Kirkwood-Buff Integrals for Aqueous Urea Solutions Based upon the Quantum Chemical Electrostatic Potential and Interaction Energies

Shuntaro Chiba,1\* Tadaomi Furuta,2 and Seishi Shimizu3

1Education Academy of Computational Life Sciences, Tokyo Institute of Technology, J3-141 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501 Japan

2School of Life Science and Technology, Tokyo Institute of Technology, B-62 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan.

3York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom

**ABSTRACT**

Cosolvents, such as urea, affect protein folding and binding, and the solubility of solutes. The modeling of cosolvents has been facilitated significantly by the rigorous Kirkwood-Buff (KB) theory of solutions, which can describe structural thermodynamics over the entire composition range of aqueous cosolvent mixtures based only on the solution density and the KB integrals (KBIs), i.e., the net excess radial distribution functions from the bulk. Using KBIs to describe solution thermodynamics has given rise to a clear guideline that an accurate prediction of KBIs is equivalent to accurate modeling of cosolvents. Taking urea as an example, here we demonstrate that an improvement in the prediction of KBIs comes from an improved reproduction of high-level quantum chemical (QC) electrostatic potential and molecular pairwise interaction energies. This rational approach to the improvement of the KBI prediction stems from a comparison of existing force fields, AMOEBA, and the generalized AMBER force field (GAFF), as well as the further optimization of the former to enable better agreement with QC interaction energies. Such improvements would pave the way towards a rational and systematic determination of the transferable force field parameters for a number of important small molecule cosolvents.

1. **Introduction**

Cosolvents, when added to water, are capable of shifting the equilibria of chemical processes, such as biomolecular conformation, protein-ligand binding, and solubility.1-5 Equilibrium shifting effect of cosolvents has been described using statistical thermodynamics in the form of the rigorous Kirkwood-Buff (KB) theory. The KB theory describes the thermodynamics of mixing, as well as the cosolvent-induced shifts in equilibria, by the use of the key quantity, the KB integral (KBI), i.e., the net excess radial distribution function (RDF) from the bulk, which can be defined for all possible combinations of molecular species in the solution. Since KBIs can be determined from experiments,6 an accurate computational modeling of liquid mixtures inevitably means an accurate reproduction of the experimental KBIs. The KB force fields (KBFF),7 widely used in computer simulations,8-15 have been developed to achieve this goal directly. Yet the process of force field optimization has been done purely by trial and error.7 Here we demonstrate an accurate modeling of cosolvents’ electrostatic potentials (ESP) and pairwise interaction energies, both of which can be calculated by high-level quantum chemistry (QC), lead to good agreement with experimental KBIs over a wide composition range.

We have chosen an aqueous urea mixture as a target, whose importance cannot be overemphasized; urea denaturation is routinely exploited in biochemical laboratories, as well as in the study of protein folding and stability;16 urea is also a well-known solubilizer.17 Despite its widespread applications, however, the precise molecular-based mechanism of the processes outlined above is still a matter of intense scientific investigation.18-20 Indeed, a recent recalculation of the transfer free energies for the side chain and the main chain analogues, and the subsequently revised picture that both the main chain and the side chain contribute to the urea-induced protein stability modifications, have demonstrated the subtlety of the problem.19-20

To have a comprehensive molecular-based understanding of how urea works, not only an extensive molecular simulation, but also an accurate force field model for urea and water are indispensable. While many force field models for urea have been developed,21-28 the prediction of the solution structure was shown to depend sensitively on the self-aggregation of urea molecules.29-30 The KBFF for urea developed by Weerasinghe and Smith has successfully overcome the aggregation problem with results agreeing with the experimental solution density and KBIs at a urea concentration of 8.02 mol L-1.30

The thusly determined KBFF for urea has been widely applied to a number of biologically and industrially important molecules.8-15 Determination of KBFFs for other important cosolvents would be greatly desirable, yet KBFFs are available only for a limited number of solvent molecules,31-39 because optimizing each atom’s van der Waals (vdW) parameters and fixed point charges7 through trial and error to reproduce experimental KBI and solution density is an enormous undertaking. Moreover, transferability of such parameters to a wider variety of molecules in general are unfortunately not clear, which severely limits the applicability of these parameters. Hence, even for aqueous urea solutions, a rational approach to force field determination, aiming at transferability and an accurate reproduction of KBIs, is indispensable.

Transferability of force field parameters has been addressed by the generalized AMBER force field (GAFF) 40-41 as well as by the polarizable force field AMOEBA,42-43 both of which can be applied systematically to small molecules.40-41, 44 Recently, several studies employed polarizable force field to improve reproducibility of KBIs;15, 45-47 some of these studies showed a good agreement between experimental and calculated KBIs. However, the force field parameters employed by these studies have been optimized by utilizing the experimental data, such as the heat of vaporization, density, gas phase dipole moment. In contrast, AMOEBA combined with a tool for automated force field determination provides a systematic way to determine force field parameters directly from quantum mechanics without the aid of these experimental values.44 As discussed later, AMOEBA is superior to GAFF because of incorporation of atomic multipole moments and induced dipole moments. However, AMOEBA alone cannot yield accurate KBIs because of its deviations in pairwise molecular interaction energies from high level quantum chemical (QC) calculations. Indeed, further optimization of pairwise molecular interaction energies leads to better agreement with experimental KBIs.

**2. Kirkwood-Buff integrals: theory and calculation**

**2.1. KBIs and thermodynamic quantities**

For the purpose of reproducing the solution structure and thermodynamics of a urea-water mixture through extensive molecular dynamics (MD) simulations, the following are key: (i) the Kirkwood-Buff integral (KBI) calculation and (ii) the relationship between KBIs and thermodynamic quantities.6 The KBI, defined in Equation 1 in terms of number fluctuations, outlines the relationship between the structure and the thermodynamics of a solution:6

(1)

where is averaged volume of the system, is the number of species , stands for the ensemble average in the grand canonical ensemble, and is the Kronecker delta, respectively. Eq. (1) implies the existence of a link between the KBI and thermodynamic quantities. This link has widely been exploited for the microscopic interpretation of thermodynamic data in terms of the KBIs. What is particularly useful is the alternative expression of KBIs via the radial distribution function (RDF) in Equation 2:6

, (2)

where the RDF, , is defined in the grand canonical ensemble between the species and as a function of (the center-of-mass (COM) distance between the two) and is Avogadro’s number, respectively. Intuitively speaking, the KBI is the overall measure of attraction or repulsion of species. If a species is overall attracted to species , the RDF exhibits peaks (higher than the bulk baseline, i.e., 1) that contribute positively to the KBI. If a species is overall repelled or excluded from species , the RDF exhibits troughs (lower than the baseline, 1) that contribute negatively to the KBI. Shorter distances always result in negative contributions due to steric repulsion between species and .48

The formulae relating the KBIs to thermodynamic quantities are useful for comparing the computationally obtained KBIs with experimental values. The KBIs have the following relationship6 to volumetric quantities, such as partial molar volume, (Equation 3) and isothermal compressibility, (Equation 4):

(3)

(4)

where and are the Boltzmann constant and temperature, respectively. (Equation 5) and (Equation 6) can be defined in terms of the KBIs and number density, of the species as:

(5)

(6)

The KBIs are not only useful in providing volumetric information but also the thermodynamics of mixing, such as the mole fraction activity coefficients of the species , and molar excess Gibbs energy of the mixture, , using Equation 7:

(7)

where is the mole fraction of species *i*.

Eqs. (3)-(7) can be inverted to yield the following expressions, useful in the calculation of the KBIs based upon experimental thermodynamic data:

(8)

(9)

(10)

Eqs. (8)-(10) were used to calculate the experimental KBIs for urea water mixture, which are shown in Table S1 of the Supporting Information. The derivation procedure of experimental and is described in Section S1 of the Supporting Information.

**2.2. Calculation of the KBIs**

Following Weerasinghe and Smith,32 as well as others,49-50 we have adopted the approximation that the KBI can be approximated by integrating between 0 to *r* (instead of 0 to ), as outlined in Equation 11:

(11)

Note also that the RDF of the NPT ensemble has been used in this approximation instead of that of the grand canonical ensemble; under this approximation, RDFs can be readily obtained using isothermal-isobaric MD simulation. Then the numerical integration of the RDF to yield KBIs, which has been a standard approach in the calculation of KBIs, has been carried out. Such integration is referred to as the running-KB integral (RKBI), in which numerical integration can be carried out by a straightforward use of the simple rectangle method, however, it suffers from inaccuracy due to numerical error arising from the finite bin size. To rectify this inaccuracy, we have adopted a direct calculation of KBIs, which follows directly from Eq. (11) and the definition of RDF, and will hereafter be referred to as the direct KBI (DKBI). To use DKBI, the number of molecule pairs with distances of less than or equal to have to be counted. The final KBIs were obtained by averaging KBIs calculated from individual MD simulations. The numerical difference between DKBI and RKBI is shown in the APPENDIX. Following Weerasinghe and Smith,32 we used averaged KBIs between 0.95 and 1.30 nm for analysis in the Results and Discussion section.

**3. Computational details**

In this study, the force field of urea was determined and optimized using the state-of-the-art computational method, the AMOEBA model, which has the advantage of being general and does not require force field adjustments through trial and error for the reproduction of the experimental KBIs. We also employed the generalized AMBER force field (GAFF), 40-41 which is a widely-used general force field. AMOEBA-based simulations were carried out using Tinker version 7.01,51 and GAFF-based simulations were carried out using AMBER 1452 with a GPU-accelerated code.53-54 Quantum chemical (QC) calculations for electrostatic potential (ESP) fitting were carried out using Gaussian 09.55

**3.1. Parameter assignment**

Here we detail how the force field parameters for urea were obtained. Electrostatics parameters of AMOEBA urea were assigned by Poltype version 1.1.3,44 which operated multiple calculation steps automatically starting with an input structure of urea. Note that ambiguities still persist regarding the precise structure of urea in the condensed phase.56-60 Here, the sp3-like pyramid structure was assumed for both nitrogen atoms. This is because planar urea has two imaginal frequencies and would not be appropriate for ESP fitting. First, the given structure was optimized at MP2/6-311G(d, p) level of theory, followed by the vibrational frequency calculation to confirm the optimized structure was not at a saddle point. The optimized structure’s density matrix at the same level of theory was inputted into the Distributed Multipole Analysis (DMA version 2.2.09),61 in order to compute partial charges, dipole moments, and quadrupole moments. These values were averaged based on the molecular symmetry, and were refitted to reproduce the ESP calculated at MP2/aug-cc-pVTZ level of theory using Potential module of TINKER with the convergence threshold of 0.1. During the refitting process, partial charges were not changed. The vdW parameter and polarizability of a dipole moment of each atom were assigned by Poltype. The valence parameters were initially assigned by Poltype and manually changed by comparing them to corresponding valence parameters of acetamide deposited in the TINKER parameter file (amoeba09.prm).42-43 The final parameter used in the present study is shown in Section S2 of the Supporting Information. We employed the AMOEBA water42 deposited in the TINKER parameter file directory for the water model.

The electrostatic parameters of the GAFF urea were assigned using the restricted electrostatic potential (RESP) fitting.62-63 The initial urea structure was optimized at HF/6-31G(d) level of theory, followed by the vibrational frequency calculation to confirm the optimized structure was not at a saddle point. Using the ESP, partial charges were computed using the Antechamber module of AMBER. The valence parameters were assigned by the Antechamber and additional improper torsion parameters were added by the Parmchk module. We employed SPC/E water64 for the water model.

**3.2. Sampling of structures**

The initial systems, covering a wide range of urea concentrations as shown in Table 1, were built to be a cubic shape using the Genbox module of GROMACS version 5.0.5.65-66 Sampling of the structures of solutions consisting of AMOEBA urea and water was carried out in the following manner. The initial system was minimized using the Minimize module of TINKER, with a convergence threshold of 41 kJ mol-1 nm-1 of a root mean square gradient, followed by a 1-ns equilibrium MD run using the Dynamics module. After equilibration, the system density and potential energy reached a plateau with the box sizes around 4.4−4.9 nm. To obtain statistically valid results, we performed several production MD runs for structure sampling with different total simulation times (as shown in Table 1) starting from different velocities, with a sampling interval of 2.5 ps. At low concentrations, statistical uncertainty in the KBI calculation was reported.32, 67 During the equilibrium and production runs we employed the Nosé-Hoover extended system isothermal-isobaric algorithm68 to integrate the equation of motion with a temperature of 298 K, pressure of 0.101325 MPa, and a time step of 0.5 fs. The vdW interaction was cut off at 1.2 nm, and the long range vdW energy correction was applied.69 The particle mesh Ewald for point multipoles algorithm70 was employed to calculate electrostatic interaction energy with a real space cutoff of 0.7 nm, 54 charge grids in the x, y, and z directions, and the β-spline interpolation order of 8 (default).

As for sampling used for the GAFF-based systems, a similar procedure was employed, i.e., the Sander module for minimization and the Pmemd module accelerated by the GPU code for MD simulation. Each individual run was performed for 20.5 ns and the last 20 ns was used for structure sampling. Relatively low calculation cost of AMBER enabled to extend simulation in order to attain statistical convergence of the results. During the MD runs, the equation of motion was integrated using the Langevin dynamics69 at 298 K, a collision frequency of 2 ps-1, and a time step of 2 fs. The lengths between hydrogen-hydrogen atoms and hydrogen-oxygen atoms were constrained using the SHAKE algorithm.71 The pressure was controlled by the Monte Carlo barostat with a target pressure of 0.101325 MPa. The vdW interaction cutoff was 1.0 nm and vdW parameters of different atom types were assigned using the Lorentz/Berthelot mixing rule. The particle mesh Ewald72-73 was employed to calculate electrostatic interaction energy with a real space cutoff of 1.0 nm, 48 charge grids in the x, y, and z directions, and the β-spline interpolation order of 4 (default).

In Section 4.4 we conducted MD simulations using the KBFF for urea.32 For these simulations, we replaced some of the parameters mentioned above to be consistent with those from the KBFF method outlined in Reference 28; i.e., the vdW interaction cutoff of 1.5 nm and the geometric mixing rule for the vdW parameter assignment were employed.32

**4. Predicting Kirkwood-Buff integrals through accurate partial charges and pairwise interaction energies**

**4.1. Better charge fitting leads to an accurate KBI prediction: AMOEBA *versus* GAFF**

Our goal is to establish a rational and systematic approach towards determining a transferable force field parameter for urea, which can reproduce experimental KBIs. To this end, it is imperative to identify the key factor, which affects the accuracy of KBI calculations. For the purpose of identifying the key factor, we start by comparing the two widely-used force fields, AMOEBA and GAFF, in the calculations of KBI.

**4.1.1. AMOEBA is superior to GAFF**

KBIs calculated using AMOEBA are closer to the experimental KBIs than the ones using GAFF, as can be clearly seen in Figure 1 (the raw data of KBIs are listed in Tables S1 (experimental) and S2 (calculated) of the Supporting Information). The difference between the experimental and the AMOEBA-calculated and are less than 12 cm3 mol-1, which is comparable to the experimental error of approximately 10 cm3 mol-1.74 As for , AMOEBA is in good agreement with the experimental values at high urea concentrations (6.20-7.83 mol L-1), whereas at lower concentrations AMOEBA reproduces the urea concentration dependence of in a qualitative manner only. These errors at lower concentrations are discussed in Section 4.2 and 4.3. Note that the calculated RDFs converge for AMOEBA and GAFF (see Figures S1 and S2 of the Supporting Information).

**4.1.2. AMOEBA, optimized for the gas-phase urea monomer, can nevertheless reproduce the thermodynamics of solution mixture reasonably well**

Indeed, even though the AMOEBA force field was not optimized to reproduce solution thermodynamic properties of aqueous urea solutions, AMOEBA can nevertheless reproduce experimental thermodynamic quantities, as demonstrated by Figure 2, in which solution density , partial molar volumes (Eq. (3)), the activity derivative (Eq. (7)), isothermal compressibility (Eq. (4)), and excess coordination numbers , are plotted against urea concentration. Figure 2 also shows that AMOEBA is superior to GAFF in terms of reproducibility of almost all the thermodynamic quantities. We also derived isothermal compressibility using the volume fluctuation,69 which gave closer values to the experimental values75 as can be seen in Figure 2. The isothermal compressibility of the neat water of SPC/E and AMOEBA were consistent with previous computationally-derived values for SPC/E76 and for AMOEBA.77-78 The discrepancy in the isothermal compressibility calculated with Eq. (4) and the volume fluctuation could be related to statistical uncertainty of these values, as reported previously.35 However, the reason for the discrepancy is not clear. Because the isothermal compressibility makes little contribution to the KBIs, this discrepancy would not affect our conclusions. The errors in the isothermal compressibility of AMOEBA values relative to experimental values would originate from AMOEBA water parameters,42 because the errors in all the urea concentrations are almost the same and would originate the error for neat water. Recently developed water models for AMOEBA,77-78 which can reproduce these experimental isothermal compressibility and other thermodynamic properties, can be used to lessen the error.

**4.2. Accurate pairwise interaction energies lead to improved KBIs**

We hypothesize that the accuracy in the KBI calculations using AMOEBA comes from the accuracy in modeling pairwise interaction energies. To examine the validity of this hypothesis, we calculated the urea-urea, urea-water, and water-water interaction energies using AMOEBA and GAFF, and compared them to the corresponding energies in the gas phase calculated using QC.

**4.2.1. Calculation details of QC pairwise energies**

The pair structures were extracted from the trajectories of the 0.51-mol L-1 solution system that were used to calculate the corresponding KBIs in Section 4.1. The dimer and monomer structures were fixed during the calculation. We extracted 500 pairs with a COM distance between 0-0.5 nm and 500 pairs with a COM distance between 0.5-0.7 nm, respectively. For the QC calculation, we employed MP2/aug-cc-pVTZ level of theory with the counterpoise correction79-80 to be consistent with the way the AMOEBA force field was derived.43

**4.2.2. AMOEBA outperformed in terms of reproduction of QC pairwise energies**

Superiority of AMOEBA in reproducing the QC pairwise interaction energies is demonstrated in Table 2, which lists the correlation coefficients, *R,* and root mean squared errors (RMSE) for AMOEBA (original) and GAFF. AMOEBA is superior in terms of all the statistical parameters, i.e. *R* came closer to 1 and RMSE to 0, to GAFF. Further, the robustness indicated by the statistical parameters for AMOEBA indicates that it reproduced pairwise interaction energies well. It is worth noting that improvements in RMSE for interaction energies (u-u, u-w, w-w) of AMOEBA compared to GAFF may seem slight, 3.80 versus 2.38, 3.72 versus 1.33, and 2.81 versus 0.877 kJ mol-1, respectively. However, these improvements are reflected significantly in the KBI calculations.

In conducting molecular simulations of a heterogeneous system, it is essential to balance solvent-solvent, solvent-solute, and solute-solute interaction energies.81 The discrepancy between experimental and calculated KBIs would be attributed to the imbalance. We believe that improving reproducibility of pairwise interaction energies would further improve KBIs and related thermodynamic quantities.

**4.2.3. Incorporating polarizable effect was essential to reproduce QC pairwise interaction energies.**

The improvement in reproducing QC interaction energies is achieved by incorporating the polarizable effect, which is realized by the presence of induced dipoles. To show the significance of induced dipole inclusion, we calculated the correlation coefficient and RMSE using AMOEBA without the induced dipole with the pair structures described above (the used trajectory was “AMOEBA original 0.5 mol L-1”). The results are included in Table 2. When induced dipoles were absent from the calculations – as in AMOEBA original and AMOEBA without the induced dipole – all relevant parameters deteriorated.

In Figure 3, differences between AMOEBA and the QC interaction energies, , of the original calculation and the one without the induced dipole are plotted as a function of the COM distances, with the RMS of the errors for urea-urea, urea-water, water-water interactions superimposed. The RMS values of using the AMOEBA "Original" were smaller than those using the AMOEBA "without induced dipole" in most cases, which suggest that incorporation of the induced dipole corrects interaction energies of adjacent molecules. In spite of such a correction, urea-urea interaction energies ranging between 0.3 to 0.4 nm were deteriorated with the incorporation of the polarizable effect, which indicates that there is a room for improving induced dipole and/or ESP parameters. However, the concentration of dimers in this range was small – 12% – making it unclear if improving the parameters can affect KBI calculations.

In addition, we confirmed that the solution density calculated employing AMOEBA without the induced dipole also significantly deviated from the experimental density (See Section S5 of the Supporting Information).

**4.3. Optimizing pairwise interaction energies makes KBIs more accurate**

We have shown that a relatively small improvement in pairwise interaction energy significantly improves the KBI calculation. Here, we show that a further improvement can be achieved by optimizing parameters for the pairwise interaction energy. We carried out the optimization only for AMOEBA, because of its superiority to GAFF in the reproduction of KBIs and other thermodynamic quantities. In the previous sections we employed the original vdW parameters, which were transferred from the existing parameters (amoeba09.prm)43 to keep their transferability. The vdW parameters assigned were the same as those of acetamide. However, further improvement could be achieved by refining the parameters to reproduce the QC pairwise interaction energy. We used 1000 urea-water structures and their energies and simultaneously optimized the vdW parameters using the Newton-Raphson method with NonlinearModelFit function in Mathematica version 10.0.82

**4.3.1. Optimization procedure of vdW parameters**

We optimized the vdW parameters to reproduce the difference between the QC interaction energy and the electrostatic term of AMOEBA interaction energy, based on the interaction energy of urea-water dimers. The vdW energy, , is described by the buffered-14-7 vdW interactions:42

where is the absolute well depth of vdW interaction potential defined between two *i* atoms, occurring at the distance, between these atoms. is the distance between atoms *i* and *j*. As for a hydrogen atom that connects to a nitrogen atom, its position is modified towards the nitrogen atom to reduce their bond distance by a reduction factor before calculating . The change in position was applied only to the vdW calculation. The reduction factor was also optimized. The original , , and the reduction factor are shown in Table 3. During the fitting we applied a set of restrictions outlined in Table 3 to maintain parameters not to deviate from the original force field. The upper and lower limits of the restriction reflect the maximum and minimum values for each element in the existing parameters, amoeba09.prm, in which only polar hydrogen parameters were considered for the hydrogen atom.

**4.3.2. Optimized vdW parameters were verified to be applicable to wide composition range**

The optimized vdW parameters gave better estimation of pairwise interaction energies as shown in the “vdW Opt (AMOEBA original 0.5 mol L-1)” section of Table 2. We first compare the results for the “original” and “vdW Opt” for the urea-water interaction, because these interaction energies were used for the optimization. The vdW optimization improved the correlation coefficient *R* from 0.983 to 0.985 and the RMSE from 1.33 kJ mol-1 to 1.27 kJ mol-1, respectively. For urea-urea interaction energy, the vdW optimization improved RMSE from 2.38 kJ mol-1 to 2.32 kJ mol-1. The improvement for urea-urea pairs validates the optimization as the urea-urea interaction energies were not part of the optimization process. The optimized vdW parameters are shown in Table 3 (vdW Opt). We also confirmed that these optimized parameters can be used at a higher concentration as outlined in Table 2, which shows improvement of interaction energies for urea-urea and urea-water dimers whose structures were extracted from the simulation at 7.83 mol L-1. Note that when employing the optimized parameters to a system consisting of compounds other than urea and water, verification of applicability of the parameters would be required.

**4.3.3. Improved pairwise interactions lead to better KBIs**

With the goal of evaluating how much KBIs and related thermodynamic quantities change with the parameter optimization, we conducted MD simulations to calculate the KBIs and related thermodynamic quantities using the same sampling procedure as described in Section 3.2. Calculated densities improved as seen in Figure 4a. RMSE of calculated densities relative to the experimental density improved to 7.1×10-3 g mL-1 from 8.7×10-3 g mL-1. In addition, the calculated KBIs improved overall as can be seen in Figure 4b. Interestingly, the optimization hardly affects the KBIs of the higher-concentration systems (6.20- and 7.83-mol L-1 systems). This is consistent with the observations of Weerasinghe and Smith30 as they found that the KBIs of urea at 8.02 mol L-1 were relatively insensitive to slight variations of vdW parameters, *σ* (±0.01 nm) and *ε* (±0.1 kJ mol-1). On the other hand, at low concentrations (0.51- to 3.1-mol L-1) improved significantly (approximately 27-54%) without deteriorating and even improving and . We conclude that reproducing QC interaction energies of urea and water molecules improves the quality of the parameters and reproduces the experimental KBIs. These findings for optimization provide an important insight for developing FF parameters for novel molecules. In addition, we observed improvements in calculated thermodynamic quantities including the densities, partial molar volumes of urea, the activity derivatives, and all the coordination numbers as shown in Figure S5 of the Supporting Information. The KBIs and the relevant thermodynamic properties are tabulated in Table S2 and S5 of the Supporting Information, respectively.

Recently Bin et al. demonstrated that inclusion of polarizability lead to improvements in calculating KBIs by MD simulation of an aqueous methanol solution, an N-methylacetamide solution, and an acetamide solution modeled with the Drude polarizable force field.46-47 However, these models had difficulties in reproducing the activity derivatives. The performance of the Drude polarizable model can be rationalized by examining pairwise interaction energies and improvements can be achieved by optimizing force field parameters to reproduce dimer integration energies as employed in this study.

**4.4. Optimized AMOEBA parameters reproduced KBIs comparable to the experimentally-derived parameters.**

It is useful to know how well the experimentally-derived parameter for urea, the KBFF for urea, 32 performs for the current systems compared to the original and optimized AMOEBA. We conducted MD simulations to derive KBIs and thermodynamic quantities using the same urea solution systems employing the procedure described in Section 3.2.

The KBFF for urea gave the best performance in reproducing solution densities relative to experimental values over the wide composition range as shown in Figure 4a. The RMSE observed was 8.5 × 10-4 g mL-1,which was ten times smaller than that of vdW-optimized AMOEBA. The KBIs were comparable to those of the vdW-optimized AMOEBA parameters as seen in Figure 4b. However, we observed that the KBIs of the KBFF urea also suffered from disagreement at lower concentrations. This might be because the parameter set was optimized to reproduce experimental KBIs at 8.02 mol L-1. The KBIs and related thermodynamic properties are given in Table S2 and Table S6 of the Supporting Information.

It should be noted that the parameters for AMOEBA original and optimized AMOEBA were determined systematically to reproduce the ESP of urea at MP2/aug-cc-pVTZ level of theory for the gas phase monomer and pairwise interaction energies for gas phase dimers, and were not optimized for the reproduction of the experimental solution density nor KBIs. The better performance of KBFF has nothing to do with reproducibility of pairwise interaction energy as confirmed in Table 2. Since the parameters of KBFF urea were determined through trial and error to reproduce the experimental density and KBIs, they describe condensed-state properties through effective potential.

Since AMOEBA parameters can be improved further to reproduce pairwise interaction energies achieving an error comparable to that of AMOEBA water dimers (1 kJ mol-1), we believe the systematic improvement would give better KBIs and thermodynamic parameters.

**5. Conclusion**

Accuracy in the computational prediction of mixed solvents can be assessed by the agreement with the experimental Kirkwood-Buff integrals (KBIs) and solution density. Taking aqueous urea mixture as an example, we have demonstrated that a better agreement with experimental KBIs can be achieved by an accurate modeling of the electrostatic potential (ESP) and molecular pairwise interaction energies calculated by high-level quantum chemical (QC) calculations.

Previously, an accurate prediction of KBIs has been achieved by the optimization of atomic partial charges and vdW parameters through trial and error30 to reproduce experimental solution density and KBIs at a certain urea concentration, which has limited the availability of KBFFs to a wider range of cosolvents as well as the transferability of atomic force field parameters. Our rational strategy will pave the way towards a set of transferable atomic force field parameters which can accurately predict KBIs.

The rationale for our approach has risen from a comparison between the two common force fields: the generalized AMBER force field (GAFF) and the polarizable force field (AMOEBA). The superiority of AMOEBA over GAFF was attributed to the former’s better modeling of the QC-calculated ESP and incorporation of polarizability (which is realized using the induced dipole). Moreover, an even better agreement with experimental KBIs was achieved through a better reproduction of the gas-phase molecular interaction energies calculated by QC.

Thus, we now have a rational benchmark for optimizing force field parameters for the prediction of KBIs and solution density: an accurate modeling of ESP with induced dipole and molecular pairwise interaction energies, both of which can be calculated in the gas phase via high-level QC.

**APPENDIX**

We calculated the KBI without numerically integrating the RDF. Useful technical information is gained by comparing the RKBI and DKBI. Figure 5 shows differences between KBIs calculated using RKBI and DKBI. The maximum error was 3 cm3 mol-1 and was smaller than experimental error74 but not negligible. Since the errors of and depend on concentration, adverse effects arise when the relationship between KBIs and concentration is studied. The errors accumulate for a short distance in all the cases with the same curve shape. In a range where RDF = 0, the difference between RKBI and DKBI originates entirely from the difference between the numerical summation of sphere shell volumes and the analytical volume of the sphere, i.e. , indicating that the RKBI would cause a significant error in the calculation of KBIs for a solute with a large radius. Accumulation of error stopped and the error reached a plateau for a range of longer distances as the numerical error decreases when RDF moves closer to unity according to Eq. (11). Other numerical integration rules, such as the trapezoidal rule, the Simpson’s rule, and the cubic spline interpolation, would alleviate the error. Note that we employed the most primitive numerical integration rule. However, since any integration method bears numerical error, the DKBI gives better estimates of KBIs. Alternatively, as has been used in recent studies,15, 49 it is useful to employ the Schnell’s sub-box method, by which KBIs could be obtained by direct analysis of molecule number fluctuations in small sub-boxes embedded in a larger reservoir without numerical integration of RDF.50

ASSOCIATED CONTENT

**Supporting Information**. Experimental thermodynamic quantities and derivatives of molar fraction activity, the parameters for urea molecule, the radial distribution functions and Kirkwood-Buff integrals, the thermodynamic quantities calculated with AMOEBA without polarizable effect. The Supporting Information is available free of charge on the http://pubs.acs.org. at DOI:

AUTHOR INFORMATION

Corresponding Author

\*E-mail: chiba.s.ac@m.titech.ac.jp. Phone: +81-3-5734-3779

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We are grateful to financial support from the FY2016 Program to Support International Publication, the Program for Promoting the Enhancement of Research Universities by MEXT.

REFERENCES

1. Timasheff, S. N., Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99*, 9721-6.

2. Timasheff, S. N., Protein Hydration, Thermodynamic Binding, and Preferential Hydration. *Biochemistry* **2002,** *41*, 13473-13482.

3. Canchi, D. R.; García, A. E., Cosolvent Effects on Protein Stability. *Annu. Rev. Phys. Chem.* **2013,** *64*, 273-293.

4. Shimizu, S.; Matubayasi, N., Hydrotropy: monomer-micelle equilibrium and minimum hydrotrope concentration. *J. Phys. Chem. B* **2014,** *118*, 10515-24.

5. Shimizu, S., Estimating hydration changes upon biomolecular reactions from osmotic stress, high pressure, and preferential hydration experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 1195-9.

6. Ben-Naim, A., *Molecular Theory of Solutions*. OUP Oxford: U.K., 2006.

7. Weerasinghe, S.; Gee, M. B.; Kang, M.; Bentenitis, N.; Smith, P. E., Developing Force Fields from the Microscopic Structure of Solutions: The Kirkwood–Buff Approach. In *Modeling Solvent Environments*, Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 55-76.

8. Klimov, D. K.; Straub, J. E.; Thirumalai, D., Aqueous urea solution destabilizes Abeta(16-22) oligomers. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 14760-5.

9. Canchi, D. R.; Paschek, D.; García, A. E., Equilibrium Study of Protein Denaturation by Urea. *J. Am. Chem. Soc.* **2010,** *132*, 2338-2344.

10. Ploetz, E. A.; Smith, P. E., Infinitely dilute partial molar properties of proteins from computer simulation. *J. Phys. Chem. B* **2014,** *118*, 12844-54.

11. Chen, F.; Smith, P. E., Theory and Computer Simulation of Solute Effects on the Surface Tension of Liquids. *J. Phys. Chem. B* **2008,** *112*, 8975-8984.

12. Kokubo, H.; Pettitt, B. M., Preferential solvation in urea solutions at different concentrations: Properties from simulation studies. *J. Phys. Chem. B* **2007,** *111*, 5233-5242.

13. Kokubo, H.; Rosgen, J.; Bolen, D. W.; Pettitt, B. M., Molecular basis of the apparent near ideality of urea solutions. *Biophys. J.* **2007,** *93*, 3392-407.

14. Mercadante, D.; Milles, S.; Fuertes, G.; Svergun, D. I.; Lemke, E. A.; Grater, F., Kirkwood-Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. *J. Phys. Chem. B* **2015,** *119*, 7975-84.

15. Ploetz, E. A.; Rustenburg, A. S.; Geerke, D. P.; Smith, P. E., To Polarize or Not to Polarize? Charge-on-Spring versus KBFF Models for Water and Methanol Bulk and Vapor-Liquid Interfacial Mixtures. *J. Chem. Theory Comput.* **2016,** *12*, 2373-87.

16. Anfinsen, C. B., Principles that Govern the Folding of Protein Chains. *Science* **1973,** *181*, 223-230.

17. Booth, J. J.; Abbott, S.; Shimizu, S., Mechanism of hydrophobic drug solubilization by small molecule hydrotropes. *J. Phys. Chem. B* **2012,** *116*, 14915-21.

18. Bolen, D. W.; Rose, G. D., Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.* **2008,** *77*, 339-62.

19. Horinek, D.; Netz, R. R., Can simulations quantitatively predict peptide transfer free energies to urea solutions? Thermodynamic concepts and force field limitations. *J. Phys. Chem. A* **2011,** *115*, 6125-36.

20. Moeser, B.; Horinek, D., Unified description of urea denaturation: backbone and side chains contribute equally in the transfer model. *J. Phys. Chem. B* **2014,** *118*, 107-14.

21. Kuharski, R. A.; Rossky, P. J., Molecular Dynamics Study of Solvation in Urea-Water Solution *J. Am. Chem. Soc.* **1984,** *106*, 5786-5793.

22. Boek, E. S.; Briels, W. J., Molecular dynamics simulations of aqueous urea solutions: Study of dimer stability and solution structure, and calculation of the total nitrogen radial distribution function GN(r). *J. Chem. Phys.* **1993,** *98*, 1422.

23. Duffy, E. M.; Severance, D. L.; Jorgensen, W. L., Urea: Potential Functions, log P, and Free Energy of Hydration. *Isr. J. Chem.* **1993,** *33*, 323-330.

24. Hernández-Cobos, J.; Ortega-Blake, I.; Bonilla-Marín, M.; Moreno-Bello, M., A refined Monte Carlo study of aqueous urea solutions. *J. Chem. Phys.* **1993,** *99*, 9122.

25. Aastrand, P. O.; Wallqvist, A.; Karlstroem, G., Molecular Dynamics Simulations of 2 m Aqueous Urea Solutions. *J. Phys. Chem.* **1994,** *98*, 8224-8233.

26. Tsai, J.; Gerstein, M.; Levitt, M., Keeping the shape but changing the charges: A simulation study of urea and its iso-steric analogs. *J. Chem. Phys.* **1996,** *104*, 9417.

27. Engkvist, O.; Price, S. L.; Stone, A. J., Developments in computational studies of crystallization and morphology applied to urea. *Phys. Chem. Chem. Phys.* **2000,** *2*, 3017-3027.

28. Kallies, B., Coupling of solvent and solute dynamics—molecular dynamics simulations of aqueous urea solutions with different intramolecular potentials. *Phys. Chem. Chem. Phys.* **2002,** *4*, 86-95.

29. Sokolić, F.; Idrissi, A.; Perera, A., Concentrated aqueous urea solutions: A molecular dynamics study of different models. *J. Chem. Phys.* **2002,** *116*, 1636.

30. Weerasinghe, S.; Smith, P. E., Cavity formation and preferential interactions in urea solutions: Dependence on urea aggregation. *J. Chem. Phys.* **2003,** *118*, 5901.

31. Weerasinghe, S.; Smith, P. E., Kirkwood–Buff derived force field for mixtures of acetone and water. *J. Chem. Phys.* **2003,** *118*, 10663.

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

33. Weerasinghe, S.; Smith, P. E., A Kirkwood–Buff derived force field for sodium chloride in water. *J. Chem. Phys.* **2003,** *119*, 11342.

34. Weerasinghe, S.; Smith, P. E., A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. *J. Chem. Phys.* **2004,** *121*, 2180-6.

35. Weerasinghe, S.; Smith, P. E., A Kirkwood−Buff Derived Force Field for Methanol and Aqueous Methanol Solutions. *J. Phys. Chem. B* **2005,** *109*, 15080-15086.

36. Kang, M.; Smith, P. E., A Kirkwood-Buff derived force field for amides. *J. Comput. Chem.* **2006,** *27*, 1477-85.

37. Bentenitis, N.; Cox, N. R.; Smith, P. E., A Kirkwood-Buff derived force field for thiols, sulfides, and disulfides. *J. Phys. Chem. B* **2009,** *113*, 12306-15.

38. Ploetz, E. A.; Smith, P. E., A Kirkwood-Buff force field for the aromatic amino acids. *Phys. Chem. Chem. Phys.* **2011,** *13*, 18154-67.

39. Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weeerasinghe, S.; Smith, P. E., A Kirkwood-Buff Derived Force Field for Aqueous Alkali Halides. *J. Chem. Theory Comput.* **2011,** *7*, 1369-1380.

40. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25*, 1157-74.

41. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006,** *25*, 247-60.

42. Ren, P.; Ponder, J. W., Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003,** *107*, 5933-5947.

43. Ren, P.; Wu, C.; Ponder, J. W., Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011,** *7*, 3143-3161.

44. Wu, J. C.; Chattree, G.; Ren, P., Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012,** *131*, 1138.

45. Zhong, Y.; Patel, S., Nonadditive Empirical Force Fields for Short-Chain Linear Alcohols: Methanol to Butanol. Hydration Free Energetics and Kirkwood-Buff Analysis Using Charge Equilibration Models. *J. Phys. Chem. B* **2010,** *114*, 11076-11092.

46. Lin, B.; Lopes, P. E.; Roux, B.; MacKerell, A. D., Jr., Kirkwood-Buff analysis of aqueous N-methylacetamide and acetamide solutions modeled by the CHARMM additive and Drude polarizable force fields. *J. Chem. Phys.* **2013,** *139*, 084509.

47. Lin, B.; He, X.; MacKerell, A. D., Jr., A comparative Kirkwood-Buff study of aqueous methanol solutions modeled by the CHARMM additive and Drude polarizable force fields. *J. Phys. Chem. B* **2013,** *117*, 10572-80.

48. Shimizu, S.; Boon, C. L., The Kirkwood-Buff theory and the effect of cosolvents on biochemical reactions. *J. Chem. Phys.* **2004,** *121*, 9147-55.

49. 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*, 1347-1355.

50. Schnell, S. K.; Liu, X.; Simon, J. M.; Bardow, A.; Bedeaux, D.; Vlugt, T. J.; Kjelstrup, S., Calculating thermodynamic properties from fluctuations at small scales. *J. Phys. Chem. B* **2011,** *115*, 10911-8.

51. Ponder, J. W., *TINKER—Software Tools for Molecular Design*, version 7.01; Ponder Research Group, Department of Chemistry, Washington University, St. Louis, MO,2014.

52. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; T.E. Cheatham, I.; Darden, T. A.; Duke, R. E.; Gohlke, H., et al. *AMBER 14*, University of California, San Francisco: 2014.

53. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013,** *9*, 3878-3888.

54. Le Grand, S.; Götz, A. W.; Walker, R. C., SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013,** *184*, 374-380.

55. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A., et al. *Gaussian 09,* Revision D.01; Gaussian, Inc.: Wallingford, CT, 2013.

56. Soper, A. K.; Castner, E. W.; Luzar, A., Impact of urea on water structure: a clue to its properties as a denaturant? *Biophys. Chem.* **2003,** *105*, 649-666.

57. Ishida, T.; Rossky, P. J.; Castner, E. W., A Theoretical investigation of the shape and hydration properties of aqueous urea: Evidence for nonplanar urea geometry. *J. Phys. Chem. B* **2004,** *108*, 17583-17590.

58. Hermida-Ramón, J. M.; Öhrn, A.; Karlström, G., Planar or Nonplanar:  What Is the Structure of Urea in Aqueous Solution? *J. Phys. Chem. B* **2007,** *111*, 11511-11515.

59. Godfrey, P. D.; Brown, R. D.; Hunter, A. N., The shape of urea. *J. Mol. Struct.* **1997,** *413–414*, 405-414.

60. Bertran, C. A.; Cirino, J. J. V.; Freitas, L. C. G., Theoretical studies of concentrated aqueous urea solutions using computacional Monte Carlo simulation. *J. Braz. Chem. Soc.* **2002,** *13*, 238-244.

61. Stone, A. J., Distributed Multipole Analysis:  Stability for Large Basis Sets. *J. Chem. Theory Comput.* **2005,** *1*, 1128-1132.

62. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993,** *97*, 10269-10280.

63. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A., Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993,** *115*, 9620-9631.

64. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The missing term in effective pair potentials. *J. Phys. Chem.* **1987,** *91*, 6269-6271.

65. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995,** *91*, 43-56.

66. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4:  Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008,** *4*, 435-447.

67. Ploetz, E. A.; Bentenitis, N.; Smith, P. E., Developing Force Fields from the Microscopic Structure of Solutions. *Fluid Phase Equilib.* **2010,** *290*, 43.

68. Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L., Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996,** *87*, 1117-1157.

69. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. Oxford University Press: 1987.

70. Sagui, C.; Pedersen, L. G.; Darden, T. A., Towards an accurate representation of electrostatics in classical force fields: efficient implementation of multipolar interactions in biomolecular simulations. *J. Chem. Phys.* **2004,** *120*, 73-87.

71. Andersen, H. C., Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983,** *52*, 24-34.

72. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N⋅ log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**.

73. 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*, 8577-8593.

74. Chitra, R.; Smith, P. E., Molecular Association in Solution:  A Kirkwood−Buff Analysis of Sodium Chloride, Ammonium Sulfate, Guanidinium Chloride, Urea, and 2,2,2-Trifluoroethanol in Water. *J. Phys. Chem. B* **2002,** *106*, 1491-1500.

75. Endo, H., The Adiabatic Compressibility of Nonelectrolyte Aqueous Solutions in Relation to the Structures of Water and Solutions. *Bull. Chem. Soc. Jpn.* **1973,** *46*, 1106-1111.

76. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014,** *5*, 3863-3871.

77. 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*, 9956-72.

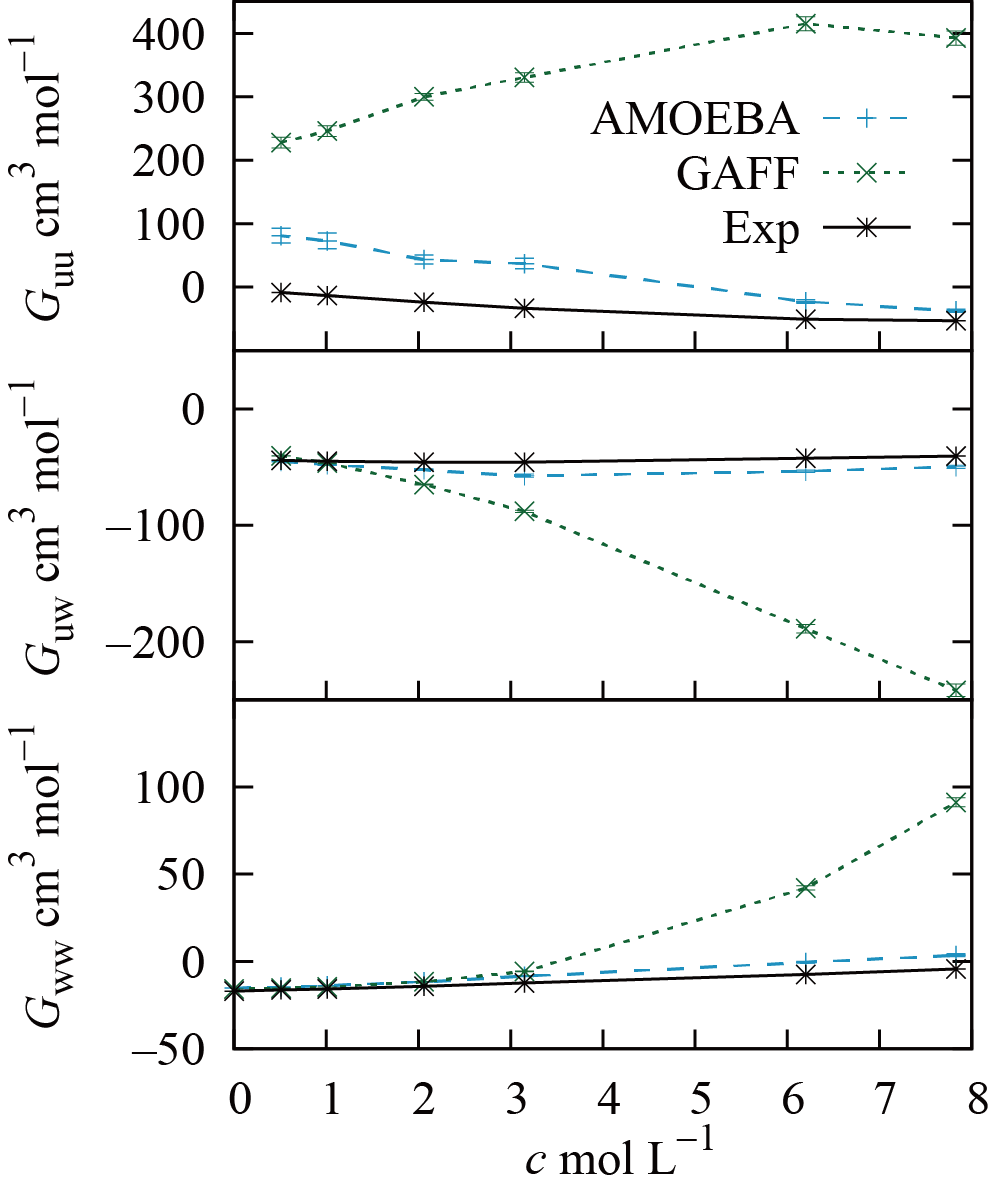
78. Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W., Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B* **2015,** *119*, 9423-37.

79. Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970,** *19*, 553-566.

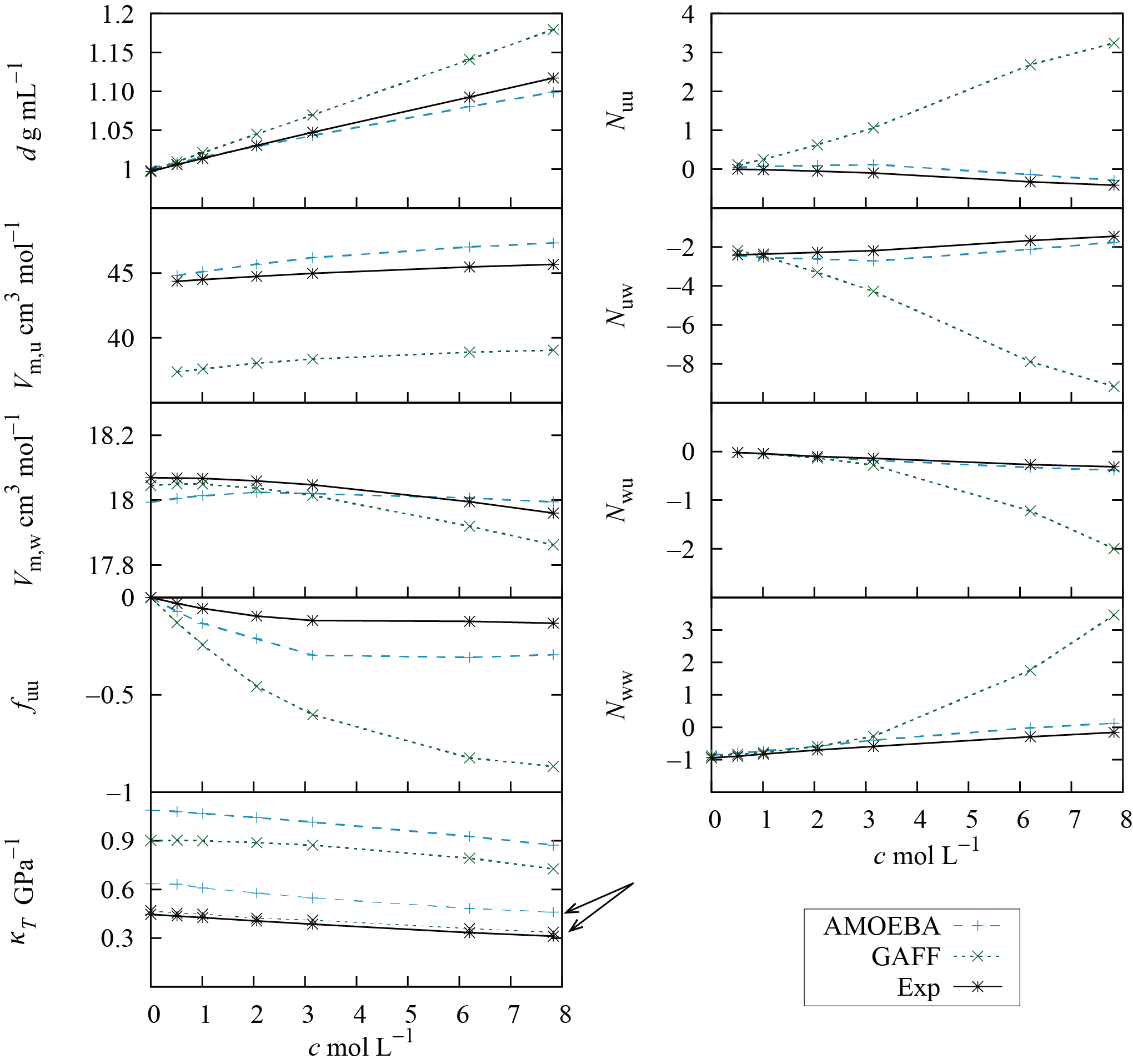
80. Simon, S. l.; Duran, M.; Dannenberg, J. J., How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *J. Chem. Phys.* **1996,** *105*, 11024.

81. Kollman, P. A., Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Acc. Chem. Res.* **1996,** *29*, 461-469.

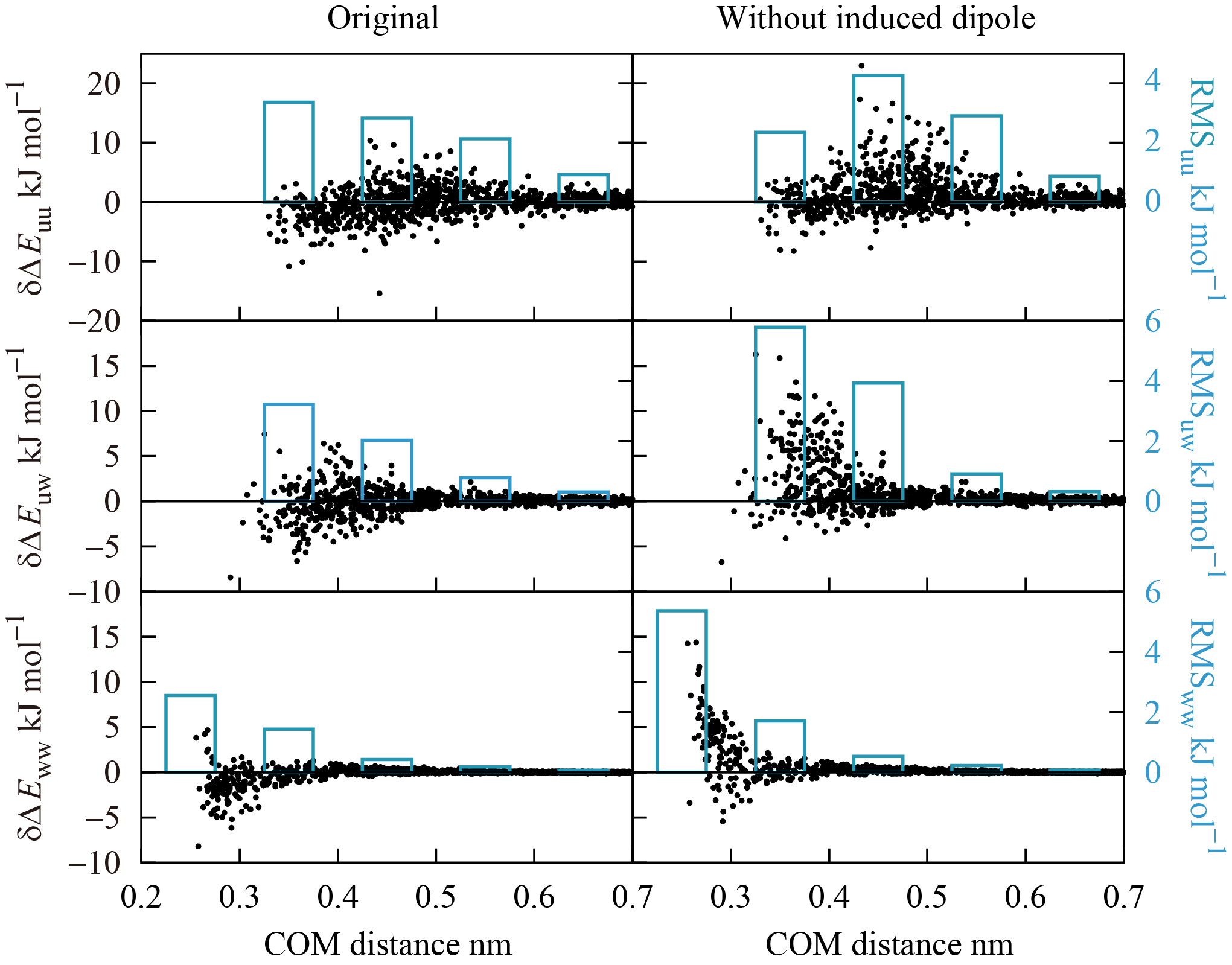
82. *Mathematica*, Version 10.0, Wolfram Research, Inc.: Champaign, IL, 2014*.*



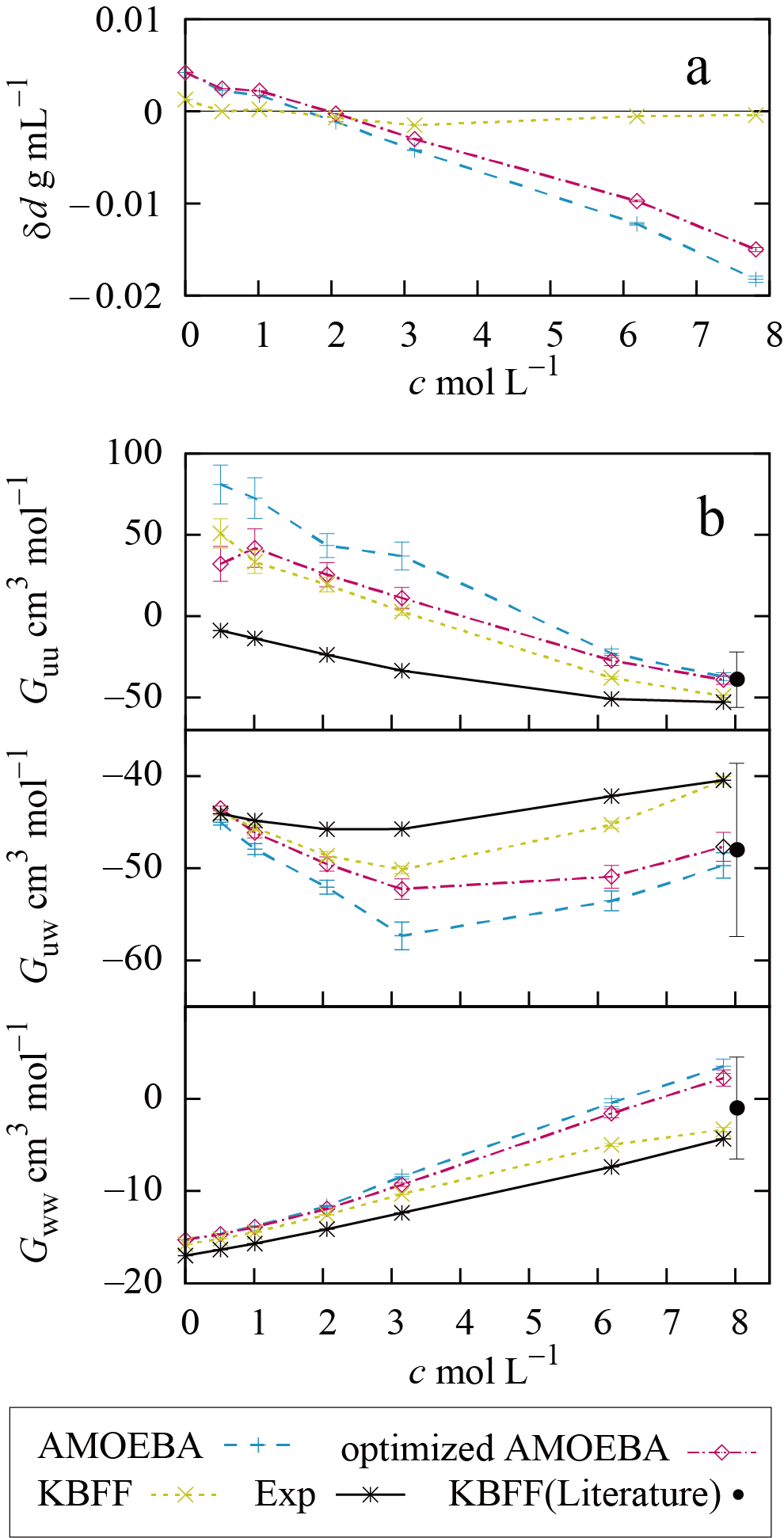
**Figure 1.** Comparison of the calculated and experimental KB integrals plotted against the molar concentration of urea. The subscripts u and w refer to urea and water, respectively. The error bar represents a standard error for a calculated value.



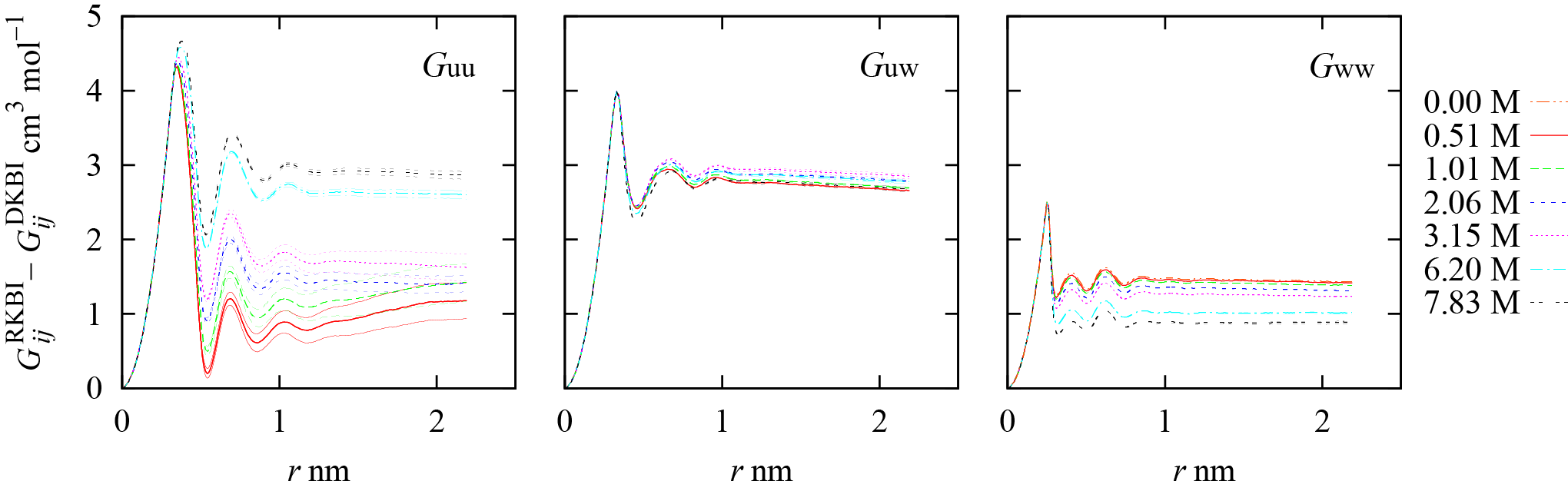
**Figure 2.** Superiority of AMOEBA in reproducing thermodynamic quantities. Solution density, , partial molar volume, , activity derivative, , isothermal compressibility, , and excess coordination numbers, , as a function of urea concentration, are plotted against molar concentrations of urea. was calculated using Eq. (4) or the volume fluctuation formula ().69 The latter is plotted with a thin line marked with an arrow.



**Figure 3.** Improvements in the reproducibility of pairwise interaction energies by incorporating the induced dipole value. Difference between pairwise interaction energies, (black dots, the left y-axis), of molecules *i* and *j* for AMOEBA and quantum chemistry calculations are plotted as a function of the center of mass (COM) distance. The root mean squares (RMS) of in COM distance for every 0.1 nm are superimposed (sky blue boxes, the right y-axis). The results for AMOEBA original (left) and AMOEBA without induced dipoles (right) are plotted.



**Figure 4.** Improvements in the vdW-optimized AMOEBA and comparison to KBFF for urea. a) The differences between experimental and calculated densities, δ*d*, are plotted against urea concentration. b) The calculated KBIs of various parameters are plotted against the urea concentration along with experimental values (Exp). The KBIs at 7.83-mol L-1 solution in this study were consistent with the literature values32 of KBFF urea at 8.02 mol L-1. The error bars represent standard errors for the calculated values.



**Figure 5.** Error caused by numerical integration of running KB integral (RKBI) method as a function of integral interval, *r*. is the difference of the KB integrals for calculations with RKBI versus direct KB integral (DKBI) methods. The errors are plotted for various concentrations with the standard error (SE) in thin lines. As for and , SEs are too small to distinguish.

**Table 1.** Simulation settings for different urea concentrations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *c*  mol L-1 a | # of Urea b | # of Water b | # of Simulations | Simulation Time (ns) c | |
| AMOEBA | GAFF |
| 0 | 0 | 3000 | 10 | 3.5 | 20.5 |
| 0.51 | 28 | 2972 | 100 | 8 | 20.5 |
| 1.01 | 56 | 2944 | 50 | 5 | 20.5 |
| 2.06 | 118 | 2882 | 30 | 5 | 20.5 |
| 3.15 | 186 | 2814 | 22 | 3.5 | 20.5 |
| 6.20 | 402 | 2598 | 22 | 3.5 | 20.5 |
| 7.83 | 536 | 2464 | 15 | 3.5 | 20.5 |

a) *c* stands for mole concentration.

b) Numbers of urea and water molecules calculated using the experimental mole concentration, *c*, and density.

c) The first 0.5 ns were not used for structure sampling.

**Table 2.** Comparison of performance of various parameters in reproducing QC interaction energies.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| parameter | | trajectories used for sampling | urea-urea | | urea-water | | water-water | |
| *R* | RMSE  kJ mol-1 | *R* | RMSE  kJ mol-1 | *R* | RMSE  kJ mol-1 |
| AMOEBA | original | AMOEBA original 0.5 mol L-1 | 0.972 | 2.38 | 0.983 | 1.33 | 0.982 | 0.88 |
| w/o induced dipole | 0.966 | 3.12 | 0.972 | 2.46 | 0.945 | 1.69 |
| vdW Opt | 0.972 | 2.32 | 0.985 | 1.27 | Same as for Original | |
| original | AMOEBA original 7.8 mol L-1 | 0.976 | 2.58 | 0.984 | 1.45 | 0.989 | 0.758 |
| vdW Opt | 0.978 | 2.44 | 0.985 | 1.41 | Same as for Original | |
| GAFF | | GAFF  0.5 mol L-1 | 0.966 | 3.80 | 0.950 | 3.20 | 0.923 | 2.60 |
| KBFF | | KBFF  0.5 mol L-1 | 0.966 | 3.96 | 0.963 | 2.15 | 0.942 | 2.93 |

*R* is a correlation coefficient for interaction energies calculated with each model versus interaction energies calculated with quantum chemistry. These interaction energies were used to calculate a root mean square error (RMSE).

**Table 3.** Original and optimized vdW parameters.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| class | original | | | vdW Opt | | | restriction | | |
| nm | kJ mol-1 | Reduction | nm | kJ mol-1 | Reduction | nm | kJ mol-1 | Reduction |
| C | 0.382 | 0.443504 |  | 0.382 | 0.422584 |  | 0.378-0.382 | 0.422584-  0.443504 |  |
| N | 0.371 | 0.46024 |  | 0.371 | 0.46024 |  | 0.371 | 0.43932-  0.46024 |  |
| H | 0.259 | 0.092048 | 0.9 | 0.27 | 0.058594828 | 0.9 | 0.12-  0.27 | 0.056484-  0.092048 | 0.9-0.91 |
| O | 0.33 | 0.468608 |  | 0.33 | 0.468608 |  | 0.33 | 0.468608 |  |

is the distance at which vdW energy between two atoms, ,is minimum with its value of . Reduction is a factor that is applied to reduce the distance between a heavy atom and the connected hydrogen atom by changing the position of the hydrogen atom. The restriction was applied during the optimization.

