Published in final edited form as:
Phys Chem Chem Phys. 2011 October 28; 13(40): 18154–18167. doi:10.1039/c1cp21883b.

A Kirkwood-Buff Force Field for the Aromatic Amino Acids

Elizabeth A. Ploetz and **Paul E. Smith***
Department of Chemistry, 213 CBC Building, Kansas State University, Manhattan, KS 66506-0401, USA.

Abstract

In a continuation of our efforts to develop a united atom non-polarizable protein force field based upon the solution theory of Kirkwood and Buff *i.e.*, the Kirkwood-Buff Force Field (KBFF) approach, we present KBFF models for the side chains of phenylalanine, tyrosine, tryptophan, and histidine, including both tautomers of neutral histidine and doubly-protonated histidine. The force fields were specifically designed to reproduce the thermodynamic properties of mixtures over the full composition range in an attempt to provide an improved description of intermolecular interactions. The models were developed by careful parameterization of the solution phase partial charges to reproduce the experimental Kirkwood-Buff integrals for mixtures of solutes representative of the amino acid sidechains in solution. The KBFF parameters and simulated thermodynamic and structural properties are presented for the following eleven binary mixtures: benzene + methanol, benzene + toluene, toluene + methanol, toluene + phenol, toluene + *p*-cresol, pyrrole + methanol, indole + methanol, pyridine + methanol, pyridine + water, histidine + water, and histidine hydrochloride + water. It is argued that the present approach and models provide a reasonable description of intermolecular interactions which ensures that the required balance between solute-solute, solute-solvent, and solvent-solvent distributions is obtained.

Introduction

Essentially all biochemical processes, and likewise many of industrial interest, take place in solution mixtures. Attempts to understand solutions have led researchers to develop, and often later disband, many theories and models over the last two centuries.[1] Despite much advancement, our present understanding of solution mixtures is far from satisfactory. One way to measure our progress in this area is by the repeated evaluation of models developed to mimic the behaviour of solution mixtures and, in particular, their experimental properties. Additionally, accurate modelling allows for the future prediction of behaviour in situations where the experimental work is prohibitive on some grounds. One of the ultimate goals is to create solution models that would only require the composition and state of the system as input, yet the models would be able to reliably output that system's thermodynamic properties. The focus of the present work lies in the development and application of such models to eventually obtain thermodynamic and structural information for peptides and small proteins in various physiological environments using classical simulation techniques.

Several force fields (FFs) for the simulation of peptides and proteins have been proposed.[2–6] However, in recent comparisons of established protein FFs a significant degree of reduced solvation, *i.e.* overestimation of solute-solute and solvent-solvent interactions, was frequently observed.[7–15] Sometimes this leads to obviously erroneous results, such as the spontaneous formation of aggregates or phase separation in systems which should be fully

*Fax: 785-532-6666; Tel: 785-532-5109; pesmith@ksu.edu.

miscible.[7–9] Unfortunately, in many studies the errors are not immediately apparent from a visual inspection of the trajectory, but can often be inferred from, for example, the rotational diffusion of proteins.[15] This imbalance in interaction strength is especially problematic when preferential interactions are of interest,[16–21] or when considering simulations of systems for which little quantitative experimental data is available for comparison.

The origin of such misbalanced interactions is often related to the approximate description of electrostatic interactions provided by non-polarizable FF electrostatic parameters.[7, 9] Consequently, polarizable FFs have been developed which, in principle, should perform better than non-polarizable FFs. However, they are computationally expensive and are not without their own set of difficulties.[10] Thus, non-polarizable FFs still remain in widespread use when simulating large systems and/or attempting to study long time events. Popular methods to develop nonpolarizable partial charges for use in condensed phases have included the use of scaled gas phase quantum mechanical partial charges based upon calculations of the restrained-electrostatic potential, the use of scaled gas phase charges from *ab initio* calculations of the minimum interaction energies and geometries between dimers, and the parameterization of partial charges to reproduce thermodynamic and structural properties of pure liquids and the free enthalpies of solvation in water and other solvents.[11] The primary disadvantage of each of these approaches is that the molecule is being studied in either the gas phase, under infinitely dilute liquid solution conditions, or as a pure liquid. Hence, because the molecule being parameterized is generally not considered in a variety of environments, the chosen charges may not optimally describe intermolecular interactions over a range of compositions, *i.e.* where solute-solute interactions become more important.

It is widely acknowledged that biomolecular force fields are imperfect and parameters are constantly being refined.[11] However, the presence of several existing force fields suggests that the biological simulation community does not need another FF, unless it has been developed with a markedly different approach. One possibility involves a focus on appropriately balancing solute-solute, solute-solvent, and solvent-solvent interactions by studying these interactions far beyond the infinitely dilute range. Naturally, an important benchmark for such a FF would be its ability to reproduce solution activities. This is the role the Kirkwood-Buff Force Field (KBFF) seeks to fill. Since the solvation process distorts the electronic structure of the solute, with different parts of a molecule becoming polarized to different extents, we believe it is extremely difficult to develop accurate FF parameters, specifically effective condensed phase charge distributions, by either simply scaling gas phase charges or just reproducing pure liquid or infinitely dilute solution properties.[22] Instead, we have developed a series of models where the best effective solution phase charges are derived empirically in order to fit experimental liquid solution data for mixtures of model compounds *over a range of liquid phase compositions*.[7, 8, 23–29] Specifically, our approach attempts to obtain an effective charge distribution that best reproduces the experimentally derived Kirkwood-Buff integrals (KBIs) for representative solution mixtures.

KB theory provides a link, through exact relationships, between the thermodynamic properties of a solution and the corresponding distributions between the various components within the solution. Consequently, the analysis of a limited set of experimental data (activities, partial molar volumes, and isothermal compressibilities) allows one to extract the molecular level distributions, albeit as integrals over the corresponding radial distribution functions, between each pair of species in the solution (see Methods). As the composition dependent integrals can also be easily extracted from computer simulations, they can then be used as target data for the development of simple solution models. This is the main aim of the KBFF approach, which we believe has several advantages, namely the following:

1. The integrals provide additional data for use during parameterization.

2. One has easy access to changes in the solute activity and thereby quantitative information concerning intermolecular distributions, which are a sensitive property of solution mixtures.[1, 30]

3. Typically one does not have to sacrifice agreement for other properties of interest in order to reproduce the integrals.[7, 8, 23–29]

A variety of solute models have been developed using this type of approach.[7, 8, 23–29] In our opinion it is important for a FF to be able to reproduce the features displayed by the experimental KBIs (see Results), because this assures that the FF is accurately balancing intermolecular interactions due to the KBIs' dependence on the solution activity. Furthermore, FFs that can reproduce the experimental KBIs could be used to gain insights into the very underlying solution microstructure that determined the KBIs. This is advantageous since the microstructure of solutions is generally difficult to determine experimentally. Thus, these FFs have the potential to significantly enhance the predictive power of solution molecular dynamics simulations.

We should note that outside of the biological FF community there has been significant work to develop FFs that reproduce solution activities, a property not normally used in the development of biological force fields. Notably the TraPPE FFs developed by the Siepmann group are specifically designed to model vapor-liquid coexistence curves.[12–14] However, while they are attempting to reproduce the activities of each component in solution, only one composition is considered for a given T and p. For biological simulations it is more important to reproduce activities over a range of compositions for a relatively narrow window of (near-physiological) temperatures and pressures. Thus while the TraPPE approach is well suited for numerous thermodynamic and engineering purposes, it is of limited use in the typical biological context.

As we continue to develop a FF specifically designed to reproduce the experimental KBIs for condensed phase solution mixtures applicable to the simulation of peptides and small proteins,[7, 8, 23–25, 27–29] models for the aromatic amino acid sidechains are required. In this work, toluene (Tol) was used as a model for the phenylalanine (Phe) sidechain, *p*-cresol (pCr) as a model for the tyrosine (Tyr) sidechain, and pyrrole (Pyrr) and indole (Ind) as models of the tryptophan (Trp) sidechain. Additionally, KBFF models for neutral histidine (His) and doubly-protonated histidine (HisH) were developed. First however, benzene (Ben) was studied as a model system to develop parameters for a fully conjugated carbon atom (C_{ar}). Then the new C_{ar} atom type was used in the parameterization of the partial charges for the other aromatics.

Beyond the mere necessity of developing these aromatic parameters to achieve completion of the KBFF for proteins, the systems studied are interesting in their own right. Aromatic molecules have unique features that make them challenging to model accurately, such as their ability to exhibit π-effects (cation-π, polar-π, and π donor-acceptor). This issue receives considerable attention due to the role π-effects frequently play in molecular recognition.[31, 32] The negative electrostatic potential on the surface of the aromatic ring and positive electrostatic potential along the ring edges allows for a favourable interaction between cations or other polar group with the face of the ring. These effects are not explicitly taken into account in any classical FFs due to a lack of orbital descriptions, but the effects undoubtedly play an important role in many of the systems the community attempts to study using these simple FFs.

Previously, the Smith group has developed models for the sidechains of other amino acids by studying their analogues solvated in water. Unfortunately, many of the solute models that are representative of the aromatic sidechains are only sparingly soluble in water, making the

study of aqueous solutions troublesome. To avoid this problem we have switched to solutions of aromatics in a solvent of methanol (MOH) or Tol where necessary. A KBFF model for MOH was previously developed by Weerasinghe and Smith from a study of aqueous solutions employing the SPC/E water model.[28, 33] Hence, in this work we have made the usual assumption that our FF was fully transferable, *i.e*. a good description of Ben + MOH solutions coupled with a good description of MOH + water solutions provides for a reasonable description of Ben + water solutions.

Methods

Kirkwood-Buff Analysis of Experimental Data

Kirkwood-Buff (KB) theory was used to extract information concerning the microstructure of binary solutions from experimental thermodynamic data available in the literature. The process by which this was achieved has been thoroughly detailed elsewhere.[7, 8, 25, 27–29, 34] Briefly, the KBIs (G_{ii}, $G_{ij} = G_{ji}$, and G_{jj}) were obtained from the composition dependent chemical potential derivatives (μ_{ij}), partial molar volumes (V_i), and isothermal compressibility (κ_T) of binary mixtures at constant pressure (p) and temperature (T) according to[1]

$$G_{ij}=kT\kappa_T-\frac{\delta_{ij}}{\rho_i}+\rho kT\frac{(1-\rho_i\overline{V}_i)(1-\rho_j\overline{V}_j)}{\rho_i\rho_j\mu_{ij}} \quad (1)$$

where δ_{ij} is the Kroenecker delta function, $\rho_i = N_i/V$ is the number density of i molecules, $\rho = \rho_i + \rho_j$, $\mu_{ij} = (\delta\mu_i/\delta N_j)_{p,T,Ni}$, k is the Boltzmann constant, T is the temperature of the liquid solution, N_i is the number of i molecules in the solution, and V is the total volume of the solution. Experimental activity data was obtained from literature values for the activity coefficients or excess molar Gibbs energies, G^E, and were fit to an analytical function to facilitate differentiation (if the fit was not already performed in the original study). Partial molar volumes were calculated from experimental data for the solution's mass density or volume as a function of composition according to[23]

$$\overline{V}_i=V_m+x_j\frac{dV_m}{dx_i} \quad (2)$$

where V_m is the molar volume of the solution, and x_i is the mole fraction of component i in the solution. For systems in which there was no experimental data available for the density or partial molar volumes as a function of composition it was assumed, based on precedence,[34] that $\overline{V}_i=V_i^\circ$, leading to $V_m+x_iV_i^\circ+x_jV_j^0$ and $V^E = 0$. Here V_i° is the molar volume of pure component i and V^E is the excess volume. It has also been shown that the value of κ_T has an insignificant effect on the calculation of the KBIs at ambient temperatures and pressures.[34] Thus, for all systems the solution κ_T was calculated according to

$$\kappa_T=\rho_i\overline{V}_i\kappa_{T,i}+\rho_j\overline{V}_j\kappa_{T,j} \quad (3)$$

where $\kappa_{T,i}$ is the isothermal compressibility of pure component i. The KBIs extracted in the above manner are most sensitive to the experimental activity or Gibbs excess data.[34, 35] Unfortunately, it is not always possible to accurately determine the appropriate errors associated with this data. In general, the errors associated with the G_{ij} values tend to increase as the concentration of i or j decreases. To a large degree the errors in both the experimental and simulated KBIs can be minimized by the use of excess coordination numbers, $N_{ij} = \rho_jG_{ij}$ (see later discussion).

Experimental Sources for Composition and Activity Data

Benzene (1) + Methanol (2): Partial molar volumes were set equal to the molar volumes of the pure components.[36] The Wilson equation was used to fit the G^E data at 308 K.[37] Toluene (1) + Methanol (2): Partial molar volumes were set equal to the molar volumes of the pure components.[36] Activity coefficients were reported directly in an experimental isothermal vapor liquid equilibria (VLE) study at 318 K and 1 bar.[38] From the activity data, G^E was calculated as a function of composition and fit using the Redlich – Kister equation with a nonlinear least squares method involving three parameters. Toluene (1) + Benzene (2): Partial molar volumes were set equal to the molar volumes of the pure components.[39] Activity data was obtained from experimental isothermal VLE data at 313 K using the Wilson equation.[40] Toluene (1) + Phenol (2) / Toluene (1) + *p*-Cresol (2): Isothermal VLE data provided a fit for V^E using the Redlich – Kister equation.[41] The NRTL G^E model was used to calculate the activity coefficients.[41] Pyrrole (1) + Methanol (2): Partial molar volumes were set equal to the molar volumes of the pure components.[39] Activity data was taken from experimental liquid-liquid equilibria (LLE) data at 298 K and fit using the NRTL equation for the ternary system methanol (1) + hexadecane (2) + pyrrole (3), where x_2 was simply set equal to zero.[39] Indole (1) + Methanol (2): Partial molar volumes were set equal to the molar volumes of the pure components.[39] Activity data at 298 K was taken from the same source as for the pyrrole + MOH system in which activity data were provided for solutions up to values of $x_{Ind} = 0.74$. Pyridine (1) + Methanol (2) / Pyridine (1) + Water (2): The V^E and G^E data were obtained from measurements of vapor pressures and densities and modelled using the Redlich – Kister power series equation.[42] Histidine (2) + Water (1) / Histidine Hydrochloride (2) + Water (1): Experimental data for His and HisH mixtures is sparse, presumably because histidine is only sparingly soluble in water. Volume data for His + water was obtained from a density versus composition study at 298 K and 1 bar.[43] Activity coefficients were obtained from a vapor pressure osmometry study at 298 K and 1 bar with a maximum concentration of $m_{His} = 0.25$.[44] We studied only the $m_{His} = 0.25$ composition with a tautomeric ratio of 15% HisA to 85% HisB.[45–49] Partial molar volumes were set equal to the molar volumes of the pure components. Activity coefficients were obtained from the same study as for His + water at 298 K and 1 bar with a maximum $m_{HisHCl} = 0.63$.[44] We studied two compositions, $m_{HisHCl} = 0.3$ and 0.6 due to the low solubility of HisHCl and the absence of a complete set of experimental data for imidazole. Analysis of the experimental and simulated data for salts is complicated by the presence of multiple ionic species in solution. We adopted the indistinguishable ion approach, which has been outlined in detail elsewhere.[8, 28, 50]

Parameter Development

The KBFF approach uses a simple Lennard-Jones 6–12 plus Coulomb potential for all nonbonded interactions. Bond, angle, and torsion parameters were adopted (with permission) from Gromos with the exception of the H-C_{ar}-C_{ar}-C_{ar} improper dihedrals for six-membered rings, which were removed in order to improve the translational temperature of the aromatics (see Results for further discussion).[51] The CH, CH2, and CH3 united atom parameters were also adopted from the Gromos force field,[51] while polar hydrogen and heteroatom parameters were taken from our previous models. The N_{py} atom type was developed following the same scheme as described in detail previously,[7, 27] in which the σ and ε parameters were determined by using the correlation between atomic size and atomic hybrid components of molecular polarizabilities.[52]

Initial studies with our existing sp^2 carbon parameters resulted in poor agreement with experiment for the liquid densities. Hence, a new aromatic carbon type was developed to reproduce several properties of pure benzene (density, crystal structure, heat of vaporization). Traditionally, we have not used heat of vaporization data for pure liquids

during our parameter development as we consider the KBFF models to be polarization corrected models in a similar fashion to the SPC/E water model,[33] which is the water model of choice. However, the polarization corrections for benzene will be small and therefore the comparison with experiment should be reliable. The effective solution phase partial atomic charges used in the Coulomb potential were parameterized to reproduce the KBIs of binary mixtures. In our development of the partial charges for the aromatics, the charge group concept was adopted in an attempt to decrease the number of variables used in the parameterization process. Starting with an initial guess for the partial charges of the molecule of interest, the simulated KBIs were compared with those extracted from the experimental data for the same system for a range of solution compositions. Simulated KBIs were calculated by integration of the center of mass (com) based radial distribution functions (rdfs), which were obtained from the simulations in the usual manner. The required KBIs are given by[30]

$$G_{ij}=4\pi\int_0^\infty[g_{ij}^{\mu VT}(r)-1]r^2dr\approx\int_0^{R_{cut}}[g_{ij}^{NpT}(r)-1]r^2dr \quad (4)$$

The simulated KBI values were obtained by carefully averaging over a range of distances in which the com rdfs were essentially unity and the KBIs had reached a reasonably constant value. This small range of distances, typically corresponding to one solvation shell, was not necessarily the same for each system or for each of the three KBIs within a specific system. Some of the systems displayed significant aggregation at some compositions, often in agreement with experiment. However, the simulation results in which the KBIs were not converged using the present system sizes were not reported because they were considered statistically unreliable. Error estimates were obtained from the averages of multiple 5 ns runs.

Once both simulated and experimental KBIs were obtained, they were compared. If the solution KBIs agreed closely with experimental values, then the parameterization process was complete for that molecule. If not, a new set of partial charges was chosen and the process was repeated. Thus, the charges on the atoms were adjusted in an iterative manner to best reproduce the KBIs for solution mixtures over a range of compositions. Once the parameterization process was complete, other thermodynamic properties of the new models were assessed. The specific parameter development for each system is discussed in the Results section. The final atom type LJ parameters are shown in Table 1 and the final partial charges are provided in Figure 1.

Molecular Dynamics Simulations

All binary mixtures were simulated with classical molecular dynamics techniques using the Gromacs program in the isothermal isobaric (*NpT*) ensemble at a pressure of 1 bar.[55] The simulation temperature for each system was chosen to match the temperature of the experimental G^E data, which corresponded to 298 K for most systems except Ben + MOH (308 K), Tol + MOH (318 K), Ben + Tol (313 K), Tol + PhOH (333 K), and Tol + pCr (333 K). All systems were simulated at mole fractions of x_1 = 0.00, 0.25, 0.50, 0.75, and 1.00. Additionally Ben + MOH, Tol + MOH, Tol + PhOH, Tol + pCr were simulated at x_1 = 0.90, and Py + water was simulated at x_1 = 0.10, as these regions displayed interesting behaviour. His + water was simulated at m_{His} = 0.25 and the HisHCl + water salt system at m_{HisHCl} = 0.30 and 0.60. All systems were simulated in a 6 nm cubic box except for His and HisHCl, which were simulated in 8 nm cubic boxes due to the low molality of the solute. The Berendsen weak coupling technique was used to modulate the temperature and pressure with relaxation times of 0.1 ps and 5.0 ps, respectively, with a 4.5×10^{-5} bar^{-1} compressibility.[56] All bonds were constrained using the Settle and LINCS algorithms for water and non-water

molecules, respectively.[57, 58] The use of bond constraints allowed for a 2 fs time step to be used for the integration of the equations of motion, which was performed using the Leap-Frog algorithm.[59] The particle-mesh-Ewald technique was used to calculate the electrostatic interactions using a convergence parameter of 2.603 nm^{-1}, cubic interpolation, a maximum fast fourier transform grid spacing of 0.12 nm for the reciprocal space sum, and tinfoil boundary conditions.[60] The simulations employed a twin range cut-off of 1.2 and 1.5 nm for the real space electrostatic and van der Waals interactions, respectively.[60] The neighbour list was updated every 10 steps. Each system was defined by a cubic simulation box into which molecules were randomly placed to generate the initial configuration for each mixture. The steepest descent method was then used to perform ≥1000 steps of energy minimization. This was followed by extensive equilibration that was continued until the intermolecular potential energy contributions and com rdfs displayed no systematic drift with time. Configurations were saved every 1.0 ps for analysis. All mixture simulations were run for at least 20 ns to ensure reasonable precision for the KBIs. We found that relatively long simulation times were necessary to yield reasonably precise data, presumably due to the large size of aromatics (slow diffusion) when compared to our previously studied molecules. Non-bonded interactions were excluded for neighbouring atoms up to six bonds away for benzene, toluene, phenol, and *p*-cresol, up to three bonds away for indole, pyrrole, and pyridine, and up to two bonds away for His and HisH. For phenol, a force constant of 7.24 kJ/mol with a multiplicity of two was used to constrain the angle between the CE-CZ-OH and CZ-OH-HH planes near 180° in the CE-CZ-OH-HH proper dihedral angle. The force constant was chosen based on the experimentally obtained values for this rotational barrier in phenol.[61–63]

Analysis of Other Simulated Solution Properties

The enthalpy of vaporization ($\Delta_{vap}H$) was calculated for the pure liquids according to

$$\Delta_{\text{vap}}H = \frac{-(E_{\text{tot}} - E_{\text{intra}})}{N} + RT \quad (5)$$

assuming that (1) the intermolecular interactions in the gas phase are negligible and (2) the intramolecular interactions are equal in the gas and liquid phases. Here, E_{tot} is the total potential energy of the molecules in the liquid phase, E_{intra} is the intramolecular potential energy of the molecules in the liquid phase, N is the number of molecules in the liquid phase, and RT is the pV work contribution from the gas phase (ideal gas approximation). The pV work term is assumed to be zero for the liquid phase (incompressible liquid approximation).[64, 65] The dielectric constant was determined from the dipole moment fluctuations using a reaction field permittivity of infinity as implemented by the g_dipoles Gromacs code.[29, 66] For liquids, which are nearly incompressible, $\Delta_{mix}V \approx 0$ thus $\Delta_{mix}H \approx \Delta_{mix}E$, where the subscript 'mix' refers to mixing throughout.[65] The enthalpy of mixing ($\Delta_{mix}H$) for each system was thus calculated according to[64]

$$\Delta_{\text{mix}}H = \frac{E_{\text{sln}}}{N_{\text{sln}}} - \sum_i x_i \frac{E_i^\circ}{N_i^\circ} \quad (6)$$

where E_i° is the total potential energy of pure i and N_i° is the number of i molecules in the pure i simulation and the subscript 'sln' refers to the solution properties. Finite difference κ_T values were obtained by performing additional 5 ns long simulations at pressures of 250 and 500 bar.[64] Finite difference isobaric thermal expansion (α_p) values were obtained from five additional 5 ns long simulations at 5 K temperature intervals encompassing the desired α_p temperature. Translational self-diffusion coefficient (D_t) values were computed using the Einstein relation from the calculation of the mean square displacement (msd) of atoms from their initial positions, as implemented by the Gromacs g_msd code.

Results and Discussion

In this section we present and describe the comparison with the available experimental data. The primary goal was a consistent set of nonbonded parameters, developed to reproduce the experimental KBIs, for simple models that may be used for future biomolecular simulations. Unfortunately, while perfect agreement with experiment was the goal, the use of relatively simple (nonpolarizable) potentials necessarily introduced a degree of approximation and consequently some systems displayed some deviation from experiment. In particular, it would be expected to be difficult to fully reproduce all of the subtle polarization effects between the extremes in composition for binary mixtures of polar and nonpolar compounds.

Benzene

Initially, we intended to use the same LJ parameters, previously developed for amides,[25] for the carbon in Ben and to simply adjust the polarity of the C_{ar}-H bonds in Ben to reproduce the experimental KBIs. We found that the KBIs were indeed very sensitive to the charge distribution. For example, in a binary mixture of Ben + MOH, the degree of MOH-MOH ($G_{MOH\text{-}MOH}$) aggregation noticeably differed depending upon the charge distribution across the C_{ar}-H bond. The difference was even observable from a simple visual inspection of the trajectories. However, the density and $\Delta_{vap}H$ of pure Ben were far from the experimental values regardless of the choice of charge. Thus, we found that it was necessary to develop a new atom type for a fully conjugated carbon, C_{ar}, to produce better agreement for these properties as well.

Starting with an initial guess for the charge distribution (C_{ar}=- 0.14 $|e|$, taken from GROMOS) the Ben LJ σ and ε of the sp^2 amide carbon type were adjusted to best reproduce the experimental liquid density of 0.8629 g/cm^3 at 308 K and $\Delta_{vap}H$ of 33.85 kJ/mol at 298 K.[36, 67] The experimental KBIs were then extracted from simulations of Ben + MOH mixtures at 308 K, analysed, and the charge distribution was parameterized.

Although Ben is not considered a polar molecule, large differences were observed in the simulated KBIs as changes were made to the charge distribution (Figure 2). We believe this was related to small changes in the point charges representing a cumulatively large change in the overall charge distribution of the molecule due to the six C_{ar}-H bonds per molecule. Additionally, the change in charge parameters would have an even more pronounced effect on Ben–Ben interactions. As can be seen in Figure 2 there was a systematic reduction in the aggregation of MOH-MOH, as was evident from the decreasing $G_{MOH\text{-}MOH}$ as the charge distribution was increased. Starting with a carbon charge of $q = -0.10\ |e|$, an incremental reduction in the KBIs occurred on scaling to $q = -0.14\ |e|$, yielding a systematic improvement across the full composition range for each of the three KBIs. This was also observed from a visual inspection of the snapshots of these mixtures in which the number of MOH that were free in solution increased as the carbon charge was increased (results not shown). It should be noted that the simulated KBIs displayed for the $q = -0.13\ |e|$ distribution in Figure 2 are not in perfect agreement with those of the final KBFF due to the iterative procedure by which the σ and ε were later adjusted.

As we have developed the KBFF models, we have repeatedly seen a similar sensitivity of the KBIs on the charge distribution across many other systems.[7, 8, 23–25, 27–29] We believe that the responsiveness of the KBIs to changes in the charge distributions is a major advantage of the current approach. The charge differences between different FFs for Ben and Phe are very small (Table 2), although an exact comparison across FFs is not straightforward due to the effects of other variables including different LJ parameters and different van der Waals and electrostatic cut-off distances.

Figure 3 indicates that Ben + MOH is a highly non-ideal system in which the large negative $G_{\text{Ben-MOH}}$ at high x_{Ben} (middle panel) and the corresponding large positive $G_{\text{MOH-MOH}}$ in the same region of the composition range (bottom panel), quantify the reduced mixing between Ben and MOH and the corresponding microscopic self-association of MOH molecules, respectively.

The final C_{ar} parameters were found to be σ = 0.381 nm, ε = 0.33 kJ/mol, $q_{c_{ar}}$=−0.13, and q_{H}=+0.13. Fortunately, this was a q very near to that used in the parameterization of σ and ε and our final simulated Ben density and $\Delta_{\text{vap}}H$ remained in good agreement with experiment despite iterating on q after choosing σ and ε. The final simulated values for the density and $\Delta_{\text{vap}}H$ for benzene were 0.862 g/cm^3 and 32.88 kJ/mol, respectively. The benzene crystal dimensions were also well reproduced with experimental (x, y, z) values of (37.049 Å, 48.664 Å, 35.264 Å) compared to our simulated (x, y, z) values of (37.271 Å, 48.955 Å, 35.475 Å), and an experimental density of 1.2119 g/cm^3 compared to the simulated density of 1.1904 g/cm^3. These results provided confidence that the parameters describing the size of an aromatic carbon, σ, and the atoms energy of most favourable interaction, ε, were appropriately chosen.[72]

In summary, the major features of this solution were accurately reproduced by MD simulations with the above parameters, but only after the partial charges were carefully parameterized to appropriately balance the solute-solute, solute-solvent, and solvent-solvent interactions.

Solution Mixtures

The simulated com rdfs for all the systems studied here are shown in Figure 4. It is noticeable that the major changes in the rdfs with composition occurred for the alcohol (major hydrogen bonding) component. The rdfs for each mixture will be discussed in more detail in the sections below. Figure 5 compares the calculated and simulated KBIs after a multiplication by the number density. The quantity $\rho_j G_{ij}$ is often referred to as an excess coordination number and is denoted as N_{ij}. As mentioned previously, the use of N_{ij} values provides a dimensionless quantity which diminishes the noise in the KBIs by effectively removing the uncertainty that is inherent at low concentrations of j in both the experimental and simulated KBI data. This uncertainty is attributed to sampling errors related to very few j molecules at high i concentration (and *vice versa*), and the relative difficulty of large aromatic molecules to move past each other at high aromatic concentrations. In addition, the N_{ij} values have a well-defined physical interpretation, *i.e.* the excess or deficit in the number of j molecules within a certain volume of space away from a central i molecule when compared to the number that would have been found had the central i not been present.

Benzene (1) + Methanol (2)

The major feature in the KBIs for Ben + MOH, as provided in Figures 3 & 5, was the increase in $G_{\text{MOH-MOH}}$ (and $N_{\text{MOH-MOH}}$) as x_{MOH} decreases. The simulations using the new models provided excellent agreement with the experimental data. This feature was also mirrored in the $g_{\text{MOH-MOH}}$ rdf displayed in Figure 4 where the first peak dramatically increased in magnitude. Noticeably, $g_{\text{Ben-Ben}}$ and $g_{\text{Ben-MOH}}$ remained comparatively constant over the full concentration range.

A large positive N_{ij}, such as that observed from the experimental peak of almost 10 for $N_{\text{MOH-MOH}}$ at $x_{\text{Ben}} \approx 0.7$, can be interpreted as an indication of excessive clustering of MOH molecules at this composition. Physically, this corresponds to larger and larger deviations from a random distribution of methanol molecules as the number of possible hydrogen bonding partners decreases. The subsequent final decrease in $N_{\text{MOH-MOH}}$

presumably indicating an increasing entropic cost of maintaining these enthalpically favourable hydrogen bonded structures. If Ben + MOH were a symmetric ideal (SI) system, the $G_{\text{MOH-MOH}}$ ($N_{\text{MOH-MOH}}$) would have been 2 cm^3mol^{-1} (0) according to[1]

$$G_{ij}^{SI}=RT\kappa_T - V_i - V_j+\rho_1 V_1^2+\rho_2 V_2^2 \quad (7)$$

The differences from the SI values were large and serve to quantify the deviation of this mixture from an ideal (essentially random) solution mixture.

A visual inspection of the $x_{\text{Ben}} = 0.90$ system revealed that MOH predominantly formed MOH-MOH chain structures over MOH-MOH cyclic structures. This was an observation that agreed with other simulations and experiments.[73–76] However, a rigorous analysis of the percentages or lifetimes of the chain and cyclic structures was not performed. A more thorough discussion of this microstructure can be found in one of our recent publications.[26]

Toluene (1) + Methanol (2)

Although there is only a small difference between the atomic structure of Ben and Tol, the corresponding experimental $N_{\text{MOH-MOH}}$ values displayed a significantly greater peak, over twice the magnitude of that in the Ben + MOH system, at $x_{\text{Tol}} \approx 0.7$. The major peak was in the same position, but much narrower than the peak in Ben + MOH. The simulations reproduced this increase in the $N_{\text{MOH-MOH}}$ maximum for the Tol + MOH, although the full increase was not reproduced quantitatively.

The charge distribution for the C_{ar}-CH3 bond in Tol was tested to determine if the best model should include polarity across the C-CH_3 bond, or if the KBIs would be better reproduced with zero charges across this bond. From fundamental chemical principles it seems reasonable that one could argue either way. The electronegativity difference between sp^2 and sp^3 carbons is minimal, supporting charges of 0.00 and 0.00 on both the C_{ar} and CH3 sites. On the contrary, our simple model for aromaticity on the Ben ring would be disturbed under such an arrangement. We tested Tol + MOH (and Ben + Tol) with the C_{ar}-CH3 charges set equal to zero, against Tol + MOH (and Ben + Tol) using the same charges for C_{ar}-CH3 as those across the C_{ar}-H of the benzene ring. The results showed a clear improvement when the polarity across the bond was retained. Thus, the only way in which the final KBFF Tol and Ben models differ from one another is in their LJ parameters as they have identical coulombic parameters.

Benzene (1) + Toluene (2)

Ben + Tol is one of the quintessential ideal systems. The Ben + Tol com rdfs and N_{ij} were not included in Figures 4 and 5 because we have discussed the excellent simulation of this ideality previously.[77] The Tol + Ben com rdfs displayed a first solvation shell at ~0.6 nm, however the $g_{\text{Ben-Ben}}$ and $g_{\text{Tol-Tol}}$ distributions had slightly different magnitudes with a small increase in the probability for finding a Ben ~0.6 nm away from a central Ben than a Tol ~0.6 nm away from a central Tol. This was likely due to the tighter packing that is possible between Ben molecules due to the absence of a methyl group when compared to Tol. The simulated G_{ij} and N_{ij} curves were also in agreement with the values predicted by Equation 7 and further supported the ideality of the system.[77]

Phenol (1) + Toluene (2) / *p*-Cresol (1) + Toluene (2)

We first attempted to use the same MOH partial charges for the C_{ar}-O3-H charge group in phenol and cresol; however, this resulted in relatively poor agreement with the experimental KBIs. After several attempts the charges which appeared to best represent the experimental

KBIs corresponded to a simple scaling of the MOH charges by a factor of 0.80. The difference in the partial charge distribution between aromatic and aliphatic alcohols may be attributed to both the aromaticity of the PhOH and pCr molecules and the large size of PhOH and pCr when compared to MOH, which allows for fewer hydrogen bonding partners to be formed in these solutions. An increase in the first peak of $g_{\text{Tol-Tol}}$ was observed as x_{Tol} approaches one. Cyclic clusters of PhOH or pCr molecules were each observed in an arrangement vaguely reminiscent of reverse micelles. A more thorough discussion of this microstructure can be found in our recent publication.[26]

Pyrrole (1) + Methanol (2) / Indole (1) + Methanol (2)

As indicated in Figure 1 we were not forced to adjust the charges around the ring to obtain good agreement with the experimental KBIs. Instead it was possible to retain the simple group concept and merely adjust the dipole across the N-H bond of indole. The same N-H dipole provided satisfactory results for both pyrrole and indole molecules. The com rdfs for these systems showed structured secondary solvation shells and a significant increase in the first solvation shell in all three com rdfs as the concentration of Pyrr or Ind was increased. The scale of these changes was much smaller than that observed for the Ben + MOH and Tol + MOH systems, so this did not necessarily suggest these systems were highly aggregating. Indeed, both the experimental and simulated excess coordination numbers did not indicate an excessive degree of association between species in these solutions.

Pyridine (1) + MOH (2) / Pyridine (1) + Water (2)

In the comparison of the Py + MOH and Py + water systems, the first solvation shell of $g_{\text{water-water}}$ exhibited a higher peak than $g_{\text{MOH-MOH}}$ at a slightly smaller radius due to the smaller excluded volume of water when compared to MOH. The Py + water N_{ij} were reproduced quite well, except at $x_{\text{Py}} = 0.1$. Experimentally, there was a high degree of self-aggregation at this composition. However, our simulations did not appear to fully capture this aggregation. In particular, the self-association of Py ($N_{\text{Py-Py}}$) was underestimated, while the solvation of Py by water ($N_{\text{Py-water}}$) was overestimated. Further attempts to improve the Py + water KBIs at low Py concentrations were unsuccessful and typically resulted in a corresponding reduction in the agreement with the Py + MOH KBIs and the Py + water $\Delta_{\text{mix}}H$ (see later).

Histidine (2) + Water (1) / HisHCl (2) + Water (1)

As mentioned previously, the experimental data for His and HisH mixtures is sparse, presumably because histidine is only sparingly soluble in water. Thus, the charge distributions were developed based upon the previous studies of Pyrr + MOH and additional studies of Py + MOH and Py + water over the full composition range, and then tested using simulations of His + water at 0.25 m_{His} and HisHCl + water at 0.3 and 0.6 m_{HisHCl}. The N-H Pyrr charge distribution obtained for use in Trp was used as an initial guess of the charges on the protonated sp^2 nitrogen (N2) present in His and HisH. The charges on HisH were then adjusted to achieve better reproduction of the experimental KBIs. Py was studied to model the charges of neutral histidine's unprotonated sp^2 nitrogen (N_{py}) and the charges on the two C_{ar}-H that flank this nitrogen. The Pyrr and Py results were then combined to develop a crude approximation for the charges in His. His + water was then simulated to ensure that the amino acids did not aggregate excessively – a minimum requirement.

The His + water and HisHCl + water systems both exhibited similar $g_{\text{water-water}}$ with a sharp first solvation shell and well-structured second and third solvation shells. $g_{\text{His-water}}$ did not display any strong peaks, while $g_{\text{HisHCl-water}}$ had a sharp first solvation shell at a distance of ~0.3 nm with a peak slightly less than 2.25. $g_{\text{His-His}}$ and $g_{\text{HisHCl-HisHCl}}$ both showed first solvation shells at a distance of ~0.7 nm with broad shoulders, however the $g_{\text{HisHCl-HisHCl}}$

peak at $m_{\mathrm{HisHCl}} = 0.3$ was ~4 and the $g_{\mathrm{His\text{-}His}}$ peak was ~2. Even in a relatively large simulation box with a length of ~8 nm, the $g_{\mathrm{HisHCl\text{-}HisHCl}}$ rdf did not reach unity. Thus the errors in $N_{\mathrm{HisHCl\text{-}HisHCl}}$ were significant and they were not reported due to their unreliability.

The N_{ij} were well reproduced in the His + water system. Although we only considered one composition, we believe that this demonstrated a modest description of neutral histidine in our simulations. The HisHCl + water N_{ij} results were also quite good, especially when considering the errors associated with the $N_{\mathrm{HisHCl\text{-}water}}$ values. Although we did not gain meaningful information on the $N_{\mathrm{HisHCl\text{-}HisHCl}}$ values, due to an unconverged $g_{\mathrm{HisHCl\text{-}HisHCl}}$, a visual inspection of the mixture did not reveal any significant self-aggregation of the amino acids with the final set of partial charges. It should be noted that ion pairing and histidine clustering were observed with other sets of trial charges, but that this clustering was not present using final charge distributions. Our analysis of the site-site coordination numbers for the His + water and HisHCl + water systems did not exhibit any interesting features (data not shown). They displayed small coordination numbers with values significantly less than one, further indicating weak interactions between His-His and HisHCl-HisHCl in the systems studied.

Thermodynamic properties

These KBFF models were designed to reproduce solution properties, with a particular focus on the experimental KBIs. However, the KBFF models developed for previous solutes have been shown to be competitive with existing FFs in their ability to reproduce other properties of solution mixtures and the properties of pure liquids.[7, 8, 23–27] Traditionally, solution properties are not necessarily well reproduced with other FFs unless, of course, a certain property was the focus of the parameterization effort.[9, 27] Here we explored the ability of the KBFF to reproduce (where data is available) the experimental $\Delta_{\mathrm{vap}}H$, α_p, κ_T, and dielectric constants (ε) of the pure liquids and the translational self-diffusion coefficient (D_{t}) and $\Delta_{\mathrm{mix}}H$ and $\Delta_{\mathrm{mix}}V^{\mathrm{E}}$ of the solutions. No values were explicitly corrected for polarization or vibrational frequency changes.

Pure liquid properties

As mentioned previously, the iterative nature of the approach used for obtaining the parameters of C_{ar}, in which the charge distribution was adjusted after the choice of the LJ σ and ε was made, resulted in small deviations from experiment for $\Delta_{\mathrm{vap}}H$. Additional errors in our results may have been due, at least in part, to the omission of polarizability and a lack of quantum corrections for vibrations, since not all the necessary data was available for calculation of these contributions. However, the SPC/E water model and the KBFF MOH model have been corrected for polarizability and have been shown to accurately reproduce the $\Delta_{\mathrm{vap}}H$ data.[29, 96] Thus, polarizability was at least implicitly accounted for in systems where one species was either water or MOH. Nevertheless, the enthalpy of vaporization data was in good agreement with experiment considering only the Ben values were used during the parameterization procedure. The values of the α_p (Table 3) were also in good agreement with experiment.

The calculated compressibilities of the pure liquids (Table 3) were consistently lower than the experimental κ_T values. The cause of this deviation was unclear. During an early iteration of the parameters for the C_{ar} of Ben, the compressibility of pure Ben was observed to be severely low (Ben $\kappa_{T,Exp}$=9.67×10^{-5} bar^{-1}, Ben $\kappa_{T,Sim}$=4.39×10^{-5} bar^{-1}), which is actually lower than the experimental and simulated values of water (water $\kappa_{T,Exp}$=4.5×10^{-5} bar^{-1}). This was related to an inability to equilibrate and/or maintain the translational temperature (T_{trans}) of the aromatic molecules. The value of T_{trans} was consistently too low

for the pure aromatics (output T_{trans}=273 K for pure Ben, with T_{Sim}=308 K). Many attempts were made to elucidate the main cause for the low T_{trans} of the pure aromatics, *e.g.* testing different temperature coupling algorithms, removing bond constraints while using a timestep of 0.5 fs, running canonical ensemble (*NVT*) simulations to eliminate the possibility of incorrect coupling between *T* and *p* induced by the timestep, *etc*. The only improvement (increase) in the values of T_{trans} and κ_T was obtained on removal of the H-C_{ar}-C_{ar}-C_{ar} improper torsional potentials in Ben and the other aromatic molecules. It should be noted that the improper torsional potentials that were removed are not required to maintain planarity in other force fields.[3] This led to an output T_{trans}=299 K (T_{Sim}=308 K) and an improved $\kappa_{T,Sim}=7.5\times10^{-5}$ bar^{-1} for Ben. Of the properties we investigated, the removal of the improper potentials only affected T_{trans} and κ_T for the pure liquids, with the properties of mixtures being essentially unaffected.

Values for the dielectric constants of the pure liquids are provided in Table 3. These values agreed moderately well with experiment, with the exclusion of pCr in which the appropriate trend in the values on going from PhOH to pCr was not observed. Simulated dielectric constants as a function of composition are shown in Figure 6 and exhibited the expected trends with composition. Unfortunately, we could not find any appropriate experimental data for comparison.

Solution properties

The $\Delta_{mix}H$ results are shown in Figure 7 for the systems in which experimental data was available. The simulated results underestimated the mixing enthalpy for Ben + MOH solutions, although the Tol + MOH values were reasonable. Unfortunately, the simulated values were not sensitive to changes in the charge distribution. Recent investigations seem to indicate that explicit polarization may be required for obtaining accurate enthalpies of mixing, and that even reasonable models can provide incorrect signs.[97, 98] The $\Delta_{mix}H$ for Py + MOH indicated that mixing of the solute and solvent was too unfavourable, while the $\Delta_{mix}H$ for Py + water was well reproduced across the full composition range. We are unsure of the reason for this disagreement; however we must point out that the $\Delta_{mix}H$ is a property that is nontrivial to model correctly, and similar deviations from experiment are common in the literature.[97, 99] Through iterative tests on the aromatic carbon's σ and ε, and through iterative tests on the partial charges of Ben, the $\Delta_{mix}H$ changes for Ben + MOH mixtures were always small. Hence, it was not clear which parameters should be modified in order for the model to reach agreement with experiment, and it appeared doubtful that a single Py model could reproduce both the water and methanol data.

In Figure 8 we provide a comparison of the excess volume of mixing for several of the solutions investigated here. The results appear to be very reasonable, especially in comparison to other FFs.[100] Previous studies have also indicated systematic deviations from experiment for the solution volume (or density) data.[100] There appears to be no such trend in the models developed here.

The experimental translational self-diffusion coefficients for Ben + MOH and Ben + Tol were available and were compared to the simulation results obtained using the KBFF models (Figure 9).[104] For Ben + MOH, $D_{t,Ben}$ was in close agreement with the experimentally observed trend that as the concentration of Ben increases $D_{t,Ben}$ decreases, except at x_{Ben} = 0.90 where we saw a small increase in $D_{t,Ben}$.[104] Experimentally, $D_{t,MOH}$ decreases as x_{Ben} increases from x_{Ben} = 0.00 to x_{Ben} ~ 0.50, and then the $D_{t,MOH}$ increases as x_{Ben} approaches 1.0.[104] Our model for Ben + MOH does not capture the increase in $D_{t,MOH}$ at low x_{MOH}, but the low simulated values seemed to agree with the presence of MOH-MOH clusters observed in our simulations and by others,[73–75, 105, 106] which one would expect to diminish the rate of diffusion of MOH. The experimental work referenced for the Ben + Tol system

describes the diffusion in this system using a single linear fit in which the values of $D_{t,Ben}$ were assumed equal to the values of $D_{t,Tol}$ over the entire composition range. Simulated values for Ben + Tol spanned the experimental values across the entire range of compositions. We note that our results were not corrected for finite size effects, which typically result in 5–10% larger diffusion coefficients.[107]

Conclusions

We have presented the Kirkwood Buff Force Field parameters for the side chains of the aromatic amino acids and have shown how simulations of these models compare with the available experimental thermodynamic and structural data in solutions over a range of compositions. The ability of our simple nonpolarizable force fields to reproduce the KBIs over the full composition range is a non-trivial result.[8, 19, 21, 25, 28, 108] For example, the ability to accurately model benzene + methanol and toluene + methanol activities over the range from pure methanol to pure benzene or pure toluene suggests that we will have a good model for phenylalanine in a variety of environments, *i.e.* as a free amino acid, as a solvent exposed residue, or as a residue buried deeply within a hydrophobic protein core. Furthermore, the ability of the classical models used in the KBFF to mimic the experimentally observed aromatic π-effects can now be tested and compared with other FFs. Finally, reasonable results were obtained for other solution properties not included in the parameter development.

The choice of a force field should depend upon which properties of a system a particular researcher wishes to probe. The overall message of this work is not meant to be that traditional force fields are fundamentally inferior to the KBFF models. We would be greatly amiss not to recognize that in many areas biological simulations using traditional force fields have certainly been successful at reproducing experimental properties, and even providing new hypotheses for experimentalists to subsequently test. Where we believe we have made an important improvement over traditional force fields is in quantifying the balance of intermolecular interactions. Before adopting the KBFF models it is always prudent to compare results from other force fields and experiment with results obtained using the KBFF for the properties of interest.[76]

Acknowledgments

The project described was supported by grant R01GM079277 from the National Institute of General Medical Sciences. EAP was supported by the KSU NSF GK-12 program under grant NSF DGE-0841414 and the NSF GRF program during this research. The content is solely the responsibility of the authors and does DGE-0841414 and the NSF GRF program during this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Science, the National Institutes of Health, or the National Science Foundation.

Notes and references

1. Ben-Naim, A. Molecular Theory of Solutions. New York: Oxford University Press, Inc.; 2006.
2. Christen M, Hunenberger P, Bakowies D, Baron R, Burgi R, Geerke D, Heinz T, Kastenholz M, Krautler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF. J. Comput. Chem. 2005; 26:1719. [PubMed: 16211540]
3. Brooks B, Brooks C, Mackerell A, Nilsson L, Petrella R, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner A, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor R, Post C, Pu J, Schaefer M, Tidor B, Venable R, Woodcock H, Wu X, Yang W, York D, Karplus M. J. Comput. Chem. 2009; 30:1545. [PubMed: 19444816]
4. Jorgensen W, Tirado-Rives J. J. Am. Chem. Soc. 1988; 110:1657.

5. Roitberg, GS.; Kolossváry, I.; Wong, KF.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, DH.; Seetin, MG.; Sagui, C.; Babin, V.; Kollman, PA. Amber 10. San Francisco: University of California; 2008.
6. Arnautova Y, Jagielska A, Scheraga H. J. Phys. Chem. B. 2006; 110:5025. [PubMed: 16526746]
7. Weerasinghe S, Smith PE. J. Chem. Phys. 2003; 118:10663.
8. Weerasinghe S, Smith PE. J. Chem. Phys. 2003; 119:11342.
9. Auffinger P, Cheatham T, Vaiana A. J. Chem. Theory Comput. 2007; 3:1851.
10. Lopes P, Roux B, Mackerell A. Theor. Chem. Acc. 2009; 124:11. [PubMed: 20577578]
11. Cramer, C. Essentials of Computational Chemistry: Theories and Models. Second edn.. West Sussex: John Wiley & Sons Ltd; 2004.
12. Chen B, Potoff J, Siepmann J. J. Phys. Chem. B. 2001; 105:3093.
13. Rai N, Siepmann J. J. Phys. Chem. B. 2007; 111:10790. [PubMed: 17713943]
14. Potoff J, Siepmann J. AIChE J. 2001; 47:1676.
15. Smith PE, van Gunsteren WF. J. Mol. Biol. 1994; 236:629. [PubMed: 7508990]
16. Ma L, Pegram L, Record M, Cui Q. Biochem. 2010; 49:1954. [PubMed: 20121154]
17. Shukla D, Shinde C, Trout B. J. Phys. Chem. B. 2009; 113:12546. [PubMed: 19697945]
18. Pierce V, Kang M, Aburi M, Weerasinghe S, Smith PE. Cell Biochem. Biophys. 2008; 50:1. [PubMed: 18043873]
19. Kang M, Smith PE. Fluid Phase Equilib. 2007; 256:14.
20. Smith PE. J. Phys. Chem. B. 2004; 108:18716.
21. Weerasinghe S, Smith PE. J. Chem. Phys. 2003; 118:5901.
22. Anslyn, EVD.; Dougherty, DA. Modern Physical Organic Chemistry. Sausalito: University Science Books; 2006.
23. Bentenitis N, Cox N, Smith PE. J. Phys. Chem. B. 2009; 113:12306. [PubMed: 19681588]
24. Gee M, Cox N, Jiao Y, Bentenitis N, Weerasinghe S, Smith PE. J. Chem. Theory Comput. 2011; 7:1369. [PubMed: 21789033]
25. Kang M, Smith PE. J. Comput. Chem. 2006; 27:1477. [PubMed: 16823811]
26. Ploetz EA, Bentenitis N, Smith PE. Fluid Phase Equilib. 2010; 290:43. [PubMed: 20161692]
27. Weerasinghe S, Smith PE. J. Phys. Chem. B. 2003; 107:3891.
28. Weerasinghe S, Smith PE. J. Chem. Phys. 2004; 121:2180. [PubMed: 15260772]
29. Weerasinghe S, Smith PE. J. Phys. Chem. B. 2005; 109:15080. [PubMed: 16852908]
30. Kirkwood J, Buff F. J. Chem. Phys. 1951; 19:774.
31. Ma J, Dougherty D. Chem. Rev. (Washington, DC, U. S.). 1997; 97:1303.
32. Meyer E, Castellano R, Diederich F. Angew. Chem. Int. Ed. 2003; 42:1210.
33. Berendsen H, Grigera J, Straatsma T. J. Phys. Chem. 1987; 91:6269.
34. Matteoli E, Lepori L. J. Chem. Phys. 1984; 80:2856.
35. Smith PE. J. Chem. Phys. 2008; 129:124509. [PubMed: 19045038]
36. Sumer K, Thompson A. J. Chem. Eng. Data. 1967; 12:489.
37. Wilson G. J. Am. Chem. Soc. 1964; 86:127.
38. Nagata I. J. Chem. Thermodyn. 1988; 20:467.
39. Park S, Kim H, Han K, Won D, Lee S, Choi M. Fluid Phase Equilib. 2001; 180:361.
40. Kassmann K, Knapp H. Berichte Der Bunsen-Gesellschaft-Phys. Chem. Chem. Phys. 1986; 90:452.
41. Klauck M, Grenner A, Taubert K, Martin A, Meinhardt R, Schmelzer J. Ind. Eng. Chem. Res. 2008; 47:5119.
42. Nakanishi K, Ashitani K, Touhara H. J. Chem. Thermodyn. 1976; 8:121.
43. Yasuda Y, Tochio N, Sakurai M, Nitta K. J. Chem. Eng. Data. 1998; 43:205.
44. Tsurko E, Neueder R, Kunz W. J. Solution Chem. 2007; 36:651.
45. Farrjones S, Wong W, Gutheil W, Bachovchin W. J. Am. Chem. Soc. 1993; 115:6813.

46. Henry B, Tekely P, Delpuech J. J. Am. Chem. Soc. 2002; 124:2025. [PubMed: 11866617]
47. Cheng F, Sun H, Zhang Y, Mukkamala D, Oldfield E. J. Am. Chem. Soc. 2005; 127:12544. [PubMed: 16144402]
48. Huang Z, Lin Z, Song C. J. Phys. Chem. A. 2007; 111:4340. [PubMed: 17474721]
49. Blomberg F, Maurer W, Ruterjans H. J. Am. Chem. Soc. 1977; 99:8149. [PubMed: 925263]
50. Chitra R, Smith PE. J. Phys. Chem. B. 2002; 106:1491.
51. van Gunsteren, WF.; Eising, AA.; Hünenberger, PH.; Krüger, P.; Mar, AE.; Scott, WRP.; Tironi, IG. Biomolecular Simulation: The Gromos96 Manual and User Guide. Zürich, Groningen: Hochschulverlag AG an der ETH Zürich and BIOMOS b.v.; 1996.
52. Miller K, Savchik J. J. Am. Chem. Soc. 1979; 101:7206.
53. Daura X, Mark A, van Gunsteren WF. J. Comput. Chem. 1998; 19:535.
54. Gee MB, Weerasinghe S, Smith PE. To be submitted. 2011
55. Hess B, Kutzner C, van der Spoel D, Lindahl E. J. Chem. Theory Comput. 2008; 4:435.
56. Berendsen H, Postma J, van Gunsteren WF, Dinola A, Haak J. J. Chem. Phys. 1984; 81:3684.
57. Hess B, Bekker H, Berendsen H, Fraaije J. J. Comput. Chem. 1997; 18:1463.
58. Miyamoto S, Kollman P. J. Comput. Chem. 1992; 13:952.
59. Hockney R, Goel S, Eastwood J. J. Comput. Phys. 1974; 14:148.
60. Darden T, York D, Pedersen L. J. Chem. Phys. 1993; 98:10089.
61. Chuchev K, Belbruno J. J. Phys. Chem. A. 2005; 109:1564. [PubMed: 16833478]
62. Larsen N, Fm Nicolais. J. Mol. Struct. 1974; 22:29.
63. Zierkiewicz W, Michalska D, Hobza P. Chem. Phys. Lett. 2004; 386:95.
64. Walser R, Mark A, van Gunsteren WF, Lauterbach M, Wipff G. J. Chem. Phys. 2000; 112:10450.
65. Seddon, JM.; Gale, JD. Thermodynamics and Statistical Mechanics. The Royal Society of Chemistry; 2001.
66. Smith PE, van Gunsteren WF. J. Chem. Phys. 1994; 100:3169.
67. Jorgensen W, Laird E, Nguyen T, Tirado-Rives J. J. Comput. Chem. 1993; 14:206.
68. Mackerell A, Bashford D, Bellott M, Dunbrack R, Evanseck J, Field M, Fischer S, Gao J, Guo H, Ha S, Joseph-Mccarthy D, Kuchnir L, Kuczera K, Lau F, Mattos C, Michnick S, Ngo T, Nguyen D, Prodhom B, Reiher W, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. J. Phys. Chem. B. 1998; 102:3586.
69. Mackerell A, Feig M, Brooks C. J. Comput. Chem. 2004; 25:1400. [PubMed: 15185334]
70. Sorin E, Pande V. Biophys. J. 2005; 88:2472. [PubMed: 15665128]
71. Oostenbrink C, Villa A, Mark A, van Gunsteren WF. J. Comput. Chem. 2004; 25:1656. [PubMed: 15264259]
72. Jeffrey G, Ruble J, Mcmullan R, Pople J. Proc. Roy. Soc. London Series A-Math. Phys. Eng. Sci. 1987; 414:47.
73. Buck U, Huisken F. Chem. Rev. (Washington, DC, U. S.). 2000; 100:3863.
74. Zoranic L, Sokolic F, Perera A. J. Chem. Phys. 2007; 127:024502. [PubMed: 17640132]
75. Wright D, Elshall M. J. Chem. Phys. 1996; 105:11199.
76. Domanska U, Zawadzki M. J. Chem. Eng. Data. 2010; 55:5413.
77. Ploetz EA, Bentenitis N, Smith PE. J. Chem. Phys. 2010; 132:164501. [PubMed: 20441282]
78. Lide, DR. CRC Handbook of Chemistry and Physics. Eighty-ninth edn.. Boca Raton: Taylor & Francis Group; 2008.
79. Moravkova L, Wagner Z, Linek J. J. Chem. Thermodyn. 2008; 40:607.
80. Mooney D, Muller-Plathe F, Kremer K. Chem. Phys. Lett. 1998; 294:135.
81. Yang C, Liu Z, Lai H, Ma P. J. Chem. Eng. Data. 2006; 51:457.
82. Jorgensen W, Mcdonald N. Theochem-J. Molec. Structure. 1998; 424:145.
83. Das A, Gadalla N, Kudchadker S, Marsh K, Rodgers A, Wilhoit R. J. Phys. Chem. Ref. Data. 1993; 22:659.

84. Mayer, V. IUPAC Data Series No. 32: Enthalpies of vaporization of organic compounds. Oxford: Blackwell Scientific Publications; 1985.
85. Pandey J, Shukla A, Tripathi N, Dubey G. Pramana-J. Phys. 1993; 40:81.
86. Tardajos G, Aicart E, Costas M, Patterson D. J. Chem. Soc.-Faraday Transactions I. 1986; 82:2977.
87. Faizullaev, SF. Optiko-akusticheskie, elektricheskie i magnitnye issledovaniya kondensirovannykh sred. Alishera Navoi: Samarkandskii Gosudarstvennyi Universitet im; 1982. p. 8
88. Alonso I, Alonso V, Mozo I, De La Fuente I, Gonzalez J, Cobos J. J. Chem. Eng. Data. 2010; 55:2505.
89. Tripathi B, Awasthi A, Pandey P. Indian J. Phys. 2010; 84:449.
90. Bhatia SCBSC, Rani R, Bhatia R. J. Chem. Eng. Data. 2011; 56:1669.
91. Cappelli C, Mennucci B, Cammi R, Rizzo A. J. Phys. Chem. B. 2005; 109:18706. [PubMed: 16853406]
92. Wypych, G. Knovel Solvents-A Properties Database. New York: ChemTech Publishing, Toronto; 2000.
93. Petrescu V. Ann. Sci. Univ. Jassy. 1940; 26:233.
94. Lide, DR. CRC Handbook of Chemistry & Physics. Boca Raton: CRC Press, Inc.; 1990–1.
95. Ahire S, Chaudhari A, Lokhande M, Mehrotra S. J. Solution Chem. 1998; 27:993.
96. Vega C, Abascal J, Conde M, Aragones J. Faraday Discuss. 2009; 141:251. [PubMed: 19227361]
97. Dai J, Li X, Zhao L, Sun H. Fluid Phase Equilib. 2010; 289:156.
98. Dai J, Wang L, Sun Y, Sun H. Fluid Phase Equilib. 2011; 301:137.
99. Lee M, van der Vegt N. J. Chem. Phys. 2005; 122:114509. [PubMed: 15836231]
100. Wensink EJW, Hoffmann AC, Van Maaren PJ, van der Spoel D. J. Chem. Phys. 2003; 119:7308.
101. Rajendran M, Renganarayanan S, Madhavan P, Srinivasan D. J. Chem. Eng. Data. 1989; 34:375.
102. Letcher T, Prasad A, Mercerchalmers J. South African Journal of Chemistry-Suid-Afrikaanse Tydskrif Vir Chemie. 1991; 44:17.
103. Chan T, Vanhook W. J. Chem. Soc.-Faraday Transactions I. 1976; 72:583.
104. Johnson P, Babb A. J. Phys. Chem. 1956; 60:14.
105. Reimschussel W, Hawlicka A. Berichte Der Bunsen-Gesellschaft-Phys. Chem. Chem. Phys. 1997:1221.
106. Pires M, Deturi V. J. Chem. Theory Comput. 2007; 3:1073.
107. Yeh I, Hummer G. J. Phys. Chem. B. 2004; 108:15873.
108. Chitra R, Smith PE. J. Chem. Phys. 2001; 115:5521.

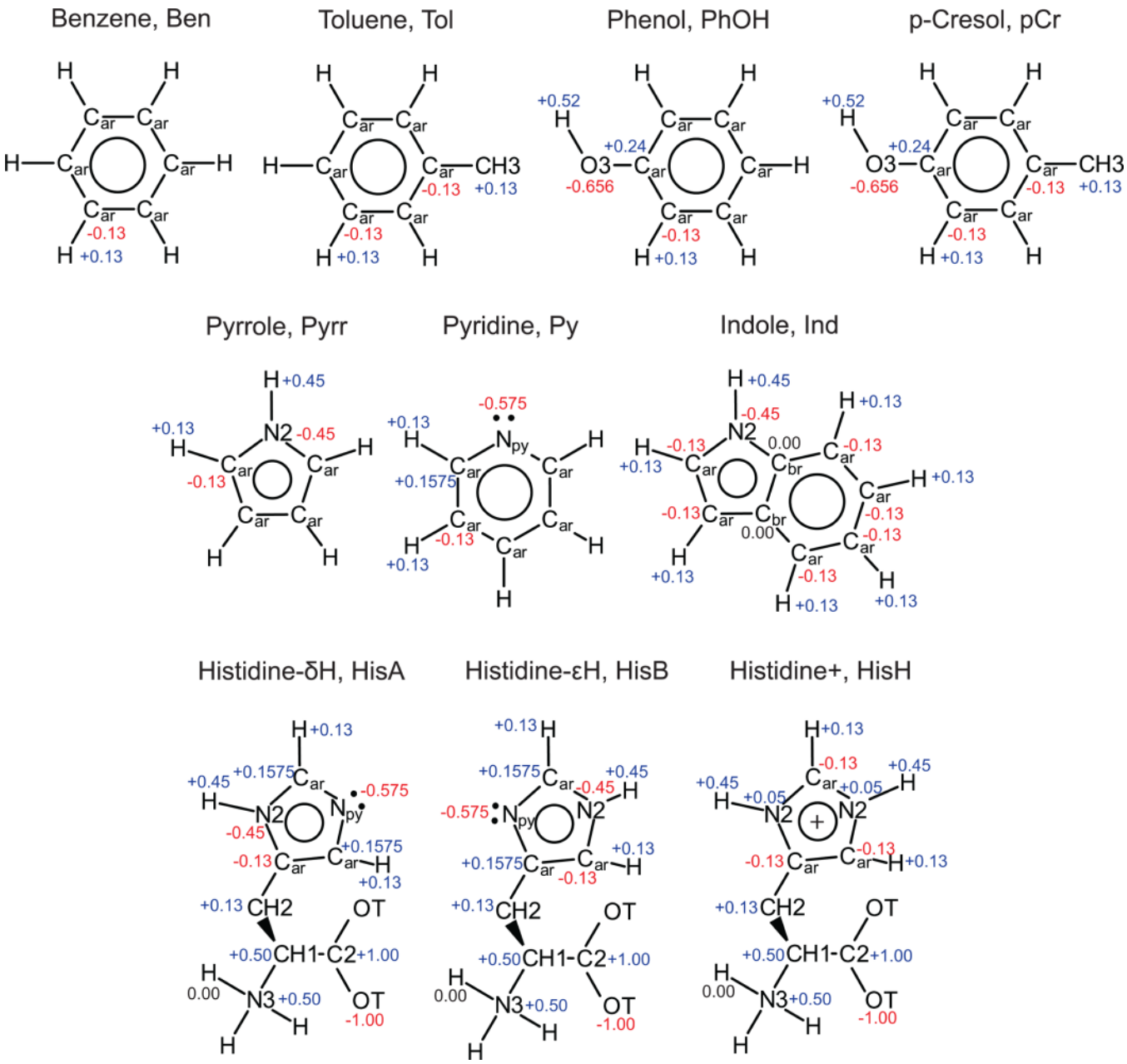

Figure 1.
Parameterized partial charges for the aromatics under study (labelled by atom type). The corresponding Lennard-Jones parameters may be found in Table 1.

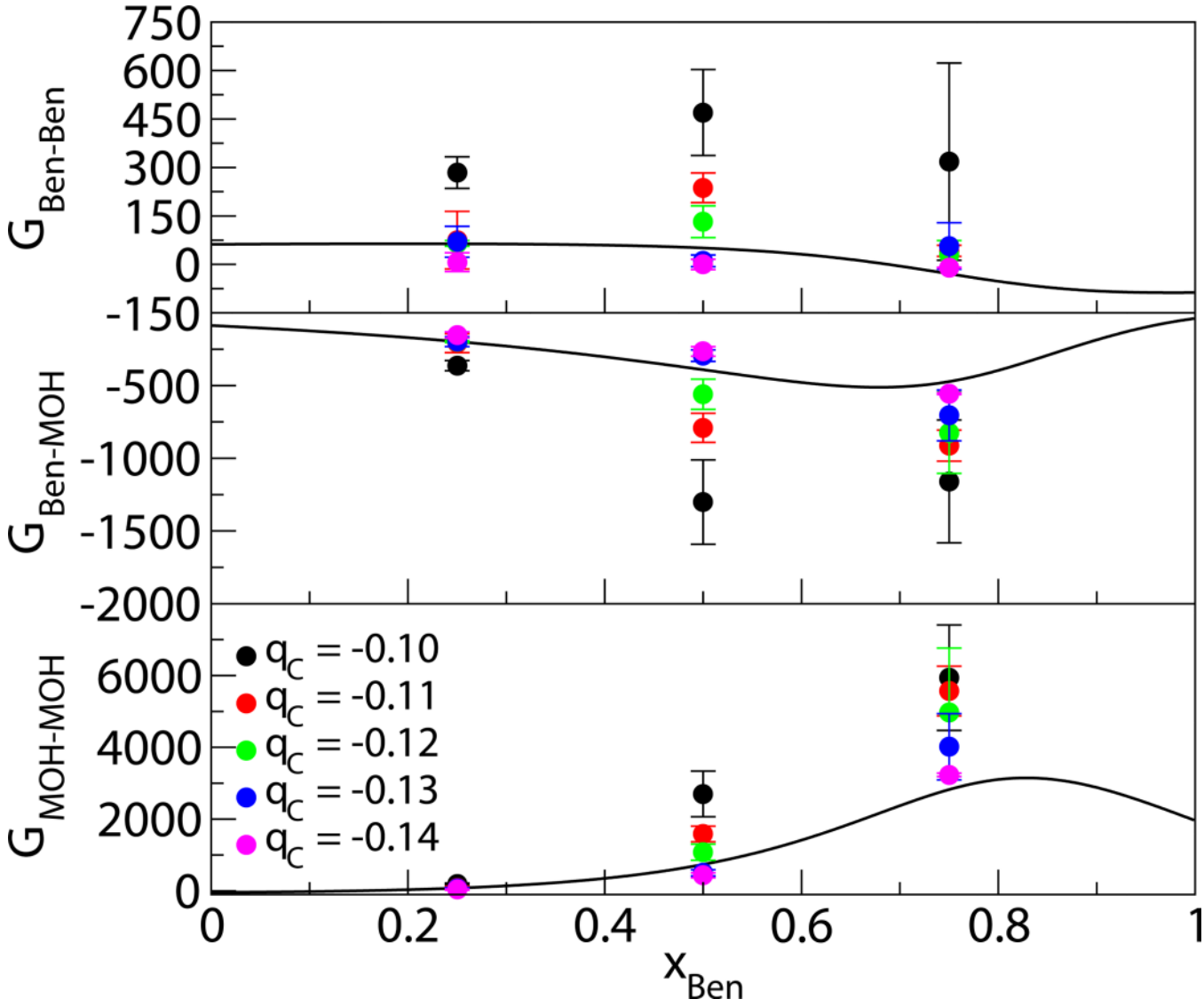

Figure 2.

Simulated KBIs (cm^3/mol) as a function of the partial charge (q_C) on the aromatic carbon. Experimental KBIs are shown as solid lines and simulated values are shown as points. Error bars show the standard deviation between multiple 5 ns block averages.

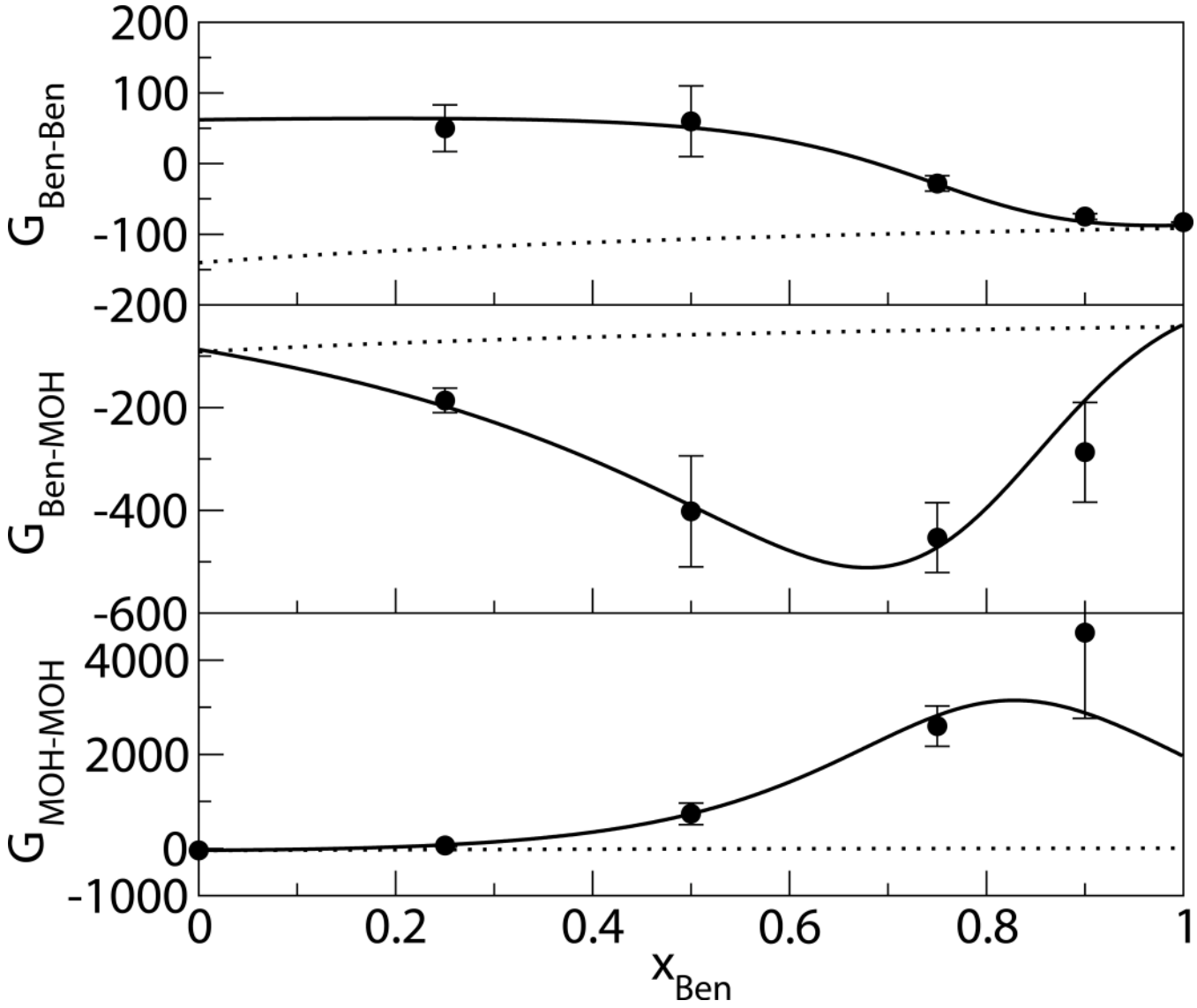

Figure 3.

The Ben+MOH KBIs (cm^3/mol) obtained from analysing the experimental data (solid lines) help quantify this system's large deviation from symmetric ideality (dotted lines). Features of this solution were closely reproduced by MD simulation (points). Error bars show the standard deviation between multiple 5 ns block averages.

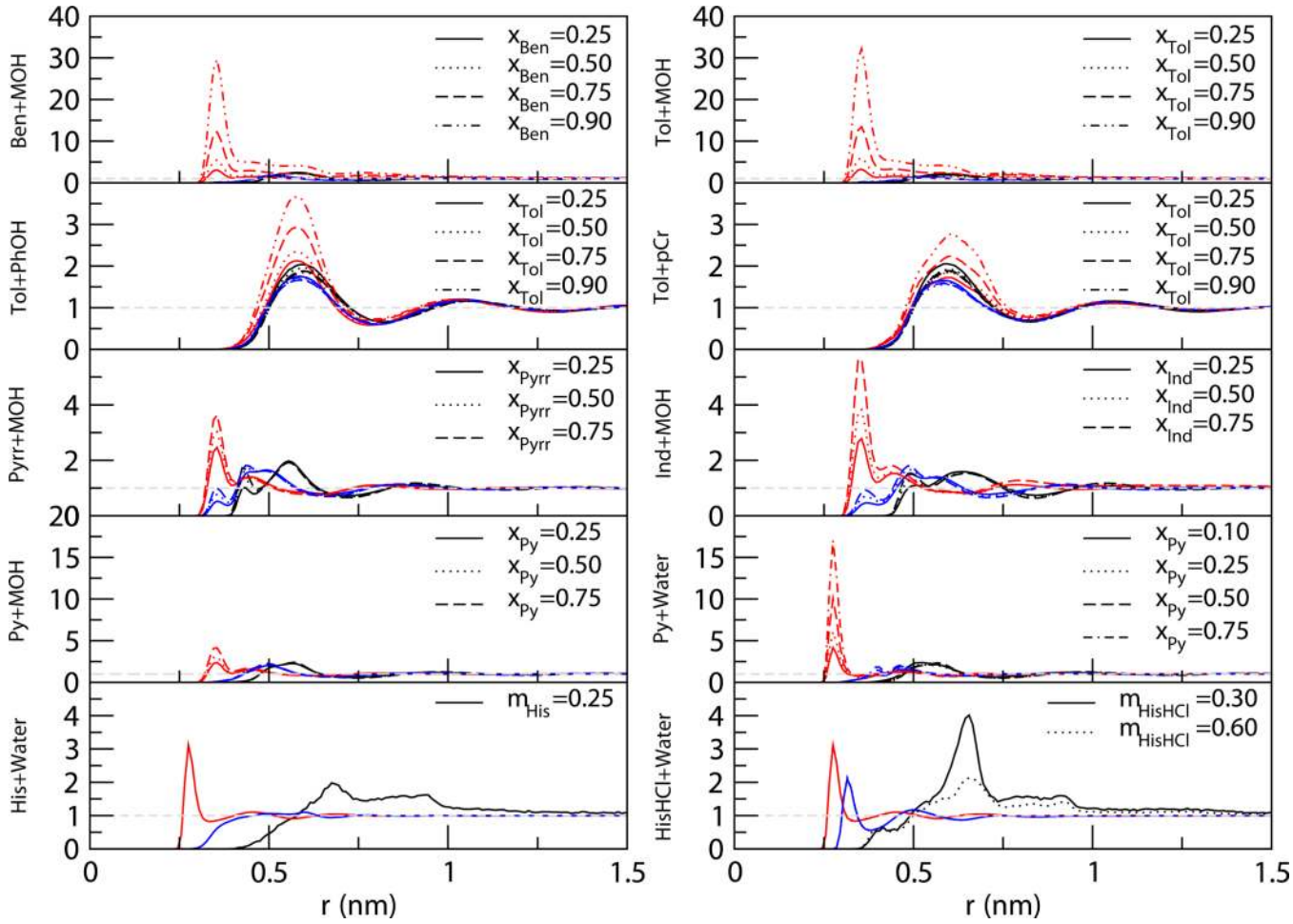

Figure 4.

Simulated center of mass (com) radial distribution functions (rdf) for the systems at each composition under study. Colours correspond to black: g_{11}, red: g_{22}, and green: g_{12}, in which systems are numbered (1) + (2) on each *y*-axis.

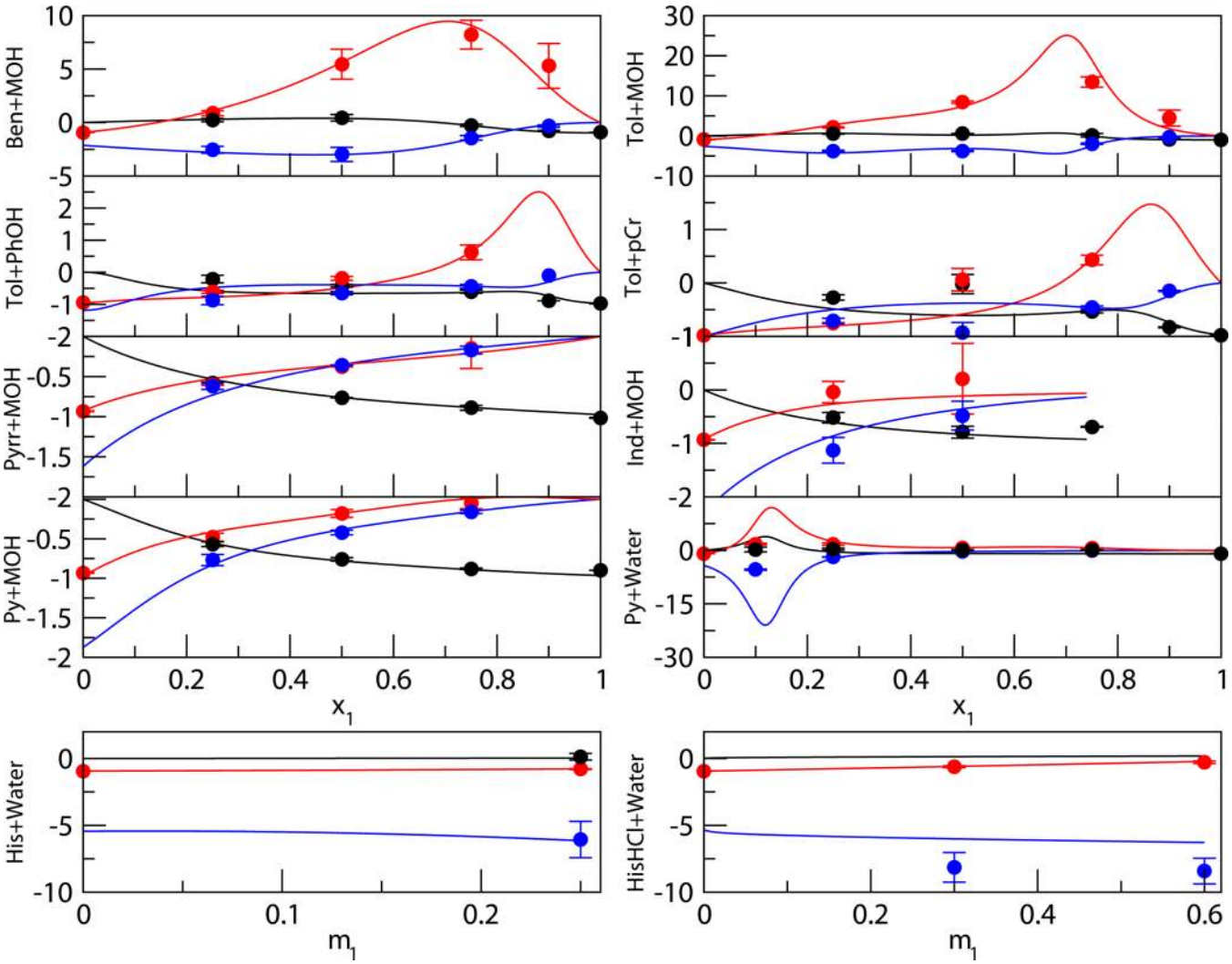

Figure 5.

Comparison of the experimental (lines) and simulated (points) excess coordination numbers (N_{ij}) for the systems under study. Colours correspond to black: N_{11}, red: N_{22}, and green: N_{12}, in which systems are numbered (1) + (2) on each *y*-axis. Error bars show the standard deviation between multiple 5 ns block averages.

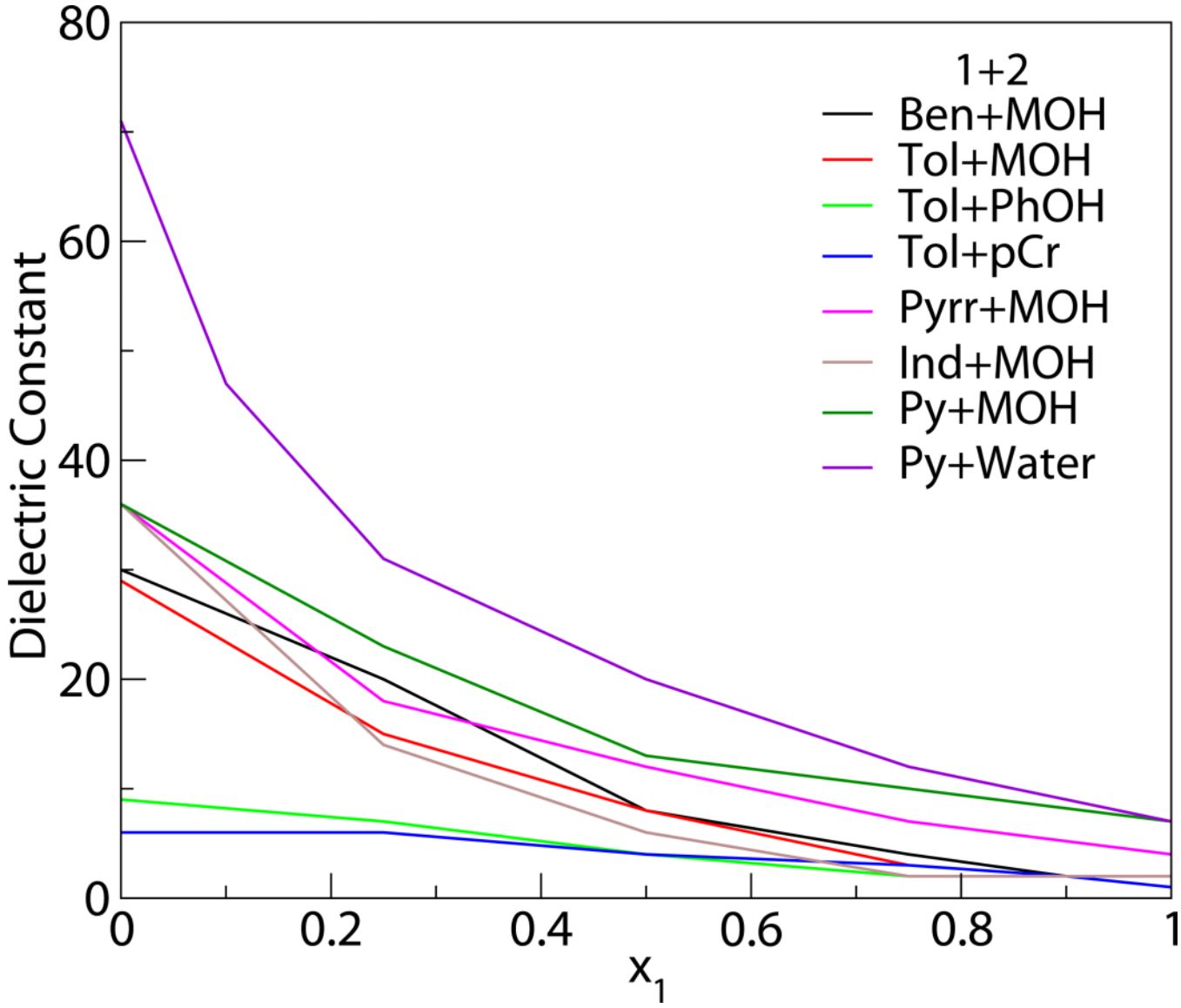

Figure 6.
Simulated dielectric constants for the systems under study.

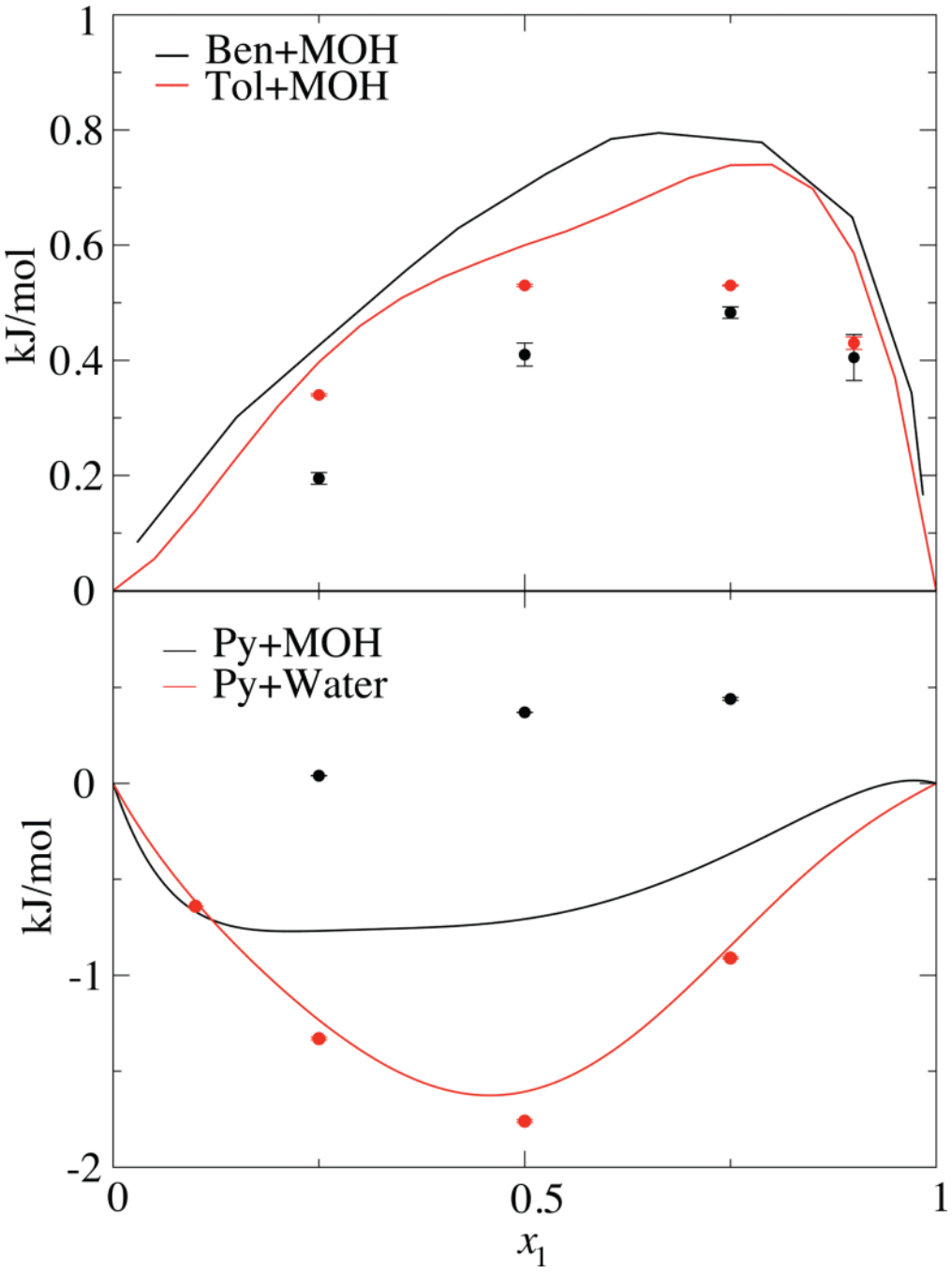

Figure 7.
Comparison of experimental (lines) and simulated (points) $\Delta_{mix}H$ (kJ/mol) for four systems. Ben + MOH T_{sim} = 308 K *vs.* T_{exp} = 303 K,[101] Tol + MOH T_{sim} = 318 K *vs.* T_{exp} = 298 K,[102] Py + MOH and Py + water $T_{sim} = T_{exp}$ = 298 K.[42, 103]

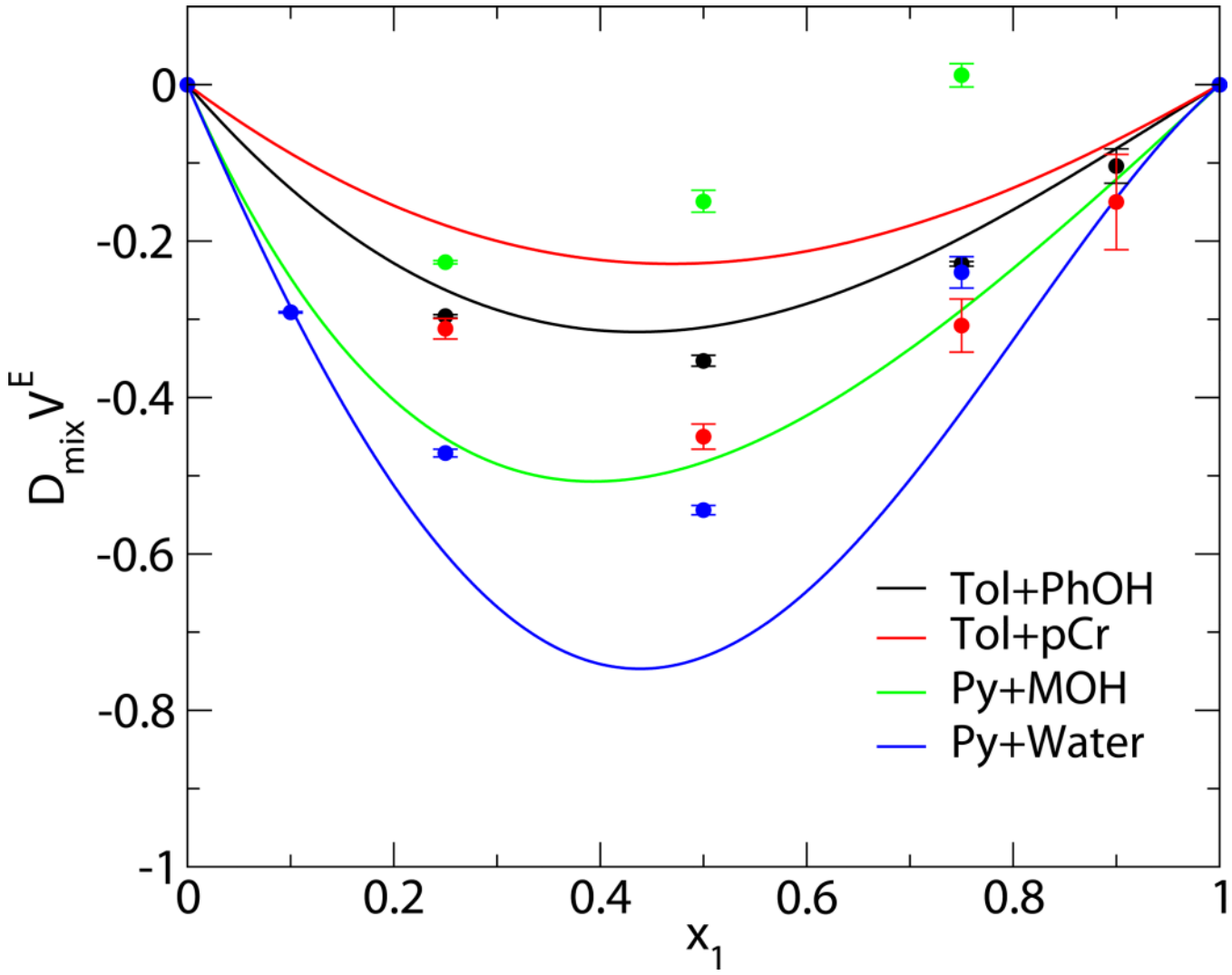

Figure 8.
Comparison of experimental (lines) and simulated (points) excess volumes of mixing (cm^3/mol) as a function of composition.

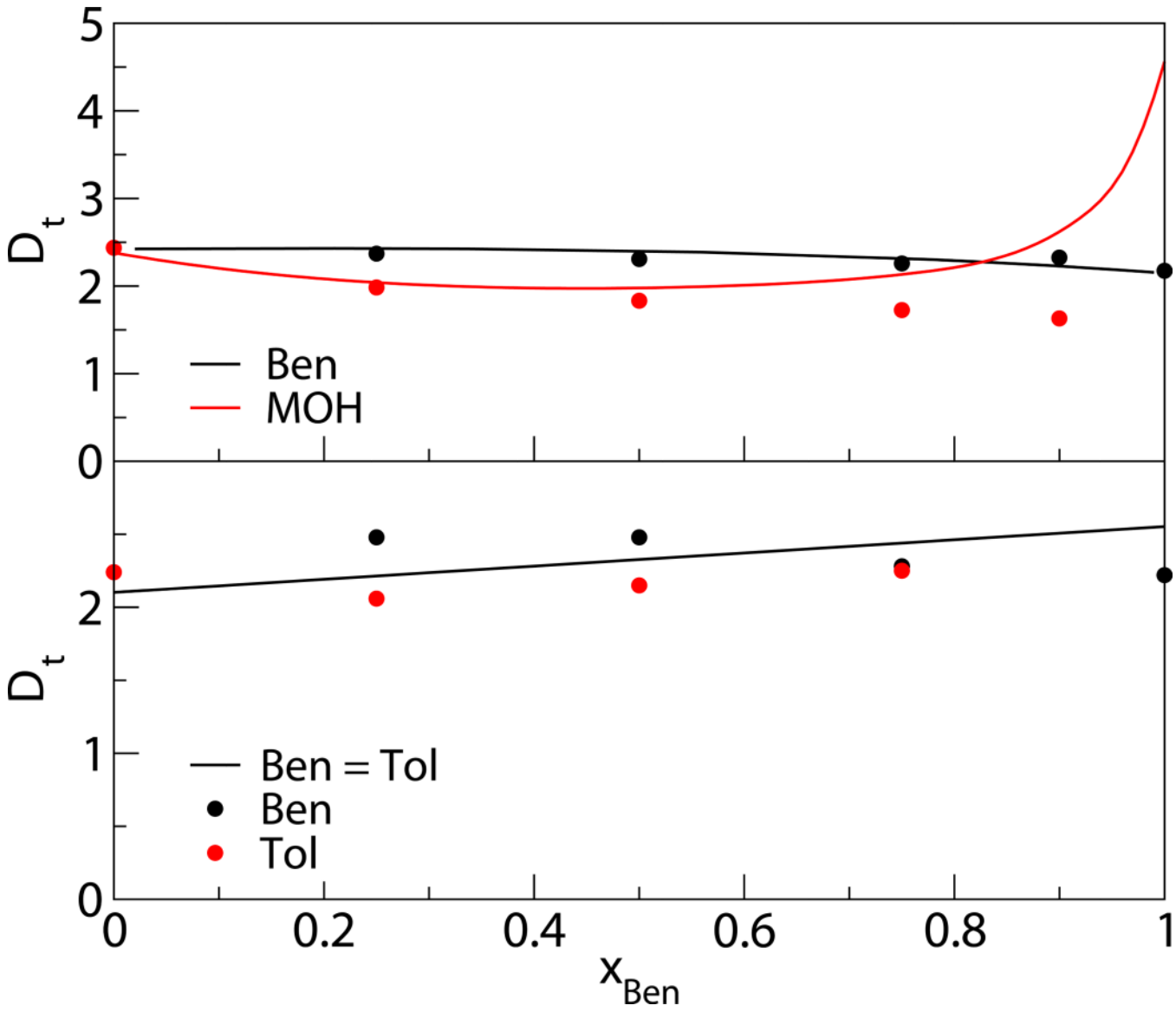

Figure 9.
Comparison of experimental (lines) and simulated (points) translational self-diffusion coefficients (D_t) for Ben + MOH (top) and Ben + Tol (bottom) ($\times 10^9$ m^2s^{-1}). Ben + MOH T_{sim} = 308 K *vs.* T_{exp} = 298 K,[104] Ben + Tol T_{sim} = 313 K *vs.* T_{exp} = 308 K.[105]

Table 1

Lennard-Jones parameters

Atom type	Description	ε (kJ/mol)	σ (nm)	Reference
H	all non-water H	0.0880	0.1580	[7]
C2	sp^2	0.3300	0.3360	[7]
CH1	united CH	0.0949	0.5019	[53]
CH2	united CH_2	0.4105	0.4070	[53]
CH3	united CH_3	0.8672	0.3748	[53]
C_{ar}	sp^2 aromatic	0.3300	0.3810	this work
C_{br}	sp^2 bridging C	0.4170	0.3770	[7]
N2	sp^2	0.5000	0.3110	[7]
N_{py}	pyridine-like	0.4530	0.3160	this work
N3	sp^3	0.5620	0.3370	[24, 54]
O3	sp^3	0.6506	0.3192	[29]
OT	terminal O	0.6047	0.3500	[24, 54]

Table 2

A Comparison of Aromatic (Hydrogen/Carbon) Partial Charges for Different Force Fields

Force Field	Ben	Phe	Reference
TraPPE-EH	+/−0.095		[13]
CHARMM27	+/−0.115	+/−0.115	[68, 69]
OPLS-AA	+/−0.115	+/−0.115	[67]
AMBER99sb	+/−0.130521	+0.0118 (CG)	[70]
		−0.1256 (CD)	
		+0.1330 (HD)	
		−0.1704 (CE)	
		+0.1430 (HE)	
		−0.1072 (CZ)	
		+0.1297 (HZ)	
GROMOS53a6	+/−0.140	+/−0.140	[71]
KBFF	+/−0.130	+/−0.130	this work

Table 3

Comparison of experimental and simulated pure liquid properties.

		ρ (g/cm^3)			$\Delta_{vap}H$ (kJ/mol)			κ_T (×10^{+5} bar^{-1})			α_p (×10^{+3} K^{-1})			ε		
	T_{melt}	$T_{Exp/Sim}$	Exp	Sim	$T_{Exp/Sim}$	Exp	Sim	T_{Exp}	Exp	Sim	T_{Exp}	Exp	Sim	$T_{Exp/Sim}$	Exp	Sim
Ben	278	308	0.863	0.862	298	33.85	32.88	298	9.67	7.5	298	1.23	1.19	298	2.2	1
Tol	178	303	0.857	0.881	298	38.01	39.16	298	9.14	6.2	298	1.07	0.97	298	2.4	1
PhOH	314	333	1.042	1.034	333	55.4	60.18	319	5.61	4.1	328	1.05	0.76	333	10.3	9
pCr	308	318	1.014	1.028	333		67.29	323	6.07	4.1	308	0.76	0.69	314	16.4	4
Pyrr	250	298	0.973	0.973	298	45.15	43.79	298	6.08	5.5	298	0.97	0.90	291	7.5	4
Py	231	298	0.978	0.993	298	40.21	40.52	298	7.04	4.4	298	1.03	0.93	298	12.0	7

T_{melt}[78]; Density (ρ): Ben[79], Tol[36], PhOH[80], pCr[81], Pyrr[39], Py[82]; Enthalpy of vaporization ($\Delta_{vap}H$): Ben[67], Tol[78], PhOH[80], Pyrr[83], Py[84]; Isothermal compressibility (κ_T): Ben[85], Tol[86], PhOH[78], pCr[87], Pyrr[76], Py[88]; Isothermal expansion (α_p): Ben[85], Tol[85], PhOH[89], pCr[90], Pyrr[76], Py[88]; Dielectric constant (ε): Ben[91], Tol[92], PhOH[80], pCr[93], Pyrr[94], Py[95]