313 K (C), respectively.

Figure 3-9



The average unsigned error corresponding to molecules with different functional groups.

Table 3-1 The scaling factor as a function of temperature for the SPC/E water

Temperature T(K)	**Density* ρ(g/cm3)**	**Dielectric Constant** Ɛ**	**Scaling Factor ζ**
273	999.84	88.2	0.9626
277	999.97	86.406	0.9621
290	998.77	81.53	0.9604
298	997.05	78.408	0.9594
310	993.33	74.69	0.9577
313	992.22	73.176	0.9574
330	984.71	67.63	0.9550

* From G. S. Kell's (1975) compilation[109]

** From D. P. Fernández's (1995) compilation[110]

Chapter 4. Solvation of Amine-grafted Silica Gel

Amine-grafted silica gel is an efficient heterogeneous catalyst for the Knoevenagel condensation and draws much attention in green chemistry for applications like heavy metal adsorption and CO_2 fixation. Despite its successful usage in diverse areas, fundamental questions remain on how the silica substrate affects the local chemical environment of the tethered amines. In this work, I use all-atom molecular dynamics simulation to investigate the solvation structures of two primary amines tethered onto a silica surface at different pHs of aqueous solutions. The atomic density profiles in the solvation shell are analyzed with a spherical harmonics expansion method for both isolated and silica-supported amines in different aqueous environments. The simulation results show that the hydration structure is influenced by the presence of surface but is rather insensitive to the surface charge. In addition, the surface effect is less prominent on tethered amine of longer chain length. Further hydrogen bonding analysis agrees with the conclusion.

4.1. Silica Gel

Recent advances in the synthesis of nanostructures offer a new paradigm for constructing catalyst systems to effectively achieve multistep reactions through spatial and temporal control of reaction pathways and transport of reactants, intermediates, and products. While interaction among the reactive species in a highly-coordinated system may be distinctively different from those in vacuum or corresponding to a bulk phase, relatively little is known on how the nanostructures affect the local chemical environment, molecular transport, and the reactivity of tethered functional groups. In this

work, I are interested in the difference between the solvation structures of freestanding amines and those tethered at a silica surface at different solution conditions. The solvent effect is often not explicit in conventional studies of surface reactions[60].

The amine functional group finds many applications in asymmetric reactions[111], heavy metal sequestering[112], and carbon dioxide fixation[113]. In particular, it is an efficient catalyst for the Knoevenagel condensation, *viz*., dehydration of aldehydes with active hydrogen compounds in a basic environment to produce ketones. The C-C bond-forming reaction follows the green chemistry principles with water as its only co-product and thus draws much research attention. In practical applications, amines are often grafted to a support to form a heterogeneous catalyst for the purposes of recycling and preventing instability and micelle formation[114]. Due to its chemical stability and synthetic simplicity, silica gel is often used as the catalyst support. The amine-functionalized silica gel catalyst system is reported to achieve fast Knoevenagel reactions at room temperature[115]. Similar catalyst systems have been demonstrated to function as adsorbents of heavy ions[116], CO_2[117], and formaldehyde vapor[118]. In this work, I limit our discussion to amine-functionalized silica systems despite the success of other catalyst supports like ZrO_2[119-120] or TiO_2[121]. Towards a rational design to improve the catalytic performance, I are interested in understanding how the silica surface affects the local chemical environment of a tethered amine molecule in comparison to that of its molecular counterpart in the bulk. Although the electronic structure of amine molecules in vacuum is well understood, the solvent and surface effect has not yet drawn much research attention[122].

Molecular dynamic simulations (MD) and integral-equation theories are commonly used for the computational studies of the structures of inhomogeneous fluids. Many high-performance packages like AMBER, GROMACS, and LAMMPS have made MD easily applicable to solving realistic problems including large biological systems[123]. One major limitation of MD is the computational cost, in particular for the accurate description of the 3-dimensional structure[124]. For example, up to 60 ns simulation time is required for sampling 3D atomic density profiles in an aqueous system with 0.5 Å of mesh resolution. Liquid-state theory provides an alternative to study the density profiles[87, 125].

In this work, I investigate the solvation structure of free and tethered amines based on MD simulations. The spherical harmonics expansion method is used for efficient sampling of the density profiles. I compare the difference in the local solvent density between silica-supported amines and corresponding amine molecules in the bulk solution. The rest of the article is structured as below. Section II provides the details of MD simulations. Section III explains the method of sampling spherical harmonics coefficients. Section IV discusses the results from MD simulations. Section V provides the conclusion.

4.2. MD Simulations of Amine-grafted Silica Gel

Recent years have witnessed many successful generalized force fields for different purposes, like the general AMBER force field (GAFF)[104], CHARMM general force field (CGenFF)[126], and the optimized potential for liquid simulations (OPLS)[127]. However, challenges remain to simulate aqueous silica interfaces reasonably accurate in order to

reproduce experimentally measured surface properties, such as the heat of immersion, contact angle and adsorption isotherm. This leads to the development of a number of specialized force fields for silica surface[128-132]. Among those, CLAYFF[128] gains popularity due to both simplicity and accuracy. CLAYFF has only two terms of bonding energy to maintain proper silanol bond and angles, one for O-H bond and the other for Si-O-H angle. And the lattice in CLAYFF is fixed to save computational time. In contrast, the force field used in this work from Emami et al[132] has full atomic mobility. Emami et al's force field claims to reproduce experimental heat of immersion in water and water contact angles and is compatible with commonly used force field such as AMBER, CHARMM, CVFF, PCFF, COMPASS and INTERFACE. According to Kroutil et al[133], both CLAYFF and Emami et al's force field yield good agreement with X-ray experimental data for water, metal ions (Na^+,Rb^+,Sr^{2+}) and negatively charged quartz (101) surface system.

From Emami's database, I prepared a slab of α-cristobalite silica. The surface contains only single silanols (Q_3) in contact with liquid water. The number density of silanol groups at the surface is 4.7 per nm^2. The slab size is $3.34\times3.48\times2.39$ nm with x-y as the facet surface. In the z-axis, the box is extended to 6.69 nm and filled with 1542 water molecules. One of the silanol groups from the surface is chemically bonded with a primary amine. Both the water molecules and amines are parameterized using the INTERFACE force field[134], which is compatible with Emami's silica model. By comparison, the primary amine without silica support is also simulated in a cubic box of 35 Å in side length. Figure 4-1 shows the scheme of the MD simulation.

To investigate the chain-length effect on the solvent structure, I compare the solvation structures for methyl (C1) amine and n-propyl (C3) amine. The pH effect on the solvation is implemented by deprotonation of surface silanol groups with the surface charge balanced by ionic species from the NaCl electrolyte. In this work, I consider two pH conditions, pH = 3 and 5, corresponding to, respectively, deprotonation of 0% and 5% of the surface silanol groups[132]. The amine groups, either tethered on the silica or free in the bulk, are assumed to be protonated at these conditions ($pK_b \approx 3-4$).

After adding water molecules in the simulation box, I first perform energy minimization for the entire system by quenching the amine configuration. Then MD runs are carried out under the NPT (isothermal-isobaric) ensemble at 298 K and 1 atm with 1 fs time step for 0.5 ns using the Nosé-Hoover style thermostat and barostat. After that, the system is equilibrated in the NVT ensemble at 298K for 1 ns, followed by 2 ns trajectory production for post-analysis under the same NVT ensemble. All MD simulations are carried out in the LAMMPS package through a velocity-Verlet integrator[135].

4.3. Sampling with Spherical Harmonics Expansion

Sampling 3D Density Profile

I follow a spherical harmonics expansion (SHE) method originally proposed by Garde and coworkers[136] to improve the efficiency of sampling for the 3D density profiles of solvent atoms. Instead of directly sampling on a 3D mesh grid, SHE allows us to calculate 1D coefficients for a limited number of spherical harmonics. In practice, the convergence time of both SHE and 3D mesh grid sampling depends on the grid spacing.

The higher the resolution is, the longer the trajectory is needed. The reduction in dimensionality from 3D to 1D significantly shortens the sampling time.

It is worth noting that the definition of spherical harmonics is slightly different in different fields. In this work, I pick the real value form defined in the geodesy:

$$\begin{aligned} Y_{nm}^{c}(\theta,\phi) &= \sqrt{2-\delta_{m0}}\sqrt{(2n+1)\frac{(n-m)!}{(n+m)!}}P_{nm}(\cos\theta)\cos(m\phi) \\ Y_{nm}^{s}(\theta,\phi) &= \sqrt{2-\delta_{m0}}\sqrt{(2n+1)\frac{(n-m)!}{(n+m)!}}P_{nm}(\cos\theta)\sin(m\phi) \end{aligned} \tag{1}$$

where δ_{m0} is the Kronecker delta function, P_{nm} is the associate Legendre polynomial, θ and ϕ are polar and azimuth components of solid angle Ω , respectively; n, m are integer indices $n = 0,1,2,....$ & $m \in [0,n]$; superscripts c and s represent the cosine and sine parts of spherical harmonics. The orthonormal condition requires

$$\frac{1}{4\pi}\int\int Y_{nm}^{x}Y_{n'm'}^{x'}d\Omega = \delta_{nn'}\delta_{mm'}\delta_{xx'} \tag{2}$$

where $d\Omega = \sin\theta d\theta d\phi$, subscript x denotes either c or s . Expanding the normalized density profile in spherical harmonics gives

$$\frac{\rho(\mathbf{r})}{\rho_b} = \sum_{n,m} C_{nm}(r)Y_{nm}^{c}(\Omega) + \sum_{n,m} S_{nm}(r)Y_{nm}^{s}(\Omega) \tag{3}$$

where ρ_b is the bulk solvent density, C_{nm} and S_{nm} are coefficients depending on the distance of an atom from the coordinate center. Apparently, S_{n0} is not of interest since $Y_{n0}^{s} \equiv 0$. As a result, for the first N primary indices ($n \le N$), I have $(N+1)^2$ coefficients.

SHE reduces the problem of sampling the 3D profile to that for sampling 1D functions C_{nm} and S_{nm}. To derive the explicit equations for sampling the coefficients in Eq.(3), I may compare it with how the 3D density profile is normally calculated by an ensemble average:

$$\rho(\mathbf{r})/\rho_b = \left\langle \sum_{i=1}^{Natom} \rho_i^{loc}(\mathbf{r}) \right\rangle / \rho_b \tag{4}$$

where ρ_i^{loc} is the instantaneous local density of atom *i*, i.e., $\rho_i^{loc}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_i)$. For a system with spherical symmetry, calculating the local density is similar to sampling the RDF:

$$g(r) = \left\langle \sum_{i=1}^{Natom} \rho_i^{loc}(r) \right\rangle / \rho_b \tag{5}$$

where the instantaneous local density is $\rho_i^{loc}(r) = \dfrac{\delta(r - r_i)}{4\pi r^2}$ with r_i being the distance between atom i and the coordinate center. Sampling the ensemble average of Dirac-delta functions requires drawing mesh grids on the radial axis and counting the number of atoms located between them. By integrating Eq.(5) over one radial mesh grid (r_M, r_{M+1}), I get

$$\int_{r_M}^{r_{M+1}} dr \int_\Omega g(r) r^2 d\Omega = \left\langle \sum_{i=1}^{Natom} \int_{r_M}^{r_{M+1}} dr \int_\Omega \frac{\delta(r_i - r)}{4\pi r^2} r^2 d\Omega \right\rangle / \rho_b \tag{6}$$

Since the interval is small, the integral of left hand side can be represented by a multiplication. Integration of the Dirac delta function on the right-hand side is simply

counting the number of atoms located in the interval. After simplification, Eq.(6) can be rewritten as

$$g(r_M) = \frac{\langle N_M \rangle}{4\pi r_M^2} / \rho_b \tag{7}$$

where N_M is the number of atom located within the radial mesh grid (r_M, r_{M+1}).

Now multiply both sides of Eq.(4) with $Y_{nm}^c(\Omega)$ and integrate over the radial mesh grid (r_M, r_{M+1}):

$$\int_{r_M}^{r_{M+1}} dr \int_\Omega \rho(\mathbf{r}) / \rho_b Y_{nm}^c(\Omega) r^2 d\Omega = \left\langle \sum_{i=1}^{Natom} \int_{r_M}^{r_{M+1}} dr \int_\Omega \delta(\mathbf{r} - \mathbf{r}_i) Y_{nm}^c(\Omega) r^2 d\Omega \right\rangle / \rho_b \tag{8}$$

By substituting SHE (Eq.(3)) and utilizing the orthonormal condition (Eq.(2)), I find

$$\int_{r_M}^{r_{M+1}} C_{nm}(r) 4\pi r^2 dr = \left\langle \sum_{i \in (r_M, r_{M+1})} Y_{nm}^c(\Omega_i) \right\rangle / \rho_b \tag{9}$$

Where the summation in the angular bracket is operated over all atoms located in the radial mesh grid. After simplification, I obtain the final equations for sampling the density profiles:

$$C_{nm}(r_M) = \frac{\left\langle \sum_{i \in (r_M, r_{M+1})} Y_{nm}^c(\Omega) \right\rangle}{4\pi r_M^2} / \rho_b \tag{10}$$

$$S_{nm}(r_M) = \frac{\left\langle \sum_{i \in (r_M, r_{M+1})} Y_{nm}^s(\Omega) \right\rangle}{4\pi r_M^2} / \rho_b \tag{11}$$

It is worth noting that the zeroth coefficient for $C_{nm}(r_M)$ is

$$C_{00}(r_M) = \frac{\left\langle \sum_{i\in(r_M, r_{M+1})} Y_{00}^{c}(\Omega_i) \right\rangle}{4\pi r_M^2} / \rho_b = \frac{\langle N_M \rangle}{4\pi r_M^2} / \rho_b \tag{12}$$

Eq.(12) indicates that $C_{00}(r_M)$ is identical to the radial distribution function (RDF) if the system is homogeneous. In most applications, SHE is truncated at N primary indices and any coefficients higher than that is assumed negligible. For systems considered in this work, I find that N = 10 is sufficient to reproduce the density profiles. To further reduce the dimensionality, I assume that pure C1 and C3 amines are solvated in water as rod-like molecules. As a result, I only need to sample the isotropic coefficients with regard to the azimuth angle (m=0). For amines tethered to silica surface, the full set of spherical harmonics coefficients are sampled.

4.4. Results and Discussions

I have sampled the solvent structure for isolated amines and silica-supported amines from the MD simulation trajectories by applying the SHE method to the atomic density profiles. Figure 4-2a shows the simulation results for methylamine (C1 amine) solvated in pure liquid water at 298 K. Water molecules are enriched near both the C terminal and the N terminal of the amine molecule but with different preferential angles. The water molecules close to the amino group are more organized than those around the methyl group because the amino group forms a hydrogen bond to water molecule. Because of the small separation between these two functional groups, the hydrogen bonding and hydrophobic hydration appears highly correlated, engendering higher density of water

molecules near the methyl group. Such correlation diminishes as the chain length increases from C1 to C3 (n-propylamine) (Panel c). In this case, water molecules are distributed more uniformly in the hydration shell and are slightly depleted from the boundary of the hydrophobic and hydrophilic domains. The segregation of hydrophilic and hydrophobic hydrations becomes more distinctive at low pH when the amines are protonated as shown in Panels c and d. At an acidic condition (here pH=3), the amino group gets protonated and attracts more water molecules due to more hydrogen bonding and ion-dipole interactions. Panel b shows that water molecules are highly packed in the first hydration shell followed by a deep void (white blue) before the next hydration shell. Interestingly, the local water density near the C-terminal is noticeably increased due to Coulombic attraction. In comparison to that near a neutral amine, more water molecules are accumulated at the pinnacle of the CH_3 group. A comparison of Panels c and d indicate that the solvation structure for the protonated amine is similar, albeit the local water density near the C-terminal domain is much reduced as the chain length increases.

I analyze the solvent structure near a silica-supported amine by first considering the orientation of a tethered amine relative to the surface normal. Taking C3 as an example, Figure 4-3 shows the probability distribution function of the angle between the main axis of amine backbone (C-N) and the normal direction of surface. At pH=3, the main axis exhibits a Gaussian-like distribution for the orientation at the surface with an average angle of 37° that is related to the local bond structure. At pH=5, however, some of the surface silanol groups become deprotonated in the less acidic solution yielding a negative surface charge density. The surface charges appear to have a strong influence on the

orientation of a tethered amine. For example, Figure 4-3 shows a bimodal distribution of the orientation when the tethered amine is close to two neighboring charge sites.

For an easy comparison of the solvation structure at the silica surface with that in the bulk, I re-cast the main axis of the amine molecule (C-N) as the new z-axis. Together with the orthogonal directions of the surface, it constitutes a z-x plane with the x-axis pointing out of the surface. The y-axis is defined by a cross product of the x and z directions. Slicing into the z-x and z-y planes along the main axis of amines provides the solvent structure in two representative directions. As shown in Figure 4-4, the silica surface restricts the accessibility of water molecules near the methyl group, resulting in a hydration structure drastically different from that in the bulk (Figure 4-2b and 2d). While the hydration structure of the protonated amine group is very much distorted, I can identify up to four hydration shells along the direction perpendicular to the surface (yellow area). For hydration of the tethered C3 amine, the longer carbon chain allows for more water molecules to bind with the N terminal. As a result, the surface exclusion effect becomes less prominent. A comparison of Figures 4(a,b) and 4(c, d) indicates that the water density in the second hydration shell for the C3 amine is slightly higher and broader than that corresponding to C1 amine. Figure 4-4(e, f) shows the solvation structure for the tethered C3 amine at pH=5. While deprotonation of silanol groups has a major influence on amine orientation (see Figure. 3), it appears that this effect is much less pronounced on the solvation structure, indicating that sovlation on surface is localized and thus is not much influenced by surrounding surface charge.

Figure 4-5 presents the radial distribution functions (RDF) of oxygen atoms near C1 and C3 amines in bulk water and at silica surface. As discussed above, the RDF curves are the same as the coefficient $C_{00}(r)$ in the spherical harmonic expansion. Consistent with the density profiles shown in Figure 4-2, the RDFs are near uniform beyond the excluded volume for non-protonated amines in the bulk water (thin red lines). The increase in the chain length shifts the position of the second peak further away from the molecular center due to the segregation of hydrophobic and hydrophilic hydrations for the C3 amine. The size effect disappears for protonated amines (black solid lines) because the local density is dominated by the strong accumulation of water molecules near the charged group. When the amine molecules are tethered at the silica surface (black dashed lines), the RDFs are similar to those in the bulk but the magnitude is significantly reduced after the second peak. In this case, the spherically averaged density does not converge to 1 because the silica surface excludes almost half of the water molecules at large distance. The RDF of oxygen atoms near the tethered C3 amine is higher than that corresponding to C1 amine, indicating that the surface effect is less prominent for longer chains. In good agreement with the local density profiles (Figure 4-4), Figure 4-6 shows that tuning pH from 5 to 3 only slightly changes the radial distribution functions.

To further investigate the surface effect, I have calculated the number of hydrogen bonds among water molecules as a function of distance from the N terminal. Following Luzar and Chandler,[137] I identify hydrogen bonds using the criterion that the O-O distance is less than 3.5 Å and the angle between intramolecular O-H bond and O-O is

less than 30°. Figure 4-7 shows the number of hydrogen bonds per water molecule at different conditions. For neutral C1 and C3 amines in the bulk (thin red lines), the number of hydrogen bonds near the solute (~3 Å) is slightly higher than that in the bulk (>6 Å). When the amines are protonated at an acidic condition, the number of hydrogen bonds near the solute (black solid lines) falls significantly. The reduction in hydrogen bonding between water molecules is a direct consequence of stronger solute-solvent interactions, in particular formation of hydrogen bonds between the charged solute and water molecules. For amines tethered at the silica surface, the number of hydrogen bonds decreases due to the surface restrictions. This surface effect is long-ranged and persists even beyond 6 Å (around 3 hydration shells). Comparing C1 and C3 amines at the surface, I see that the reduction in hydrogen bonding is more prominent for the smaller amine.

Finally, Figure 4-8 compares the hydrogen bonding for tethered C3 amine at pH=3 and 5. In agreement with the local density profiles, it seems that the local hydrogen bonding is rather insensitive to long-range ion-ion interactions introduced by surface protonation.

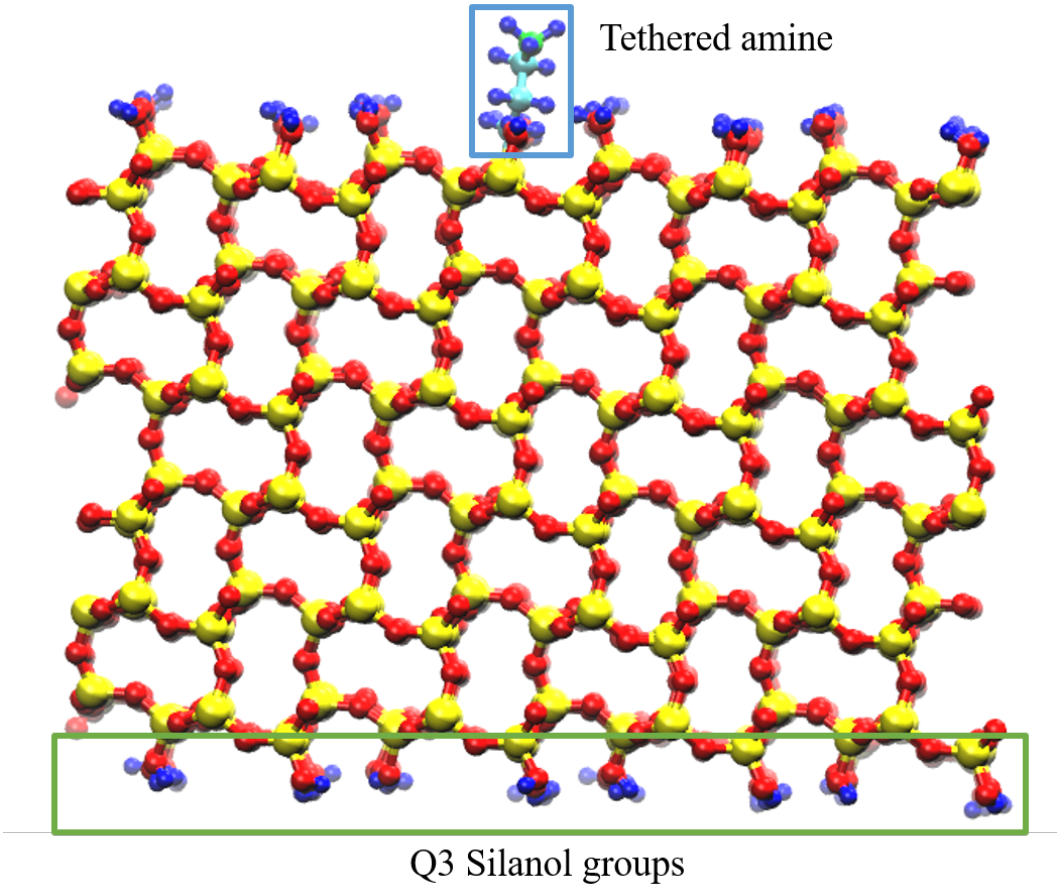
4.5. Conclusions

Amines have found broad applications in industrial reactions and are commonplace in both prokaryotic and eukaryotic cells for the regulation of template-dependent biosynthesis. The physicochemical properties of amines tethered on a surface are expected to be different from those in aqueous solutions. In this work, I have analyzed the hydration structures of amine molecules in the bulk and their analogs tethered at a

silica surface using classical molecular dynamics (MD) simulations. The 3D solvent structures have been effectively sampled through the spherical harmonics expansion method. I find that hydrophobic and hydrophilic hydrations are strongly correlated for methylamine in bulk water and such correlation quickly diminishes as the chain length increases.

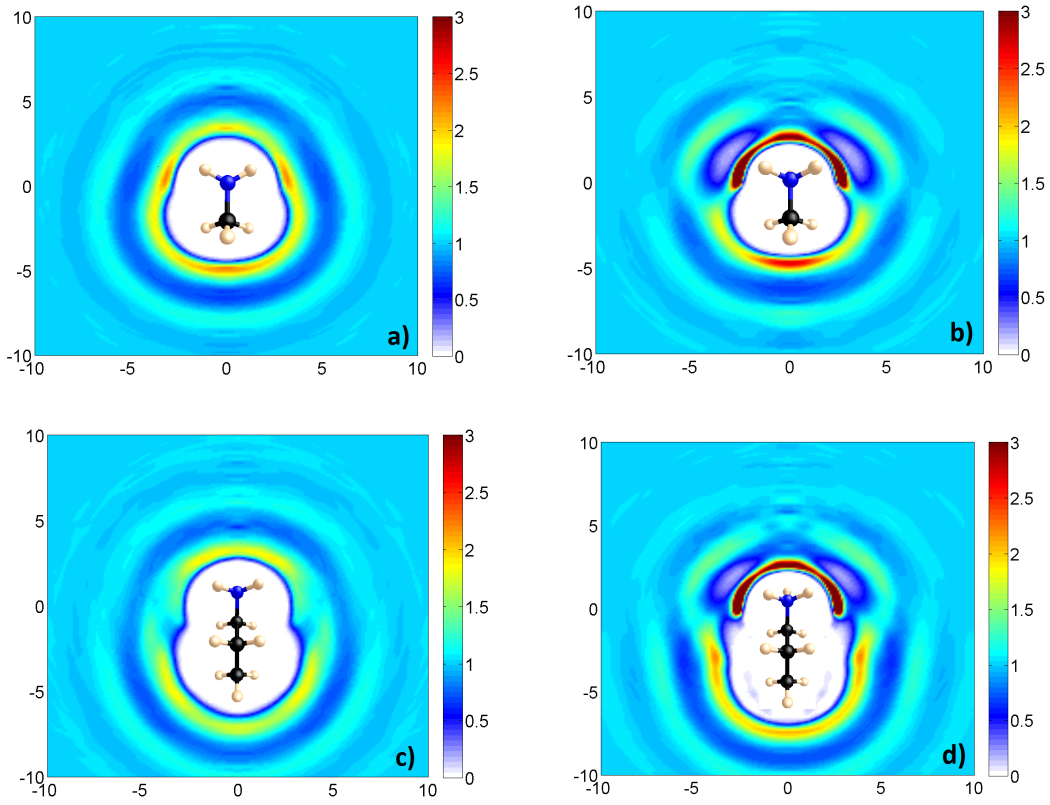
While the hydration structure of the protonated nitrogen atom is virtually independent of the amine size, protonation has a strong impact on water distribution near the hydrophobic group. Compared with amine hydration in pure water, the silica surface has a profound influence on the local distribution of water molecules near a tethered amine. The surface almost completely excludes water molecules from the hydrophobic group and promotes the formation of additional solvation shells along the direction perpendicular to the surface. While the surface charging has strong effects on the orientation of the tethered amine, its influence on the hydration structure is relatively insignificant, suggesting the hydration is localized on the surface and thus not much affected by surface charge.. By analyzing the number of hydrogen bonds per water molecule near free and tethered amines, I find that the silica surface also interrupts the hydrogen-bonding network. The surface effect becomes less significant as the chain length increases. Because the solvation structure could potentially play an important role in chemical reactivity and transportation, it is our hope that the results reported in this work would be useful to further fundamental understanding of confinement effects on surface properties of amine groups.

Figure 4-1



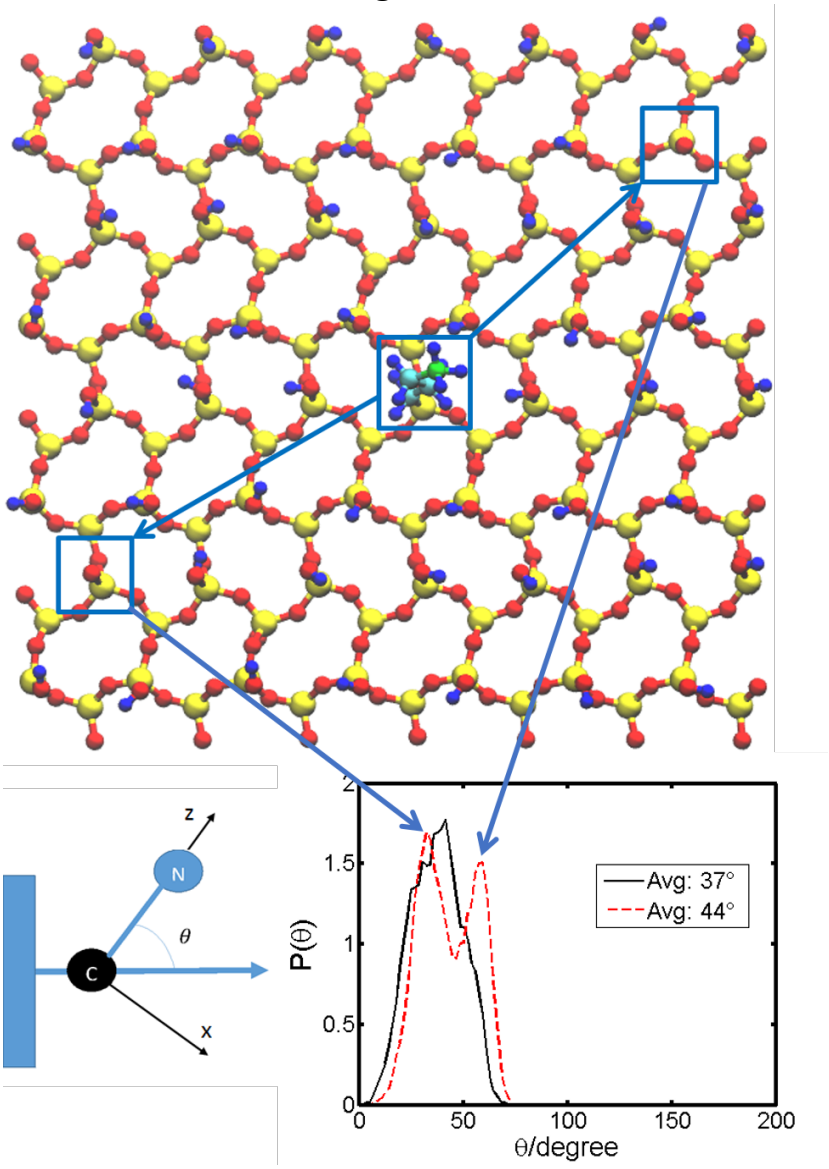
Schematic of n-propylamine with its carbon end tethered at a silica surface (vicinal Q3). Silicon: yellow spheres, oxygen: red spheres, hydrogen: blue spheres, carbon: teal spheres. Water molecules are not shown.

Figure 4-2



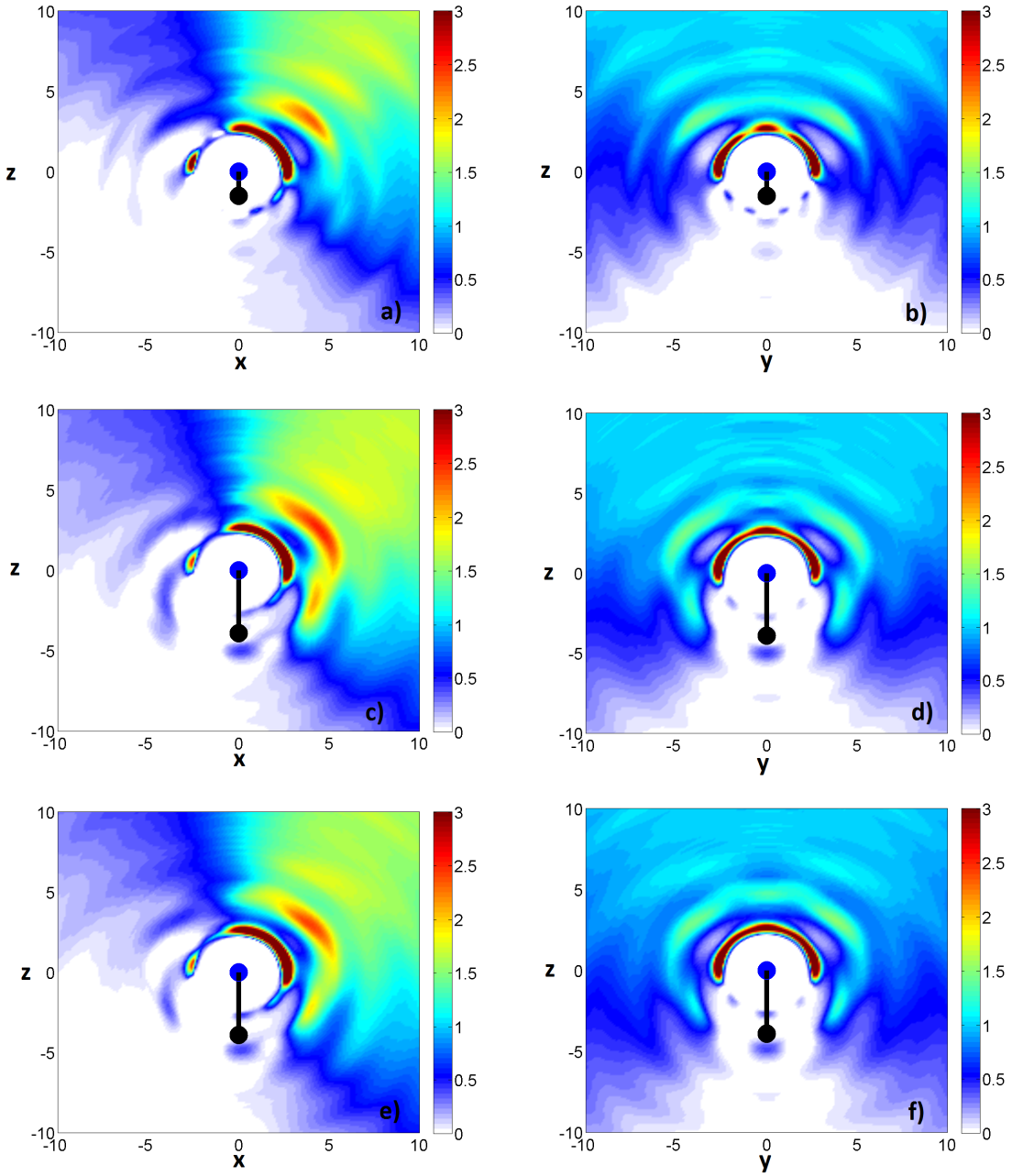
Solvent structure near a) a C1 amine in pure liquid water, b) a C1 amine in an aqueous solution at pH=3, c) a C3 amine in pure liquid water, d) a C3 amine in an aqueous solution at pH=3. The color in the side bar gives the reduced density of oxygen in water (Atom N is marked as blue and C as black for the amine).

Figure 4-3



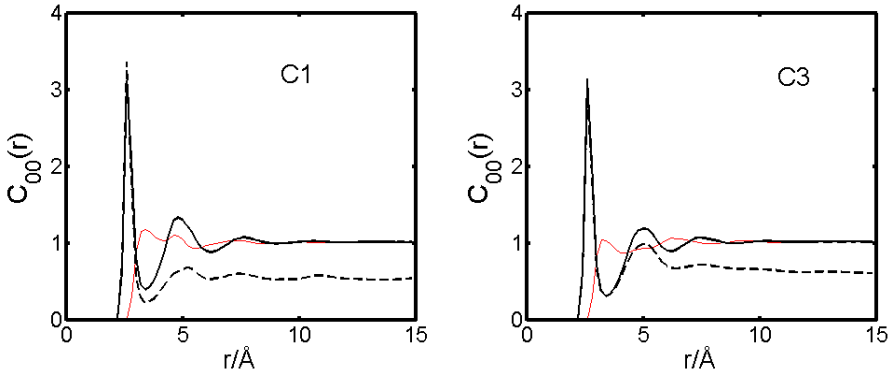
Orientation distribution for C3 amine tethered on a silica surface (black solid: pH=3, red dashed: pH=5); two boxed locations (bottom left and top right) show deprotonated silanol charged sites that alter the orientation of the tethered amine (middle).

Figure 4-4



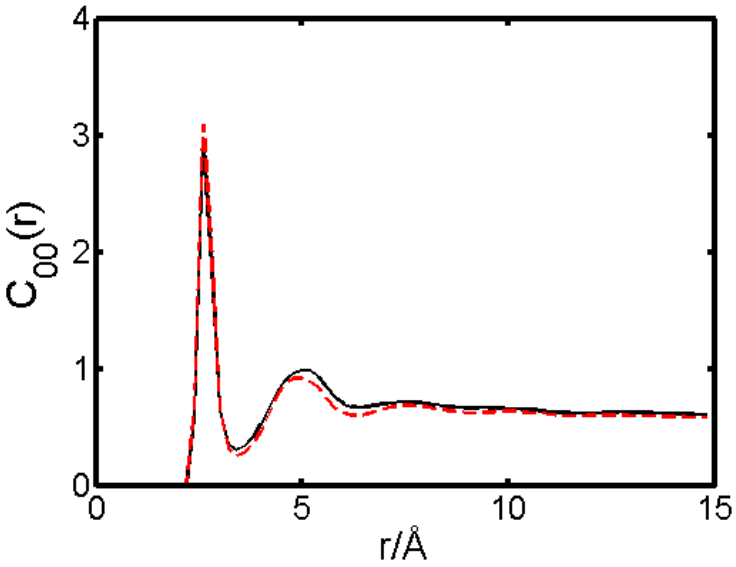
Density profiles of oxygen atoms of water molecule near a silica-supported amine a) and b) C1 amine at pH=3, c) and d) C3 amine at pH=3, e) and f) C3 amine at pH=5. All panels on the left side are the atomic density profiles in the x-z plane, and those on the right side are the corresponding results in the y-z plane. The coordination system is explained in the main text.

Figure 4-5



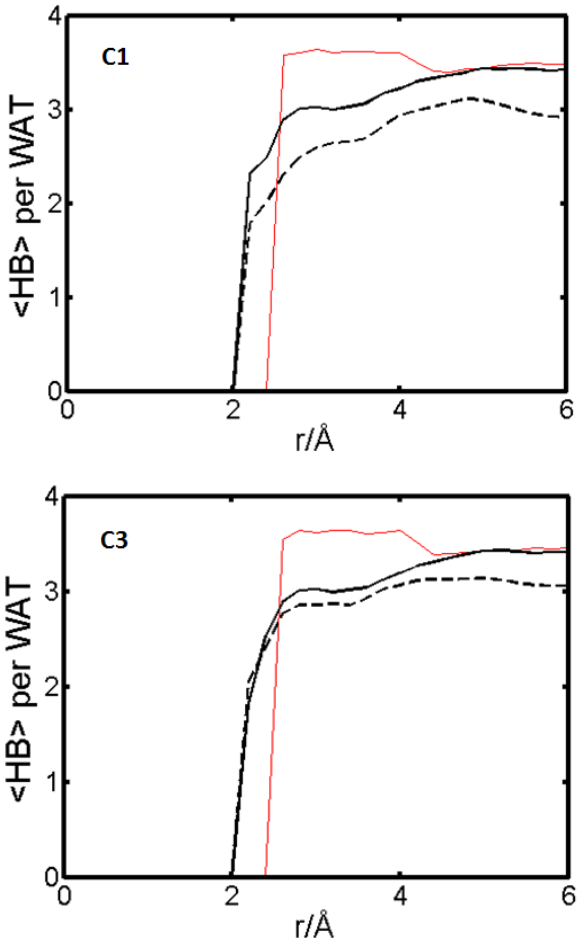
The radial distribution functions of oxygen atom near an amine molecule in the bulk and that tethered at the silica surface. Black solids: protonated in the bulk, thin red lines: neutral in the bulk, black dashed lines: protonated amine at the silica surface.

Figure 4-6



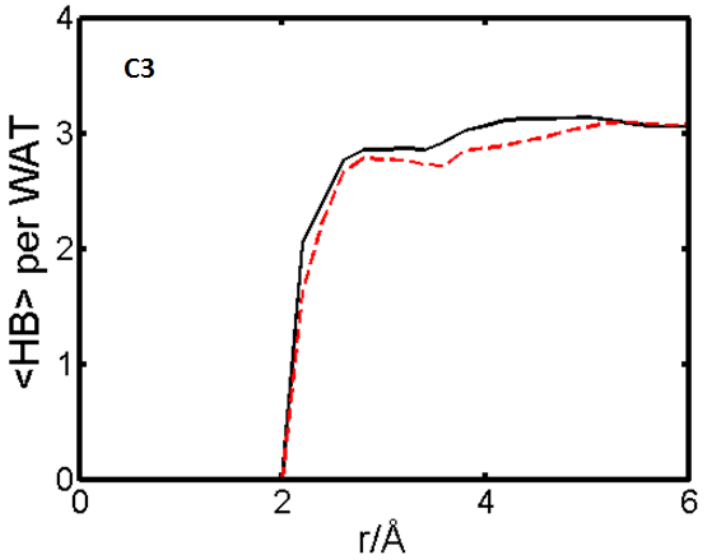
The radial distribution functions of oxygen atom near a propylamine tethered at the silica surface (solid line: pH=3, dashed line: pH=5)

Figure 4-7



The number of hydrogen bonds per water molecules near hydrated amines (black solid: protonated amine in bulk water, red thin: neutral amine in bulk water, black dashed: protonated amine at the silica surface pH=3). Here *r* represents the distance from the oxygen site of each water molecule to the N terminal of the amine.

Figure 4-8



The pH effect on hydrogen bonding among water molecules in the solvation shell of tethered propylamine (solid line: pH=3, dashed line: pH=5).

Chapter 5. Hybrid Method for Ion Solvation

This chapter proposes a hybrid molecular density functional theory (MDFT) for studying solvation by cooperating accurate direct correlation functions of the bulk solvent with molecular dynamic simulation for the *atomic* density profiles of the solvent molecules. The new computational procedure requires only the 3-dimensional (3D) atomic density profiles of the solvent molecules near the solute as the input, and offers an accurate yet efficient alternative to the thermodynamic integration or perturbation methods for simulation of solvation free energies. The hybrid MDFT is fully compatible with conventional atomistic force fields and thus can be easily integrated with standard simulation packages. Illustrative examples are given for predicting the self-solvation free energies of water and hydration free energies of ions.

5.1. Introduction of Hybrid Method

In a recent work[138], Zhao proposed a theoretical procedure to predict the solvation free energy in liquid water by combining molecular dynamics (MD) simulation and the classical density functional theory (DFT). The work was built upon the framework of molecular DFT (MDFT) that incorporates the fundamental measure theory (FMT) for the excluded volume effects and a homogeneous reference fluid approximation (HRFA) introduced by Borgis and coworkers for the van der Waals attraction and long-range electrostatic interactions.[12, 139-141] Different from the conventional implementation of DFT[5, 78, 142], they used MD simulation to generate the density profiles of the solvent molecules surrounding each solute, and the solvation free energy is then calculated from the change in the grand potential of the solvent owing to the insertion of a solute

molecule. Because the DFT links the microscopic structure (at equilibrium state) directly with thermodynamic variables, such a hybrid method eliminates lengthy simulation of intermediate states and thus offers a computationally much efficient alternative to thermodynamic integration and perturbation methods in molecular simulation of solvation free energies. This hybrid method predicts the solvation free energies of halide and alkali ions in excellent agreement with the simulation/experimental results.[138]

Whereas the hybrid method combines the merits of DFT and molecular simulation and is promising for broad applications, several challenges remain to be addressed for systems without spherical symmetry. First, their version of MDFT relies on a rigid model for solvent molecules and the solvent density profile is generally expressed as a 6-dimensional function of the molecular position and orientation (for a linear model of the solvent molecules, the dimensionality reduces to 5). An accurate evaluation of the 6-dimensional density profile by molecular simulation imposes a severe numerical constraint and limits application of hybrid MDFT to large biological systems. Second, the direct correlation function embedded in their MDFT functional depends on the separation as well as the relative orientations between two solvent molecules, i.e., it is a 12-dimensional function at a given thermodynamic condition for the solvent like water. For a conventional three-site model for water molecules, we need a 6-dimensional function to characterize the position and orientation of each molecule. Evaluation of the solvation free energy involves a double convolution of the high-dimensional direct correlation function with the 6-dimensional solvent density profile that demands large computer memory and thus further hampers its many practical potentials.

The dimensionality of the DFT calculations can be drastically reduced by using the Chandler-McCoy-Singer (CMS) theory for molecular systems. Unlike the MDFT, the CMS theory is formulated in terms of the atomic (site) density profiles and the site-site direction correlation functions are one-dimensional depending only on the corresponding pair distances. With a hypothesis on the universality of the bridge functional for the excluded-volume effects[14] and the site-site direct correlation functions of the pure solvent evaluated from MD simulation[101, 143], we extended the CMS theory into a 3D molecular density functional theory (3D-MDFT) for solvation calculations.[144] With the bridge functional represented by the modified fundamental measure theory, we have demonstrated that the 3D-MDFT enables us to predict the solvation free energy and the solvent structure simultaneously. Similar to the CMS theory, the 3D-MDFT theory is capable of treating both rigid and chain-like molecular fluids.[10, 145-146] Besides, it does not involve angular variables, and thus is simpler to implement and computationally more efficient in comparison with the high-dimensional approach.

Whereas the 3D-MDFT is computationally efficient, we found that the universal bridge functional is not applicable to the site-site interactions and formulation of such contributions remains a theoretical challenge. Besides, numerical convergence can be an issue in minimization of the free energy with respect to the site densities. Motivated from the virtues of the site description and MD simulation, here we introduce a hybrid 3D-MDFT/MD protocol to predict solvation free energy. Because the 3D-MDFT is based on the density profiles of individual atoms (site) and the site-site direct correlation functions of the bulk solvent, the new computational procedure requires sampling of only 3-

dimensional density profiles and evaluations of 3-dimensional integrals pertinent to the atomic positions and the one-dimensional direct correlation functions. The reduction of dimensionality drastically improves the computational efficiency and the combination with MD also avoids the inconvenient numeric iteration treatment to get the density profile in 3D-MDFT, thus the hybrid method opens the potential to solvation of large molecular systems. The method is first tested for self-solvation calculations of rigid TIP3P and is then applied in ion solvation. The computational procedure can be readily extended to other more complex solvation systems.

5.2. Derivations of Hybrid Method

The solvation free energy is defined as the reversible work to transfer a solute molecule at a given position in vacuum into a given position in a pure solvent at fixed temperature and pressure or equivalently, at fixed solvent temperature and chemical potential. The solvation free energy is immaterial to the solute translational degree of freedom. Within the framework of the classical density functional theory[5, 78, 142], the solvation free energy is given by the change in the grand potential of the solvent due to the transfer of the solute molecule

$$\Delta G_{sol}[\rho_i(\mathbf{r});\mu,V,T] = \Omega[\rho_i(\mathbf{r});\mu,V,T] - \Omega[\rho_i^{\,b};\mu,V,T] \qquad (1)$$

where $\rho_i(\mathbf{r})$ stands for the local density profile of i site (atom) of the solvent molecule near the dissolved solute molecule. If the solvent is represented by the TIP3P model, i = O or H designating for oxygen or hydrogen atoms, respectively. In Eq.(1), $\rho_i^{\,b}$ is the average atomic density of the pure solvent; μ and T are, respectively, the solvent

chemical potential and absolute temperature; and V is the system volume. In numerical implementation of Eq.(1), the system volume should be sufficiently large such that the properties of water molecules remote from the solute are the same as those corresponding to the pure solvent in the bulk. For solvation free energy calculations, parameters (μ,V,T) are fixed and thus, for the sake of simplicity, they are suppressed in the equations below.

The solvated system can be regarded as an inhomogeneous system of the solvent subject to an external potential due to the solvent-solute interaction. The grand potential of such inhomogeneous system is given by

$$F[\rho_i(\mathbf{r})] = F^{id}[\rho_i(\mathbf{r})] + F^{ex}[\rho_i(\mathbf{r})] - \sum_i \int d\mathbf{r}\rho_i(\mathbf{r})\psi_i(\mathbf{r}) . \qquad (2)$$

Here $F^{id}[\rho_i(\mathbf{r})]$ represents the ideal part of the intrinsic Helmholtz energy functional, i.e., it is the same as that of an ideal-gas system of the solvent *molecules* with the same temperature, chemical potential and inhomogeneous site density profiles. The second term $F^{ex}[\rho_i(\mathbf{r})]$ denotes the excess Helmholtz energy that arises from the intermolecular interactions. The one-body potential $\psi_i(\mathbf{r})$ follows its usual definition, i.e., $\psi_i(\mathbf{r}) = \mu_{i-b} - V_i^{ext}(\mathbf{r})$, where $V_i^{ext}(\mathbf{r})$ is the external potential exerting on site i due to the presence of the solute molecule, and μ_{i-b} stands for a nominal chemical potential of site i. The nominal chemical potentials of individual sites satisfy the stoichiometric relation[13]

$$\mu = \sum_i \mu_i \qquad (3)$$

As shown below, the nominal chemical potential of each site is irrelevant to the final expression for the solvation free energy functional.

The excess Helmholtz energy functional can be formally expressed relative to that of the bulk solvent at the system temperature and chemical potential plus additional terms due to the density inhomogeneity[5, 15]

$$F^{ex}[\rho_i(\mathbf{r})] = F^{ex}[\rho_i^{\,b}] + \sum_i \mu_{i-b}^{ex} \int d\mathbf{r} \Delta\rho_i(\mathbf{r}) - \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r} \int d\mathbf{r}' \Delta\rho_i(\mathbf{r}) \Delta\rho_j(\mathbf{r}') c_{ij}^{(2)}(|\mathbf{r}-\mathbf{r}'|) + F_B[\rho_i(\mathbf{r})] \tag{4}$$

where k_B stands for the Boltzmann constant, and $\Delta\rho_i(\mathbf{r}) = \rho_i(\mathbf{r}) - \rho_i^b$ is the deviation of local site density to bulk density. In Eq.(11), μ_{i-b}^{ex} represents the nominal excess chemical potential of site i in the bulk solvent, and $c_{ij}^{(2)}(r)$ is the bulk site-site direct correlation function. The final term in Eq.(11), $F_B[\rho_i(\mathbf{r})]$, consists of a definition of the bridge functional, i.e., it accounts for all contributions to the excess Helmholtz free energy beyond the quadratic expansion.

Substituting Eq.(11) into Eq.(2), and thereafter into Eq.(1), yields a close form for the solvation free energy:

$$\Delta G_{sol}[\rho_i(\mathbf{r})] = \Delta F^{id}[\rho_i(\mathbf{r})] - \sum_i \int d\mathbf{r} \rho_i(\mathbf{r})[\psi_i(\mathbf{r}) - \mu_{i-b}] - \sum_i \mu_{i-b}^{id} \int d\mathbf{r}[\rho_i(\mathbf{r}) - \rho_i^{\,b}] - \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r} \int d\mathbf{r}' \Delta\rho_i(\mathbf{r}) \Delta\rho_j(\mathbf{r}') c_{ij}^{(2)}(|\mathbf{r}-\mathbf{r}'|) + F_B[\rho_i(\mathbf{r})] \tag{5}$$

where $\Delta F^{id}[\rho_i(\mathbf{r})] \equiv F^{id}[\rho_i(\mathbf{r})] - F^{id}[\rho_i^b]$. In deriving Eq.(5), we have used the identity $\mu_{i-b}^{ex} = \mu_{i-b} - \mu_{i-b}^{id}$.

To find an explicit expression for ΔF^{id}, we can consider the solvation in a reference ideal-gas system that has the same site density $\rho_i(\mathbf{r})$ as that of the real system under consideration (see Figure 1(a)(c) for illustration of both systems). We denote the corresponding one-body potential in such ideal reference system as $\psi_i^{IG}(\mathbf{r})$ (in corresponding to $\psi_i(\mathbf{r})$ in Eq.(2)), and then the ideal solvation free energy functional can be expressed as, similar to Eq.(5),

$$F_{IG}[\rho_i(\mathbf{r})] = \Delta F_{IG}^{id}[\rho_i(\mathbf{r})] + \\ -\sum_i \int d\mathbf{r}\rho_i(\mathbf{r})[\psi_i^{IG}(\mathbf{r}) - \mu_{i-b}^{id}] - \sum_i \mu_{i-b}^{id} \int d\mathbf{r}[\rho_i(\mathbf{r}) - \rho_i^b] \quad (6)$$

In comparison to Eq.(5), Eq.(6) does not contain the site-site direct correlations and the bridge functional due to the absence of intermolecular interactions. On the other hand, starting from the partition function, we can prove the solvation free energy in the ideal-gas reference system is fully determined by the atomic density profiles[144]

$$\Delta G_{sol}^{IG} = -\frac{k_B T}{M} \sum_i \int d\mathbf{r} \Delta\rho_i(\mathbf{r}) = -\Delta N k_B T \quad (7)$$

where M is the total number of sites within the solvent molecule. For the TIP3P model, $M = 3$. In the above equation, ΔN represents the difference in the average number of solvent *molecules* before and after the solvation.

Although the ideal reference system shares the same bulk and inhomogeneous solvent densities as the real system, Eq.(5) and (6) describe the solvation free energies of two unrelated solutes: one corresponds to that of the real system under consideration and the other is for an undefined solute that gives rises to $\psi_i^{IG}(\mathbf{r})$. Nevertheless, the real and ideal systems consist of the same molecule species and share the same density profiles and thus they have the same ideal part of the intrinsic Helmholtz energy functional:

$$\Delta F^{id}[\rho_i(\mathbf{r})] = \Delta F_{IG}{}^{id}[\rho_i(\mathbf{r})] \tag{8}$$

Eq.(8) holds for an arbitrary set of atomic density profiles $\{\rho_i(\mathbf{r})\}_{i=O,H}$.

With the help of Eq.(8), we can subtract Eq.(6) from Eq.(5). After some rearrangement, we derive

$$\begin{aligned} F[\rho_i(\mathbf{r})] = F_{IG}[\rho_i(\mathbf{r})] - \sum_i \int d\mathbf{r}\rho_i(\mathbf{r})[\psi_i(\mathbf{r}) - \mu_{i-b}] + \sum_i \int d\mathbf{r}\rho_i(\mathbf{r})[\psi_i^{IG}(\mathbf{r}) - \mu_{i-b}^{id}] \\ - \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r} \int d\mathbf{r}' \Delta\rho_i(\mathbf{r}) \Delta\rho_j(\mathbf{r}') c_{ij}^{(2)}(|\mathbf{r} - \mathbf{r}'|) + F_B[\rho_i(\mathbf{r})] \end{aligned} \tag{9}$$

The bridge potential $B_i(\mathbf{r})$ is introduced as the functional derivative of F_B with respect to $\rho_i(\mathbf{r})$

$$B_i(\mathbf{r}) = \frac{\delta F_B}{\delta \rho_i(\mathbf{r})} \tag{10}$$

At the equilibrium condition, the formula of solvation free energy is

$$\Delta G_{sol} = -\Delta N k_B T + \frac{k_B T}{2} \sum_{i,j} \int d\mathbf{r} d\mathbf{r}' \left(\rho_i(\mathbf{r}) + \rho_i^b\right) c_{ij}^{(2)} \Delta\rho_j(\mathbf{r}') + \Gamma\left[\rho_i(\mathbf{r})\right] \tag{11}$$

where $\Gamma\left[\rho_i(\mathbf{r})\right]=F_B\left[\rho_i(\mathbf{r})\right]-\sum_i\int d\mathbf{r}\rho_i(\mathbf{r})B_i(\mathbf{r})$. As expected, the solvation free energy is independent of the nominal excess chemical potential of individual sites.

Whereas Eq.(11) is formally exact, the generalized bridge functional $\Gamma[\rho_i(\mathbf{r})]$ associated with the bridge functional $F_B[\rho_i(\mathbf{r})]$ is generally not accessible. By simply dropping the bridge functional, Eq.(11) reduces to the solvation free energy expression from the CMS theory[13, 147]

$$\Delta G_{sol}^{HNC}=-\Delta N k_B T+\frac{k_B T}{2}\sum_{i,j}\int d\mathbf{r}d\mathbf{r}'\left(\rho_i(\mathbf{r})+\rho_i^b\right)c_{ij}^{(2)}\Delta\rho_j(\mathbf{r}') \tag{12}$$

Because the CMS theory is equivalent to truncate the functional expansion of excess Helmholtz energy in 3D-MDFT up to the second order, Eq.(12) is also referred to quadratic expansion approximation (QEA). To account for higher order terms, bridge term is approximated with a reference hard sphere system according to the universal ansatz.

$$\Gamma\left[\rho_i(\mathbf{r})\right]\simeq\Gamma^{HS}\left[\rho_i(\mathbf{r});\sigma_i\right] \tag{13}$$

where σ_i is the input parameter, representing the reference hard sphere diameter of site i. For TIP3P water model, reference hard sphere diameter of hydrogen site σ_H is fixed as 1Å, and σ_O is calibrated against experiment values.

For the hybrid site DFT proposed here, we calculate the site density profile from simulation, thus the minimization issue of 3D-MDFT is irrelevant.

5.3. Efficient Sampling of 3D Density Profiles

Standard grid-method to sample 3D density profiles[136, 148] in MD simulation is time-consuming. Here, we utilize spherical harmonics (SH) expansion method to boost the calculation efficiency. The density profile around a tagged molecule can be expressed as ensemble average of the Dirac delta functions

$$\rho_i(\mathbf{r}) = \left\langle \sum_c \delta\left(\mathbf{r} - \mathbf{r}_c^i\right) \right\rangle \tag{14}$$

where $\langle \dots \rangle$ is the ensemble average, $\delta(\mathbf{r} - \mathbf{r}_i)$ is the Dirac delta function and stands for local density of site i in molecule c. Expanding density profile into spherical harmonics yields

$$\rho_i(\mathbf{r}) = \sum_{nm} \rho_{nm}^i(r) Y_{nm}(\theta,\phi) \tag{15}$$

where (r,θ,ϕ) is the spherical coordinate, Y_{nm} is spherical harmonics, n,m are indices ($-n \le m \le n$) , and ρ_{nm}^i is corresponding coefficients. Insert Eq.(15) into (14), and perform a integration $\int_r^{r+\Delta r} r^2 dr \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \bar{Y}_{nm}(\theta,\phi)\cos\theta$ on both sides gives

$$\int_r^{r+\Delta r} \rho_{nm}^i(r) r^2 dr = \left\langle \sum_{N_r} \bar{Y}_{nm}\left(\theta_c^i, \phi_c^i\right) \right\rangle \tag{16}$$

where N_r stands for the number of site i located in the shell $r_c^i \in (r, r+\Delta r)$. After simplification, Eq.(16) becomes

$$\rho_{nm}(r) = \frac{\left\langle \sum_{N_r} \bar{Y}_{nm}\left(\theta_c^i, \phi_c^i\right) \right\rangle}{r^2 \Delta r} \tag{17}$$

which provides a way to reduce 3D sampling into 1D. In theory, sampling an infinite number of SH coefficients is unrealistic and unnecessary. In fact, only coefficients of n smaller than 10 are sampled and the summation in Eq.(15) converges within the accuracy of later computation.

5.4. Water Self-solvation

This hybrid method is first tested with water self-solvation. The atomic density profiles are extracted from MD simulation of a bulk TIP3P water model by using the efficient sampling method discussed in the previous section. Figure 5-1 shows the average radial distribution of oxygen (OW) and hydrogen (HW) sites. Since the distance is from the center of the tagged molecule instead of its oxygen site, curves are different from oxygen-oxygen and oxygen-hydrogen radial distribution function in the literature.

Two peaks of OW curve represent two different type of surrounding oxygen connected by hydrogen bonds. 3D density profile of oxygen is plotted in Figure 5-2. For better demonstration, only reduced density higher than 2 is visible. The area shows the first hydration shell structure of the tagged water molecule. Contour maps are also presented and clearly show 4 hydrogen-bonding sites. Similarly Figure 5-3 provides the 3D density information of hydrogen site. Unlike oxygen site whose reduced density can reach over 30, hydrogen sites are less packed.

With the 3D density and the reference hard sphere diameter σ_0 (2.65 Å, close to the Lennard-Johns parameter 3.15Å) as input into Eq.(11), DFT reproduces the experiment value of self-water solvation free energy -6.32 kcal/mol. The result shows that the hybrid method works for water self-solvation with a reasonable choice of reference hard sphere diameter σ_0.

5.5. Ion Solvation

Self-consistent DFT calculation of SFE for ions are challenging due to the high packing density from Coulomb interactions. When the peak is high in density profile, calculations become more sensitive and require more iterations for convergence. However, the hybrid method circumvents the iteration step and thus provides a useful tool in charged system. Here, we investigate SFE of 34 ions, 9 monovalent ions from Cheatham[149], 22 divalent ions from Babu[150], and 4 trivalent Lanthanide ions from Reinhoudt[151].

Each ion is first solvated in a TIP3P water box of 30Å length in size, and then equilibrated in a NPT ensemble at room temperature and 1 bar for 5 ns, followed by another 5 ns in NVT ensemble. A total of 10,000 configurations are sampled in the following trajectory production step of 10 ns in NVT ensemble. In the simulations, Nose-Hoover style barostat and thermostat are used. All simulations are carried in GROMACS.

Due to symmetry, only radial distribution is sampled to calculate the density profile for ions. In order to calculate SFE from hybrid method, a reasonable value for the reference hard sphere diameter σ_0 needs to be chosen. The choice for each set of ions

based on its charge is shown in Table 5-1. It is worth noting that σ_0 does not necessarily have to be the same. For different ions, the reference hard sphere systems can vary. For ions of the same charge, the nuance may be small and thus the same σ_0 may be chosen. However, this difference can no longer be ignored for ions of different charges.

With density profile and σ_0 as input, we calculated the DFT prediction and HNC prediction of SFE for ions. To be noted, the Born correction term has been added to both theoretical predictions due to limited size effect.

$$\Delta G_{sol}^{born} = -322\left(1-\frac{1}{\epsilon}\right)/\left(2r_{cut}\right) \tag{18}$$

where ϵ is the dielectric constant of water model, r_{cut} is the half the simulation size (15 Å), and the unit for free energy is kcal/mol, Å for distance.

Results for monovalent ions are presented in Figure 5-4. For simplicity of comparison, the negative sign for SFE is omitted. Both DFT and HNC predictions are close to experimental values. With increasing size of ions, SFE decreases for both cations and anions. DFT overestimates SFE for monovalent cations, while HNC underestimates SFE for anions. It shows the higher order term $\Gamma\left[\rho_i\right]$ is not import for monovalent ions.

The situation is different for divalent and trivalent ions. Without $\Gamma\left[\rho_i\right]$, HNC considerably underestimates SFE. In fact, the higher order term contributes more and more with increasing ion charge. And it can easily reaches to 200-300 kcal/mol contribution. This demonstrates the importance of high-order terms in DFT calculation.

On the other side, the reference hard sphere system can actually captures the bridge functional even for charged system. However, DFT predictions are also not perfect. The prediction error can reach ~20% for Co2+. This is due to the strong inhomogeneity of water density near ions.

5.6. Conclusions

In this Chapter, I propose a hybrid method by combing MDFT and MD simulations and test its performance for solvation free energies calculations. Different from an earlier version of MDFT, only 3D site density profiles are required, instead of the 6-dimensional molecular density profile. The dimensionality reduction opens up the opportunities to applications of the hybrid method to complex biological systems.

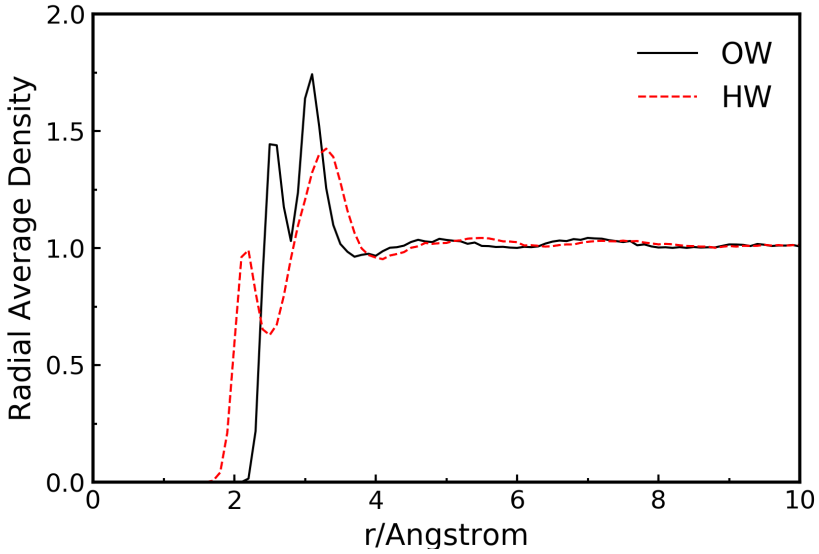
The hybrid method is first tested with water self-solvation free energy. With reasonable choice of a reference hard-sphere diameter, MDFT reproduces the experimental value. The good performance demonstrates the feasibility of hybrid method.

Further, the hybrid method is applied to ions of different charges. From fitting with experimental data, the reference hard-sphere diameter varies with the ion charge. DFT captures how SFE decreases with increasing ion size for monovalent ions. In addition, the bridge functional has much smaller contribution to SFE for monovalent ions than that for divalent or trivalent ions. In fact, it can make up 200-300 kcal/mol to SFE. Thus, it demonstrates that reference hard-sphere system can be applied to charged systems and its contribution cannot be ignored especially for systems of large charges.

Table 5-1. Hard sphere diameter for different systems

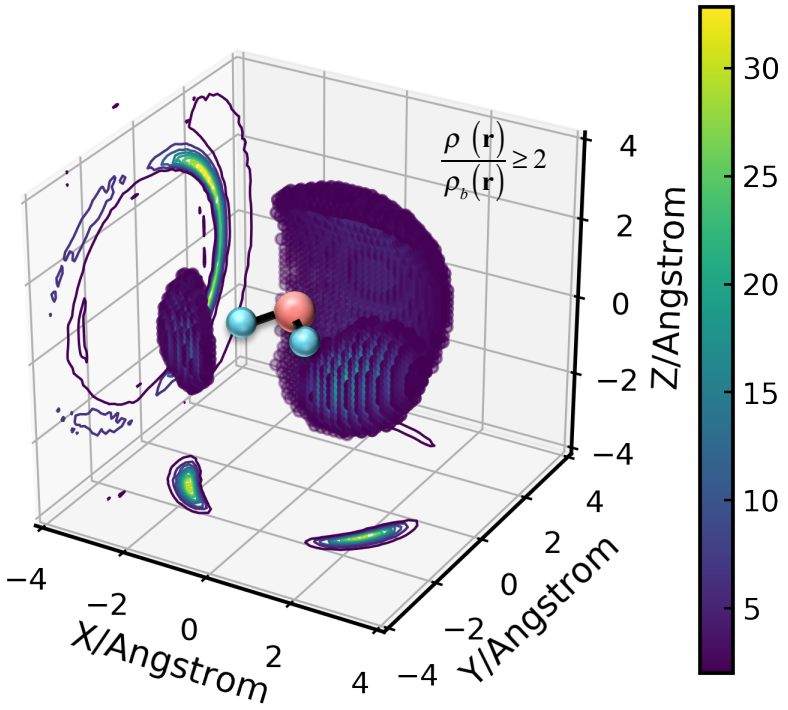
HS Diameter(Å)	O	H
water	2.65	1.0
mono-valent	2.68	1.0
di-valent	2.84	1.0
tri-valent	2.84	1.0

Figure 5-1



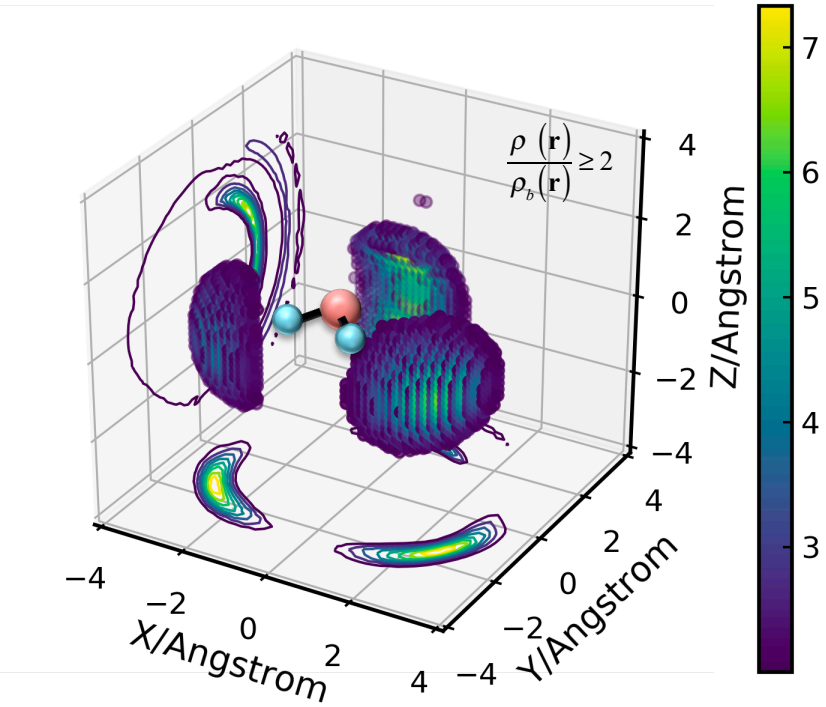
Radial average density of site O, H from the center of tagged water molecule

Figure 5-2



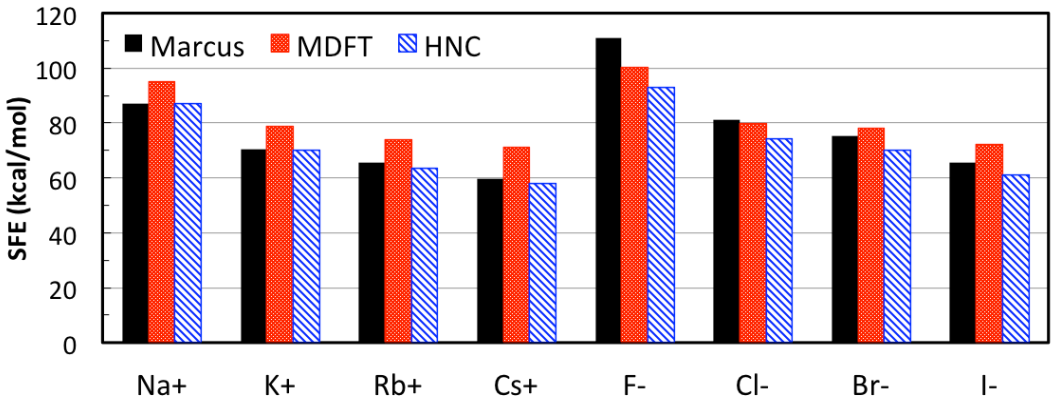
3D density and contour map of oxygen sites

Figure 5-3



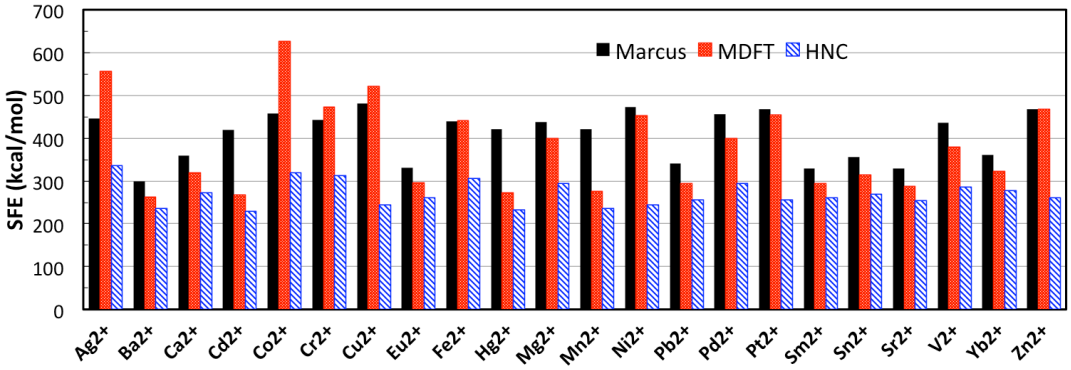
3D density and contour map of hydrogen sites

Figure 5-4



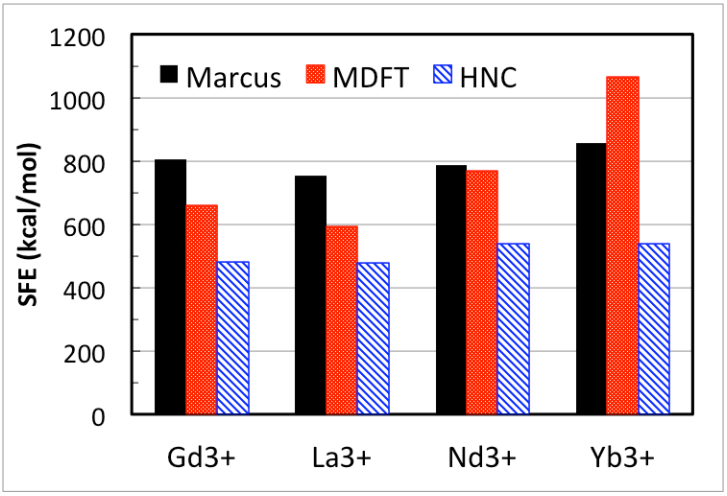
Comparison of solvation free energies for monovalent ions

Figure 5-5



Comparison of solvation free energies for divalent ions

Figure 5-6



Comparison of solvation free energies for trivalent ions

Chapter 6. Conclusions

In this dissertation, I have investigated solvation systems by applying molecular dynamics (MD) simulations and molecular density functional theory (MDFT).

To refine MDFT predictions, I have developed new numerical methods to sample site-site direction correlation functions (DCF) from MD. With the new DCF as input, the mean error of solvation free energy (SFE) predictions for 504 small solutes improve from 1.04 kcal/mol to 0.66 kcal/mol when compared with MD simulation results. In addition, the new method is not limited to 3-site or 4-site water models but all molecules without special symmetries.

In cases when the solvation system is not at room temperature, for example, the storage of highly explosive nitro-toluenes, I extended the MDFT calculation by using DCFs from the integral-equation theory and calibrated according to experimental data along with 197 solutes at 277K, 298K and 313K. The mean error is slightly higher than 1 kcal/mol and is comparable with MD simulation results. Molecules with peroxide functional groups have the highest prediction error, which is consistent with the fact that no peroxide models have been systematically validated with available thermodynamic properties.

In studying the hydration structure of amine-grafted silica gels, MD simulation provides direct evidence of strong surface effect on grafted amines. The presence of surface can not only repel water due to volume exclusion, but also affect the hydrogen-bonding network. This surface effect is long-ranged and can take place even after three hydration shells. For amines of longer chain length, the surface effect is less prominent.

Even though changes in pH can alter the orientation of grafted-amine on the surface, the surface effect is relatively invariant to pH.

References

1. Marini, A.; Munoz-Losa, A.; Biancardi, A.; Mennucci, B., What Is Solvatochromism? *J Phys Chem B* **2010**, *114*, 17128-17135.
2. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E., How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys J* **2011**, *100*, L47-L49.
3. Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J., Solvent Effects .5. Influence of Cavity Shape, Truncation of Electrostatics, and Electron Correlation Ab Initio Reaction Field Calculations. *J Phys Chem-Us* **1996**, *100*, 16098-16104.
4. Yu, Y. X.; Wu, J. Z., Structures of Hard-Sphere Fluids from a Modified Fundamental-Measure Theory. *J. Chem. Phys.* **2002**, *117*, 10156-10164.
5. Wu, J. Z.; Li, Z. D., Density Functional Theory for Complex Fluids. *Annu. Rev. Phys. Chem.* **2007**, *58*, 85-112.
6. Wu, J. Z., *Density Functional Theory for Liquid Structure and Thermodynamics*; Springer-Verlag Berlin, 2009; Vol. 131, p 1-73.
7. Evans, R., *Density Functionals in the Theory of Nonuniform Fluids*; Marcel Dekker: New York, 1992, p 85-175.
8. Jeanmairet, G.; Levesque, M.; Sergiievskyi, V. P.; Borgis, D., Classical Density Functional Theory to Tackle Solvation in Molecular Liquids. In *Computational Trends in Solvation and Transport in Liquids*, Sutmann, G.; Grotendorst, J.; Gompper, G.; Marx, D., Eds. Forschungszentrum Jülich GmbH: 2015; Vol. 28, pp 187-208.
9. Liu, Y.; Wu, J. Z., Long-Range Angular Correlations in Liquid Water. *Journal of Chemical Physics* **2013**, *139*, 041103.
10. Seok, C.; Oxtoby, D. W., Nucleation in N-Alkanes: A Density-Functional Approach. *J. Chem. Phys.* **1998**, *109*, 7982-7990.
11. Gray, C. G.; Gubbins, K. E., *Theory of Molecular Fluids: Fundamentals* Oxford University Press: New York, 1985; Vol. Volume 1.
12. Ramirez, R.; Gebauer, R.; Mareschal, M.; Borgis, D., Density Functional Theory of Solvation in a Polar Solvent: Extracting the Functional from Homogeneous Solvent Simulations. *Phys. Rev. E* **2002**, *66*.
13. Chandler, D.; Mccoy, J. D.; Singer, S. J., Density Functional Theory of Nonuniform Polyatomic Systems .1. General Formulation. *Journal of Chemical Physics* **1986**, *85*, 5971-5976.
14. Rosenfeld, Y., Free-Energy Model for Inhomogeneous Fluid Mixtures - Yukawa-Charged Hard-Spheres, General Interactions, and Plasmas. *Journal of Chemical Physics* **1993**, *98*, 8126-8148.
15. Hansen, J. P.; McDonald, I. R., *Theory of Simple Liquids*, Third ed.; Academic Press: London, 2006.
16. Rosenfeld, Y., Free Energy Model for Inhomogeneous Fluid Mixtures: Yukawa-Charged Hard Spheres, General Interactions, and Plasmas. *J. Chem. Phys.* **1993**, *98*, 8126-8148.

17. Rosenfeld, Y., Structure and Effective Interactions in Multi-Component Hard-Sphere Liquids: The Fundamental-Measure Density Functional Approach. *J. Phys. Condens. Matter* **2002**, *14*, 9141-9152.
18. Rosenfeld, Y., Structure and Effective Interactions in Multi-Component Hard-Sphere Liquids: The Fundamental-Measure Density Functional Approach. *Journal of Physics-Condensed Matter* **2002**, *14*, 9141-9152.
19. Yu, Y. X.; Wu, J. Z., A Fundamental-Measure Theory for Inhomogeneous Associating Fluids. *J. Chem. Phys.* **2002**, *116*, 7094-7103.
20. Yu, Y. X.; Wu, J. Z.; Xin, Y. X.; Gao, G. H., Structures and Correlation Functions of Multicomponent and Polydisperse Hard-Sphere Mixtures from a Density Functional Theory. *J. Chem. Phys.* **2004**, *121*, 1535-1541.
21. Chandler, D., Derivation of an Integral-Equation for Pair Correlation-Functions in Molecular Fluids. *Journal of Chemical Physics* **1973**, *59*, 2742-2746.
22. Chandler, D.; Mccoy, J. D.; Singer, S. J., Density Functional Theory of Nonuniform Polyatomic Systems .2. Rational Closures for Integral-Equations. *J Chem Phys* **1986**, *85*, 5977-5982.
23. Ichiye, T.; Chandler, D., Hypernetted Chain Closure Reference Interaction Site Method Theory of Structure and Thermodynamics for Alkanes in Water. *Journal of Physical Chemistry* **1988**, *92*, 5257-5261.
24. Raineri, F. O.; Stell, G., Dielectrically Nontrivial Closures for the Rism Integral Equation. *Journal of Physical Chemistry B* **2001**, *105*, 11880-11892.
25. Zhao, S. L.; Liu, Y.; Liu, H. L.; Wu, J. Z., Site-Site Direct Correlation Functions for Three Popular Molecular Models of Liquid Water. *J Chem Phys* **2013**, *139*.
26. Chuev, G. N.; Vyalov, I.; Georgi, N., Extraction of Site-Site Bridge Functions and Effective Pair Potentials from Simulations of Polar Molecular Liquids. *Journal of Computational Chemistry* **2014**, *35*, 1010-1023.
27. Marucho, M.; Kelley, C. T.; Pettitt, B. M., Solutions of the Optimized Closure Integral Equation Theory: Heteronuclear Polyatomic Fluids. *J Chem Theory Comput* **2008**, *4*, 385-396.
28. Kovalenko, A., Multiscale Modeling of Solvation in Chemical and Biological Nanosystems and in Nanoporous Materials. *Pure Appl Chem* **2013**, *85*, 159-199.
29. Hirata, F., *Molecular Theory of Solvation* Kluwer Academic Publishers: Dordrecht, 2003.
30. Zhao, S.; Wu, J., An Efficient Method for Accurate Evaluation of the Site-Site Direct Correlation Functions of Molecular Fluids. *Molecular Physics* **2011**, *109*, 2553-2564.
31. Berne, B., *Statistical Mechanics: Part A: Equilibrium Techniques*; Springer US, 2012.
32. Chandler, D.; Andersen, H. C., Optimized Cluster Expansions for Classical Fluids .2. Theory of Molecular Liquids. *J. Chem. Phys.* **1972**, *57*, 1930-+.
33. Hirata, F.; Rossky, P. J.; Pettitt, B. M., The Interionic Potential of Mean Force in a Molecular Polar-Solvent from an Extended Rism Equation. *Journal of Chemical Physics* **1983**, *78*, 4133-4144.

34. Sullivan, D. E.; Gray, C. G., Evaluation of Angular-Correlation Parameters and the Dielectric-Constant in the Rism Approximation. *Molecular Physics* **1981**, *42*, 443-454.
35. Chuev, G. N.; Vyalov, I.; Georgi, N., Extraction of Atom-Atom Bridge and Direct Correlation Functions from Molecular Simulations: A Test for Ambient Water. *Chem Phys Lett* **2013**, *561*, 175-178.
36. Perkyns, J. S.; Pettitt, B. M., A Dielectrically Consistent Interaction Site Theory for Solvent Electrolyte Mixtures. *Chem Phys Lett* **1992**, *190*, 626-630.
37. Salacuse, J. J.; Denton, A. R.; Egelstaff, P. A., Finite-Size Effects in Molecular Dynamics Simulations: Static Structure Factor and Compressibility .1. Theoretical Method. *Physical Review E* **1996**, *53*, 2382-2389.
38. Overduin, S. D.; Patey, G. N., Understanding the Structure Factor and Isothermal Compressibility of Ambient Water in Terms of Local Structural Environments. *Journal of Physical Chemistry B* **2012**, *116*, 12014-12020.
39. Yu, H. A.; Roux, B.; Karplus, M., Solvation Thermodynamics - an Approach from Analytic Temperature Derivatives. *J Chem Phys* **1990**, *92*, 5020-5032.
40. Fu, J.; Wu, J. Z., Predicting Thermodynamic Data from First Principles: Hydration Free Energies for Common Organic Molecules. *Fluid Phase Equilibria* **2016**, *407*, 304–313.
41. Fu, J.; Liu, Y.; Wu, J. Z., Molecular Density Functional Theory for Multiscale Modeling of Hydration Free Energy. *Chemical Engineering Science* **2015**, *126*, 370–382.
42. Liu, Y.; Fu, J.; Wu, J. Z., High-Throughput Prediction of the Hydration Free Energies of Small Molecules from a Classical Density Functional Theory. *The Journal of Physical Chemistry Letters* **2013**, *4*, 3687−3691.
43. Liu, Y.; Zhao, S. L.; Wu, J. Z., A Site Density Functional Theory for Water: Application to Solvation of Amino Acid Side Chains. *J. Chem. Theory Comput.* **2013**, *9*, 1896-1908.
44. Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A., Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350-358.
45. Andersen, H. C.; Chandler, D., Optimized Cluster Expansions for Classical Fluids. I. General Theory and Variational Formulation of the Mean Spherical Model and Hard Sphere Percus‐Yevick Equations. *The Journal of Chemical Physics* **1972**, *57*, 1918-1929.
46. Cummings, P. T.; Stell, G., Interaction Site Models for Molecular Fluids. *Molecular Physics* **1982**, *46*, 383-426.
47. Holovko, M.; Kovalenko, A.; Hirata, F., Partial Molar Volume of Nonionic Surfactants in Aqueous Solution Studied by the Kb/3d-Rism-Kh Theory. *J Mol Liq* **2016**, *217*, 103-111.
48. Soper, A. K.; Phillips, M. G., A New Determination of the Structure of Water at 25-Degrees-C. *Chemical Physics* **1986**, *107*, 47-60.
49. Levy, Y.; Onuchic, J. N., Water Mediation in Protein Folding and Molecular Recognition. *Annu Rev Bioph Biom* **2006**, *35*, 389-415.

50. Frauenfelder, H.; Fenimore, P. W.; Chen, G.; McMahon, B. H., Protein Folding Is Slaved to Solvent Motions. *P Natl Acad Sci USA* **2006**, *103*, 15469-15472.
51. Litwinienko, G.; Ingold, K. U., Solvent Effects on the Rates and Mechanisms of Reaction of Phenols with Free Radicals. *Accounts Chem Res* **2007**, *40*, 222-230.
52. Timasheff, S. N., The Control of Protein Stability and Association by Weak-Interactions with Water - How Do Solvents Affect These Processes. *Annu Rev Bioph Biom* **1993**, *22*, 67-97.
53. Liptay, W., Electrochromism and Solvatochromism. *Angew Chem Int Edit* **1969**, *8*, 177-&.
54. Hansen, N.; van Gunsteren, W. F., Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632-2647.
55. Hirata, F., *Molecular Theory of Solvation*; Kluwer Academic Publishers: Dordrecht ; Boston, 2003, p x, 358 p.
56. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Uniform Treatment of Solute-Solvent Dispersion in the Ground and Excited Electronic States of the Solute Based on a Solvation Model with State-Specific Polarizability. *J. Chem. Theory Comput.* **2013**, *9*, 3649-3659.
57. Truhlar, D. G., Inverse Solvent Design. *Nat. Chem.* **2013**, *5*, 902-903.
58. Cramer, C. J.; Truhlar, D. G., Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem Rev* **1999**, *99*, 2161-2200.
59. Roux, B.; Simonson, T., Implicit Solvent Models. *Biophys. Chem.* **1999**, *78*, 1-20.
60. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem Rev* **2005**, *105*, 2999-3093.
61. Peeples, C. A.; Schreckenbach, G., Implementation of the Sm12 Solvation Model into Adf and Comparison with Cosmo. *J. Chem. Theory Comput.* **2016**, *12*, 4033-4041.
62. Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G., Predicting Aqueous Free Energies of Solvation as Functions of Temperature. *J Phys Chem B* **2006**, *110*, 5665-5675.
63. Andreussi, O.; Dabo, I.; Marzari, N., Revised Self-Consistent Continuum Solvation in Electronic-Structure Calculations. *J Chem Phys* **2012**, *136*.
64. Sundararaman, R.; Schwarz, K. A.; Letchworth-Weaver, K.; Arias, T. A., Spicing up Continuum Solvation Models with Salsa: The Spherically Averaged Liquid Susceptibility Ansatz. *J Chem Phys* **2015**, *142*.
65. Kollman, P., Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem Rev* **1993**, *93*, 2395-2417.
66. Mobley, D. L.; Klimovich, P. V., Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J Chem Phys* **2012**, *137*.
67. de Ruiter, A.; Oostenbrink, C., Free Energy Calculations of Protein-Ligand Interactions. *Curr Opin Chem Biol* **2011**, *15*, 547-552.
68. Brandsdal, B. O.; Osterberg, F.; Almlof, M.; Feierberg, I.; Luzhkov, V. B.; Aqvist, J., Free Energy Calculations and Ligand Binding. *Adv. Protein Chem.* **2003**, *66*, 123-58.
69. Jiang, W.; Roux, B., Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (Fep/H-Remd) for Absolute Ligand Binding Free Energy Calculations. *J Chem Theory Comput* **2010**, *6*, 2559-2565.

70. Lv, C.; Zheng, L.; Yang, W., Generalized Essential Energy Space Random Walks to More Effectively Accelerate Solute Sampling in Aqueous Environment. *J. Chem. Phys.* **2012**, *136*, 044103.
71. Lee, S.; Chen, M.; Yang, W.; Richards, N. G., Sampling Long Time Scale Protein Motions: Osrw Simulation of Active Site Loop Conformational Free Energies in Formyl-Coa:Oxalate Coa Transferase. *J. Am. Chem. Soc.* **2010**, *132*, 7252-7253.
72. Zheng, L.; Chen, M.; Yang, W., Simultaneous Escaping of Explicit and Hidden Free Energy Barriers: Application of the Orthogonal Space Random Walk Strategy in Generalized Ensemble Based Conformational Sampling. *J. Chem. Phys.* **2009**, *130*, 234105.
73. Zheng, L.; Chen, M.; Yang, W., Random Walk in Orthogonal Space to Achieve Efficient Free-Energy Simulation of Complex Systems. *Proc Natl Acad Sci U S A* **2008**, *105*, 20227-32.
74. Shirts, M. R.; Mobley, D. L., An Introduction to Best Practices in Free Energy Calculations. *Methods Mol Biol* **2013**, *924*, 271-311.
75. Duffy, E. M.; Jorgensen, W. L., Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000**, *122*, 2878-2888.
76. Paschek, D., Temperature Dependence of the Hydrophobic Hydration and Interaction of Simple Solutes: An Examination of Five Popular Water Models. *J Chem Phys* **2004**, *120*, 6674-6690.
77. Henderson, D., *Fundamentals of Inhomogeneous Fluids*; M. Dekker: New York, 1992, p x, 606 p.
78. Wu, J. Z., Density Functional Theory for Chemical Engineering: From Capillarity to Soft Materials. *AlChE J.* **2006**, *52*, 1169-1193.
79. Wu, J. Z., Density Functional Theory for Liquid Structure and Thermodynamics. In *Molecular Therodynamics of Complex Systems*, Lu, X.; Hu, Y., Eds. 2009; Vol. 131, pp 1-73.
80. Borgis, D.; Gendre, L.; Ramirez, R., Molecular Density Functional Theory: Application to Solvation and Electron-Transfer Thermodynamics in Polar Solvents. *J. Phys. Chem. B* **2012**, *116*, 2504-2512.
81. Kovalenko, A.; Kobryn, A. E.; Gusarov, S.; Lyubimova, O.; Liu, X. J.; Blinov, N.; Yoshida, M., Molecular Theory of Solvation for Supramolecules and Soft Matter Structures: Application to Ligand Binding, Ion Channels, and Oligomeric Polyelectrolyte Gelators. *Soft Matter* **2012**, *8*, 1508-1520.
82. Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V., Towards a Universal Method for Calculating Hydration Free Energies: A 3d Reference Interaction Site Model with Partial Molar Volume Correction. *J. Phys. Condens. Matter* **2010**, *22*.
83. Sergiievskyi, V. P.; Jeanmairet, G.; Leyesque, M.; Borgis, D., Fast Computation of Solvation Free Energies with Molecular Density Functional Theory: Thermodynamic-Ensemble Partial Molar Volume Corrections. *J. Phys. Chem. Lett.* **2014**, *5*, 1935-1942.
84. Chong, S. H.; Ham, S., Thermodynamic-Ensemble Independence of Solvation Free Energy. *J. Chem. Theory Comput.* **2015**, *11*, 378-380.

85. Misin, M.; Fedorov, M. V.; Palmer, D. S., Communication: Accurate Hydration Free Energies at a Wide Range of Temperatures from 3d-Rism. *J. Chem. Phys.* **2015**, *142*.
86. Liu, Y.; Fu, J.; Wu, J. Z., High-Throughput Prediction of the Hydration Free Energies of Small Molecules from a Classical Density Functional Theory. *J Phys Chem Lett* **2013**, *4*, 3687-3691.
87. Fu, J.; Liu, Y.; Wu, J. X., Molecular Density Functional Theory for Multiscale Modeling of Hydration Free Energy. *Chem Eng Sci* **2015**, *126*, 370-382.
88. Fu, J.; Liu, Y.; Wu, J. Z., Fast Prediction of Hydration Free Energies for Sampl4 Blind Test from a Classical Density Functional Theory. *J. Comput. Aided Mol. Des.* **2014**, *28*, 299-304.
89. Borgis, D.; Levy, N.; Marchi, M., Computing the Electrostatic Free-Energy of Complex Molecules: The Variational Coulomb Field Approximation. *J. Chem. Phys.* **2003**, *119*, 3516-3528.
90. Jeanmairet, G.; Levesque, M.; Sergiievskyi, V.; Borgis, D., Classical Density Functional Theory to Tackle Solvation in Molecular Liquids. Computational Trends in Solvation and Transport in Liquids. In *Computational Trends in Solvation and Transport in Liquids*, Sutmann, G.; Grotendorst, J.; Gompper, G.; Marx, D., Eds. Schriften des Forschungszentrums Jülich: 2015; Vol. 28.
91. Jeanmairet, G.; Levy, N.; Levesque, M.; Borgis, D., Molecular Density Functional Theory of Water Including Density–Polarization Coupling. *J. Phys.: Condens. Matter* **2016**, *28*, 244005
92. Liu, Y.; Wu, J. Z., Communication: Long-Range Angular Correlations in Liquid Water. *J. Chem. Phys.* **2013**, *139*.
93. Ahmed, A.; Sandler, S. I., Temperature-Dependent Physicochemical Properties and Solvation Thermodynamics of Nitrotoluenes from Solvation Free Energies. *J. Chem. Eng. Data* **2015**, *60*, 16-27.
94. Staudinger, J.; Roberts, P. V., A Critical Compilation of Henry's Law Constant Temperature Dependence Relations for Organic Compounds in Dilute Aqueous Solutions. *Chemosphere* **2001**, *44*, 561-576.
95. Zhao, S. L.; Liu, Y.; Liu, H. L.; Wu, J. Z., Site-Site Direct Correlation Functions for Three Popular Molecular Models of Liquid Water. *J. Chem. Phys.* **2013**, *139*, Artn 064509.
96. Jin, Z. H.; Wu, J. Z., Hybrid Mc-Dft Method for Studying Multidimensional Entropic Forces. *J. Phys. Chem. B* **2011**, *115*, 1450-1460.
97. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269-6271.
98. Barker, J. A.; Henderso.D, Perturbation Theory and Equation of State for Fluids - Square-Well Potential. *J. Chem. Phys.* **1967**, *47*, 2856-&.
99. Sheng, S.; Wu, J., Direct Correlation Functions for Three-Site and Four-Site Water Models. *Mol. Phys.* **2016**, 1-13.
100. Fu, J.; Wu, J. Z., Toward High-Throughput Predictions of the Hydration Free Energies of Small Organic Molecules from First Principles. *Fluid Phase Equilib.* **2016**, *407*, 304-313.

101. Zhao, S. L.; Wu, J. Z., An Efficient Method for Accurate Evaluation of the Site-Site Direct Correlation Functions of Molecular Fluids. *Mol. Phys.* **2011**, *109*, 2553-2564.
102. Blum, L.; Hoye, J. S., Mean Spherical Model for Asymmetric Electrolytes .2. Thermodynamic Properties and Pair Correlation-Function. *J. Phys. Chem.* **1977**, *81*, 1311-1317.
103. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, Efficient Generation of High-Quality Atomic Charges. Am1-Bcc Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132-146.
104. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25*, 1157-1174.
105. Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247-260.
106. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and Testing of the Opls All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* **1996**, *118*, 11225-11236.
107. Neese, F., The Orca Program System. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73-78.
108. Li, J. P.; Fu, J.; Huang, X.; Lu, D. N.; Wu, J. Z., Predicting Hydration Free Energies of Amphetamine-Type Stimulants with a Customized Molecular Model. *J. Phys. Condens. Matter* **2016**, *28*, Artn 344001.
109. Kell, G. S., Precise Representation of Volume Properties of Water at One Atmosphere. *Eng. Tech. Appl. Sci.* **1979**, 16-16.
110. Fernandez, D. P.; Mulev, Y.; Goodwin, A. R. H.; Sengers, J. M. H. L., A Database for the Static Dielectric-Constant of Water and Steam. *J Phys Chem Ref Data* **1995**, *24*, 33-69.
111. Kumar, I.; Ramaraju, P.; Mir, N. A., Asymmetric Trienamine Catalysis: New Opportunities in Amine Catalysis. *Org Biomol Chem* **2013**, *11*, 709-716.
112. Huang, X. P.; Chang, X. J.; He, Q.; Cui, Y. M.; Zhai, Y. H.; Jiang, N., Tris(2-Aminoethyl) Amine Functionalized Silica Gel for Solid-Phase Extraction and Preconcentration of Cr(Iii), Cd(Ii) and Pb(Ii) from Waters. *J Hazard Mater* **2008**, *157*, 154-160.
113. Rochelle, G. T., Amine Scrubbing for Co2 Capture. *Science* **2009**, *325*, 1652-1654.
114. Angeletti, E.; Canepa, C.; Martinetti, G.; Venturello, P., Amino-Groups Immobilized on Silica-Gel - an Efficient and Reusable Heterogeneous Catalyst for the Knoevenagel Condensation. *J Chem Soc Perk T 1* **1989**, 105-107.
115. Isobe, K.; Hoshi, T.; Suzuki, T.; Hagiwara, H., Knoevenagel Reaction in Water Catalyzed by Amine Supported on Silica Gel. *Mol Divers* **2005**, *9*, 317-320.
116. Manu, V.; Mody, H. M.; Bajaj, H. C.; Jasra, R. V., Adsorption of Cu2+ on Amino Functionalized Silica Gel with Different Loading. *Ind Eng Chem Res* **2009**, *48*, 8954-8960.

117. Knowles, G. P.; Graham, J. V.; Delaney, S. W.; Chaffee, A. L., Aminopropyl-Functionalized Mesoporous Silicas as Co2 Adsorbents. *Fuel Process Technol* **2005**, *86*, 1435-1448.
118. Saeung, S.; Boonamnuayvitaya, V., Adsorption of Formaldehyde Vapor by Amine-Functionalized Mesoporous Silica Materials. *J Environ Sci-China* **2008**, *20*, 379-384.
119. Mercera, P. D. L.; Vanommen, J. G.; Doesburg, E. B. M.; Burggraaf, A. J.; Ross, J. R. H., Zirconia as a Support for Catalysts - Evolution of the Texture and Structure on Calcination in Air. *Appl Catal* **1990**, *57*, 127-148.
120. Mercera, P. D. L.; Vanommen, J. G.; Doesburg, E. B. M.; Burggraaf, A. J.; Ross, J. R. H., Zirconia as a Support for Catalysts - Influence of Additives on the Thermal-Stability of the Porous Texture of Monoclinic Zirconia. *Appl Catal* **1991**, *71*, 363-391.
121. Hadjiivanov, K. I.; Klissurski, D. G., Surface Chemistry of Titania (Anatase) and Titania-Supported Catalysts. *Chem Soc Rev* **1996**, *25*, 61-+.
122. Dyson, P. J.; Jessop, P. G., Solvent Effects in Catalysis: Rational Improvements of Catalysts Via Manipulation of Solvent Interactions. *Catal Sci Technol* **2016**, *6*, 3302-3316.
123. Niethammer, C., et al., Ls1 Mardyn: The Massively Parallel Molecular Dynamics Code for Large Systems. *J Chem Theory Comput* **2014**, *10*, 4455-4464.
124. Stumpe, M. C.; Blinov, N.; Wishart, D.; Kovalenko, A.; Pande, V. S., Calculation of Local Water Densities in Biological Systems: A Comparison of Molecular Dynamics Simulations and the 3d-Rism-Kh Molecular Theory of Solvation. *J Phys Chem B* **2011**, *115*, 319-328.
125. Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V., Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. *Chem Rev* **2015**, *115*, 6312-6356.
126. Vanommeslaeghe, K., et al., Charmm General Force Field: A Force Field for Drug-Like Molecules Compatible with the Charmm All-Atom Additive Biological Force Fields. *J Comput Chem* **2010**, *31*, 671-690.
127. Jorgensen, W. L.; Tiradorives, J., The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *J Am Chem Soc* **1988**, *110*, 1657-1666.
128. Cygan, R. T.; Liang, J. J.; Kalinichev, A. G., Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field. *J Phys Chem B* **2004**, *108*, 1255-1266.
129. Hill, J. R.; Sauer, J., Molecular Mechanics Potential for Silica and Zeolite Catalysts Based on Ab-Initio Calculations .1. Dense and Microporous Silica. *J Phys Chem-Us* **1994**, *98*, 1238-1244.
130. Cruz-Chu, E. R.; Aksimentiev, A.; Schulten, K., Water-Silica Force Field for Simulating Nanodevices. *J Phys Chem B* **2006**, *110*, 21497-21508.
131. Lopes, P. E. M.; Murashov, V.; Tazi, M.; Demchuk, E.; MacKerell, A. D., Development of an Empirical Force Field for Silica. Application to the Quartz-Water Interface. *J Phys Chem B* **2006**, *110*, 2782-2792.

132. Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H., Force Field and a Surface Model Database for Silica to Simulate Interfacial Properties in Atomic Resolution. *Chemistry of Materials* **2014**, *26*, 2647-2658.
133. Kroutil, O.; Chval, Z.; Skelton, A. A.; Predota, M., Computer Simulations of Quartz (101)-Water Interface over a Range of Ph Values. *J Phys Chem C* **2015**, *119*, 9274-9286.
134. Heinz, H.; Lin, T. J.; Mishra, R. K.; Emami, F. S., Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The Interface Force Field. *Langmuir* **2013**, *29*, 1754-1765.
135. Plimpton, S., Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J Comput Phys* **1995**, *117*, 1-19.
136. Parimal, S.; Cramer, S. M.; Garde, S., Application of a Spherical Harmonics Expansion Approach for Calculating Ligand Density Distributions around Proteins. *J Phys Chem B* **2014**, *118*, 13066-13076.
137. Luzar, A.; Chandler, D., Hydrogen-Bond Kinetics in Liquid Water. *Nature* **1996**, *379*, 55-57.
138. Zhao, S. L.; Jin, Z. H.; Wu, J. Z., New Theoretical Method for Rapid Prediction of Solvation Free Energy in Water. *J. Phys. Chem. B* **2011**, *115*, 6971-6975.
139. Levesque, M.; Vuilleumier, R.; Borgis, D., Scalar Fundamental Measure Theory for Hard Spheres in Three Dimensions: Application to Hydrophobic Solvation. *J. Chem. Phys.* **2012**, *137*, 034115.
140. Ramirez, R.; Borgis, D., Density Functional Theory of Solvation and Its Relation to Implicit Solvent Models. *J. Phys. Chem. B* **2005**, *109*, 6754-6763.
141. Zhao, S. L.; Ramirez, R.; Vuilleumier, R.; Borgis, D., Molecular Density Functional Theory of Solvation: From Polar Solvents to Water. *J. Chem. Phys.* **2011**, *134*.
142. Wu, J., *Molecular Thermodynamics of Complex Systems*; Springer: Berlin, 2009; Vol. 131, p 1-73.
143. Zhao, S. L.; Liu, Y.; Liu, H. L.; Wu, J. Z., Site-Site Direct Correlation Functions for Three Popular Models of Liquid Water *J. Chem. Theo. Comput. (submitted)* **2013**.
144. Liu, Y.; Zhao, S. L.; Wu, J. Z., A New Site Density Functional Theory for Water: Application to Solvation of Amino-Acid Side Chains. *J. Chem. Theory Comput. (Accepted)* **2013**.
145. Talanquer, V.; Oxtoby, D. W., Nucleation of Pores in Amphiphile Bilayers. *J. Chem. Phys.* **2003**, *118*, 872-877.
146. Talanquer, V.; Oxtoby, D. W., Nucleation in a Slit Pore. *J. Chem. Phys.* **2001**, *114*, 2793-2801.
147. Hooper, J. B.; McCoy, J. D.; Curro, J. G., Density Functional Theory of Simple Polymers in a Slit Pore. I. Theory and Efficient Algorithm. *J. Chem. Phys.* **2000**, *112*, 3090-3093.
148. Sheng, S. J.; Fu, J.; Wong, B. M.; Wu, J. Z., Solvation Structure of Surface-Supported Amine Fragments: A Molecular Dynamics Study. *J Phys Chem C* **2017**, *121*, 22156-22163.

149. Joung, I. S.; Cheatham, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J Phys Chem B* **2008**, *112*, 9020-9041.
150. Babu, C. S.; Lim, C., Empirical Force Fields for Biologically Active Divalent Metal Cations in Water. *J Phys Chem A* **2006**, *110*, 691-699.
151. van Veggel, F. C. J. M.; Reinhoudt, D. N., New, Accurate Lennard-Jones Parameters for Trivalent Lanthanide Ions, Tested on [18]Crown-6. *Chem-Eur J* **1999**, *5*, 90-95.