THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

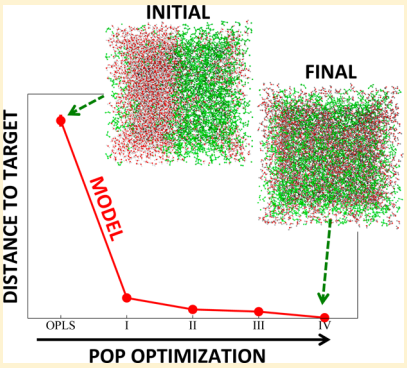
pubs.acs.org/JPCB

Optimizing Potentials for a Liquid Mixture: A New Force Field for a *tert*-Butanol and Water Solution

Michele Di Pierro,[†] Mauro L. Mugnai,[‡] and Ron Elber*[,†,‡]

[†]Institute for Computational Engineering and Sciences and [‡]Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States

ABSTRACT: A technology for optimization of potential parameters from condensed-phase simulations (POP) is discussed and illustrated. It is based on direct calculations of the derivatives of macroscopic observables with respect to the potential parameters. The derivatives are used in a local minimization scheme, comparing simulated and experimental data. In particular, we show that the Newton trust region protocol allows for more accurate and robust optimization. We apply the newly developed technology to study the liquid mixture of *tert*-butanol and water. We are able to obtain, after four iterations, the correct phase behavior and accurately predict the value of the Kirkwood Buff (KB) integrals. We further illustrate that a potential that is determined solely by KB information, or the pair correlation function, is not necessarily unique.



INTRODUCTION

Molecular dynamics (MD) is a useful tool to study molecular mechanisms in materials science and biophysics. Advancements in computer power and simulation techniques continuously raise the bar of what is possible to model; systems of millions of atoms can be studied with MD,[1] and the longest simulations can reach milliseconds.[2] New interesting applications are investigated, and new simulation challenges are found.

As the world of molecular simulations grows in size and complexity, there is a growing demand for more accurate force fields capable of recovering subtle physical phenomena that are difficult to reproduce with simplified interaction models. The recent and continuous increase of simulation lengths allows us to compute converged statistical averages to be compared to experimental data that were inaccessible for simulations in the past.[3]

The functional form of MD force fields remained essentially the same for decades. The energy function consists of bonding and nonbonding terms. The bonded interactions consist of two-, three-, and four-body interactions (respectively, bonds, angles, and torsions), and the nonbonded interactions are a sum of electrostatic forces between fixed-point charges placed on atom centers, and hard-core and dispersion forces modeled by Lennard-Jones (LJ) interaction.

This functional form showed robustness and transferability, and is the method of choice of most simulation software. While many quantitative and qualitative observations support the validity of such MD force fields, it is certainly possible to improve their functional form. There are many ongoing efforts in this direction; the addition of polarization terms[4] and the addition of statistical potentials[5] are examples of such efforts. In this paper we consider the standard functional form for the MD force field, and we focus on the process of the choice of the parameters optimized against condensed-phase simulations. This idea was put forward by Jorgensen in the OPLS force field.[6] Our contribution is in the design of an automated refinement algorithm that, in principle, can handle a large number of parameters. Our algorithm is not restricted to a particular choice of a functional form; however, our software[7] is designed to work with the functional form and the data structure of the MOIL program.[8]

Choosing the optimal set of parameters is important and attracted a considerable amount of manual and partially automated work in the past. Widely used parameter sets (OPLS,[6] AMBER,[9] CHARMm[10]) have been subject to continuous updates and refinements; some updates are improvements on biopolymer models (peptides[11]), while others are additions of parameters for some new small molecules.[12]

The development of force field parameters for a small molecule typically involves multiple stages. It involves quantum mechanical calculations (usually in the gas phase) to fit molecular mechanics parameters and condensed-phase calculations to fit parameters so that thermodynamic properties can be reproduced.

Sometimes, properties of liquid mixtures are not well reproduced by the parameters developed to describe a single-component system. It is therefore desired to address liquid mixtures more directly and to consider theories and algorithms tailored for these systems. There are a few examples of theories that capture properties of solutions in a relatively small number of parameters. The Kirkwood–Buff (KB) integrals[13] summarize

Special Issue: William L. Jorgensen Festschrift

Received: June 1, 2014
Revised: July 22, 2014
Published: July 26, 2014

   dx.doi.org/10.1021/jp505401m | *J. Phys. Chem. B* 2015, 119, 836–849

a set of experimental observables characterizing liquid mixtures and are useful targets of optimization of potential parameters.[12,14,15]

Di Pierro and Elber have recently published an automated method to refine parameters of force fields using as targets experimental observables that can be predicted with computational statistical mechanics. A prime advantage of the technology is its ability to handle the coupled optimization of a large number of potential parameters. We named our algorithm POP (Parameter OPtimization),[7] and we illustrated its effectiveness on hundreds of parameters, exploring potentials for peptide folding in aqueous solutions. Independently, and at the same time, Wang et al. published a similar algorithm and used it to refine a potential for liquid water,[16] illustrating the general applicability of the method.

In the present paper, we combine POP and the observables of KB theory to optimize the potential for liquid mixtures. We also discuss enhancements to the original POP algorithm that enable faster and more accurate convergence to the desired set of parameters.

We use the POP method to improve the current force field for *tert*-butanol (TBA) in aqueous solution (see Figure 1). Our

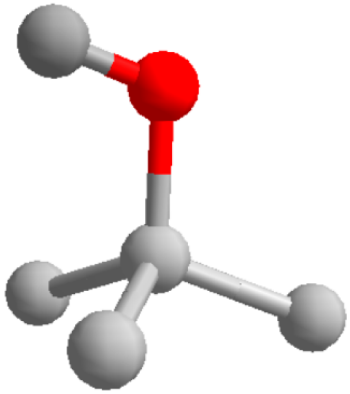


Figure 1. United-atom model of TBA; note that each methyl group is represented by a single united atom.

starting point is the OPLS united-atom (OPLSUA) parameters for TBA[17] and the TIP3P[18] water model. We develop a new set of parameters only for TBA. We retain the same water model that was tested comprehensively by now on a very large number of systems. We seek a set of TBA parameters that better reproduce the KB integrals estimated from experiments[19] over a range of different concentrations. While optimization for TBA–water mixtures according to KB integrals have been done in the past,[15] the present study is automated, producing high-quality potentials, and makes it possible to address questions about the uniqueness of the results.

This paper proceeds as follows. In the Methods Section, we revisit the theory of POP and introduce an improved optimization algorithm. We then discuss the KB theory and its application in MD simulations; last, we focus on its use in the context of the POP algorithm. In the Results and Discussion section, we develop a new force field for liquid mixtures of TBA and water. Discussion and conclusions are left for the final section.

METHODS SECTION

POP Algorithm. We denote an experimental measurement of an observable O by O_{exp}. The measured quantity corresponds to the (canonical) ensemble average of a certain function of the phase space (positions are collectively indicated by $\mathbf{R}$ and momenta by $\mathbf{P}$) that may or may not depend on the force field parameters π

$$\langle O\rangle_{(N,V,T,\pi)} = \frac{\int d\mathbf{R}\, d\mathbf{P}\cdot O(\mathbf{R},\mathbf{P},\pi)e^{-\beta H(\mathbf{R},\mathbf{P},\pi)}}{\int d\mathbf{R}\, d\mathbf{P}\cdot e^{-\beta H(\mathbf{R},\mathbf{P},\pi)}} \tag{1}$$

The ensemble average of the observable always depends on the set of parameters π through the exponential weight, and of course, it depends on the macroscopic constraints of the system (number of particles N, volume V, and temperature of the thermal reservoir in contact with the system T).

In practice, the ensemble average above can be substituted by a time average over a trajectory

$$\langle O\rangle_{(N,V,T,\pi)} = \lim_{t\to\infty}\frac{1}{t}\int_0^t O(t')\,dt' \tag{2}$$

provided that the system is ergodic and that the dynamics reproduces the canonical sampling (e.g., isokinetic dynamics[20]).

One way to validate the results of a simulation is to measure how much computed observables differ from experimental measurements. Here, we optimize the parameters in the MD force field in order minimize the discrepancy between computed and experimental observables.

Given N_O experimental observations, we define our target function to be

$$\Theta(\pi) = \sum_{i=1}^{N_O} w_i[\langle O_i\rangle_\pi - O_{exp,i}]^2 \tag{3}$$

where w_i are constant weights which are ones in the present example. Other choices of the target function are possible, provided that the target function is differentiable. Ideally, the target function Θ should be zero. The optimal set of parameters is π^*, such that

$$\pi^* = \arg\min \Theta(\pi) \tag{4}$$

We minimize the target function using a trust region Newton method.[22] To do so, we need the gradient vector and the Hessian matrix (or some approximation of it[23]) in parameter space.

We have shown in a previous paper that any derivative of the target function can be calculated as a single ensemble average;[7] the calculation is analytical and affected only by the statistical error associated with the ensemble average.

The gradient vector is

$$\nabla_\pi\Theta(\pi) = 2\sum_{i=1}^{N}[\langle O_i\rangle_\pi - O_{exp,i}]\nabla_\pi\langle O_i\rangle_\pi$$
$$\nabla_\pi\langle O_i\rangle_\pi = \langle\nabla_\pi O_i\rangle_\pi - \beta[\langle\nabla_\pi H\cdot O_i\rangle_\pi - \langle\nabla_\pi H\rangle_\pi\langle O_i\rangle_\pi] \tag{5}$$

while the Hessian matrix is

$$\nabla\nabla_\pi^T\Theta(\pi) = 2\sum_{i=1}^{N}[\langle O_i\rangle_\pi - O_{exp,i}]\nabla\nabla_\pi^T\langle O_i\rangle_\pi + 2\sum_{i=1}^{N}\nabla_\pi\langle O_i\rangle_\pi\nabla_\pi^T\langle O_i\rangle_\pi$$
$$\begin{aligned}\nabla\nabla_\pi^T\langle O\rangle_\pi = {}& \langle\nabla\nabla_\pi^T O\rangle_\pi + \beta\{-\langle\nabla_\pi H\nabla_\pi^T O\rangle_\pi + \langle\nabla_\pi H\rangle_\pi\langle\nabla_\pi^T O\rangle_\pi\\ & - \langle\nabla_\pi O\nabla_\pi^T H\rangle_\pi - \langle O\nabla\nabla_\pi^T H\rangle_\pi + \langle\nabla_\pi O\rangle_\pi\langle\nabla_\pi^T H\rangle_\pi\\ & + \langle O\rangle_\pi\langle\nabla\nabla_\pi^T H\rangle_\pi\} + \beta^2\{\langle\nabla_\pi HO\nabla_\pi^T H\rangle_\pi\\ & - 2\langle\nabla_\pi H\rangle_\pi\langle O\nabla_\pi^T H\rangle_\pi + 2\langle O\rangle_\pi\langle\nabla_\pi H\rangle_\pi\langle\nabla_\pi^T H\rangle\\ & - \langle O\rangle_\pi\langle\nabla_\pi H\nabla_\pi^T H\rangle_\pi\}\end{aligned} \tag{6}$$

Note that the Hessian matrix, while symmetric by construction, is in general indefinite.

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836–849

Using the gradient and the Hessian (calculated for a given parameter set π_0), we can build a quadratic model for the target function $\Theta(\pi)$ in a neighborhood of the point π_0; the quadratic model $m(p)$ is a function of the displacement vector $p = \pi - \pi_0$. The quadratic model is accurate in a neighborhood of π_0; we characterize this region of the parameter space by the space contained in a spherical domain of radius Δ. Later on, we will explain how the radius can be iteratively updated.

We minimize the target function by iteratively updating the parameter set

$$\pi_{k+1} = \pi_k + p_k \tag{7}$$

where the increment p_k is chosen by solving the subproblem

$$p_k = \arg\min m_k(p) \qquad \text{s.t.} \| p \| \leq \Delta_k$$
$$m_k(p) = \Theta(\pi_k) + \nabla^T\Theta(\pi_k)p + \frac{1}{2}p^T\nabla\nabla^T\Theta(\pi_k)p \tag{8}$$

The subproblem can be solved in approximated way[23] or exactly; here, given the dimension of the parameter space and having calculated the Hessian, we find the exact solution of the subproblem following the method of Moré and Sorensen[22] (Appendix A).

One specific problem of full space optimization of MD force fields lies in the range of values of different parameters; some parameter ranges of values are in the hundreds of thousands (e.g., some van der Waals parameters), while others are of order one (e.g., torsion coefficients). The difference of several orders of magnitude presents a significant challenge for simple optimization algorithms like steepest descent. In order to make adjustments that are homogeneous, we introduce a scaling matrix D_k such that every element of the vector $\gamma_k = D_k\pi_k$ is of order one. The matrix D_k is a diagonal matrix

$$\begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \cdots & \\ & & & d_n \end{bmatrix} \qquad d_i = \begin{cases} 1/|(\pi_k)_i| & \text{if } |(\pi_k)_i| > 1 \\ 1 & \text{if } |(\pi_k)_i| \leq 1 \end{cases} \tag{9}$$

In this way, all of the parameters are scaled to be in the range of -1 to 1.

We can now solve the subproblem in an elliptical trust region

$$p_k = \arg\min m_k(p) \qquad \text{s.t.} \| Dp \| \leq \Delta_k$$
$$m_k(p) = \Theta(\pi_k) + \nabla^T\Theta(\pi_k)p + \frac{1}{2}p^T\nabla\nabla^T\Theta(\pi_k)p \tag{10}$$

The parameters that we optimize include the atomic (partial) charges within a molecule. An obvious constraint on the space of the optimization is the preservation of the molecular charge, that is, the total molecular charge must not change upon optimization. This is in contrast to other parameters such as bond length. We impose charge conservation as a linear constraint. We write the total charge as $Q = \sum_{l=1}^{m} n_l q_l$, where q_l is the partial charge associated with the atom type l, there are n_l atoms of type l, and the total number of atom types is m. Keeping the total charge constant means satisfying the linear constraint $Q_k = Q_0$, where k, as before, is the iteration index. Exploiting the linearity of the constraint, we can fix the total charge by projecting the increment p_k on the hyperplane of constant charge defined by the constraint Q

$$p_k' = p_k - \frac{p_k \cdot \nabla_\pi Q}{\nabla_\pi Q \cdot \nabla_\pi Q}\nabla_\pi Q \tag{11}$$

We now outline the strategy to update the trust region radius Δ_k.[23] The choice of radius is based on the agreement between the model function m_k and the target function Θ. Given a step p_k at iteration k, such agreement can be measured by the following ratio

$$\rho_k = \frac{\text{actual reduction}}{\text{predicted reduction}} = \frac{\Theta(\pi_k) - \Theta(\pi_k + p_k)}{m_k(0) - m_k(p_k)} \tag{12}$$

The predicted reduction is always positive; therefore, if ρ_k is negative, then $\Theta(\pi_k + p_k)$ is greater than the current value $\Theta(\pi_k)$, and the step must be rejected.

If ρ_k is close to 1, there is good agreement between the model and target function; therefore, it is safe to expand the trust region.

If ρ_k is positive but not close to 1, we do not alter the trust region; if ρ_k is close to 0, then we shrink the trust region (see Appendix A for more details).

KB Theory. The KB theory of fluid mixtures[13] relates some integrals of the pair correlation functions (microscopic observables) computed in the grand-canonical ensemble to derivatives of the chemical potential, isothermal compressibility, and partial molar volumes (macroscopic quantities). A detailed derivation of the theory can be found elsewhere.[13,24] Here, we present the key concepts for our application. The use of KB to optimize potential parameters was put forward by Smith.[25] The optimization using Newton−Raphson in a trusted region is our contribution.

Let us consider a binary mixture of two chemical species, chemical species S1 and chemical species S2. The symbols "A" and "B" can either be S1 or S2. Let us define the following pair correlation function

$$g_{AB}^{\mu_1\mu_2 VT}(\vec{r}_1, \vec{r}_2) = \frac{\rho_{AB}^{\mu_1\mu_2 VT}(\vec{r}_1, \vec{r}_2)}{\rho_A^{\mu_1\mu_2 VT}(\vec{r}_1)\rho_B^{\mu_1\mu_2 VT}(\vec{r}_2)} = \frac{\langle \sum_{i\in A}\sum_{j\in B}\delta(\vec{r}_1 - \vec{r}_i)\delta(\vec{r}_2 - \vec{r}_j)\rangle_{\mu_1\mu_2 VT}}{\langle \sum_{i\in A}\delta(\vec{r}_1 - \vec{r}_i)\rangle_{\mu_1\mu_2 VT}\langle \sum_{j\in B}\delta(\vec{r}_2 - \vec{r}_j)\rangle_{\mu_1\mu_2 VT}} \tag{13}$$

The averages are performed in the grand-canonical ensemble (holding fixed the reservoir temperature T, the volume V, and the chemical potentials of the two chemical species S1 and S2, μ_1 and μ_2, respectively). This function expresses the joint probability of finding the center of mass of a molecule of species A (we indicate its position by $\vec{r}_i$) at $\vec{r}_1$ and the center of mass of a molecule of species B (we indicate its position by $\vec{r}_j$) at $\vec{r}_2$, relative to the probability of the two independent events. Note that, even though only the centers of mass appear in eq 13, the theory is general; it does not require spherically symmetric molecules. Indeed, the internal degrees of freedom of A and B molecules, as well as their overall orientation, are accounted for in the ensemble average.[13,24]

Let us assume that the system is homogeneous and therefore that the probability of finding a molecule in a specific place is constant anywhere in the system. In this case, we can rewrite eq 13 as

$$g_{AB}^{\mu_1\mu_2 VT}(\vec{r}_1 - \vec{r}_2) = \frac{\left\langle \sum_{i\in A}\sum_{j\in B}\frac{1}{V}\delta(\vec{r}_1 - \vec{r}_2 - \vec{r}_i + \vec{r}_j)\right\rangle_{\mu_1\mu_2 VT}}{\left\langle \sum_{i\in A}\frac{1}{V}\right\rangle_{\mu_1\mu_2 VT}\left\langle \sum_{j\in B}\frac{1}{V}\right\rangle_{\mu_1\mu_2 VT}} = \frac{\frac{1}{V}\langle \sum_{i\in A}\sum_{j\in B}\delta(\vec{r}_1 - \vec{r}_2 - \vec{r}_i + \vec{r}_j)\rangle_{\mu_1\mu_2 VT}}{\frac{\langle N_A\rangle_{\mu_1\mu_2 VT}}{V}\frac{\langle N_B\rangle_{\mu_1\mu_2 VT}}{V}} = \frac{\rho_{AB}^{\mu_1\mu_2 VT}(\vec{r}_1 - \vec{r}_2)}{\rho_A^{\mu_1\mu_2 VT}\rho_B^{\mu_1\mu_2 VT}} \quad (14)$$

Let us now define $\vec{r} = \vec{r}_1 - \vec{r}_2$ and $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$. If the probability of finding the center of mass of B-type molecules around a molecule of species A depends only on their distance and not the orientation of the vector that connects them (i.e., the system, averaged over its internal degrees of freedom and the orientations of the molecules, is isotropic), then we can rewrite eq 14

$$g_{AB}^{\mu_1\mu_2 VT}(r) = \frac{\rho_{AB}^{\mu_1\mu_2 VT}(r)}{\rho_A^{\mu_1\mu_2 VT}\rho_B^{\mu_1\mu_2 VT}} = \frac{1}{V}\frac{\left\langle \sum_{i\in A}\sum_{j\in B}\frac{1}{4\pi r^2}\delta(r - r_{ij})\right\rangle_{\mu_1\mu_2 VT}}{\rho_A^{\mu_1\mu_2 VT}\rho_B^{\mu_1\mu_2 VT}} \quad (15)$$

where we transformed the Dirac's delta from Cartesian coordinates to polar coordinates.

The key object in the KB theory is the so-called KB integral

$$G_{AB}^{\mu_1\mu_2 VT} = \int_V d\vec{r}\ [g_{AB}^{\mu_1\mu_2 VT}(r) - 1] \quad (16)$$

The meaning of this quantity becomes clearer if we rewrite eq 16 as

$$\rho_B^{\mu_1\mu_2 VT} G_{AB}^{\mu_1\mu_2 VT} = \int_V d\vec{r}\left[\frac{\rho_{AB}^{\mu_1\mu_2 VT}(r)}{\rho_A^{\mu_1\mu_2 VT}} - \rho_B^{\mu_1\mu_2 VT}\right] \quad (17)$$

The left-hand side of eq 17 is the so-called excess coordination number. The integrand on the right-hand side of eq 17 has two terms: first the conditional probability of finding a molecule of species B around a molecule of species A and second the probability of finding the molecule of species B. The integral gives the excess (or shortage) of molecules of species B in volume V around a molecule of species A with respect to the average number of B-type molecules in the same volume V. Obviously, at large distances (for small solutes, typically a few nanometers), the correlation between A-type and B-type molecules is lost (i.e., $\rho_{AB}^{\mu_1\mu_2 PVT}(r) \xrightarrow[r\to\infty]{} \rho_A^{\mu_1\mu_2 VT}\rho_B^{\mu_1\mu_2 VT}$), and the integrand of eq 17 is zero. This means that the KB integral carries local, microscopic information that can be evaluated with a MD simulation.

At the same time, it is possible to show[13,24] that the KB integral (eq 16) is equal to

$$G_{AB}^{\mu_1\mu_2 VT} = V\left[\frac{\langle N_A N_B\rangle_{\mu_1\mu_2 VT} - \langle N_A\rangle_{\mu_1\mu_2 VT}\langle N_B\rangle_{\mu_1\mu_2 VT}}{\langle N_A\rangle_{\mu_1\mu_2 VT}\langle N_B\rangle_{\mu_1\mu_2 VT}} - \frac{\delta_{AB}}{\langle N_A\rangle_{\mu_1\mu_2 VT}}\right] \quad (18)$$

where N_A and N_B are the number of molecules of type A and B, respectively, and δ_{AB} is the usual Kronecker's delta.

This equation expresses the connection with thermodynamics. The fluctuation of numbers of particles in the system is a macroscopic object, and it can be expressed in terms of derivatives of the chemical potential of A-type molecules with respect to the number of particles of species B, the isothermal compressibility, and the partial molar volumes of the two species.[13,24] These quantities can be measured experimentally. It is also possible to extract the KB integrals from these thermodynamical quantities.[26]

Therefore, KB theory provides a useful protocol to analyze MD simulation and connect the results with measurable quantities; it requires extracting from the trajectory the pair correlation function and its integral, which are routinely computed. Indeed, it has been found in numerous applications in recent years, particularly in the context of force field parametrization.[12,14,15,27]

Nevertheless, there are some caveats. First of all, MD simulations are performed at a constant number of particles, therefore, in the canonical, not in the grand-canonical ensemble. The connection between the KB integral and the measurable quantities relies upon eq 18, which in the canonical ensemble would be[24]

$$G_{AB}^{N_1N_2VT} = -V\frac{\delta_{AB}}{N_A} \quad (19)$$

where N_1 and N_2 are the fixed number of particles of species S1 and S2, respectively.

Obviously, in this case, the connection with the chemical potential would be lost. How can we compute a grand-canonical average from a canonical simulation? A possible way is to compute the KB integral in a volume V' that is much smaller than the volume of the system. In such a volume the number of molecules fluctuates. The rest of the system acts as a molecular reservoir. This procedure is correct as long as the pair correlation function in eq 16 decays to 1 within V'. This leads to the second caveat; sometimes the pair correlation function decays to 1 very slowly, and we need to account properly for its long tail. A careless truncation may have a bad impact on the evaluation of the KB integral. To understand why, let us consider the case in which the KB integral in eq 16 is computed in a spherical domain of radius R_C

$$G_{AB}^{\mu_1\mu_2 VT}(R_C) = \int_0^{R_C}[g_{AB}^{\mu_1\mu_2 VT}(r) - 1]4\pi r^2\, dr \quad (20)$$

If at R_C the pair correlation function is not 1, we are neglecting a contribution to the integral that might potentially be very large as it is multiplied by the square of the radius.

Ganguly and van der Vegt[28] have investigated these caveats and proposed empirical corrections to the KB integrals to alleviate these problems. Others[29] have carried out the calculation of KB integrals using the adaptive resolution scheme for MD.[30] In our case, we decided to compute the integral of the pair correlation in the NVT ensemble without corrections, but we computed the integral of the pair correlation function up to 20 Å, which is large compared with what is commonly found in the literature.[12,14,15,28] In this way, we can check whether the KB integral has indeed reached a plateau. To ensure that the system can be considered grand-canonical, we run the simulations in cubic systems of roughly a 65 Å box length. This ensures that the reservoir is around 7 times larger than the volume V' in which the pair correlation function is computed.

KB Observable Functions. In this paper, we use the method described above to optimize the MD force field using the three KB integrals of a binary liquid mixture as a target for optimization. We write an observable function associated with the KB integrals. We derive the observable function from the definition of the KB integral; using eq 15 for the pair correlation function but computed in the canonical ensemble, we obtain

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

Table 1. Molar Concentration of TBA, Density, Volumes, and Number of Molecules of Each One of the Systems Simulated[a]

	concentrations					
TBA molar concentration	0.04	0.10	0.14	0.17	0.20	0.30
density g/cm^3	0.9707	0.9357	0.9146	0.9010	0.8836	0.8577
volume Å^3	64.079^3	64.393^3	64.723^3	64.987^3	65.255^3	66.170^3
number of TBA molecules	304	637	808	919	1000	1289
number of water molecules	7290	5733	4968	4488	4096	3006

[a]The parameters were chosen such that the densities were within 1% of the experimental values reported in ref 31 at 308.15 K.

$$G_{AB}^{(N,V,T,\pi)}(R_C) = \int_0^{R_C} [g_{AB}(r) - 1]4\pi r^2 \, dr = \left\langle \int_0^{R_C} \sum_{\substack{i\in A \\ j\in B}} \frac{1}{\rho_A \rho_B V} \delta(|\vec{r}_i - \vec{r}_j| - r) \, dr \right\rangle_{(N,V,T,\pi)} - \frac{4}{3}\pi R_C^3 \tag{21}$$

The function inside of the canonical ensemble average is our observable; in this case, it only depends on the set of coordinates **R**. We define our observable functions to be

$$O_1(\mathbf{R}) = \int_0^{R_C} \sum_{\substack{i\in S1 \\ j\in S2}} \frac{1}{\rho_1 \rho_2 V} \delta(|\vec{r}_i - \vec{r}_j| - r) \, dr - \frac{4}{3}\pi R_C^3$$

$$O_2(\mathbf{R}) = \int_0^{R_C} \sum_{\substack{i\in S1 \\ j\in S1}} \frac{1}{\rho_1 \rho_1 V} \delta(|\vec{r}_i - \vec{r}_j| - r) \, dr - \frac{4}{3}\pi R_C^3$$

$$O_3(\mathbf{R}) = \int_0^{R_C} \sum_{\substack{i\in S2 \\ j\in S2}} \frac{1}{\rho_2 \rho_2 V} \delta(|\vec{r}_i - \vec{r}_j| - r) \, dr - \frac{4}{3}\pi R_C^3 \tag{22}$$

where $\rho_1 = N_1/V$ and $\rho_2 = N_2/V$.

Let us define

$$\overline{(N_B(R_C)|A)}(\mathbf{R}) = \frac{1}{N_A} \sum_{\substack{i\in A \\ j\in B}} \int_0^{R_C} \delta(|\vec{r}_i - \vec{r}_j| - r) \, dr$$

that is, the number of molecules of species B that are at a distance less than R_C from a molecule of species A averaged on all of the molecules of species A. Similarly, we can define

$$\overline{(N_A(R_C))} = \rho_A \frac{4}{3}\pi R_C^3$$

the average number of molecules in a spherical volume of radius R_C. We can rewrite eq 22 as

$$O_1(\mathbf{R}) = \frac{1}{\rho_2}[\overline{(N_2(R_C)|S1)}(\mathbf{R}) - \overline{(N_2(R_C))}] = \frac{1}{\rho_1}[\overline{(N_1(R_C)|S2)}(\mathbf{R}) - \overline{(N_1(R_C))}]$$

$$O_2(\mathbf{R}) = \frac{1}{\rho_1}[\overline{(N_1(R_C)|S1)}(\mathbf{R}) - \overline{(N_1(R_C))}]$$

$$O_3(\mathbf{R}) = \frac{1}{\rho_2}[\overline{(N_2(R_C)|S2)}(\mathbf{R}) - \overline{(N_2(R_C))}] \tag{23}$$

Therefore, the observables in eq 23 measure the excess (or shortage) of a molecule of type A around a molecule of type B compared to the average in the system. The observables do not carry explicit parameter dependence, which simplifies the expressions for the gradient and the Hessian. While the observable function depends explicitly only on the coordinates, it is clear that its ensemble average depends on the composition of the liquid mixture. The ensemble average is a function of the variables

$$(N_1, N_2, V, T, \pi)$$

if we introduce the molar fraction

$$x = \frac{N_1}{N_1 + N_2} \qquad N = N_1 + N_2$$

the same set of variables can be written as

$$(N, x, V, T, \pi)$$

In the following, we are only interested in changes in parameters and molar fractions, and we will therefore drop the other variables; clearly, it is to be intended that experiment and computation are performed under the same conditions.

For multiple mole fractions N_x, the target function is

$$\Theta(\pi) = \sum_{j=1}^{N_x} \sum_{i=1}^{3} [\langle O_i \rangle_{(x_j,\pi)} - O_{\exp,i}(x_j)]^2 \tag{24}$$

COMPUTATIONAL DETAILS

Optimization. For the MD simulations, mixtures of TBA and water were prepared at the mole fractions of TBA in Table 1. The model used for water was TIP3P,[18] and it remained fixed throughout the process of potential refinement. The starting model chosen for TBA was the OPLSUA.[17] The system was prepared to have density consistent with that of experiment (error of about 1%;[31] see Table 1). Particle Mesh Ewald (PME)[32] was used to account for long-range electrostatic interactions with a grid of 64 × 64 × 64. Short-range electrostatic interactions were calculated by real space summation up to a cutoff of 9.5 Å; the same cutoff was used for van der Waals interactions. Periodic boundary conditions were applied. The equations of motion were integrated using the multiple time step integrator RESPA[33] with a time step of 1 fs. Short-range forces were updated every femtosecond, while long-range interactions were calculated every 4 fs according to the protocol in MOIL described in ref 8. The sampling in the *NVT* ensemble was enforced by rescaling the velocities (isokinetic ensemble[21]). The temperature was set to be 300 K in all of the simulations. The experimental results used in the target function are those in ref 19.

An iteration of the optimization of the potential parameters includes a series of MD simulations to collect a sample of structures. An ensemble of structures computed with a particular force field is analyzed to calculate the new parameter set. The new parameter set and potential are used in a successive MD simulation, from which we collect new structures that are analyzed again.

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

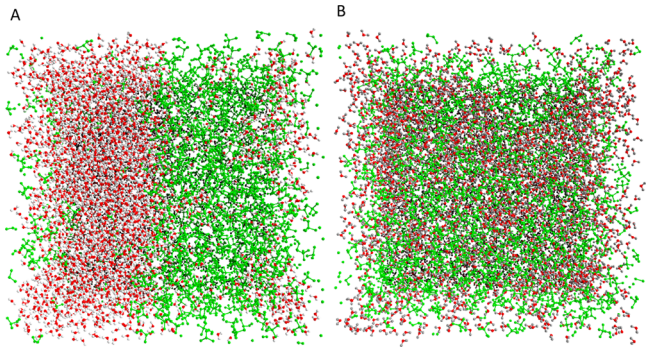


Figure 2. Snapshots of the simulation box; TBA is in green. On the left, we show an equilibrium snapshot of the mixture of TBA and water at 0.20 TBA mole fraction using the OPLSUA force field. Phase separation is evident by visual inspection. On the right is the same system after equilibration using POP4ff; by visual inspection, the solution is mixed.

For the first three iterations, the target function used was

$$\Theta_{\mathrm{I-III}}(\pi) = \sum_{i=1}^{3} [\langle O_i \rangle_{(x,\pi)} - O_{\mathrm{exp},i}(x)]^2$$

at the single mole fraction of TBA of $x = 0.2$.

For the last iteration, we used the target function

$$\Theta_{\mathrm{IV}}(\pi) = \sum_{j=1}^{2} \sum_{i=1}^{3} [\langle O_i \rangle_{(x_j,\pi)} - O_{\mathrm{exp},i}(x_j)]^2$$

with mole fractions of TBA of $x_1 = 0.04$ and $x_2 = 0.10$.

The optimization process was stopped at the IV iteration, where the experimental observables matched the simulated quantities within acceptable error bars.

All of the MD simulations were performed using the software package MOIL in its GPU variant.[8] The analysis of the structures, including the calculation of the gradient and Hessian and the updating of the parameter set, was performed using the software POP[7] included in MOIL.

The KB integrals were calculated up to a cutoff distance of 20 Å. For each iteration, a simulation of 60 ns was performed. We discarded the initial equilibration phase (10 ns), and from the last 50 ns, we collected 4990 equally spaced (in time) structures. The structures were used in the calculations for parameter optimization by POP.

Validation. To validate our potential, we examined the performance of the newly developed model over a range of mixtures at different concentrations. We prepared six systems at mole fractions for which experimental results for the Kirkwood integrals are known[19] (see Table 1). Each of the systems was simulated for 60 ns with the same setup as that described in the optimization paragraph, and the experimental observables were compared to the simulations.

Force Fields. The starting force field for TBA is OPLSUA.[17] In this force field, TBA is composed of six particles because each of the methyl groups is treated as a single particle without internal degrees of freedom. Here, we report the OPLSUA force field and the optimized force field, which we will refer to as POP4ff.

The complete potential is a sum of bonding and nonbonding terms

$$U_{\mathrm{total}} = \sum_{\mathrm{b}} U_{\mathrm{b}} + \sum_{\theta} U_{\theta} + \sum_{\phi} U_{\phi} + \sum_{\mathrm{LJ}} U_{\mathrm{LJ}} + \sum_{\mathrm{elec}} U_{\mathrm{elec}}$$

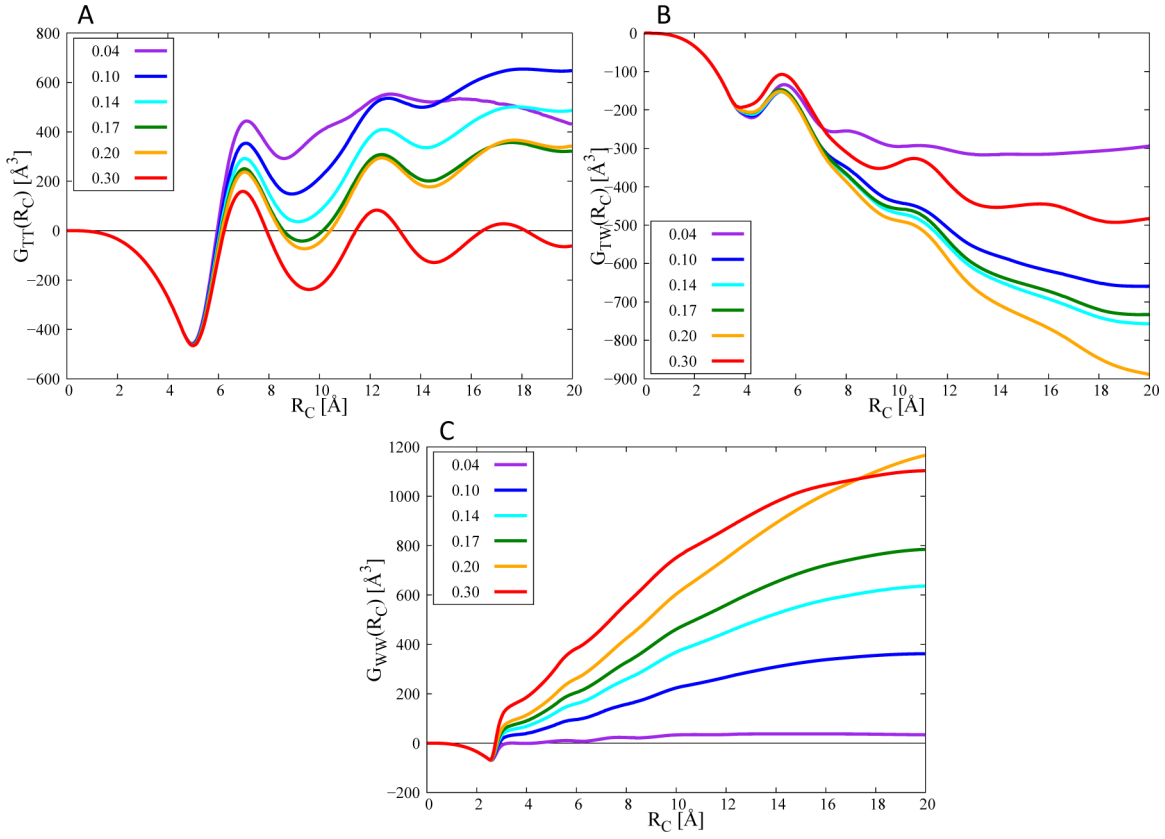


Figure 3. The value of the KB integral of (A) TBA−TBA as a function of R_C (see eq 21), (B) TBA−water, and (C) water−water. The mole fractions in Table 1 are displayed in different colors. At 20 Å, most of the curves are close to a plateau, indicating convergence.

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

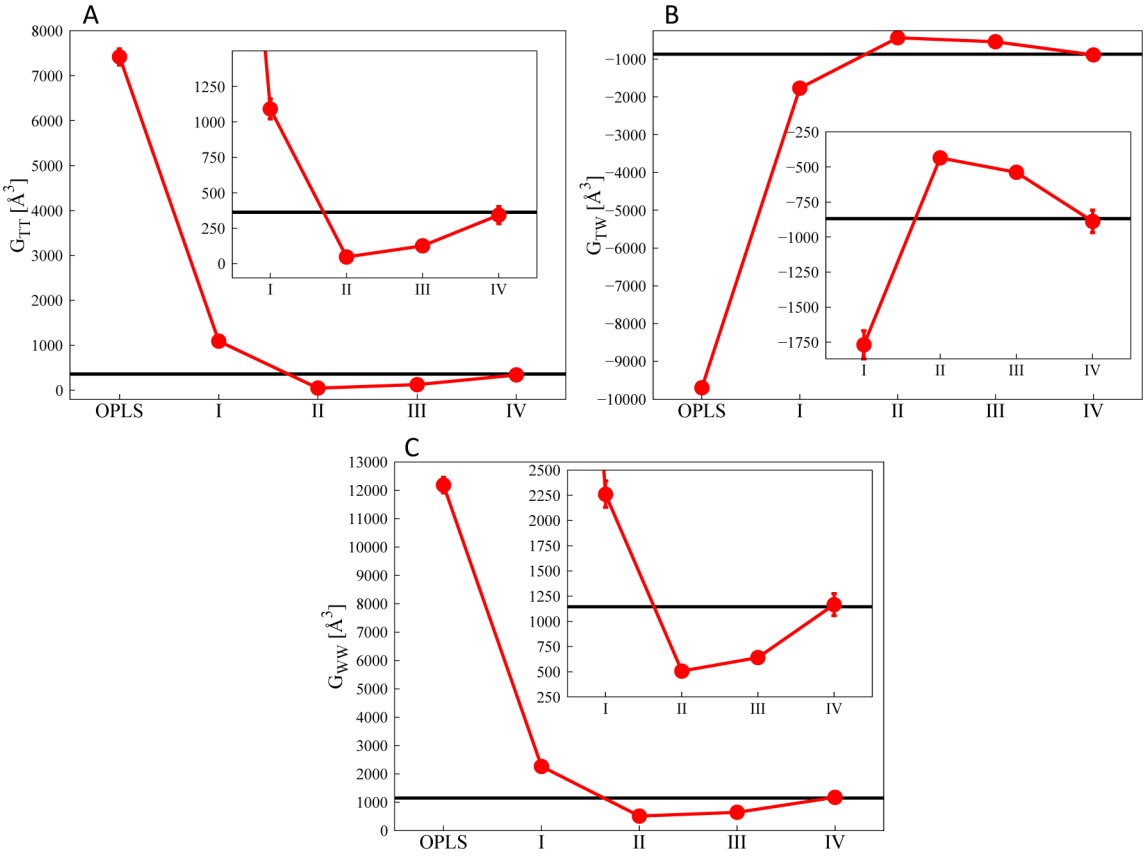


Figure 4. (A) KB integral for TBA−TBA at a mole fraction of TBA 0.2 as a function of the optimization iteration. The first data point corresponds to the KB integral calculated with the force field OPLSUA.[17] The data point IV corresponds to the final force field POP4ff, and the experimental value[19] is represented by the black horizontal line. In the inset, the same data are shown with a magnified scale for the last four data points. The error bars, computed with block analysis,[21] are sometimes below the size of the point. (B,C) Same data for TBA−water and water−water; even in this case, the final force field reproduces the experimental value.

The functional form for bonded terms (bonds, angles, and torsions) is

$$U_b = \frac{k_b}{2}(r - r_0)^2 \qquad U_\theta = \frac{k_\theta}{2}(\theta - \theta_0)^2$$

$$U_\phi = \sum_n K_n \cos(n\phi + \delta_n)$$

The nonbonded terms (LJ and electrostatic) are

$$U_{LJ} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad U_{elec} = k_{el}\frac{q_i q_j}{r_{ij}}$$

with combination rules

$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j} \qquad \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$$

Our software utilizes the equivalent formulation

$$U_{LJ} = \left(\frac{A_i A_j}{r_{ij}}\right)^{12} - \left(\frac{B_i B_j}{r_{ij}}\right)^{6}$$

Some of the results will be presented with respect to parameters A and B. Angles and bonds parameters were not optimized and are therefore shared between OPLSUA and POP4ff.

The torsions parameters were optimized in POP4ff but are very similar to the ones of OPLSUA; POP4ff final parameters were $(K_1, K_2, K_3) = (0.0001, -0.0003, 0.3258)$, while OPLSUA parameters were $(K_1, K_2, K_3) = (0, 0, 0.325)$. Hence, in practice, only the amplitude of the three-fold rotation is different from zero.

RESULTS AND DISCUSSION

TBA is a tertiary alcohol. Unlike the other butyl alcohols, TBA is miscible in water at any proportion and any temperature;[34] it is also the largest monohydric alcohol to be fully soluble.[35] TBA−water mixtures exhibit many anomalous physical properties. Solutions of TBA in water show an anomalously large volume contraction, indicating that the trimethyl groups must be somehow easily accommodated into the water structure.[36] There is evidence that TBA when added to water in solvating peptides behaves as a helix promoter.[37]

The peculiar characteristics of TBA motivate us to apply our newly developed procedure to investigate it and improve current potentials.

We first simulated the mixture of TBA and water using the parameters of the OPLSUA force field. After equilibration, phase separation is evident by visual inspection, as illustrated in Figure 2. This is a known effect as force fields optimized to reproduce pure liquid properties often exhibit too much self-aggregation when observed in solution.[38] This system was the starting point of our optimization. In the same figure, we show the same system once equilibrated with the optimized force field of the fourth iteration; we will refer to this force field as POP4ff.

In the optimization, we used KB integral extracted from small-angle X-ray scattering.[19] In calculating the KB integrals from MD

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

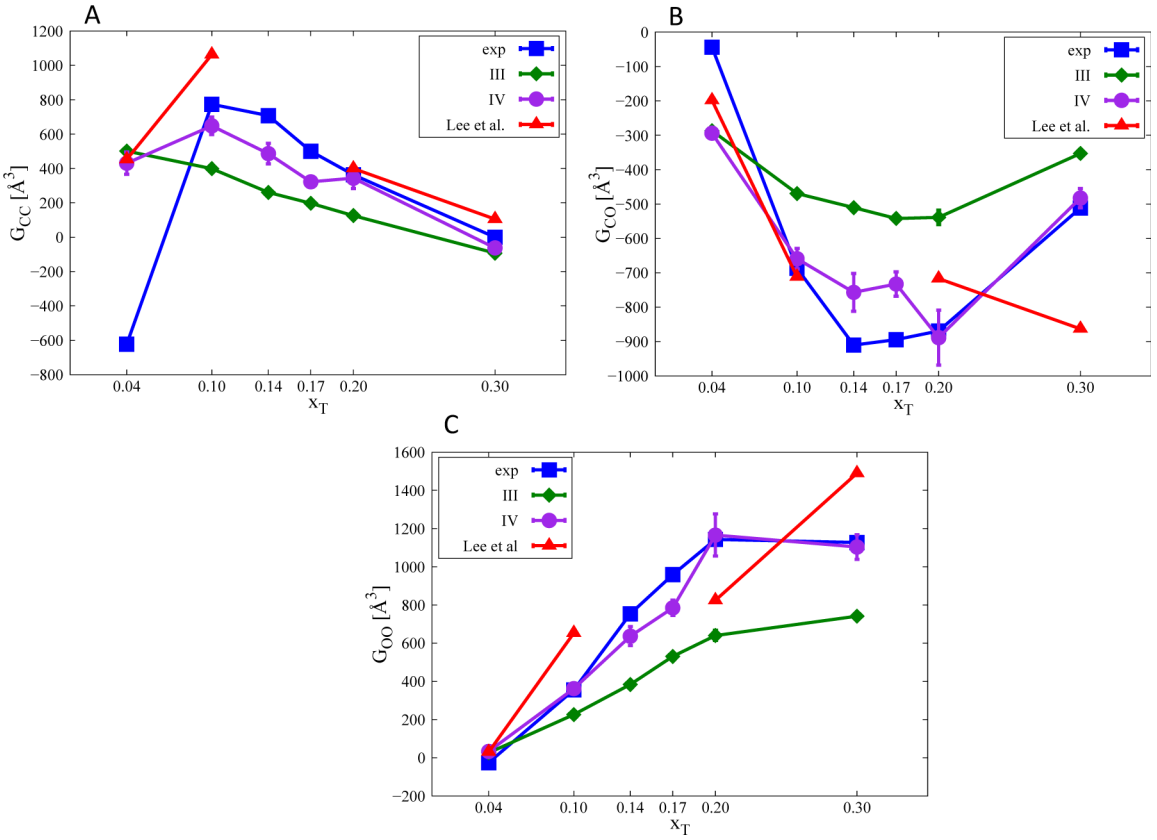


Figure 5. KB integrals as a function of TBA mole fraction. (A) TBA−TBA; (B) TBA−water; (C) water−water. The blue line shows the experimental results from ref 19, the red line shows results from Lee and Van Der Vegt potential,[15] the green line shows computational results from force field POP3ff, and the purple line shows computational results from force field POP4ff.

data, we assumed the position of the oxygen atom to be the position of water molecule; similarly, the position of the TBA molecule was assumed to be the one of the central carbon.

In Figure 3, we show the KB integrals for POP4ff as a function of the cutoff R_C for all of the concentrations tested in our simulations. At a distance of 20 Å the integrals are approaching a plateau, suggesting that we are close to the region where the integral is converged.

Lee and Van der Vegt[15] already used the KB theory to develop a force field for TBA, obtaining good results. They used the LJ parameters from GROMOS[39] and SPC[40] as a water model. They tuned the dipole moment of the TBA molecule so that they could better reproduce the KB integrals over a range of concentrations. Their protocol, while successful, shows the typical limitation of current force field development; they could adjust only a few parameters at the time. The choice of those parameters was left to chemical intuition. We repeated the optimization of the force field in an automated procedure using KB integrals as the optimization target for the POP algorithm. All of the parameters (excluding bonds and angles) of the model were subject to automated optimization following the gradient. The TBA molecule, in the united-atom model, is composed of three torsions (of one torsion type) and six atoms (of four different atom types). In OPLSUA, polar hydrogen atoms have a zero van der Waals radius. We kept this convention, and we did not optimize those parameters. The total number of parameters under optimization was 13.

Four iterations of the optimization procedure were conducted. The progression in approaching the experimental values through the optimization iterations is shown in Figure 4.

The first three iterations were conducted using as a target the KB integrals with the mole fraction of TBA at 0.2. The first optimization step significantly improved the original OPLSUA force field, as shown in Figure 4. Iterations II and III yielded smaller but significant improvements.

After the third iteration, the optimization procedure produced only minor improvements. We therefore decided to use a more informative target function. Inspection of the results over a broader range of concentrations suggested using as targets the KB integrals at TBA mole fractions of 0.04 and 0.10. The experimental KB integral of species TBA−TBA exhibits a global minimum at a TBA mole fraction of 0.04 and a global maximum at a TBA mole fraction of 0.10 (see squares in Figure 5). This feature is missing in the force field obtained after the third iteration (Figure 5). Iteration IV produced a significant improvement over the whole concentration range. The force field produced by the fourth iteration (POP4ff) reproduces very well the three KB integrals over a wide concentration range (0.04−0.3), outperforming both OPLSUA and the force field developed by Lee and Van Der Vegt. The KB integrals for the different force fields as functions of the concentration are shown in Figure 5. Note that only TBA mole fractions of 0.2, 0.04, and 0.1 were used at any time in the optimization, leaving us ample data for meaningful testing.

The pair correlation functions of TBA-TBA, TBA−water and water−water are shown in Figure 6. It is clear that the system is now well mixed because long-range correlations are absent. The pair correlation functions of TBA−TBA and TBA−water of OPLSUA (yellow in Figure 6) and POP4ff (purple in Figure 6) are significantly different; OPLSUA shows two nearby peaks,

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

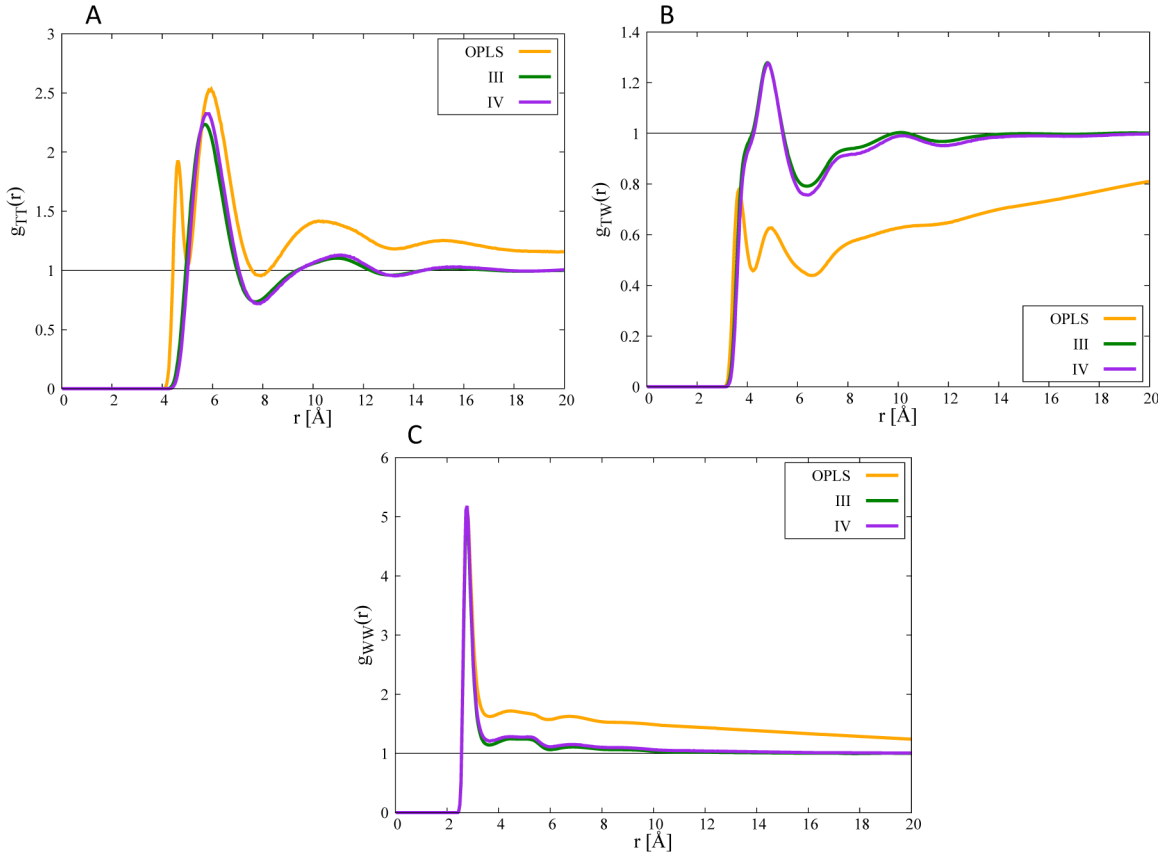


Figure 6. (A) Pair correlation function for species TBA−TBA; the yellow curve is the pair correlation function computed with the OPLS force field, the green curve is the pair correlation function computed with POP3ff, and the purple curve is the pair correlation function computed with force field POP4ff. (B,C) The same information as (A) for species TBA−water and water−water. All of the results were obtained at a TBA mole fraction of 0.20.

Table 2. Bonded parameters for United-Atom TBA for the Force Fields OPLSUA and POP4ff[a]

bonds	k_b (kcal/mol Å^2)		r_0 (Å)
O−H	553.0		0.945
C−O	320.0		1.430
C−CH_3	268		1.530
angles	k_θ (kcal/mol)		θ_0
CH_3 −C−CH_3	63.0		112.00
H−O−C	55.0		108.50
O−C−CH_3	80.0		108.00
torsions	K_1 (kcal/mol)	K_2 (kcal/mol)	K_3 (kcal/mol)
H−O−C− CH_3	0	0	0.325

[a]Bonded terms are the standard OPLS force field. Angles and bonds were not optimized; the change in torsions parameters was found to be small during the calculations, and their adjustment is ignored.

Table 3. Nonbonded Parameters for United-Atom TBA for the Force Fields OPLSUA and POP4ff

atom Type	q (e)	ε (kcal/mol)	σ (Å)
OPLS Parameters			
C	0.265	0.050	3.800
CH_3	0	0.160	3.910
O	−0.700	0.170	3.070
H	0.435	0	0
POP Parameters			
C	0.04670	0.0372	3.9899
CH_3	0.10900	0.1271	4.0618
O	−0.58197	0.1566	3.1104
H	0.20827	0	0

whereas POP4ff shows a single smooth peak. The pair correlation functions of force fields POP3ff (green in Figure 6) and POP4ff deviate only slightly. We did not include any information about the shape of the pair correlation functions in the target function; this change is a byproduct of the optimization procedure. Whether this is correct or not is difficult to say because we do not know the pair correlation function from experiment, only its integral, which hides such features.

The different interaction types that are optimized at once do not contribute in the same way to the optimization. The sensitivity vector g is defined as the gradient in parameter space of the target function in eq 24; it is a local feature of the target function, and it changes as the parameter set is improved by the optimization process. To provide useful information and following the procedure described in the Methods Section, we multiply the gradient by the scaling matrix D^{-1}, enforcing homogeneity in parameter space (see eq 9). We also normalize the sensitivity vector. In Table 4, we report the 13 different components of the normalized and scaled sensitivity obtained after POP analysis of the simulations carried out with the OPLSUA parameters, that is

$$\frac{g}{|g|} = \frac{\nabla^T \Theta(\pi_{\mathrm{OPLSUA}}) D^{-1}}{|\nabla^T \Theta(\pi_{\mathrm{OPLSUA}}) D^{-1}|} \tag{25}$$

Initially, the KB integrals are insensitive to the torsions parameters, mildly sensitive to the LJ parameters (as was also

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

Table 4. Scaled and Normalized Sensitivities[a] of the 13 Parameters Used in the Optimization

Torsion		
K_1	K_1	K_1
1.9×10^{-5}	1.9×10^{-5}	1.9×10^{-5}
van der Waals A		
C	CH_3	O
-6.5×10^{-3}	-4.2×10^{-2}	-3.0×10^{-2}
van der Waals B		
C	CH_3	O
2.7×10^{-2}	1.1×10^{-1}	8.6×10^{-2}
Charges		
C	CH_3	O	H
0.079	−0.86	0.36	0.31

[a]See eq 25.

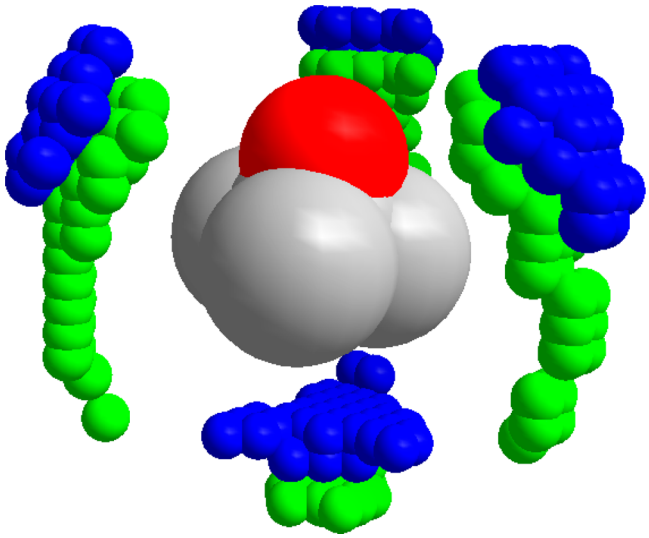


Figure 7. Regions with the highest density of TBA (green) and water (blue) around a central TBA molecule. The densities are measured on a grid of size 0.5 Å. As before, we assumed the position of the TBA molecule to be the position of the central carbon and the position of the oxygen to represent the center of the water molecule. We color by green the cells that have a density of TBA larger than 85% of the maximum density of TBA measured. We color by blue the cells that have a density of water larger than 75% of the maximum density of water. The methyl groups are well hydrated, as shown by the presence of a region of space with the high density of water just under those groups. Methyl groups of TBA are also found close to the hydroxyl group of other TBA molecules. The presence of both water and TBA around both the hydroxyl group and the methyl groups indicates the absence of hydrophobic interactions between methyl groups.

noted by Lee and van der Vegt[15] by direct testing), and highly sensitive to the charge distribution.

It has been noted[38] that to properly mimic the behavior of liquid mixtures, tuning only the dipole moment of the solute is not sufficient; it is required to find a solute charge distribution that represents higher-order moments of the charge distribution. Indeed, we find that this distribution is the most important feature affecting the KB integrals.

The dipole moment of TBA for OPLSUA is 2.28 D, and in our optimized force fields, it is 3.20 D. The difference in dipole moment is the result of a redistribution of partial charges involving all of the atoms; in OPLSUA, the methyl groups are neutral, and the positive charge is on the central carbon, while in our potential, the central carbon is almost neutral, and each of the methyl groups carry a small positive charge (see Table 3).

After the third iteration and using a target function that includes information on two concentrations, the sensitivity is significantly different. After the adjustment of the charge distribution, the improved force field shows the highest sensitivity to the LJ parameters. The last optimization step was indeed mainly a readjustment of the LJ parameters.

Finally, we extracted the angular dependence of the distribution of TBA and water around a central TBA molecule (see Figure 7). The blue dots represent regions with high water density and the green dots regions with high TBA density. The figure is roughly symmetric for rotations of 120° around the C−O axis of TBA. Because of steric repulsion, the high densities are in the grooves between these atoms. A region with high density of water is situated just under the methyl groups; contrary to what could be intuitive, the methyl groups, which are usually considered hydrophobic, are found to be well hydrated. Around the hydroxyl group, TBA tends to stay closer than water. Our model of TBA does not show hydrophobic interactions between methyl groups.

CONCLUSIONS

We provided a simple systematic procedure to optimize force fields to reproduce properties of liquid mixtures connected to the KB integrals.

We made a useful improvement to the original POP algorithm. The first version of POP was using gradient descent as a minimization algorithm. The introduction of the trust region Newton algorithm has improved the performance of POP in many ways. First of all, the Newton algorithm is known to have better convergence properties. Also, the concept of the trust region provides an easy and efficient way to assess the quality of the quadratic model used in the minimization. Finally, the use of a hyperelliptical trust region takes into account the different scales of magnitude present in the parameter set and allows modifications to the parameters that are homogeneous on a relative scale. We remark that the results shown in this paper were achieved with just four iterations of parameter adjustments.

We developed a new force field for TBA that approximates the behavior of mixtures of TBA with water better than force fields currently available, as shown by comparison to the experimental KB values. In the first three iterations of the optimization, we included in the target function only the KB integrals at one concentration (0.20 mole fraction of TBA). These iterations showed larger sensitivity toward the partial charges of TBA. In the last step of the optimization, we included in the target function the KB integrals of two lower concentrations (0.04 and 0.10 mole fraction of TBA). In this case, the charges, already optimized, did not change significantly, while most of the sensitivity was to LJ parameters. Even though our algorithm allowed for changes in torsional parameters, they remained essentially unchanged from the original OPLSUA parameters. Lee and Van der Vegt[15] observed in the past that the KB integrals depend more on a partial charge distribution than on LJ parameters. Similar observations were made in the context of urea parametrization with KB integrals.[12] Our results for the first three iterations confirm, strengthen, and quantify these previous observations (see Table 3 for sensitivities to different force field parameters) as they are the consequence of an automated optimization of all of the parameters at once. Nevertheless, we also observed that a fourth step of optimization of LJ parameters

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

was necessary to improve the force field beyond what we obtained from the first three iterations (see Figure 5).

Lee and Van Der Vegt[15] used the GROMOS[39] force field, which has a more general (and complex) type of LJ interaction. The pair interaction parameters A_{ij} and B_{ij} are not separable to single-atom parameters (i.e., A_{ij}=A_i·A_j) but depend instead on both indices. We illustrate here that the decomposable presentation, with a smaller number of parameters, works well. As a practical consequence, we note that separating the parameters describing pairs of interactions to products of two single-particle parameters makes it easier to apply the Ewald sum for LJ interactions and to obtain a more accurate description of long-range forces.

Other differences between the POP4ff force field and the force field developed by Lee and Van De Vegt[15] lie in the values of the parameters that in some cases are strikingly different; for example, the radius for the central carbon is ∼6 Å in the force field that they used, while in POP4ff, the same carbon atom has a dimension that deviates only slightly from the original OPLS atom type (∼4 Å). The charge distribution is also very different because they chose to keep the methyl groups neutral.

The fact that such diverse force fields produce similar results is a warning sign against optimizing liquid potentials solely according to KB data. The results are unlikely to be unique. This is also reflected in the comparison of two different potentials (POP3ff and control) reported in Appendix B. Therefore, when we refine the parameters using KB integrals, we need either to make sure that the changes to the force field are minimal or to use a larger pool of observables to ensure compatibility toward other observables.

Finally, we want to stress two important features of our optimization method. First, our method minimizes the target function with minimal changes to the parameter set. This is an important feature, given that force field parameters were already optimized extensively in the past. If we need to adjust these parameters against a new set of data, small adjustments are to be preferred because they are less likely to perturb results of previous refinement of the current set of potential parameters. Second, our method does not require larger simulation time if the set of observables used for the optimization is increased. The change in the parameters is obtained as the postprocessed analysis of one MD trajectory. This will make it easier and faster to optimize the force field against a larger pool of observables whenever such a large pool is available.

APPENDIX A: ALGORITHMS

We describe the trust region Newton method (TRNM) used in POP; for more details, see refs 22 and 23. At iteration k, p_k is the increment to the parameter set, and ρ_k is the following ratio

$$\rho_k = \frac{\text{actual reduction}}{\text{predicted reduction}} = \frac{\Theta(\pi_k) - \Theta(\pi_k + p_k)}{m_k(0) - m_k(p_k)}$$

The following algorithm prescribes how to iteratively update the trust region radius Δ_k.

Algorithm (*Trust Region*)

Given $\Delta_{\max} > 0, \Delta_0 \in (0, \Delta_{\max})$ and $\eta \in \left[0, \frac{1}{4}\right)$

for $k = 0,1,2,\ldots$

Find p_k *solving the quadratic sub-problem;*

Evaluate ρ_k*;*

if $\rho_k < \frac{1}{4}$

$\Delta_{k+1} = \frac{1}{4}\|p_k\|$;

else

if $\rho_k > \frac{3}{4}$ *and* $\|p_k\| = \Delta_k$

$\Delta_{k+1} = \min(2\Delta_k, \Delta_{\max})$;

else

$\Delta_{k+1} = \Delta_k$;

if $\rho_k > \eta$

$\pi_{k+1} = \pi_k + p_k$;

else

$\pi_{k+1} = \pi_k$;

end(for)

η is a tuning parameter (kept fixed through the iterations) that is used to optimize the performance of the algorithm; in our case, given the small number of iterations performed, we could not study the efficiency of the algorithm for several values of η. In practice, we used the value $\eta = 0$.

As defined in the Methods Section, our quadratic subproblem is the following

$$p_k = \arg\min m_k(p) \qquad \text{s.t.} \|Dp\| \le \Delta_k$$

$$m_k(p) = \Theta(\pi_k) + \nabla^T\Theta(\pi_k)p + \frac{1}{2}p^T\nabla\nabla^T\Theta(\pi_k)p$$

We first cast this elliptical trust region problem in the canonical form of an equivalent spherical trust region problem in the variable $\gamma_k = Dp_k$

$$\gamma_k = \arg\min m_k(\gamma) \qquad \text{s.t.} \|\gamma\| \le \Delta_k$$

$$m_k(\gamma) = \Theta(\pi_k) + \nabla^T\Theta(\pi_k)D^{-1}\gamma + \frac{1}{2}\gamma^T D^{-T}\nabla\nabla^T\Theta(\pi_k)D^{-1}\gamma$$

Then we apply the inverse transformation to obtain $p_k = D^{-1}\gamma_k$.

For a simpler notation, let us define $B = D^{-T}\nabla\nabla^T\Theta(\pi_k)D^{-1}$ and $g^T = \nabla^T\Theta(\pi_k)D^{-1}$; we recall that the matrix B is symmetric by construction.

The following theorem gives a precise characterization of the solution of spherical trust region problem.

Theorem:

The vector γ^* is a global solution of the trust region problem

$$\min_{\gamma\in\mathbb{R}^n} m(\gamma) = f + g^T\gamma + \frac{1}{2}\gamma^T B\gamma \qquad \text{s.t.} \|\gamma\| \le \Delta$$

Table 5. Nonbonded Parameters for POP3ff[a]

atom type	q (e)	ε (kcal/mol)	σ (Å)
C	0.04670 (0.45702)	0.0499 (0.1155)	3.8010 (4.0471)
CH_3	0.10920 (0)	0.1595 (0.0645)	3.9119 (4.3646)
O	−0.58180 (−0.56038)	0.1696 (0.5831)	3.0695 (2.5736)
H	0.20750 (0.10336)	0 (0)	0 (0)

[a]In brackets are the parameters for control.

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

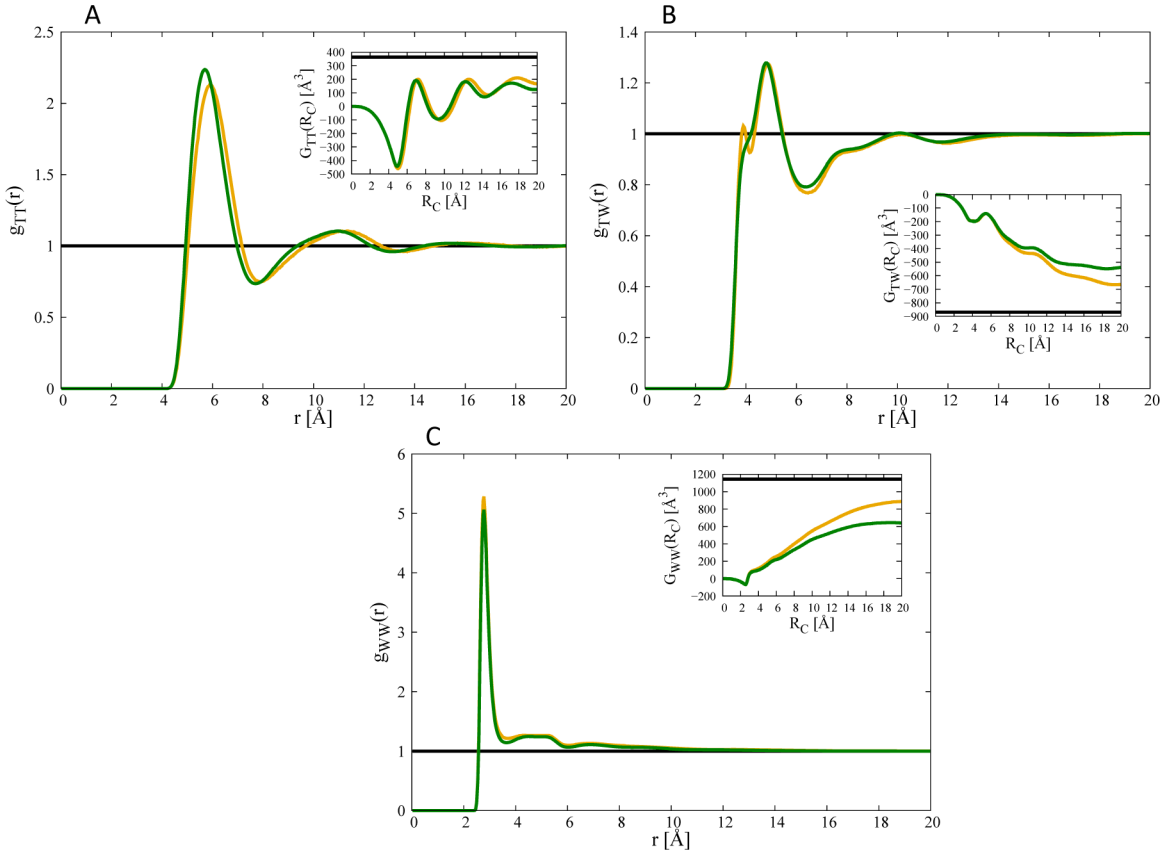


Figure 8. Pair correlation function of (A) TBA−TBA, (B) TBA−water, and (C) water−water. The green line represents POP3ff and the yellow line the control force field. In the insets, the KBIs are reported as a function of the cutoff distance. In black is shown the experimental value.

if and only if $\|\gamma^*\| \leq \Delta$ and there is a scalar $\lambda \geq 0$ such that the following conditions are satisfied:

$$(B + \lambda I)\gamma^* = -g$$

$$\lambda(\Delta - \|\gamma^*\|) = 0$$

$(B + \lambda I)$ is positive semidefinite

To find the solution for γ^*, we need to find the scalar λ that satisfies the conditions in the theorem and then solve the linear system.

An iterative algorithm that performs both the tasks is the following:

Algorithm (*Exact Trust Region*)

Given $\lambda^{(0)}, \Delta > 0$

for $l = 0,1,2,...$

$$\text{Factor } B + \lambda^{(l)} I = R^T R$$
$$\text{Solve } R^T R\gamma_l = -g,\ R^T q_l = \gamma_l$$
$$\text{Set } \lambda^{(l+1)} = \lambda^{(l)} + \left(\frac{\|\gamma_l\|}{\|q_l\|}\right)^2 \left(\frac{\|\gamma_l\| - \Delta}{\Delta}\right)$$

end(for)

The initial $\lambda^{(0)}$ is set to be zero if the matrix B is positive definite. If the matrix B is indefinite, it is instead set to be $\lambda^{(0)} = -\min\{\lambda_1^-, \lambda_2^-, ...\} + 0.00001$, where $\{\lambda_1^-, \lambda_2^-, ...\}$ are the negative eigenvalues of B. In this way, $B + \lambda^{(0)} I$ is always positive definite, and it is therefore possible to perform its Cholesky factorization.[41]

APPENDIX B

On the Uniqueness of the Potential Derived from KB Integrals

The optimization as described may depend on the initial conditions, producing potentials that are consistent with the experimental observables that we examined but not unique. To explore the uniqueness of the potential, we conducted another optimization (we will refer to it as the "control") starting from another set of parameters and for slightly different conditions. Bond and angle parameters were left as is. The initial torsion parameters were the same as those in OPLSUA (see Table 2). Also, the initial charges were chosen to be the OPLSUA (see Table 3). The initial LJ parameters for the central carbon of TBA were set to (ε, σ) = (0.14450 kcal/mol, 3.96 Å), while the initial ones of the methyl groups were set to (ε, σ) = (0.2940 kcal/mol, 3.73 Å). The initial LJ parameters for the oxygen and hydrogen of TBA's hydroxyl group were chosen to be the OPLSUA ones (see Table 3). We ran the same system as the one in Table 1 for a molar fraction of TBA of 0.20 but with a box size slightly larger (65.505^3 Å^3), corresponding to the average one found by Lee and van der Vegt for the same system.[15]

The first run of this system showed a phase separation and the presence of a vacuum bubble in the periodic box. The optimization was carried out with three iterations. At the beginning of each iteration, we ran 5 ns of simulation using the replica exchange algorithm,[42,43] using 50 replicas equally spaced between 300 and 422.5 K. After 1 ns of equilibration, we collected 4000 structures, equally spaced in time and ran POP to optimize the force field parameters. In this case, we left the methyl groups uncharged, as it is in OPLSUA. At the second

dx.doi.org/10.1021/jp505401m | J. Phys. Chem. B 2015, 119, 836−849

iteration, our protocol gave us a negative charge for the TBA hydrogen of the hydroxyl group. We decided that such a result was unphysical; therefore, we discarded the optimization of the charge of the hydrogen atom at the second and third iterations. The final nonbonded parameters that we derived with this method are reported in the Table 5.

The charge distribution results in a dipole moment of 3.03 D, slightly lower than the one found for the POP3ff and POP4ff (3.20 D). The torsion parameters of the control optimization changed more than the POP3ff and POP4ff parameters. The final result was $(K_1, K_2, K_3) = (0.0019, -0.0032, 0.5580)$.

Figure 8 shows the pair correlation functions and the integral of the pair correlation functions for POP3ff and the result of the control optimization.

We optimize a potential for TBA using two different protocols. The initial conditions for the parameters were different, and a constraint on the charges of the methyl groups was applied in only one case. Nevertheless the KB integrals were computed at similar accuracy in both cases (see insets). The shape of the pair correlation functions is remarkably close. The largest qualitative difference between the control (yellow) and the POP3ff (green) is perhaps in the first peak of the TBA−TIP3 pair correlation function for the control, which is missing in the POP3ff.

Both potentials are capable of reproducing the KBIs with comparable accuracy, and the pair correlation functions obtained with these two potentials are remarkably similar. This shows that the different potential can produce not only similar KB integrals but also similar pair correlation functions.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by NIH Grant GM59796 and Welch Grant F-1783 to R.E. Texas Advanced Computing Center (TACC) is gratefully acknowledged for resources used in these calculations.

REFERENCES

(1) Zhao, G. P.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J. Y.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; et al. Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics. *Nature* **2013**, *497* (7451), 643−646.

(2) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330* (6002), 341−346.

(3) Lange, O. F.; van der Spoel, D.; de Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* **2010**, *99* (2), 647−655.

(4) (a) Shi, Y.; Xia, Z.; Zhang, J. J.; Best, R.; Wu, C. J.; Ponder, J. W.; Ren, P. Y. Polarizable Atomic Multipole-Based Amoeba Force Field for Proteins. *J. Chem. Theory Comput* **2013**, *9* (9), 4046−4063. (b) Savelyev, A.; MacKerell, A. D. J. All-Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model. *J. Comput. Chem.* **2014**, *35*, 1219−1239.

(5) Mackerell, A. D. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25* (13), 1584−1604.

(6) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.

(7) Di Pierro, M.; Elber, R. Automated Optimization of Potential Parameters. *J. Chem. Theory Comput* **2013**, *9* (8), 3311−3320.

(8) Ruymgaart, A. P.; Cardenas, A. E.; Elber, R. MOIL-Opt: Energy-Conserving Molecular Dynamics on a GPU/CPU System. *J. Chem. Theory Comput* **2011**, *7* (10), 3072−3082.

(9) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.

(10) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMm: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545−1614.

(11) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113* (26), 9004−9015.

(12) Weerasinghe, S.; Smith, P. E. A Kirkwood−Buff Derived Force Field for Mixtures of Urea and Water. *J. Phys. Chem. B* **2003**, *107* (16), 3891−3898.

(13) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. 1. *J. Chem. Phys.* **1951**, *19* (6), 774−777.

(14) Weerasinghe, S.; Smith, P. E. Kirkwood−Buff Derived Force Field for Mixtures of Acetone and Water. *J. Chem. Phys.* **2003**, *118* (23), 10663−10670.

(15) Lee, M. E.; van der Vegt, N. F. A. A New Force Field for Atomistic Simulations of Aqueous Tertiary Butanol Solutions. *J. Chem. Phys.* **2005**, *122* (11), 114509.

(16) Wang, L. P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013**, *117* (34), 9956−9972.

(17) Jorgensen, W. L. Optimized Intermolecular Potential Functions for Liquid Alcohols. *J. Phys. Chem.* **1986**, *90* (7), 1276−1284.

(18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.

(19) Nishikawa, K.; Kodera, Y.; Iijima, T. Fluctuations in the Particle Number and Concentration and the Kirkwood−Buff Parameters of *tert*-Butyl Alcohol and Water Mixtures Studied by Small-Angle X-ray Scattering. *J. Phys. Chem.* **1987**, *91* (13), 3694−3699.

(20) Abrams, J. B.; Tuckerman, M. E.; Martyna, G. J. Equilibrium Statistical Mechanics, Non-Hamiltonian Molecular Dynamics, and Novel Applications from Resonance-Free Timesteps to Adiabatic Free Energy Dynamics. In *Lecture Notes in Physics, "Computer Simulations in Condensed Matter: From Material to Chemical Biology"*; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer, 2006; Vol. *703*, pp 139−192.

(21) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2002; Vol. *22*, p 638.

(22) More, J. J.; Sorensen, D. C. Computing a Trust Region Step. *Siam J. Sci. Stat. Comput.* **1983**, *4* (3), 553−572.

(23) Nocedal, J.; Wright, S. J. *Numerical Optimization*; Springer: New York, 1999; Vol. *20*, p 636.

(24) Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum Press: New York, 1992; Vol. *28*, p 697.

(25) Chitra, R.; Smith, P. E. Properties of 2,2,2-Trifluoroethallol and Water Mixtures. *J. Chem. Phys.* **2001**, *114* (1), 426−435.

(26) Bennaim, A. Inversion of Kirkwood−Buff Theory of Solutions — Application to Water−Ethanol System. *J. Chem. Phys.* **1977**, *67* (11), 4884−4890.

(27) (a) Weerasinghe, S.; Smith, P. E. A Kirkwood−Buff Derived Force Field for Sodium Chloride in Water. *J. Chem. Phys.* **2003**, *119* (21), 11342−11349. (b) Weerasinghe, S.; Smith, P. E. A Kirkwood−Buff Derived Force Field for the Simulation of Aqueous Guanidinium Chloride Solutions. *J. Chem. Phys.* **2004**, *121* (5), 2180−2186. (c) Weerasinghe, S.; Smith, P. E. A Kirkwood−Buff Derived Force Field for Methanol and Aqueous Methanol Solutions. *J. Phys. Chem. B* **2005**, *109* (31), 15080−15086.

(28) Ganguly, P.; van der Vegt, N. F. A. Convergence of Sampling Kirkwood−Buff Integrals of Aqueous Solutions with Molecular

Dynamics Simulations. *J. Chem. Theory Comput* **2013**, *9* (3), 1347−1355.
(29) Mukherji, D.; van der Vegt, N. F. A.; Kremer, K.; Delle Site, L. Kirkwood−Buff Analysis of Liquid Mixtures in an Open Boundary Simulation. *J. Chem. Theory Comput* **2012**, *8* (2), 375−379.
(30) Fritsch, S.; Poblete, S.; Junghans, C.; Ciccotti, G.; Delle Site, L.; Kremer, K. Adaptive Resolution Molecular Dynamics Simulation through Coupling to an Internal Particle Reservoir. *Phys. Rev. Lett.* **2012**, *108* (17), 170602.
(31) Egorov, G. I.; Makarov, D. M. Densities and Volume Properties of (Water Plus *tert*-Butanol) over the Temperature Range of (274.15 to 348.15) K at Pressure of 0.1 MPa. *J. Chem. Thermodyn* **2011**, *43* (3), 430−441.
(32) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577−8593.
(33) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular-Dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990−2001.
(34) Nishikawa, K.; Hayashi, H.; Iijima, T. Temperature-Dependence of the Concentration Fluctuation, the Kirkwood−Buff Parameters, and the Correlation Length of *tert*-Butyl Alcohol and Water Mixtures Studied by Small-Angle X-ray Scattering. *J. Phys. Chem.* **1989**, *93* (17), 6559−6565.
(35) Visser, C. D.; Perron, G.; Desnoyers, J. E. Heat-Capacities, Volumes, and Expansibilities of *tert*-Butyl Alcohol−Water Mixtures from 6 to 65 °C. *Can. J. Chem.* **1977**, *55* (5), 856−862.
(36) Kusalik, P. G.; Lyubartsev, A. P.; Bergman, D. L.; Laaksonen, A. Computer Simulation Study of *tert*-Butyl Alcohol. 2. Structure in Aqueous Solution. *J. Phys. Chem. B* **2000**, *104* (40), 9533−9539.
(37) Gallardo, I. F.; Webb, L. J. Demonstration of α-Helical Structure of Peptides Tethered to Gold Surfaces Using Surface Infrared and Circular Dichroic Spectroscopies. *Langmuir* **2012**, *28* (7), 3510−3515.
(38) Ploetz, E. A.; Bentenitis, N.; Smith, P. E. Developing Force Fields from the Microscopic Structure of Solutions. *Fluid Phase Equilib.* **2010**, *290* (1−2), 43−47.
(39) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53a5 and 53a6. *J. Comput. Chem.* **2004**, *25* (13), 1656−1676.
(40) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Ralation to Protein Hydration. In *Intermolecular Forces*; Springer: The Netherlands, 1981.
(41) Trefethen, L. N.; Bau, D. *Numerical Linear Algebra*; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1997; Vol. *12*, p 361.
(42) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141−151.
(43) Marinari, E.; Parisi, G. Simulated Tempering — A New Monte-Carlo Scheme. *Europhys. Lett.* **1992**, *19* (6), 451−458.