Published in final edited form as:

J Phys Chem B. 2010 September 2; 114(34): 11076–11092. doi:10.1021/jp101597r.

Non-additive Empirical Force Fields for Short-Chain Linear Alcohols: Methanol to Butanol. Hydration Free Energetics and Kirkwood-Buff Analysis Using Charge Equilibration Models

Yang Zhong and Sandeep Patel*

Department of Chemistry and Biochemistry, 238 Brown Lab, University of Delaware, Newark, Delaware 19716, USA

Abstract

Building upon the nonadditive electrostatic force field for alcohols based on the CHARMM charge equilibration (CHEQ) formalism, we introduce atom-pair specific solute-solvent Lennard-Jones (LJ) parameters for alcohol-water interaction force fields targeting improved agreement with experimental hydration free energies of a series of small molecule linear alcohols as well as ab initio water-alcohol geometries and energetics. We consider short-chain, linear alcohols from methanol to butanol as they are canonical small-molecule organic model compounds to represent the hydroxyl chemical functionality for parameterizing biomolecular force fields for proteins. We discuss molecular dynamics simulations of dilute aqueous solutions of methanol and ethanol in TIP4P-FQ water, with particular discussion of solution densities, structure defined in radial distribution functions, electrostatic properties (dipole moment distributions), hydrogen bonding patterns of water, as well as a Kirkwood-Buff (KB) integral analysis. Calculation of the latter provides an assessment of how well classical force fields parameterized to at least semi-quantitatively match experimental hydration free energies capture the microscopic structures of dilute alcohol solutions; the latter translate into macroscopic thermodynamic properties through the application of KB analysis. We find that the CHEQ alcohol force fields of this work semi-quantitatively match experimental KB integrals for methanol and ethanol mole fractions of 0.1 and 0.2. The force field combination qualitatively captures the concentration dependence of the alcohol-alcohol and water-water KB integrals, but due to inadequacies in the representation of the microscopic structures in such systems (which cannot be parameterized in any systematic fashion), *a priori* quantitative description of alcohol-water KB integrals remains elusive.

I. INTRODUCTION

Short-chain alcohols such as methanol, ethanol, propanol, and butanol have long served as model compounds for developing classical force fields for proteins. As one of the most well-studied systems of classical molecular dynamics simulations, empirical force fields were constructed for alcohol molecules, the majority of which are expressed as fixed-charge potentials^{1–6}. While additive potential functions work reasonably well in reproducing bulk condensed-phase properties, they prohibit accurate investigations of local electrostatic changes. In response, polarizable alcohol force fields incorporating explicit electronic induction effects have been proposed. Gao et al.⁷ reported a polarizable intermolecular potential function for alcohols by defining atomic dipoles. Gonzalez et al.⁸ developed an ethanol force field with a point molecular polarizability. An additional dipole polarizable model of methanol was developed by Caldwell et al.⁹. Noskov et al. introduced Drude particles to carry charges and allowed the molecular dipole to respond to the field¹⁰. Patel et

*Corresponding author. sapatel@udel.edu.

al.^{11,12} contributed polarizable force fields for methanol and ethanol employing the charge equilibration formalism. A large number of molecular simulation studies have been performed on alcohol aqueous solutions, considered as ideal systems for studying amphiphilic hydration, employing existing polarizable, or nonpolarizable, alcohol force fields.^{10,13–22}

From the point of view of force field development, simple alcohols offer computationally tractable model compounds for constructing interaction potentials for biological systems. An important consideration in developing classical force fields is the interaction between a chemical functional group and the major biological solvent, water. Along with structures and energetics of complexes of model compounds with water, an important condensed phase energetic property is the hydration free energy, ΔG_{hyd} . We note that the hydration free energy has been considered as a key property in recent force field development work^{16,23,24}. Values for various alcohol models have been reported and applied as part of small molecule test sets to assess force fields. Shirts et al.²⁵ used recent versions of the OPLS-AA 1996⁴, CHARMM22²⁶, and AMBER(ff94)²⁷ parameter sets along with the TIP3P solvent model²⁸ to perform thermodynamic Integration (TI) and obtained ΔG_{hyd} for methanol and ethanol along with thirteen other amino acid analogues. To account for Lennard-Jones (LJ) interactions beyond a specified spherical cutoff distance, they introduced an analytic long range correction (LRC)²⁵ and reported that the free energies of hydration for those popular computational models are less favorable than experiment by 0.47 – 1.06 kcal/mol. They also claimed that ΔG_{hyd} can be highly dependent on simulation protocols. The authors later extended the study by examining the dependence of hydration free energies on several different water models²⁹. The water models considered were the original TIP3P²⁸, TIP4P²⁸, SPC³⁰, SPC/E³¹, TIP4P-Ew³², and a sparsely-used model developed by Sun and Kollman named the TIP3P-MOD³³. A previous study of methanol and ethanol aqueous solutions by Deng and Roux³⁴ using the CHARMM22 force field^{26,27} with a spherical solvent boundary potential (SSBP)³⁵ found ΔG_{hyd} of -4.88 and -4.08 kcal/mol, comparing to experimental results of -5.11 and -5.01 kcal/mol³⁶, respectively. Hess and van der Vegt employed AMBER99³⁷, GROMOS 53A6²³ and OPLS-AA^{4,38,39} force fields and five different water models to compute the hydration thermodynamic properties of 13 amino acid side chain analogues, including methanol and ethanol, and found greater influence of the water model than the solute force field on prediction of solute hydration free energy⁴⁰. Mobley *et al.* explored the effect of partial charges on hydration free energies with fixed-charge force fields and found that the discrepancy with experimental free energies grows with the molecular polarity⁴¹. A recent extensive test on 504 neutral small organic compounds using the general Amber force field⁴² showed an root-mean-squared (RMS) error of 1.24 kcal/mol and a correlation coefficient of 0.89 between the computed and experimental hydration free energy⁴³. Similar effort had also been made in the current polarizable force field. Anisimov *et al.*¹⁶ developed a polarizable empirical force field for the primary and secondary alcohol series based on the classical Drude oscillator. The CHARMM22²⁶ results in their study were too favorable compared to experimental data by 4-19% and the polarizable model tended to enhance the effect. The authors integrated free energies of hydration during parameter optimization and added *explicit*, pair-specific atom type $O_{alcohol}-O_{water}$ solute-solvent non-bond interaction terms, which appeared to improve agreement of gas-phase alcohol-water interactions and ΔG_{hyd} . In a more recent, detailed study, Baker and coworkers²⁴ parameterized pair-specific LJ parameters between heavy atoms of alkanes, alcohols and ethers and the oxygen atom of the SWM4-NDP water model⁴⁴ for the CHARMM⁴⁵ Drude polarizable force field. Based on a Weeks-Chandler-Anderson⁴⁶ (WCA) decomposition of the Lennard-Jones interaction between solute and solvent, the authors attributed the main source of systematic error in the hydration free energies to the dispersion interaction²⁴. Moreover, the authors determined that the limitations in the LJ combining rules were at the heart of the systematic deviations of computed hydration free energies from experimental

values. The authors also proposed a preliminary improved combining rule based on their fitted pair-specific LJ parameters²⁴. Thus, a subset of specific solute-solvent interactions were determined explicitly, while the remainder of the set were obtained from traditional application of empirical combining rules to generate solute-solvent interactions based on the non-bond interaction parameters (size and well-depth) of the homo-dimer atomic pair. With the latter idea being an underlying premise of this work, we proceed to extend the existing model (CHARMM charge equilibration force field, CHEQ) by introducing explicit hetero-atom type non-bond interactions without resorting to a predefined combining rule. Admittedly, this effectively reduces the number of constraints on the parameter determination problem. We further comment on this idea in Section III.

Nonadditive electrostatic force fields based on the charge equilibration formalism (CHEQ) predicted -5.3 ± 0.04 kcal/mol for ΔG_{hyd} of methanol using a potential of mean force (PMF) approach, -5.6 ± 0.2 kcal/mol for TI method¹⁴ and -5.70 ± 0.23 kcal/mol for ethanol using TI (experimental hydration free energies for methanol and ethanol are -5.11 and -5.01 ³⁶, respectively). In this paper, we discuss introduction of *specific* solute-solvent Lennard-Jones interaction terms between atom types associated with alcohol and water molecules in order to improve the existing alcohol charge equilibration (CHEQ) force fields; particular focus aims to reproduce experimental hydration free energies while maintaining the bulk liquid properties reported previously.^{11,12} The new force field simultaneously scales down the overly favorable gas-phase alcohol-water interactions of the earlier force field combinations (that use mixing rules)^{11,12}.

We also investigate solution properties (density, dipole moment distribution, radial distribution functions (RDFs) and hydrogen bond pattern) and changes caused by the solute-solvent LJ parameters. Finally, we apply Kirkwood-Buff analysis⁴⁷ to assess the macroscopic thermodynamics predicted from the microscopic structure defined by this combination of polarizable models; in Section III I we discuss our computed Kirkwood-Buff (KB) integrals, excess coordination numbers, and activity derivatives.

II. THEORY AND METHODS

A. Water and Alcohol Force Fields

All simulations used the CHARMM force field based on a charge equilibration formalism (CHEQ) for alcohol molecules (propanol and butanol adopted the ethanol parameters¹² and methanol force field was developed by Patel and Brooks¹¹) and the TIP4P-FQ water model⁴⁸. CHEQ has been applied to various systems over the past two decades^{11,12,48–53}. The method is founded on Sanderson's idea of electronegativity equalization^{54,55} and formulated by Yang and Parr⁵⁶ based on the density functional theory. This translates to the redistribution of charge among constituent atoms within each molecule so as to equalize the electronegativity (chemical potential) at each point.

In the charge equilibration formalism, the electrostatic energy, E_{elec} , of a molecule with N atoms is:

$$E_{elec} = \sum_{i=1}^N \chi_i^0 Q_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \eta_{ij} Q_i Q_j + \sum_{i=1}^N \Phi_i Q_i \quad (1)$$

where Q_i 's are atomic partial charges, and χ 's are atom electronegativities defining the directionality of charge flow, and the η 's are the atom hardnesses representing a resistance to electron flow to or from an atom. We assign individual atomic hardness values, η_i according

to the work of Patel and Brooks^{11,12} and derive heterogeneous elements, η_{ij} , based on the combining rule:

$$\eta_{ij}(r_{ij}, \eta_i, \eta_j) = \frac{\frac{1}{2}(\eta_i + \eta_j)}{\sqrt{1.0 + \frac{1}{4}(\eta_i + \eta_j)^2} r_{ij}^2} \quad (2)$$

where r_{ij} is the separation between atoms (or more generally sites) i and j . This local

screened coulomb potential has the correct $\frac{1}{r_{ij}}$ limiting behavior for separations greater than about 2.5 Å. This interaction is computed for 1-2, 1-3, and 1-4 sites (sites included in bonds, angles, and dihedrals); sites separated by 5 or more sites interact via a standard Coulomb interaction. In the case of interacting molecules, the interaction between sites on different molecules is also of the Coulomb form. We note that the form of this combining rule efficiently recapitulates the separation-dependence of the two-electron Coulomb interaction between unit charges on different nuclei at small separations⁵⁷. Finally, any external potential arising from electrostatic interactions between molecules or between spatially-separated sites within a molecule (i.e., sites not involved in bond, angle, and dihedral interactions) is represented by the potential Φ in the last term of Eq 1. With respect to charge dynamics, an extended Lagrangian formalism is used to propagate the charges in time with the constraint that the total charge on a molecule is conserved. Extended Lagrangian methods are now routinely applied as efficient means for propagating electronic degrees of freedom for polarizable force fields¹⁶. The fictitious charge dynamics, analogous to wavefunction dynamics in Car-Parrinello (CP) type methods⁵⁸, is determined by a charge

mass (adiabaticity parameter in CP dynamics). The units for this mass are $\left(\frac{\text{energy} \cdot \text{time}^2}{\text{charge}^2}\right)$. The charges are propagated based on forces arising from the difference between the average electronegativity of the molecule and the instantaneous electronegativity at an atomic site.

The non-electrostatic non-bond interaction, based on the pair separation distance r_{ij} between atoms i and j , is of the Lennard-Jones type (standard in the CHARMM⁴⁵ implementation):

$$E(r_{ij}) = \sum_{ij} \epsilon_{ij} \left(\frac{R_{min,ij}^{12}}{r_{ij}^{12}} - 2 \frac{R_{min,ij}^6}{r_{ij}^6} \right) \quad (3)$$

where the summation runs over all atom pairs not involved in bond, angle, and dihedral interactions. The ϵ_{ij} is the Lennard-Jones interaction well-depth, and $R_{min,ij}$ is the location of the minimum. In the canonical approach, the ϵ_{ij} and $R_{min,ij}$ are determined using the atom type ϵ_i , ϵ_j and $R_{min,i}$, $R_{min,j}$ parameters combined through an analytic relation such as the Lorentz-Berthelot (L-B) mixing rule. This protocol applies a geometric mean (Berthelot) rule for the ϵ_{ij} , $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$, and an arithmetic mean rule (Lorentz) for the

$$R_{min,ij}, R_{min,ij} = \frac{1}{2} (R_{min,i} + R_{min,j}).$$

Though this mixing rule is based on physical arguments, it is not unique as there exist alternate combining rules such as the Kong⁵⁹ and Waldman-Hagler⁶⁰ rules which link the well depths, ϵ_{ij} to the relative sizes of interacting atoms. This effectively goes beyond the usual L-B mixing rule approximation and affords more parametric flexibility for defining interactions between different atom types. Indeed, certain deficiencies in the commonly applied L-B mixing rules have been reported in the recent literature^{24,61}. There still remains

significant work to define the accuracy of combining rules across a broad range of functional group types or to develop and rationalize novel combining rules. Work in this direction is reflected in recent studies addressing non-polar, alkane-type fluids⁶² and more recently a broader class of non-polar and polar chemical functionalities²⁴.

B. Simulation Details

Hydration free energies of alcohols in model dilute aqueous solutions are calculated using a cubic box of 216 TIP4P-FQ water molecules⁴⁸ and one alcohol molecule via thermodynamic integration (TI) as discussed previously^{14,15,63}. We first electrostatically decouple the alcohol molecule from the solvent before decoupling the Lennard-Jones (LJ) interactions. Double wide sampling⁶⁴ is employed for the electrostatic decoupling and the LJ decoupling. The electrostatic component of the free energy of hydration is computed with 200 ps of molecular dynamics sampling for methanol and ethanol (50 ps of equilibration and 150 ps of sampling) at constant pressure and temperature (NPT) at 298 K and 1 atm pressure at each λ - coupling parameter (details of the pressure and temperature control are given below). Each λ - window of the LJ decoupling is sampled for 300 ps, the first 50 ps of which is considered as the equilibration. The free energies of hydration for the original CHEQ model (using L-B combination rules) are calculated using 21 λ - windows for the electrostatic decoupling and 21 λ - windows for the LJ decoupling to be consistent with previous work on methanol and ethanol¹⁴; each window is sampled for the same amount of time as described above. For larger alcohol molecules, we extend the total simulation time by twice for propanol and 3 times for butanol for both LJ and electrostatic decoupling and use 11 windows for both electrostatic or LJ decoupling. Equilibrium charge distributions on each molecule are reversibly sampled as we electrostatically decouple the solute. The separation shifted scaling method of Zacharias et al.⁶⁵ is employed to ensure more uniform sampling along the LJ decoupling path. The hydration free energy is determined by the average of ten independent runs performed with different random seeds. The uncertainty in the TI calculations is determined from the standard deviation between these ten independent replicates.

The long-range correction (LRC) accounting for the LJ interactions between a *single* solute molecule and solvent molecules separated by radial distances greater than a specified cutoff distance is estimated for each solute molecule using an analytic repulsion-dispersion correction suggested by Shirts et al²⁹:

$$E_{LRC}(r_{ij}) = \sum_i \sum_j 16 \pi \rho \epsilon_{ij} \int_{r=r_{on}}^{\infty} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] (1 - S(r_{ij})) r_{ij}^2 dr \quad (4)$$

where ρ is the number density of the solvent molecules, ϵ_{ij} is the LJ well-depth (as discussed above), σ_{ij} solute-solvent effective LJ diameter, and r_{on} is the radial distance from an atom at which the LJ switching function takes effect. The solute-solvent LJ parameters used in the above equation are either those with specific values (as determined from the fitting procedure of this work) or L-B combining-rule based values; atom type pairs for which specific values are used are shown in Table I. The double summation runs over the atoms of a single solute and single solvent molecule; the solvent density accounts for the number of solvent atoms necessary to be included in the correction. For the hydration free energy results shown in Table II, the LRC values given use the final (after fitting) atom pair specific LJ values. Since the CHARMM non-bond force field functional form uses the location of the minimum of the potential energy function between two atoms, R_{min} , which differs from the parameter, σ (the Lennard-Jones diameter), appropriate transforms between R_{min} and σ are included in the calculation. The LJ switching function, $S(r_{ij})$, used in CHARMM⁶⁶ is:

$$S(r_{ij}) = \begin{cases} 1, & r_{ij} \leq r_{on} \\ \frac{(r_{off}^2 - r_{ij}^2)^2 (r_{off}^2 + 2r_{ij}^2 - 3r_{on}^2)}{(r_{off}^2 - r_{on}^2)^3}, & r_{on} < r_{ij} \leq r_{off} \\ 0, & r_{ij} > r_{off} \end{cases} \quad (5)$$

This function (cubic in r_{ij}^2) satisfies $S(r_{on}) = 1$, $S(r_{off}) = 0$, $\left(\frac{dS}{dr}\right)_{r_{on}} = 0$ and $\left(\frac{dS}{dr}\right)_{r_{off}} = 0$, giving a continuous potential energy and force.

We later present data on gas-phase water-alcohol dimer properties. As part of the gas-phase calculations, we first performed steepest descent minimization (SD) with angle and dihedral constraints based on the corresponding values in the ROH and BIS structures in the work of Anisimov et al.¹⁶ The minimized geometries are used as starting structures of MP2/aug-cc-pvDZ optimizations. Then force field alcohol-water dimerization energies and hydrogen bond lengths are computed for MP2/aug-cc-pvDZ optimized structures with both structure and charge minimization. Necessary for the CHEQ approach is charge normalization over relevant molecular units/segments. For methanol and ethanol, charge is normalized (i.e., conserved) over the entire molecule. Propanol and butanol are treated using two charge normalization units. The first includes the hydroxyl group (OH) and the next two methylene units (i.e., an ethanol-like subunit within the molecule); the second spans the rest of the molecule. The choice of charge normalization units in the larger alcohols propanol and butanol is based on the fact that we use the pure ethanol *electrostatic* parameters to extend to larger charge flow lengths. In much the same spirit of our earlier work on alkanes and applications to lipid bilayers⁶⁷, we use the ethanol electrostatic model as a fundamental unit to limit charge transfer; thus, by analogy to Reference⁶⁷, we chose the charge normalization group as we have.

To analyze solution properties at low concentrations of methanol, ethanol, propanol and butanol (some properties are computed only for methanol and ethanol as there are both experimental and prior simulation studies to compare our results with), ten nanosecond simulations for aqueous alcohol solutions at mole fractions 0.1 and 0.2 were performed using cubic boxes of varying numbers of water and alcohol molecules. System sizes varied with concentration; a cubic cell contains 24 (54) methanol and 216 water molecules to simulate solution of 0.1 (0.2) mole fraction. Ethanol-water mixtures used 28 or 63 alcohol molecules with 250 water for each concentration. A Verlet leapfrog integrator with timestep of 0.0005 picoseconds (0.5 femtoseconds) generated the trajectories in the constant pressure-temperature ensemble (NPT). Constant pressure was maintained via a Langevin piston method⁶⁸ and a constant temperature of 298 K was maintained with a Hoover thermostat⁶⁹. Non-bonded interactions were switched to zero via a switching function from 10 to 11 Å for all the alcohol solutions studied. The Smooth Particle Mesh Ewald⁷⁰ method with 20 grid points in each dimension and screening parameter of $\kappa = 0.33$ are used to describe long-range electrostatic interactions. Charge degrees of freedom were maintained at 1 K using a single Nose-Hoover bath. For simulations of aqueous methanol and ethanol solutions,

charge degrees of freedom were assigned masses of 0.000085 and $0.000025 \frac{\text{kcal}}{\text{mol}} \frac{ps^2}{e^2}$, for alcohol and water, respectively. There are no differences in equilibrium properties of pure water or aqueous alcohol solutions using the different charge masses for water and the timestep of 0.5 fs. Methanol, ethanol, and water maintain charge normalization over individual molecules with no intermolecular charge transfer. Data from the first 0.25 nanoseconds under constant temperature and pressure at 298K and 1 bar was considered as equilibration and not included when computing solution properties.

For the Kirkwood-Buff analysis, we performed simulations using larger system sizes following Smith and coworkers^{71–76} in order to sample radial distribution functions at longer separations and generate KB integrals, G_{ij} . We perform ten-nanosecond molecular dynamics simulations for systems of size eight times larger than the original cubic boxes. The first 125 ps are considered equilibration. Nonbonded interactions were switched on from 10 to 11 Å. The PME method uses 40 grid points for methanol solutions and 50 for ethanol solutions. All the other parameters are adopted from simulations of the original cubic boxes (see above).

C. Parameterization Strategy for Atom-Pair Specific Non-Bond Interaction Parameters

We adopt the use of atom-pair specific Lennard-Jones non-bond parameters; these parameters are to be used for cross interactions between the atom types of the solute and water. For the remaining paper, we refer to the force field of this work as CHEQ-PSP, CHEQ Pair Specific Parameters.

We reiterate that the goal of the present study is to improve the prediction of hydration free energies of small molecule alcohols in TIP4P-FQ⁴⁸ solvent using charge equilibration force fields for solute and solvent. The original combination of polarizable alcohol and water force fields gave rise to a systematic overestimation of hydration free energies relative to experiment. This is demonstrated by the values of hydration free energies under the column labeled "orig" in Table II. The data shows that the deviation from experiment increases with increasing length of alkyl chain (errors for propanol and butanol are much higher than that of methanol and ethanol). This suggests that additional alkyl groups contribute to the overestimation of hydration free energy in a systematic manner as the length of the alkyl chain increases. Furthermore, our initial attempts to improve accuracy of predicted hydration free energies using only the alcohol hydroxyl oxygen and water oxygen pair-specific interactions were unsuccessful in reproducing the qualitative experimental trend of decreasing hydration free energies (less favorable hydration) with increasing alcohol alkyl chain length (predicted hydration free energies of propanol and butanol more favorable than that of methanol and ethanol). This contrasts with the behavior reported for the Drude alcohol force fields¹⁶. The specific cause of these different behaviors is difficult to assess. One possible explanation may lie in the nature of treatment of polarization in the two formalisms. The Drude model allows for additive polarization effects. Each heavy atom to which a Drude particle is attached contributes linearly to the total polarization. In the case of CHEQ force fields, addition of units to a molecule (i.e., methylene or hydroxyl groups, or larger units) leads to non-linear scaling of polarizability since charge is allowed to distribute over a larger spatial extent. The scaling of polarizability in CHEQ methods is known to be superlinear^{77–82}. Increase in polarization gives rise to enhanced induced dipole - induced dipole interactions leading to more favorable hydration free energies with increasing alkyl chain length. Thus, in the case of the alcohol series, the additional less-favorable interactions introduced via pair-specific LJ interactions between solute and solvent may be required, *in part* to offset the polarizability scaling effects. We note that this is conjectural at this point, but our attempts to modulate hydration free energies using a small palette of pair-specific LJ interactions did not produce practical solutions. Furthermore, limiting the spatial extent for charge redistribution via selection of appropriate charge normalization units, provides a means for controlling the superlinear scaling^{79,81,82}.

This being the case, we chose to include the cross-interactions between alcohol alkyl carbon atom types (CT2 and CT3 in Figure 1) and water atom types in our fitting protocol. For the present, we aim to parameterize one set of pair-specific parameters to apply for alcohol-water solution systems studied (potentially appropriate to other aqueous alcohol systems). We optimize the Lennard-Jones parameters between alcohol alkyl carbon atoms (CT3 and CT2), hydroxyl group atoms (OH1 and H), and water atoms (OT and HT); the atom types

are shown in Figure 1. Table I shows the specific alcohol-water atom types for which pair-specific Lennard-Jones parameters are determined without resort to any combining rule. There are in total sixteen optimized parameters.

The optimization process for the sixteen atom-pair specific Lennard-Jones non-bond parameters is demonstrated schematically in Figure 2. We begin with the original CHEQ alcohol parameters developed for ethanol¹² and use the L-B combining rule to generate the initial solute-solvent interactions. One parameter p_i out of the sixteen pair-specific LJ parameters $\{p_1 \dots p_{16}\}$ is modified first to reproduce the experimental hydration free energy of butanol (ΔG_{hyd} error < 10%) in the uncertainty range of the simulation result), which was poorly predicted based on the CHEQ model using the L-B combining rule. We comment in further detail on our rationale for starting with butanol in Section III. Then the parameter set containing the optimized parameter p_i is assessed by comparing to the *ab initio* gas-phase butanol-water hetero-dimer interactions, experimental free energy of hydration of methanol, ethanol, and propanol, and the *ab initio* dimerization energies and geometries in vacuum for methanol, ethanol and propanol, sequentially. If any of the tests fail (ΔG_{hyd} error or the gas-phase alcohol as proton donor dimer property error is greater than 10 %; the proton acceptor heterodimer property tolerance is 20 % to lower the the rejection ratio and this is considered reasonable as discussed in Section III B), the next parameter p_{i+1} is varied for the new parameter set, p_1 to p_i of which have been changed previously, and the process, beginning with testing of butanol hydration free energy, repeats. The parameter of interest in each iteration would be increased or decreased by a small amount to reveal the free energy dependence of this parameter and this information is used to guide the direction of optimizing this parameter. We note that there is no unique way of changing parameters during parameterization; experience and observation learned from unsuccessful parameter sets provides information such as dependence and sensitivity of parameters on the fitting property and efficiently helps optimization. An alternative approach to generate a broader palette of test parameters for the search is to perturb via free energy perturbation between parameter sets using an appropriate thermodynamic cycle; this is the approach taken by Baker et al²⁴. To speed up the parameterization process, we perform short thermodynamic integration simulations (one tenth of simulation time described in Section II B) to estimate the hydration free energies. As the final assessment of the parameter set, longer free energy calculations are performed as described in Section II B to confirm successful parameter sets, and the final hydration free energy results are displayed in Table II.

III. RESULTS AND DISCUSSION

The mixing of the original CHEQ alcohol and TIP4P-FQ⁴⁸ polarizable force fields using traditional combination rules (i.e., L-B mixing rules) were found to overestimate the methanol-water and ethanol-water hydration energies^{14,15} as well as overestimate the gas-phase water-alcohol dimer energies. Previous studies have noted the overestimation of interaction well-depths using the L-B mixing rules (though we take care to note that those studies almost exclusively focus on non-polar systems such as alkanes or perfluoroalkanes, or Lennard-Jones spheres and mixtures⁶¹). This work adds explicit Lennard-Jones interaction terms between alcohol and water atom types (Table I) without resorting to empirical combination rules, aiming to reproduce the experimental value of hydration free energies while maintaining the good agreement of the pure alcohol force fields with experimental properties of the pure liquids. We further explore gas-phase alcohol-water interactions, dilute solution densities, dipole moment distributions, and distribution functions and hydrogen bond patterns of alcohol-water mixtures using the force field with the solute-solvent LJ corrections. Finally, we discuss the nature of low concentration aqueous methanol and ethanol solutions via Kirkwood-Buff analysis.

A. Free Energies of Hydration

Table II compares the CHARMM22¹⁶, Drude¹⁶, CHEQ and experimental free energies of solvation for the alcohols studied. The free energies include a long range correction(LRC) for the truncated LJ interactions. The additive force field (CHARMM22²⁶) was shown to predict more favorable free energies of hydration than the experimental data as did the original Drude. The systematic trend for experiment was reduced by using solute-solvent $O_{alcohol}-O_{water}$ LJ terms not derived from combination rules for the Drude model¹⁶.

We choose to begin with butanol hydration free energy as a target for parameterization for two reasons. As discussed in Section II C, due to the superlinear scaling of polarizability within the CHEQ formalism^{77–82}, it is becoming evident that straightforward transferability of *small* molecule parameters to larger analogues is difficult⁶⁷. In light of the present situation, we take the approach used by Davis et al⁶⁷ to parameterize initially to a larger molecule (in this case butanol), and then apply the parameters to smaller molecules proximal to the original molecule in the analogue series (i.e, methanol, ethanol, propanol). Though not done in this study, one can consider extending to larger molecules as well. Second, the original non-bond parameters for ethanol were not transferable to the larger alcohols due to superlinear scaling of molecular polarizability with system size in charge equilibration schemes⁶³. The hydration free energy of butanol exhibited the high deviation from experiment using free energy calculations based on molecular dynamics and thermodynamic integration with parameters determined by fitting to properties of smaller alcohols (i.e. ethanol). Since we were not able to determine a non-bond parameter set that predicted butanol hydration free energy adequately, we chose to begin by fitting butanol free energies and work towards the smaller alcohols. We reiterate that this is similar to the protocol followed by Davis et al⁶⁷.

We note that the same specific alcohol-water LJ parameters are applied for methanol even though the methanol force field reported has different LJ parameters.¹¹ According to studies by Zhong et al.^{14,15}, the ΔG_{hyd} for methanol in TIP4P-FQ⁴⁸ is -5.6(0.2) kcal/mol and that of ethanol is -5.7(0.2) kcal/mol based on the alcohol CHEQ model without any solute-solvent LJ corrections; these are too favorable compared to the experimental values. After applying the solute-solvent LJ terms, the average hydration free energies deviate slightly (< 6.4%) from the experimental values but the systematic overestimation is eliminated and even reversed for methanol, so that the experimental values are all within the uncertainty ranges of CHEQ-PSP values. Meanwhile, the improved agreement between the experimental data and the CHEQ-PSP results using identical parameters for different alcohols lends support to the transferability of the atom pair specific interactions. Table I shows the specific atom pairs and associated interaction parameters from the fitting of this work.

We note that the uncertainties presented in Table II are higher than those reported for the Drude oscillator models^{16,24}. We offer several reasons for this difference, all with the common theme of differences in amount and nature of sampling. The Drude force field study decomposes the solute-solvent interactions into electrostatic, repulsion, and dispersion, the latter two decomposed via the WCA⁴⁶ scheme. The present work decouples the Lennard-Jones interaction, including repulsion and dispersion, simultaneously. Thus, the effective sampling time for the Drude case may be longer than that for the current work (we cannot say definitively what the exact difference in sampling time is). Furthermore, one can consider the WCA scheme as a slower decoupling process with greater overlap of structure and energy distributions between adjacent stages (i.e., Hamiltonians) prescribed for the decoupling process; greater overlap effectively contributes to reduced uncertainty. Furthermore, we use constant temperature and pressure simulations (NPT) for decoupling, as opposed to the constant temperature and volume (NVT simulations used in the Drude studies. Due to the added requirement of now sampling over the volume fluctuations in the

NPT ensemble, the uncertainty estimates increase relative to those for sampling in the NVT ensemble. Thus, we believe these factors contribute to higher apparent uncertainties in the values of hydration free energies reported in this work. Longer sampling, more λ -windows, and incorporation of alternative maximum likelihood estimators, such as done in the Bennett Acceptance Ratio Method^{25,29,83,84}, would lower uncertainties.

Before proceeding, we comment on the use of pair-specific LJ terms. In the context of the present study introducing non-bond interactions between solute and solvent in order to better reproduce experimental condensed phase properties (in this case hydration free energies), we acknowledge that perhaps polarization over a broad range of chemical functionalities should allow nominally for transferability from gas-phase to condensed phase environments. Unfortunately, due to the empirical nature of the treatment of polarizability even in such force fields, it is difficult to assess the nature of transferability. Polarizable empirical force fields include a repulsion-dispersion term in addition to the non-additive electrostatic term. In general, this repulsion-dispersion term is determined based on gas-phase interactions of model compounds with water as well as condensed phase properties of the pure fluid. The work of MacKerell^{16,44,85–87} and coworkers as well as AMBER development efforts^{9,88–91} demonstrate the widespread use of this protocol. Thus, it appears at this point that there is no universal philosophy of polarizable force field development that relies solely on gas-phase (*ab initio*) parameterizations which can then be applied directly (i.e., without further modification) to the condensed phase; this latter is the ideal scenario, and the inability of current approaches to allow this route perhaps speaks to an important void in our understanding of how to incorporate the necessary physics (specifically the coupling of LJ and electrostatic interactions) in a classical formalism. We note that the origins of reduced molecular polarizability in the condensed phase have been ascribed to Pauli exclusion⁹², heterogeneity of electric fields within molecular volumes⁹³, and formation of hydrogen bonded structures of polar compounds with water⁹⁴. The idea of the dependence of molecular polarizability on local environment has also been studied in detail using information theoretic methods for partitioning the electron density in the condensed phase via approaches such as Hirshfeld partitioning^{95–97}.

Though superficially the use of explicit solute-solvent interaction parameters without combination rules may appear counterintuitive, we believe that there are plausible physical reasons for adopting this approach. Since most empirical force fields to date have been developed based on pure liquid, condensed phase properties, the nature of how the electrostatic properties of the solute (i.e., charge distribution, dipole moment, etc.) change from a pure liquid to an aqueous environment are not considered. As a primary concern, we consider that transferring an alcohol molecule from its pure bulk liquid to aqueous solution will lead to changes in the partial charges on atomic sites. Traditionally, such a change of charge would entail modification of the repulsion-dispersion parameters of a non-polarizable force field. The electrostatics and non-bond interactions are in this sense coupled. Thus, using solute-solvent, atom pair-specific interactions between solute and water is one approach, admittedly simplified, to account for this coupling. In the ideal case, one would implement, within a classical formulation, a functional form that would couple the electrostatic and repulsion-dispersion interactions. Such studies are rare in the literature; the problem is rather difficult. Chen and Siepmann⁹⁸ have reported such studies in treating vapor-liquid coexistence; Bauer et al⁹⁹ have also recently explored models to explicitly couple electrostatics and non-bond interactions for pure water models. Moreover, similar sentiment has been expressed by the work of MacKerell and coworkers, particularly in the context of polarizable force fields for which transferability is much more sensitive to the coupling of electrostatics and repulsion-dispersion¹⁶. Furthermore, the effects of coupling between electrostatic and non-bond dispersion-repulsion parameters will be less (though non-zero) for non-polar molecules where dispersion interactions are dominant.

B. Gas-Phase Water-Alcohol Interactions

We next consider gas-phase alcohol-water interaction energies and geometries (shown in Table III) based on the alcohol CHEQ-PSP model. *ab initio* geometries are optimized using MP2/aug-cc-pvDZ calculations. Alcohol as *Proton Acceptor* (PA) geometry optimizations start from the BIS structures suggested by Anisimov et al.¹⁶. For further reference, we also list results of Fileti et al.¹⁰⁰, who computed dimer geometries using MP2/aug-cc-pvDZ optimizations and performed single-point calculations at the CCSD(T) level. Here we consider alcohol molecules only in the *trans* form; the difference in relative energy between *trans* and *gauche* heterodimers is negligible based on Fileti's study of ethanol-water dimers¹⁰⁰.

Table III shows corresponding alcohol-water gas-phase dimer interaction energies and hydrogen bond distances using four force fields: 1. the additive CHARMM22 (results taken from Anisimov et al.¹⁶), 2. the Drude with solute-solvent corrections¹⁶, 3. the original CHEQ^{11,12}, and 4. CHEQ-PSP (current work). Consistent with the large negative hydration free energies, the gas-phase alcohol-water interactions are significantly more favorable and bond lengths are smaller compared to the QM results using the additive CHARMM22 force field. The Drude model improves on the prediction of interaction energies by explicitly accounting for electrostatic induction effects and the same trend is observed for the CHEQ models in the cases where the alcohol is the proton donor. With solute-solvent LJ parameters, the alcohol CHEQ-PSP force field predicts gas-phase dimer energies and geometries more consistent with *ab initio* calculations. The explicit solute-solvent interactions contribute a more repulsive element to the interaction energy, thus pushing the two species apart and lowering the interaction energy.

When water donates proton (PA structures), the dimer energies are still too favorable based on the current model (with solute-solvent interactions), but we consider it reasonable given the lack of lone pairs and anisotropic polarizability on oxygen for the current CHEQ-PSP force field. The same effect was observed with the Drude oscillator model without lone pairs due to limitations in the anisotropy of the electrostatic representation¹⁰¹. The study of Anisimov et al.¹⁶ extended the alcohol Drude model to include lone pairs on the oxygen atom and assigned anisotropic polarizability on the oxygen atom, yielding less favorable alcohol-water gas-phase interactions with better agreement with the *ab initio* values. The further use of pair specific LJ interactions also contributed to the improved agreement of water-alcohol dimer properties relative to the reference quantum mechanical data. Moreover, the impact of the conformation of alcohol-water cluster as well as the reference *ab initio* level of theory should also be noted¹⁰⁰. Previous studies for methanol-water dimers using the original CHEQ force field with standard combination rules for water-alcohol LJ interactions demonstrated good agreement with various *ab initio* calculations¹⁴.

C. Validity of the Approach and Nature of Pair-Specific Lennard-Jones Parameters

During the optimization process, 16 pair-specific LJ parameters are optimized to fit 20 properties, including hydration free energies of the four alcohol molecules and 8 gas-phase alcohol-water dimerization energies and 8 gas-phase alcohol-water hydrogen bond lengths. Even though the number of optimized parameters is smaller than that of reference properties, we applied the modified, sample size-corrected Akaike Information Criterion (AICc)¹⁰² to assess the risk of over-fitting. The method has been used similarly during parameterization and validation of the Drude oscillator pair-specific LJ parameter force field²⁴. We adopt the AICc equation used in that work^{24,102}:

$$AIC_c = 2k + n \ln \left(\frac{RSS}{n} \right) + \frac{2k(k+1)}{n-k-1} \quad (6)$$

where k is the number of free parameters in a model, RSS is the Residual Sum of Squares, and n is the number of observations. AICc assesses a penalty for introducing free parameters, k , into a model, and rewards higher quality of fit reflected in smaller RSS values. Of two AICc values for two statistical models evaluated against the same reference data, a lower AICc value suggests a better performing model. Here, we compare two models: the CHEQ-orig model uses the L-B combination rules for the solute-solvent nonbond interactions and no free parameter applies ($k = 0$); the CHEQ-PSP model has 16 pair-specific LJ parameters, $k = 16$, as stated above. Due to the ambiguity of rigorously accounting for the relative contributions to RSS of properties differing in magnitude and dimension²⁴, we compute AICc values for the three different classes of target properties. For the hydration free energies, 4 observations contribute to RSS (methanol, ethanol, propanol and butanol hydration free energy), giving $\text{AICc}^{\text{CHEQ-orig}} = 2.46$ and $\text{AICc}^{\text{CHEQ-PSP}} = -20.19$. Considering 8 vacuum alcohol-water dimerization energies, $\text{AICc}^{\text{CHEQ-orig}} = -1.64$ and $\text{AICc}^{\text{CHEQ-PSP}} = -34.08$. Based on 8 gas-phase alcohol-water hetero-dimer hydrogen bond lengths, we found $\text{AICc}^{\text{CHEQ-orig}} = -31.87$ and $\text{AICc}^{\text{CHEQ-PSP}} = -65.89$. Here, the RSS computed for gas-phase dimer properties are referenced to the MP2/aug-cc-pvDZ results (see Section III B for more details). All three types of properties reflect consistently improved performance for the model with optimized specific solute-solvent nonbond parameters (lower AICc values for the CHEQ-PSP model), suggesting that the model proposed in this work is balanced between the improvement in quality of fit to targeted reference data and the increased parametric flexibility upon introduction of free parameters. We note that the AICc test can be formulated in terms of the χ^2 measure¹⁰². However, since we are not able to ascribe an absolute variance to the predicted gas-phase dimer properties (energies and hydrogen bond lengths), we choose to apply the RSS version of the method.

Table I shows the final parameters. Figure 3 shows the final optimized parameters along with those derived from different published combination rules and those determined for the Drude oscillator alcohol force field¹⁶. We present expressions for several combining rules to which we compare the present results further below.

The arithmetic and geometric mean combining rules are:

$$\text{Arithmetic mean: } A_{ij} = \frac{(A_i + A_j)}{2} \quad (7)$$

$$\text{Geometric mean: } A_{ij} = \sqrt{A_i A_j} \quad (8)$$

where A_i represents generally the ϵ_i and $R_{\min,i}$ parameters.

$$\text{Waldman-Hagler rule}^{60}: \epsilon_{ij} = \sqrt{\epsilon_{ii} \epsilon_{jj}} \left(\frac{2 R_{\min,ii}^3 R_{\min,jj}^3}{R_{\min,ii}^6 + R_{\min,jj}^6} \right), R_{\min,ij} = \left(\frac{R_{\min,ii}^6 + R_{\min,jj}^6}{2} \right)^{1/6} \quad (9)$$

$$\text{Fender - Halsey harmonic - mean rule}^{103}: \epsilon_{ij} = \frac{2 \epsilon_{ii} \epsilon_{jj}}{(\epsilon_{ii} + \epsilon_{jj})} \quad (10)$$

$$\text{Halgren HHG mean rule}^{104}: \epsilon_{ij} = \frac{4 \epsilon_{ii} \epsilon_{jj}}{(\sqrt{\epsilon_{ii}} + \sqrt{\epsilon_{jj}})^2} \quad (11)$$

$$\text{Halgren cubic - mean rule}^{104}: R_{min,ij} = \frac{R_{min,ii}^3 + R_{min,jj}^3}{R_{min,ii}^2 + R_{min,jj}^2} \quad (12)$$

The top panel of Figure 3 shows the values of the Lennard-Jones well-depth, ϵ_{ij} , and the bottom panel the $R_{min,ij}$ values for the atom type pairs for which specific LJ interactions were determined in this study. Values using other combining rules are also shown.

We observe that the L-B combining rule, which applies a geometric mean rule for ϵ_{ij} and an arithmetic mean rule for $R_{min,ij}$, leads to parameters that are closest to those determined in this study. This suggests that the L-B rule is a good starting point for further parameterization, though the straightforward application of this combining rule does not yield optimum parameters. The differences between the parameters based on L-B rule and the final optimized values are on the order of 10 % (Table I compares the specific solute-solvent LJ parameters for the CHEQ-orig model and CHEQ-PSP model). Meanwhile we note the deviation of parameters determined in this study compared to other combining rules as well as the Drude values¹⁶. The latter differences may be caused by different water models (Drude SWM4-NDP water has different LJ parameter on oxygen and water hydrogen has no contribution to the nonbond interactions), different treatment of polarization effects, and different parameterization protocols. Interestingly, however, the size parameter, $R_{min,ij}$ determined for CHEQ-PSP and the Drude alcohol force field are surprisingly similar to one another, as well as closely related to the arithmetic mean combining rule.

D. Density

Densities are computed for 0.1 and 0.2 mole fraction aqueous methanol and ethanol solutions using the original CHEQ force field and the CHEQ-PSP of this work (Table IV). The densities are computed from the average system volume over ten nanoseconds of simulation with the first 0.25 nanoseconds taken as equilibration time; autocorrelation analysis is used to obtain statistically relevant averages and uncertainties. The addition of pair-specific LJ parameters shows little effect on densities of the dilute aqueous solutions studied as the density changes induced are smaller than the uncertainty in measurements. The methanol solution densities decrease slightly towards the experimental values compared to the results based on the original force field combination^{14,15}, and ethanol solutions show slightly higher densities with the solute-solvent interaction correction.

E. Dipole Moment Distributions

We further explore the effect of explicit solute-solvent LJ terms on condensed phase electrostatic environment by computing the dipole moment distributions for solute and water at 0.1 and 0.2 mole fraction as shown in Figure 4. Also shown for methanol and ethanol (first two rows) are the results using the original force field combination (labeled as *orig* in Figure 4); the third and fourth rows show results for propanol and butanol. The column on the left shows the alcohol dipole moment distributions, and that on the right those for water in the corresponding solution. For the four alcohols studied, the dipole moments of both alcohol and water shift to lower values with increasing alcohol concentration. This suggests a less polar environment for both the solutes and the solvent at higher concentrations, which

is consistent with the dipole distribution dependence on concentration reported for methanol-water¹⁴ and ethanol-water^{10,15}. The most probable values of dipole moment shift to lower values going from methanol to butanol; this is attributed to a greater influence of the larger, less polar alkyl groups. The change in water dipole distributions, on the other hand, is indistinguishable between different alcohol solutions. Addressing the effect of explicit non-bond interactions between solutes and water, the alcohol dipole moments are reduced with the CHEQ-PSP interaction model. This is because the solute-solvent interaction distances are effectively increased (as will be discussed in the next section), thus lowering the polarization response of the alcohol (and to a lesser extent the water); this is reflected in the lower dipole moment. The alcohol is affected to a greater degree as it is more polarizable, and thus more sensitive to external electrostatic field (in this case represented by the solvent field at short distances). This implies weaker electrostatic interactions involving alcohol molecules, in accord with the less favorable hydration energies and gas-phase dimerization energies.

F. Radial distribution functions

To probe the structure of alcohol aqueous solutions, we have computed $H_{\text{alcohol-O}_{\text{water}}}$, $O_{\text{alcohol-H}_{\text{water}}}$ and $O_{\text{alcohol-O}_{\text{water}}}$ radial distribution functions (RDFs) for each of the four alcohol aqueous solutions with mole fraction of 0.1 and 0.2 (Figure 5). The first peak of RDFs are consistently higher for higher concentrations, indicating a stronger hydrogen bonding environment with increasing alcohol mole fraction. No significant difference for the RDFs corresponding We also include the corresponding RDFs obtained from dynamics trajectories based on the original force fields (dashed curves in Figure 5, panels a, b, and c for methanol and ethanol solutions), the first peaks of which are higher than those using the new CHEQ-PSP model. The increase of the first solvation shell of alcohol-water hydrogen bonding suggests stronger solute-solvent interaction when the specific solute-solvent LJ terms are not applied. Furthermore, the distance of the first-peak maximum has increased using the current CHEQ-PSP force fields; this is consistent with increased water-alcohol gas-phase dimer bond lengths (Table II).

To compare with previous neutron diffraction study¹⁰⁵ on a 1:9 molar ratio methanol-water mixture, we also computed RDF's for methanol carbon and water oxygen, C-Ow, (not shown here). The location of the first peak based on neutron diffraction data is 3.7 Å, corresponding to approximately 10 water molecules around the methyl group¹⁰⁵. The C-Ow RDF based on the original CHEQ model shows the first peak at 3.45 Å with 15.5 water molecules in the first solvation shell; the number of water molecule is slightly reduced to 15.0 with the same first peak position when pair-specific parameters are applied. This suggests that the less favorable solute-solvent interactions in the CHEQ-PSP force field improve the agreement of RDF's for this pair type with experiment.

G. Hydrogen Bond Pattern

In previous work^{15,106}, we investigated the nature of water hydrogen bonding in dilute methanol and ethanol solutions by computing excess hydrogen bond patterns and profiles. We present a similar analysis here to determine the extent of the changes to hydrogen bonding patterns arising from the corrected interaction model. In order to compare with the hydrogen bonding results based on the original force field combination, we adopt the same definition of a hydrogen bond: two molecules are considered to be hydrogen bonded when their H...O distance is less than 2.4 Å and the O-H...O angle is larger than 150°. As suggested by Noskov et al.¹⁰, we apply the same criteria for all alcohol mixtures analyzed. The excess hydrogen bond is defined as the deviation of hydrogen bond number in solution from the value predicted for the ideal, noninteracting alcohol-water mixture. As used by

Noskov et al.¹⁰ the ideal number of H-bonds per molecule is determined by a linear relation between the ideal number of hydrogen bonds, n_{Hb}^{id} , and alcohol concentration, X_{alco} :

$$n_{Hb}^{id} = n_{Hb}^{wat} - X_{alco} \times (n_{Hb}^{wat} - n_{Hb}^{alco}) \quad (13)$$

where $n_{Hb}^{wat} = 3.04$ is the average hydrogen bond per molecule in pure water; $n_{Hb}^{alco} = 2.58$ for methanol and 2.46 for ethanol, representing the average number of hydrogen bonds in corresponding liquid alcohol. We use identical n_{Hb}^w and n_{Hb}^a values for both force fields since the explicit solute-solvent LJ parameters added in the new model only affect cross interactions in aqueous solutions.

To visualize water hydrogen bonding patterns near alcohol molecules, we examine the H-bond excess of water molecules in a slice (0.5 Å thick) of the C-O-H plane of the reference methanol and the C2-C1-O plane of the centered ethanol up to a distance of 5 Å (6 Å for ethanol) from the carbon atom (the ethylene carbon C1 for ethanol) at $X_{alco} = 0.1$ (see Figure 6). We observe two positive hydrogen bond excess regions for water molecules around the hydroxyl group (red regions in Figure 6), corresponding to the two hydrogen bonding positions of the centered alcohol molecule when water acts as a proton donor and acceptor. Surrounding the alkyl group of alcohol molecules, significant amount of water molecules form less hydrogen bonds than the ideal average bulk hydrogen number n_{Hb}^{id} . As the separation between water and alcohol increases, the effect from the polar or nonpolar group of the central alcohol molecule diminishes and the hydrogen bonding environment becomes more homogeneous and bulk-like. Similar hydrogen patterns are found in our previous studies for molecular dynamics trajectories based on the force fields without solute-solvent LJ terms^{15,106} Qualitatively, water molecules in the methanol solution exhibit less H-bond excess than the ones in ethanol-water mixtures (more blue in the upper panel of Figure 6).

For further quantitative analysis of the impact of the solute-solvent interaction correction on hydrogen bonding patterns, we consider the excess number of water H-bonds at given radial distance R from the nearest methanol/ethanol C or O atom as¹⁰:

$$N_{Hb}(R) = \int_0^R n_{Hb}^{excess}(r) dr = \int_0^R [n_{Hb}(r) - n_{Hb}^{id} \times n_{wat}(r)] dr \quad (14)$$

where $n_{Hb}(r)$ is the total number of H-bonds for water molecules whose oxygen atom is in the shell of radius r with a thickness of dr , and $n_{wat}(r)$ is the number of water in the shell.

The excess H-bonds $n_{Hb}^{excess}(r)$ and the running integral $N_{Hb}(R)$ in Figure 7(for methanol) and Figure 8(for ethanol) are calculated from the dynamics trajectories of $X_{alco} = 0.1$ and 0.2 solutions based on the CHEQ force fields with and without solute-solvent interaction corrections. The dotted vertical lines represent the positions of the first peak and the minimum between the first and second coordination shells for the corresponding RDFs. As illustrated in Figure 6, negative H-bond excess only appears within the first hydration shell of C_{alco} (second vertical lines of panel a and c in Figure 7 and Figure 8). On the other hand, the dominant positive contribution of hydrogen bond excess comes from water molecules with smaller $R_{O_{alco}-O_{wat}}$ than the position of the $O_{alco}-O_{wat}$ RDFs first peak (first vertical lines of panels b and d in Figure 7 and Figure 8), confirming the preference of water hydrogen bonding when the alcohol hydroxyl group is near. With regard to hydrophobic hydration, the corrected force field, nominally capturing hydration free energetics more

faithfully compared to the original model combination, still finds no evidence of H-bond excess near the nonpolar moieties as the "iceberg" model¹⁰⁷. The total positive H-bonding excesses ($N_{Hb}(R)$) predicted by the original model decrease and even become negative in methanol aqueous solutions, because the chance of a water hydrogen bonding to an alcohol molecule is reduced by the modified solute-solvent parameters in the current work as suggested by the smaller first peaks of RDFs in Figure 5. Consistent with a previous study tangentially exploring the bulk region of methanol-water interfacial systems¹⁰⁶ as well as bulk aqueous methanol solutions using the original model (Figure 7 panel c d), we find higher excess in higher methanol concentration solution ($X_{alco} = 0.2$) for the current force field (Figure 7 panel a b). For ethanol-water mixtures (Figure 8), the same concentration dependence is observed for both the original force field (ref¹⁵ and Figure 8 panel c d) and the CHEQ-PSP model (Figure 8 panel a b).

Meanwhile, in accordance with the higher first hydration peak in Figure 5, the current force field predicts a higher H-bond excess for waters vicinal to ethanol rather than those near methanol, suggesting an environment more supportive to hydrogen-bonding in ethanol-water mixtures. This reflects the stronger ethanol-water interaction in both the proton accepting and donating configurations compared to the analogous methanol-water interactions (Table III).

The influence of local electrostatics are connected to the hydrogen bonding pattern by computing the water average dipole moment as a function of the the distance between water oxygen and the closest alcohol oxygen ($R_{O_{alco}-O_{wat}}$) at $X_{alco}=0.1$ and $X_{alco}=0.2$ as shown in Figure 9 (Panels a and b show results of this work for methanol and ethanol, respectively; panels c and d show results using the original force fields without specific solute-solvent interactions). As elaborated in previous work on ethanol-water solutions¹⁵, water molecules are found to have higher dipole moment within the first hydration shell of alcohol oxygen where the positive excess is observed in Figure 7 and Figure 8. The minimum in the dipole moment profile coincides with the position of most negative H-bond excess, and the average water dipole moment gradually increases with the $O_{alco} - O_{wat}$ separation, corresponding to the second positive excess region in panels b and d of Figure 7 and Figure 8.

Figure 7 and Figure 8 lower panels show more negative excess minimum values at $X_{alco}=0.2$ than the ones at $X_{alco}=0.1$, consistent with the smaller minimum dipole moments for methanol and ethanol (Figure 9) in more concentrated solutions. Based on Figure 7 and Figure 8, there is very little difference in the dipole moment profiles between the original and CHEQ-PSP models (as also seen in Figure 4).

H. Hydrogen-Bond Mediated Clustering

In a previous study, we showed that ethanol-water mixtures displayed similar clustering/percolation behavior macroscopically as other studies based on the CHEQ polarizable model¹⁵. Here we adopt the same method to analyze molecular segregation in available concentrations of methanol and ethanol aqueous solutions modeled with the CHEQ-PSP force field and compare to the initial model without solute-solvent LJ corrections. This analysis has bearing on the discussion of the Kirkwood-Buff results to be addressed in the following section.

Clustering statistics were analyzed using a custom FORTRAN program for analyzing molecular segregation in different trajectories based on an algorithm adopted from the work of Geiger et al.¹⁰⁸ In brief, we consider a molecule as belonging to a cluster when it is hydrogen bonded to any member of that cluster ($d_{OH} < 2.4$ and $\angle OH \cdots O > 150.0^\circ$).

Table V shows the percentages of alcohol or water molecules present in clusters based on the model with and without corrections. Both force fields suggest that at small concentrations ($X_{alco} = 0.1$ and $X_{alco} = 0.2$), the majority of water molecules are present in a water hydrogen-bond network and the fraction of alcohol molecules forming hydrogen bonding clusters increases with concentration. At a given concentration, the CHEQ-PSP force field allows more alcohol molecules to aggregate in clusters because alcohol-water interactions are reduced. On the other hand, specific solute-solvent LJ terms added suggest very little negative effect on water clustering which may be caused by the increased alcohol clusters perturbing the water percolation structure.

To assess clustering microscopically, Figure 10 displays the distribution of cluster size for alcohol and water molecules. Solid lines without symbols represent the methanol CHEQ-PSP force field at $X_a = 0.1$, solid lines with open circles the methanol CHEQ-PSP at $X_a = 0.2$, dashed lines with open squares the original force field at $X_a = 0.1$, and dashed lines with open diamonds the original force field at $X_a = 0.2$. Similar to the observation for ethanol-water mixtures in previous work¹⁵, the probability of finding a certain size of methanol or ethanol cluster decreases as the cluster size increases; water percolation is observed as the for the dilute solutions ($X_{alco}=0.1$ and 0.2) studied. Under the same concentration, alcohol molecules are more likely to be present in larger clusters using the CHEQ-PSP force field, which is consistent with the higher percentage of alcohol molecules as members of clusters (Table V). Water cluster distributions are indistinguishable between the two force fields and water percolation is observed in the high probability of large water clusters (narrow distributions on the right side of Figure 10). This observation is consistent with the neutron diffraction experiments results, which suggests that water percolates when $X_{meth} < 0.54$ and methanol percolates when $X_{meth} > 0.27$ ¹⁸.

I. Kirkwood-Buff (KB) Integrals

Kirkwood-Buff (KB) theory, first developed in 1951⁴⁷, has been used extensively to analyze macroscopic thermodynamic properties of solutions since Ben-Naim applied the KB integral(G_{ij}):

$$G_{ij} = \int_0^\infty (g_{ij}^{\mu VT} - 1) 4\pi r^2 dr \quad (15)$$

for binary solution mixtures¹⁰⁹. In Eq 15, g_{ij} is the center-of-mass based RDF between species i and j . Specifically, the integral quantifies the change in the distribution of "j" molecules surrounding a central "i" molecule from a reference system in which the "j" molecules are randomly distributed in an equal volume element of bulk solution¹¹⁰. From a practical standpoint, a related caveat pertains to the sensitivity of KB integrals to small deviations from a bulk-like distribution due to the r^2 weighting factor arising from the volume element Jacobian. Furthermore, the RDF in the Kirkwood-Buff context is interpreted as the potential of mean force (PMF), $W_{ij}(r) = -k_B T \ln(g_{ij}(r))$, since the distribution function corresponds to averaging over all configurations of all other "i" and "j" molecules excluding the i - j pair under consideration. The KB integral can be used to examine solution behavior including the *local* composition, preferential solvation, and microstructure^{111–116}. The connection of atomistic structural properties to macroscopic thermodynamics is one of the most powerful capabilities of this theoretical formalism. An important quantity in the KB approach is the excess coordination number $N_{ij} = \rho_j G_{ij}$, where ρ_j is the density of species "j" in the mixture. N_{ij} provides information about the local change of "i" and "j" molecule distributions about an "i" molecule upon introducing an "i" molecule. Positive N_{ij} indicates a locally enhanced density of "j" molecules around an "i"

molecule relative to that in the bulk environment (i.e., at large separations). This can be interpreted as favorable interaction between the two species¹¹⁰.

Recent extensions of Kirkwood-Buff integrals to force field development have been demonstrated in extensive work by P. E. Smith and coworkers over the last decade^{71–76}. Kirkwood-Buff force fields have been developed for a wide variety of small molecules taken as analogues for peptides and proteins. We follow in the present work much of the protocol outlined in earlier work by Smith and coworkers; our intent is to assess the ability of the current charge equilibration force field for alcohols to reproduce experimentally-derived Kirkwood-Buff integrals and excess coordination numbers. We emphasize that the present CHEQ model has not been parameterized to reproduce experimental KB integrals or excess coordination numbers, which themselves appear to suffer from non-trivial uncertainties arising from differences in experimental technique used to determine them or to numerical analysis of experimental thermodynamic and scattering (SAXS¹¹⁷) data.

A KB analysis for the experimental data on methanol-water mixtures was reported by Weerasinghe et al⁷⁵ who developed a force field to reproduce the experimental KB integrals. To apply the KB theory on simulation data, the following assumption was invoked for the KB integral:

$$G_{ij} = \int_0^\infty (g_{ij}^{\mu VT} - 1) 4\pi r^2 dr \approx \int_0^R (g_{ij}^{NPT} - 1) 4\pi r^2 dr \quad (16)$$

where R is a cutoff value where the RDFs become essentially unity. By evaluating the integral up to various final R values, we can compute the reported KB integrals as averages between values of R from 15 Å to 20 Å (15 to 19.8 Å for methanol-water mixtures at 0.1 mole fraction); this is the protocol defined by Smith and coworkers⁷⁵. Here we use 20 Å as the cutoff value R in Eq(16), which is approximately one-half the cubic system cell size. Figure 11 displays center of mass RDFs (solid lines for $X_a = 0.1$, dashed curves for $X_a = 0.2$).

The RDFs show very little differences between the two alcohol mole fractions studied. For both methanol concentrations, the RDFs suggest strong water-water and alcohol-alcohol interactions; for ethanol solutions, the alcohol-water and alcohol-alcohol interactions appear to be more balanced. Figure 12 shows the corresponding potentials of mean force providing further insight into the energetics of association in the mixtures. For methanol solutions, the self-interaction PMFs (panels a and c of Figure 12) show much lower free energy wells (at least of factor 2 times more favorable) for the self-interactions relative to the methanol-water interaction comprised of two almost equally favorable free energy minima (Figure 12 panel b). For ethanol solutions, the ethanol-ethanol and ethanol-water interactions in condensed phase are almost balanced (energy well of -0.25 kcal/mol for both cases). In fact, both PMFs show a broad well of -0.25 kcal/mole out to about 6 Å separation. In all cases, the water-water self interaction shows a deep well of -0.75 to -0.85 kcal/mole with relatively no barrier to association. The alcohol-alcohol energetics show less deeper wells, with slightly larger barriers to association, though the magnitudes of these are sufficiently overcome by thermal fluctuations. The relative energetics of interaction between the methanol and ethanol solutions also is reflected in the hydrogen bonding patterns discussed above. Specifically, we noted that the ethanol-water solutions accommodate greater hydrogen bonding for water relative to the methanol-water solutions. This in part arises from the more balanced ethanol-ethanol and ethanol-water association energetics.

Figure 13 shows the KB integrals after integration of the RDFs to the cutoff distance R . All the RDFs are essentially unity beyond 10 Å but the KB integrals converge more slowly because small fluctuations in the RDF about unity are magnified by the volume factor.

Table VI shows the final KB integrals; these are calculated by averaging values between 15 and 20 Å (15 to 19.8 Å for methanol solution at 0.1 mole fraction) for the CHEQ-PSP force field. As suggested in previous studies^{75,118}, excess coordination numbers N_{ij} are computed to damp inherent fluctuations in G_{ij} integrals at low concentrations (Table VI). For methanol-water mixtures, the CHEQ-PSP force field predicts the N_{ww} value for methanol-water mixtures in good agreement with the experiment⁷⁵. The experimental N_{mw} is well reproduced for the 0.1 mole fraction solution, but slightly overestimated by the current force field for $X_{meth} = 0.2$. It is not surprising to find that N_{mm} deviates the most from the experimental values, especially for the higher concentration, because the methanol-methanol interaction in the original CHEQ model was found to be more favorable, leading to an overestimation of bulk methanol density¹¹. Meanwhile, less methanol molecules at low concentrations can cause higher uncertainties for RDFs and introduce error in the excess number as well as KB integral computed. We observe that Weerasinghe's KB-derived force field produces lower solution densities for low methanol mole fractions than the experimental densities, suggesting that molecular interactions (solute-solvent and solute-solute) are underestimated by the force field. Experimental N_{mw} ⁷⁵ increases roughly from -1.5 to -1.2 for methanol mole fraction 0.1 to 0.2 and such concentration dependence is qualitatively reproduced in the simulated N_{mw} . Although we observe different trend in experimental data for N_{mm} , the KB integral concentration dependence from small-angle X-ray scattering (SAXS)¹¹⁷ is reproduced. Our predicted G_{mw} and G_{ww} agree well with SAXS data while G_{mm} is less negative, which can be attributed to the overly favorable methanol-methanol interaction and larger solvation shell in RDFs as discussed above. For ethanol aqueous solutions, the KB integrals from SAXS¹¹⁹ are semi-quantitatively reproduced by the CHEQ-PSP model and experimental concentration dependence of KB integrals is well matched with the exception of G_{ew} . The less negative G_{ew} at higher ethanol concentration may be attributed to the stronger solute-solvent interaction of the force field suggested by the gas-phase interactions (Table III) and hydration free energy (Table II) giving rise to greater solvation (i.e. greater positive area under peaks) in the RDFs.

We note in passing that by using the computed KB integrals, we are able to determine other solution properties including activity derivatives¹²⁰. We find for the activity derivative,

$$f_{cc} = \left(\frac{\partial \ln f_c}{\partial \ln x_c} \right)_{p,T} = - \frac{x_c \rho_w \Delta G}{1 + x_c \rho_w \Delta G} \quad (17)$$

where $\Delta G = G_{cc} + G_{ww} - 2G_{cw}$. The activity derivatives for methanol-water mixture at 0.1 and 0.2 mole fraction, -0.04 and -0.17, well reproduce the experimental data¹²¹ and agree with the KB force field of Weerasinghe et al⁷⁵.

Summarizing, we observe that despite capturing the solute-solute and solvent-solvent interaction energetics accurately (to the extent that the pure liquid properties of solute and solvent reproduce well a series of experimental pure liquid condensed phase data), along with the solute-solvent interaction in condensed phase as demonstrated by reasonably good agreement of hydration free energies of solute in solvent, without explicitly fitting to KB integrals (and excess coordination numbers), we are not able to quantitatively and broadly reproduce the necessary local microstructure and its concentration dependence underlying observed macroscopic thermodynamics. Whether this is due to a missing physics in the models or the fact that the KB integrals were not included in the fitting protocol remains an

outstanding question and studies continue to address this question. We note that such fitting represents a daunting challenge, but the results would prove quite valuable. We finally note that the observed systematic deviations of computed KB integrals from experiment are self-consistently explained by the enhanced microscopic structure (vis-a-vis, clustering) resulting from the simulations.

We note that KB integrals are not routinely computed using most force fields; thus, an exhaustive comparison is not possible at this time. However, we reiterate that though we see qualitative agreement with experimental data relating to excess coordination numbers and KB integrals (as well as the concentration dependence of these properties), the force field presented in this work was not explicitly parameterized to reproduce these values quantitatively. This is unlike the KB integral based force fields derived in the work of Smith and coworkers over the years, which have demonstrated the successful application of KB integrals in force field development for biological systems, particularly with the perspective of chemical functional groups acting as proxies for biological building blocks (amino acids, nucleic acids) being solutes in aqueous solutions^{71–76}.

IV. SUMMARY AND CONCLUSIONS

We present a modified charge equilibration force field for short-chain linear alcohols extending the existing CHARMM charge equilibration force field for alcohols^{11,12}. We add explicit solute-solvent Lennard-Jones interactions between atoms of the alcohol and solvent represented here with the TIP4-FQ water model. The interactions are added to improve the hydration free energies of methanol, ethanol, propanol, and butanol in water. Hydration free energy quantifies the effects of molecular interactions and serves as a useful property for parameter verification. Addition of solute-solvent LJ terms effectively improves the agreement of gas-phase alcohol-water dimer interactions and hydrogen bond lengths with select *ab initio* results. The resulting weaker solute-solvent interaction also influences solution densities, dipole distributions, RDFs, and hydrogen bonding patterns though no qualitative changes are found for such solution properties. This is a satisfying result in that we can improve the description of hydration free energetics without dramatically perturbing existing solution properties. We applied Kirkwood-Buff analysis to molecular dynamics trajectories of 0.1 and 0.2 mole fraction aqueous methanol and ethanol solutions using the modified force field and achieved semi-quantitative agreement of KB integrals and excess coordination numbers, and composition dependence, with experimental values with the exception of the more concentrated ethanol solution.

Acknowledgments

The authors gratefully acknowledge financial support for this work from the National Institutes of Health (COBRE, 5P20RR017716-07) awarded to the Department of Chemistry and Biochemistry at the University of Delaware.

References

1. Wensink EJW, Hoffmann AC, van Maaren PJ, van der Spoel D. J. Chem. Phys. 2003; 119:7308.
2. Jorgensen WL. J. Phys. Chem. 1986; 90:1276.
3. Allinger NL, Chen KH, Lii JH, Durkin KA. J. Comp. Chem. 2003; 24:1447. [PubMed: 12868110]
4. Jorgensen WL, Maxwell DS, Tirado-Rives J. J. Am. Chem. Soc. 1996; 118:11225.
5. Gonzalez MA, Bermejo FJ, Enciso E, Cabrillo C. Philos. Mag. 2004; 84:1599.
6. Saiz L, Padro JA, Guardia E. J. Phys. Chem. B. 1997; 101:78.
7. Gao J, Habibollahzadeh D, Shao L. J. Phys. Chem. 1995; 99:16460.
8. Gonzalez MA, Enciso E, Bermejo FJ, Bee M. J. Chem. Phys. 1999; 110:8045.
9. Caldwell JW, Kollman PA. J. Phys. Chem. 1995; 99:6208.

10. Noskov SY, Lamoureux G, Roux B. *J. Phys. Chem. B.* 2005; 109:6705. [PubMed: 16851754]
11. Patel S, Brooks CL III. *J. Chem. Phys.* 2005; 122:024508. [PubMed: 15638599]
12. Patel S, Brooks CL III. *J. Chem. Phys.* 2005; 123:164502. [PubMed: 16268707]
13. Yu H, Geerke DP, Liu H, van Gunsteren W. *J. Comp. Chem.* 2006; 27:1494. [PubMed: 16838298]
14. Zhong Y, Warren GL, Patel S. *J. Comp. Chem.* 2008; 29:1142. [PubMed: 18074339]
15. Zhong Y, Patel S. *J. Phys. Chem. B.* 2009; 113:767. [PubMed: 19115819]
16. Anisimov VM, Vorobyov IV, Roux B, MacKerell AD Jr. *J. Chem. Theory Comput.* 2007; 3:1927.
17. Dixit S, Crain J, Poon WCK, Finney JL, Soper AK. *Nature.* 2002; 416:829. [PubMed: 11976678]
18. Dougan L, Bates SP, Hargreaves R, Fox JP. *J. Chem. Phys.* 2004; 121:6456. [PubMed: 15446945]
19. Allison SK, Fox JP, Hargreaves R, Bates SP. *Phys. Rev. B.* 2005; 71:024201.
20. Roney AB, Space B, Castner EW, Napoleon RL, Moore PB. *J. Phys. Chem. B.* 2004; 108:7389.
21. Fidle J, Rodger PM. *J. Phys. Chem. B.* 1999; 103:7695.
22. Kusalik PG, Lyubartsev AP, Bergman DL, Laaksonen A. *J. Phys. Chem.* 2000; 1104:9533.
23. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. *J. Comp. Chem.* 2004; 25:1656. [PubMed: 15264259]
24. Baker CM, Lopes PEM, Zhu X, Roux B, MacKerell AD Jr. *J. Chem. Theory Comput.* 2010; 4:1181.
25. Shirts MR, Pitera JW, Swope WC, Pande VS. *J. Chem. Phys.* 2003; 119:5740.
26. MacKerell AD Jr. et al. *J. Phys. Chem. B.* 1998; 102:3586.
27. Cornell WD, Cieplak P, Bayly CR, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. *J. Am. Chem. Soc.* 1995; 117:5179.
28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J. Chem. Phys.* 1983; 79:926.
29. Shirts MR, Pande VS. *J. Chem. Phys.* 2005; 122:134508. [PubMed: 15847482]
30. Berweger CD, van Gunsteren WF, Muller-Plathe F. *Chem. Phys. Lett.* 1995; 232:429.
31. Berendsen HJC, Grigera JR, Straatsma TP. *J. Phys. Chem. B.* 1987; 91:6269.
32. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. *J. Chem. Phys.* 2004; 120:9665. [PubMed: 15267980]
33. Sun Y, Kollman PA. *J. Comp. Chem.* 1995; 16:1164.
34. Deng Y, Roux B. *J. Phys. Chem. B.* 2004; 108:16567.
35. Beglov D, Roux B. *J. Chem. Phys.* 1994; 100:9050.
36. Kelly CP, Cramer CJ, Truhlar DG. *J. Chem. Theory Comput.* 2005; 1:1133.
37. Wang J, Cieplak P, Kollman PA. *J. Comp. Chem.* 2000; 21:1049.
38. Rizzo RC, Jorgensen WL. *J. Am. Chem. Soc.* 1999; 121:4827.
39. Price ML, Ostrovsky D, Jorgensen WL. *J. Comp. Chem.* 2001; 22:1340.
40. Hess B, van der Vegt NFA. *J. Phys. Chem. B.* 2006; 110:17616. [PubMed: 16942107]
41. Mobley DL, Dumont E, Chodera JD, Dill KA. *J. Phys. Chem. B.* 2007; 111:2242. [PubMed: 17291029]
42. Wang J, Wolf WR, Caldwell JW, Kollman PA, Case DA. *J. Comp. Chem.* 2004; 25:1157. [PubMed: 15116359]
43. Mobley DL, Christopher IB, Cooper MD, Shirts MR, Dill KA. *J. Chem. Theory Comput.* 2009; 5:350.
44. Lamoureux G, Harder E, Vorobyov IV, Roux B, MacKerell AD Jr. *Chem. Phys. Lett.* 2006; 418:245.
45. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. *J. Chem. Phys.* 1983; 4:187.
46. Weeks JD, Chandler D, Andersen HC. *J. Chem. Phys.* 1971; 54:5237.
47. Kirkwood JG, Buff FP. *J. Chem. Phys.* 1951; 19:774.
48. Rick SW, Stuart SJ, Berne BJ. *J. Chem. Phys.* 1994; 101:6141.
49. Patel S, Brooks CL III. *J. Comp. Chem.* 2004; 25:1. [PubMed: 14634989]
50. Patel S, MacKerell AD Jr, Brooks CL III. *J. Comp. Chem.* 2004; 25:1504. [PubMed: 15224394]

51. Patel S, Brooks CL III. *Mol Simul.* 2004; 32:231.
52. Rick, SW.; Stuart, SJ. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations.. In: Lipkowitz, KB.; Boyd, DB., editors. *Reviews in Computational Chemistry.* Wiley; New York: 2002. p. 89
53. Rick SW, Stuart SJ, Bader JS, Berne BJ. *J. Mol. Liq.* 1995; 31:65.
54. Tsuzuki S, Uchimaru T, Matsumura K, Mikami M, Tanabe K. *J. Chem. Phys.* 1999; 110:11906.
55. Stockman PA, Blake GA, Lovas FJ, Suenram RD. *J. Chem. Phys.* 1997; 107:3782.
56. Parr, RG.; Yang, W. *Density-Functional Theory of Atoms and Molecules.* Oxford University Press; 1989.
57. Rappe AK, Goddard WA III. *J. Phys. Chem.* 1991; 95:3358–3363.
58. Car R, Parrinello M. *Phys. Rev. Lett.* 1985; 55:2471. [PubMed: 10032153]
59. Kong CL. *J. Chem. Phys.* 1973; 59:2464.
60. Waldman M, Hagler AT. *J. Comp. Chem.* 1993; 14:1077.
61. Delhommelle J, Millie P. *Mol. Phys.* 2001; 99:619–625.
62. Song W, Rossky PJ, Maroncelli M. *J. Chem. Phys.* 2003; 119:9145–9162.
63. Warren GL, Patel S. *J. Chem. Phys.* 2007; 127:17705614.
64. Jorgensen WL, Ravimohan C. *J. Chem. Phys.* 1985; 83:3050–3054.
65. Zacharias M, Straatsma TP, McCammon JA. *J. Chem. Phys.* 1994; 100:9025.
66. Steinbach PJ, Brooks BR. *J. Comp. Chem.* 1994; 15:667.
67. Davis JE, Warren GL, Patel S. *J. Phys. Chem. B.* 2008; 112:8298–8310. [PubMed: 18570394]
68. Allen, MP.; Tildesley, DJ. *Computer Simulation of Liquids.* Clarendon Press; 1987.
69. Nose S. *Mol. Phys.* 1984; 52:255.
70. Darden T, York D, Pedersen L. *J. Chem. Phys.* 1993; 98:10089.
71. Weerasinghe S, Smith PE. *J. Phys. Chem. B.* 2003; 107:3891.
72. Weerasinghe S, Smith PE. *J. Chem. Phys.* 2003; 118:10663.
73. Weerasinghe S, Smith PE. *J. Chem. Phys.* 2003; 119:11342.
74. Weerasinghe S, Smith PE. *J. Chem. Phys.* 2004; 121:2180. [PubMed: 15260772]
75. Weerasinghe S, Smith PE. *J. Phys. Chem. B.* 2005; 109:15080. [PubMed: 16852908]
76. Kang M, Smith PE. *J. Comp. Chem.* 2006; 27:1477. [PubMed: 16823811]
77. Nistor RA, Polihronov JG, Muser MH, Mosey NJ. *J. Chem. Phys.* 2006; 125:094108. [PubMed: 16965073]
78. Chelli R, Pagliai M, Procacci P, Cardini G, Schettino V. *J. Chem. Phys.* 2005; 122:074504. [PubMed: 15743251]
79. Chelli R, Procacci P. *J. Chem. Phys.* 2002; 117:9175.
80. Chen J, Martinez TJ. *Chem. Phys. Lett.* 2007; 438:315.
81. Warren GL, Davis JE, Patel S. *J. Chem. Phys.* 2008; 128:144110. [PubMed: 18412426]
82. Shimizu K, Chaimovich H, Farah JPS, Dias LG, Bostick DL. *J. Phys. Chem. B.* 2004; 108:4171.
83. Bennett CH. *J. Chem. Phys.* 1976; 22:245.
84. Shirts MR, Pande VS. *J. Chem. Phys.* 2005; 122:144107. [PubMed: 15847516]
85. Lamoureux G, MacKerell AD Jr. Roux B. *J. Chem. Phys.* 2003; 119:9851.
86. Vorobyov IV, Anisimov VM, MacKerell AD Jr. *J. Phys. Chem. B.* 2005; 109:18988. [PubMed: 16853445]
87. Lopes PE, Lamoureux G, Roux B, MacKerell AD Jr. *J. Phys. Chem. B.* 2007; 111:2873. [PubMed: 17388420]
88. Cieplak P, Caldwell J, Kollman PA. *J. Comp. Chem.* 2001; 22:1048.
89. Caldwell J, Dang LX, Kollman PA. *J. Am. Chem. Soc.* 1990; 112:9144.
90. Dang LX, Rice JE, Caldwell J, Kollman PA. *J. Am. Chem. Soc.* 1991; 113:2481.
91. Caldwell J, Kollman PA. *J. Am. Chem. Soc.* 1995; 117:4177.
92. Kaminski GA, Stern HA, Berne BJ, Friesner RA. *J. Phys. Chem. A.* 2004; 108:621–627.
93. Schropp B, Tavan P. *J. Phys. Chem. B.* 2008; 112:6233–6240. [PubMed: 18198859]

94. Botek E, Giribet C, de Azua MR, Negri RM, Bernik D. J. Phys. Chem. A. 2008; 112:6992–6998. [PubMed: 18593142]
95. Krishtal A, Senet P, Yang M, van Alsenoy C. J. Chem. Phys. 2006; 125:034312.
96. Krishtal A, Senet P, van Alsenoy C. J. Chem. Thoery Comput. 2008; 4:426–434.
97. Bultinck P, van Alsenoy C, Ayers P, Carbo-Dorca R. J. Chem. Phys. 2007; 126:144111. [PubMed: 17444705]
98. Chen B, Xing J, Siepmann JI. J. Phys. Chem. B. 2000; 104:2391.
99. Bauer B, Patel S. J. Chem. Phys. 2009
100. Fileti EE, Chaudhuri P, Canuto S. Chem. Phys. Lett. 2004; 400:494.
101. Harder E, Anisimov VM, Vorobyov IV, Lopes PEM, Noskov SY, MacKerell AD Jr. Roux B. J. Chem. Thoery Comput. 2006; 2:1587.
102. Akaike H. J. Econometrics. 1981; 16:3.
103. Fender BEF, Halsey GD Jr. J. Chem. Phys. 1962; 36:1881.
104. Halgren TA. J. Am. Chem. Soc. 1992; 114:7827.
105. Soper AK, Finney JL. Phys. Rev. Lett. 1993; 71:4346. [PubMed: 10055223]
106. Patel S, Zhong Y, Bauer BA, Davis JE. J. Phys. Chem. B. 2009; 113:9241. [PubMed: 19569724]
107. Frank HS, Evans MW. J. Chem. Phys. 1945; 13:507.
108. Geiger A, Stanley HE. Phys. Rev. Lett. 1982; 49:1895.
109. Ben-Naim A. J. Chem. Phys. 1977; 67:4884.
110. Pierce V, Kang M, Aburi M, Weerasinghe S, Smith PE. Cell Biochemistry and Biophysics. 2008; 50:1–22. [PubMed: 18043873]
111. Ben-Naim A. Pure Appl. Chem. 1990; 62:25.
112. Pfund DM, Lee LL, Cochran HD. Fluid Phase Equilib. 1988; 39:161.
113. Rubio RG, Prolongo MG, Diaz Pena M, Renuncio JAR. J. Phys. Chem. 1987; 91:1177.
114. Ben-Naim A. Cell Biophys. 1988; 12:3694.
115. Matteoli, E.; Mansoori, GA. Fluctuation Theory of Mixtures. Tylor & Francis; New York: 1990.
116. Newman KE. Chem. Soc. Rev. 1994; 23:31.
117. Donkersloot MCA. J. Solution Chem. 1979; 8:293.
118. Shulgin I, Ruckenstein E. J. Phys. Chem. B. 1999; 103:2496.
119. Nishikawa K, Iijima T. J. Phys. Chem. 1993; 97:10824.
120. Ben-Naim, A. Statistical Thermodynamics for Chemists and Biochemists. Plenum Press; New York: 1992.
121. Butler JAV, Thomson DW, MacLennan WH. J. Chem. Soc. 1933:674.
122. Washburn, EW., editor. International Critical Tables of Numerical Data, Physics, Chemistry and Technology. Knovel; New York: 2003.

FIG. 1.
Atom notations for alcohol and water studied in this work.

FIG. 2.
Optimization process for atom-pair specific LJ nonbond parameters.

FIG. 3.

Lennard-Jones parameters based on different combination rules.

a: $A_{ij} = (A_i + A_j)/2$, $A = \epsilon$ in the upper panel and $A = R_{min}$ in the lower panel.

b: $A_{ij} = \sqrt{A_i A_j}$, $A = \epsilon$ in the upper panel and $A = R_{min}$ in the lower panel.

c: $\epsilon_{ij} = \sqrt{\epsilon_{ii} \epsilon_{jj}} \frac{2Rmin_{ii}^3 Rmin_{jj}^3}{Rmin_{ii}^6 + Rmin_{jj}^6}$, ref⁶⁰.

d: $\epsilon_{ij} = 2\epsilon_{ii}\epsilon_{jj}/(\epsilon_{ii} + \epsilon_{jj})$, ref¹⁰³.

e: $\epsilon_{ij} = 4\epsilon_{ii}\epsilon_{jj}/(\sqrt{\epsilon_{ii}} + \sqrt{\epsilon_{jj}})^2$, ref¹⁰⁴.

f: ref¹⁶.

g: this work.

h: $Rmin_{ij} = \left(\frac{Rmin_{ii}^6 + Rmin_{jj}^6}{2} \right)^{1/6}$, ref⁶⁰.

i: $Rmin_{ij} = \frac{Rmin_{ii}^3 + Rmin_{jj}^3}{Rmin_{ii}^2 + Rmin_{jj}^2}$, ref¹⁰⁴.

FIG. 4.

Dipole moment distributions of alcohol (left panels) and water molecules (right panels) in solutions. For methanol and ethanol, panels for the force field combination of the current work and using the LB combination rule are shown in separate panels. Solid curves represent data for $X_a = 0.1$ and dashed curves for $X_a = 0.2$.

FIG. 5.

Radial distribution functions of alcohols in aqueous solutions. Solid curves for $X_a = 0.2$ are shifted upwards by 2.0 (methanol and ethanol) or 0.5 (propanol and butanol) Angstroms. Dashed curves in the panels for methanol-water and ethanol-water RDFs represent the radial distribution functions obtained from MD simulations using the original force field of Patel and Brooks^{11,12}

FIG. 6. Hydrogen bonding excess around carbon atom of the closest methanol(top panel)/ ethanol(bottom panel) molecule for alcohol mole fraction 0.1

Methanol-Water

FIG. 7.

Excess hydrogen bond profile based on $n_{Hb}(r)$ (solid curves for $X_a = 0.1$, dashed curves for $X_a = 0.2$) and the running integral $N_{Hb}(R)$ (open circles for $X_a = 0.1$, open squares for $X_a = 0.2$) as a function of separation between the nearest methanol carbon C_{alco} (panels a and c) or oxygen O_{alco} and water oxygen O_{wat} panels (b and d). See text for Equations for the $n_{Hb}(r)$ and $N_{Hb}(R)$. Panels a and b show data for the current CHEQ-PSP force field; panels c and d show data for the original force field of Patel and Brooks¹¹. (** = Reference¹¹)

FIG. 8.

Excess hydrogen bond profile based on $n_{Hb}(r)$ (solid curves for $X_a = 0.1$, dashed curves for $X_a = 0.2$) and the running integral $N_{Hb}(R)$ (open circles for $X_a = 0.1$, open squares for $X_a = 0.2$) as a function of separation between the nearest ethanol carbon C_{alco} (panels a and c) or oxygen O_{alco} and water oxygen O_{wat} (panels b and d). See text for Equations for the $n_{Hb}(r)$ and $N_{Hb}(R)$. Panels a and b show data for the current CHEQ-PSP force field; panels c and d show data for the original force field of Patel and Brooks¹². (** = Reference¹²)

FIG. 9.

Average water dipole moment as a function of the distance between the nearest methanol (panels a and c) and ethanol (panels b and d) oxygen O_{alco} and water oxygen O_{wat} . Solid curves represent $X_a = 0.1$, dashed curves $X_a = 0.2$. Panels a and b correspond to the results using the force field of this work; panels c and d to those using the original force field of Patel and Brooks^{11,12} (** = Reference¹¹) (** = Reference¹²)

FIG. 10. Cluster size versus probability. Panels a and c show methanol and ethanol cluster probabilities, respectively. Panels b and d show water cluster probabilities for methanol and ethanol solutions, respectively.

FIG. 11.

The center of mass RDFs for KB analysis. Panels a, b, and c show data for methanol-water solution; panels d, e, and f for ethanol-water solutions. Solid curves represent data for $X_a = 0.1$ and dashed curves for $X_a = 0.2$

FIG. 12.

Potentials of mean force corresponding to RDF's of Figure 8. Panels a, b, and c show data for methanol-water solution; panels d, e, and f for ethanol-water solutions. Dashed curves represent data for $X_a = 0.1$ and dotted curves for $X_a = 0.2$

FIG. 13. Kirkwood-Buff integrals. Panels a, b, and c show data for methanol-water solution; panels d, e, and f for ethanol-water solutions. Solid curves represent data for $X_a = 0.1$ and dashed curves for $X_a = 0.2$

TABLE I

LJ parameters between alcohol and water atoms in CHEQ-orig (based on the L-B combination rules) and CHEQ-PSP models.

Atoms	$\epsilon_{ij}^{CHEQ-orig}$	$\epsilon_{ij}^{CHEQ-PSP}$	$R_{min,ij}^{CHEQ-orig}$	$R_{min,ij}^{CHEQ-PSP}$
OH1-OT	-0.2239	-0.2339	3.5229	3.5429
OH1-HT	-0.0897	-0.0797	1.9745	1.9945
H-OT	-0.1147	-0.1047	1.9974	1.9974
H-HT	-0.0460	-0.0060	0.4490	0.4490
CT2-OT	-0.1415	-0.1452	3.7879	3.5347
CT2-HT	-0.0567	-0.0582	2.2395	2.2240
CT3-OT	-0.1415	-0.1452	3.7879	3.6294
CT3-HT	-0.0567	-0.0582	2.2395	2.2240
OH1-OT ^{meth}	-0.2141	-0.2339	3.5029	3.5429
OH1-HT ^{meth}	-0.0858	-0.0797	1.9545	1.9945
CT3-OT ^{meth}	-0.1445	-0.1452	3.7729	3.6294
CT3-HT ^{meth}	-0.0579	-0.0582	2.2245	2.2238

TABLE II

Free Energy of Hydration, kcal/mol

Model	C22 ^a		Drude ^a		CHEQ-orig		CHEQ-PSP			Expt.	
	ΔG		ΔG		LRC	ΔG	LRC	ΔG_{chg}	ΔG_{rdw}	ΔG	ΔG
MeOH	-4.98 (0.08)		-5.20 (0.19)		-0.29	-5.97 (0.45)	-0.28	-5.62 (0.24)	1.09 (0.34)	-4.80 (0.42)	-5.11 ^b
EtOH	-5.34 (0.12)		-4.97 (0.13)		-0.45	-6.03 (0.45)	-0.40	-6.65 (0.28)	1.72 (0.35)	-5.33 (0.45)	-5.01 ^b
1-ProH	-5.33 (0.24)		-4.85 (0.15)		-0.61	-6.53 (0.47)	-0.52	-6.76 (0.17)	2.14 (0.30)	-5.14 (0.34)	-4.83 ^b
1-BuOH	-5.60 (0.21)		-4.67 (0.23)		-0.77	-6.37 (0.36)	-0.64	-6.66 (0.12)	2.50 (0.37)	-4.80 (0.39)	-4.72 ^b

^aNonpolarizable and Drude data are from ref¹⁶.^bExperimental data from ref³⁶.

TABLE III

Alcohol-Water Gas-phase Dimer Energies and Hydrogen Bond Distances

Property	CHARMM22 ^a	Drude ^a	CHEQ-orig	CHEQ-PSP	MP2 ^b	MP2 ^c	CCSD(T) ^c
MeOH							
PA ΔE_{dimer}	6.23	4.74	6.07	5.57	5.22	5.15	5.02
$R_{O...H}$	1.84	1.90	1.80	1.88	1.97	1.90	–
PD ΔE_{dimer}	6.21	4.45	4.74	4.44	4.51	4.48	4.45
$R_{O...H}$	1.82	1.94	1.90	1.96	2.01	1.94	–
EtOH							
PA ΔE_{dimer}	6.54	5.52	6.85	6.64	5.55	5.49	5.35
$R_{O...H}$	1.86	1.89	1.81	1.84	1.97	1.90	–
PD ΔE_{dimer}	6.31	4.63	4.92	4.77	4.43	4.27	4.28
$R_{O...H}$	1.82	1.93	1.92	1.94	2.02	1.95	–
1-ProH							
PA ΔE_{dimer}	6.39	5.30	6.80	6.55	5.52	–	–
$R_{O...H}$	1.86	1.89	1.81	1.85	1.97	–	–
PD ΔE_{dimer}	6.35	4.71	4.84	4.68	4.48	–	–
$R_{O...H}$	1.82	1.93	1.92	1.94	2.01	–	–
1-BuOH							
PA ΔE_{dimer}	–	–	7.00	6.81	5.68	–	–
$R_{O...H}$	–	–	1.80	1.83	1.97	–	–
PD ΔE_{dimer}	–	–	4.96	4.80	4.43	–	–
$R_{O...H}$	–	–	1.91	1.94	2.02	–	–

^a results for nonpolarizable and Drude model reported in ref¹⁶.^b MP2 calculations using MP2/aug-cc-pvDZ basis. PA geometry optimizations starts from the BIS structures¹⁶.^c Fileti et al.¹⁰⁰ obtained the dimer geometries by MP2/aug-cc-pvDZ optimizations and performed single-point calculations at the CCSD(T) level.

TABLE IV

Simulated and Experimental Density

X	$\rho^{CHEQ-orig}$ (g/L)	$\rho^{CHEQ-PSP}$ (g/L)	ρ^{expa} (g/L)
MeOH			
0.1	989.58(13.74)	987.77(14.03)	963.52
0.2	973.48(12.72)	968.97(12.93)	947.18
EtOH			
0.1	975.06(12.12)	981.03(12.33)	966.39
0.2	949.58(11.21)	954.81(11.60)	931.48

^a Density experimental data from ref¹²².

TABLE V

Percentage of Molecules Forming Clusters

X	alcohol cluster ^{PSP%}	alcohol cluster ^{orig%}	water cluster ^{PSP%}	water cluster ^{orig%}
MeOH				
0.1	33.68	11.91	98.51	98.58
0.2	51.54	25.01	97.30	97.35
EtOH				
0.1	16.56	9.37	98.74	98.79
0.2	31.70	20.39	97.61	97.72

TABLE VI

Kirkwood-Buff Integrals (in unit of cm^3/mol) and Excess Coordination Numbers of Simulated Methanol-Water Mixtures

X_{dco}	G_{aa}	G_{aw}	G_{ww}	N_{aa}	N_{aw}	N_{ww}
MeOH						
0.1	-26.06 (0.97)	-34.06 (0.49)	-19.92 (0.41)	-0.13 (0.01)	-1.56 (0.02)	-0.91 (0.02)
0.1 ^{expa}	-58	-32	-12	-0.25	-1.50	-0.60
0.2	-6.84 (0.80)	-35.86 (1.74)	-8.74 (1.51)	-0.06 (0.01)	-1.34 (0.06)	-0.33 (0.01)
0.2 ^{expa}	-42	-38	-8	-0.40	-1.20	-0.25
EtOH						
0.1	-48.09 (0.96)	-41.29 (1.54)	-10.55 (0.35)	-0.23 (0.01)	-1.75 (0.07)	-0.45 (0.01)
0.1 ^{expb}	-56.9	-51.0	-7.6			
0.2	-44.65 (0.96)	-33.18 (3.82)	-3.66 (3.00)	-0.36 (0.01)	-1.07 (0.12)	-0.12 (0.10)
0.2 ^{expb}	-25.5	-74.0	23.1			

^aMethanol solution experimental KB integrals are estimated based on SAXS work of ref¹¹⁷. Experimental excess coordination numbers are read from Figure 3 of ref⁷⁵.

^bEthanol solution SAXS data from ref¹¹⁹.