Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules

Pengyu Ren,*,[†] Chuanjie Wu,[‡] and Jay W. Ponder*,[‡]

[†]Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States

[‡]Department of Chemistry and Department of Biochemistry & Molecular Biophysics, Washington University, St. Louis, Missouri 63130, United States

Ⓢ *Supporting Information*

ABSTRACT: An empirical potential based on permanent atomic multipoles and atomic induced dipoles is reported for alkanes, alcohols, amines, sulfides, aldehydes, carboxylic acids, amides, aromatics, and other small organic molecules. Permanent atomic multipole moments through quadrupole moments have been derived from gas phase *ab initio* molecular orbital calculations. The van der Waals parameters are obtained by fitting to gas phase homodimer QM energies and structures, as well as experimental densities and heats of vaporization of neat liquids. As a validation, the hydrogen bonding energies and structures of gas phase heterodimers with water are evaluated using the resulting potential. For 32 homo- and heterodimers, the association energy agrees with *ab initio* results to within 0.4 kcal/mol. The RMS deviation of the hydrogen bond distance from QM optimized geometry is less than 0.06 Å. In addition, liquid self-diffusion and static dielectric constants computed from a molecular dynamics simulation are consistent with experimental values. The force field is also used to compute the solvation free energy of 27 compounds not included in the parametrization process, with a RMS error of 0.69 kcal/mol. The results obtained in this study suggest that the AMOEBA force field performs well across different environments and phases. The key algorithms involved in the electrostatic model and a protocol for developing parameters are detailed to facilitate extension to additional molecular systems.

■ INTRODUCTION

Organic molecules are the basic constituents of biology and material science. Modeling studies involving organic compounds are widely used in many areas such as physical chemistry, biological structure and function, and nanotechnology. Progress in quantum chemistry and the availability of fast computers has empowered the routine study of small molecules with high levels of *ab initio* theory and large basis sets. However, first principles statistical thermodynamics sampling techniques are still not practical for use with most high-level QM methods. Thus, molecular modeling based on empirical potentials is widely used for theoretical inquiries into microscopic and macroscopic phenomena across chemistry and biology. Atom-based force field models such as MM3,[1] AMBER,[2] CHARMM,[3] OPLS,[4] and GROMOS[5] have been developed for a wide range of organic compounds and biomacromolcules. These models describe electrostatic interactions with fixed point charges on atoms and treat van der Waals interactions via Lennard-Jones potentials or other simple functions. Numerous studies have shown that many of the physical properties and structures of organic molecules can be adequately reproduced with current fixed charge force fields. Increases in computing power have enabled the simulation of larger molecular systems and more precise investigation of their properties. However, there are acknowledged shortcomings of the current generation of fixed charge potentials. They assume the atomic charges derived from training systems are approximately transferable to systems in different chemical environments. Explicit accounting of many-body effects is required for a general potential to capture the electrostatic response to different molecular environments: homo- or heterogeneous, low or high dielectric, nonpolar or highly polarizable.

Polarization effects were initially used in the description of molecular refractivity and other chemical phenomena nearly 100 years ago.[6] Early in the era of modern computational chemistry, polarization was applied to the study of enzymatic reactions[7] and incorporated into prototype molecular dynamics algorithms.[8] Recently, there have been increasing efforts toward developing polarizable force fields for molecular simulation, based on a variety of empirical models for induction, such as classical induced dipoles,[2,9−22] fluctuating charges,[23−30] and Drude oscillators.[9,31−35] Detailed discussions of the various polarization models can be found in recent reviews of polarizable force field development.[36−40] The performance of different approaches in accounting for polarization has been compared in the study of ion and small molecule interactions.[41,42] The modeling of neat organic liquids, including alcohols, acids, amides, and aromatics, has also been reported using polarizable potentials.[11,22,35,43−50]

Restriction to fixed atomic point charges constrains the flexibility of a model in representing the electrostatic potential around a molecule[51,52] and thus limits the accuracy of the treatment of molecular interactions. Improvement can be achieved by adding extra charge sites, typically at bond centers or lone pair positions. For example, the TIPxP series of water models, TIP3P,[53] TIP4P,[53] and TIP5P,[54] adopts increasing numbers of charge sites. Recently, the extra site approach was introduced into a Drude

oscillator-based polarizable model as a way to address the anisotropy in atomic charge distribution due to lone pair electrons.[50] Alternatively, one can directly incorporate higher order moments, such as dipole and quadrupole moments, at the atomic centers to improve the representation of the charge distribution. The convergence advantage of using multipoles distributed over atomic sites, as opposed to a single molecule-centered set of moments, has been discussed in the literature.[55,56] Over two decades ago, Buckingham and Fowler[57,58] were the first to apply distributed multipole moments to structural modeling of small molecule complexes. Their proposed intermolecular potential, consisting of hard sphere repulsion and atomic multipole-based electrostatics, was able to reproduce a number of experimental equilibrium geometries and orientational preferences. Recently, coarse-grained potentials with point multipoles have been used successfully in modeling hydrogen-bonded molecular liquids.[59,60]

The AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) force field was initially developed for water.[18,20] The current study reports the extension of the AMOEBA model to organic compounds including alkanes, alcohols, amines, sulfides, aldehydes, carboxylic acids, amides, and aromatics. A cornerstone of the AMOEBA force field is an improved electrostatic potential based on atomic multipoles and classical induced dipole moments. The atomic multipole moments are obtained from high-level *ab initio* calculations on gas phase monomers. An empirical atomic dipole induction model describes the many-body polarization effects important in clusters and condensed phase environments. A small, consistent set of atomic polarizability parameters is used to treat intermolecular polarization as well as intramolecular polarization between functional group fragments. van der Waals (vdW) parameters are refined via gas-phase homodimer molecular orbital calculations and molecular dynamics simulation of liquid properties. Additional gas-phase and liquid-phase computations, including hydrogen bonding in gas-phase heterodimers, the dielectric and diffusion constants of neat liquids, and hydration free energies of organic compounds have been utilized to validate the resulting force field.

■ METHODS

Potential Energy Model. The interaction energy among atoms is expressed as

$$U = U_{bond} + U_{angle} + U_{b-a} + U_{oop} + U_{torsion}$$
$$+ U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind} \qquad (1)$$

where the first five terms describe the short-range valence interactions: bond stretching, angle bending, bond—angle cross term, out-of-plane bending, and torsional rotation. The last three terms are the nonbonded interactions: van der Waals, permanent electrostatic, and induced electrostatic contributions. The individual terms for these interactions have been described in detail in a previous publication.[61] Some additional methodology, introduced to treat electrostatic polarization in molecular systems beyond water, will be detailed below. Polarization effects in AMOEBA are treated via Thole's interactive induction model that utilizes distributed atomic polarizability.[62,63] According to this interactive induction scheme, induced dipoles produced at the atomic centers mutually polarize all other sites. A damping function is used at short-range to eliminate the polarization

catastrophe and results in correct anisotropy of the molecular response (i.e., diagonal components of the molecular polarizability tensor) starting from isotropic atomic polarizabilities. Thole damping is achieved by screening of pairwise atomic multipole interactions and is equivalent to replacing a point multipole moment with a smeared charge distribution.[13] The damping function for charges is given by

$$\rho = \frac{3a}{4\pi} \exp(-au^3) \qquad (2)$$

where $u = r_{ij}/(\alpha_i \alpha_j)^{1/6}$ is the effective distance as a function of interatomic distance r_{ij} and the atomic polarizabilities of atoms i (α_i) and j (α_j). The coefficient a is the dimensionless width of the smeared charge distribution and controls the damping strength. The corresponding damping functions for charge, dipole, and quadrupole interactions were reported previously.[18]

The Thole model is able to reproduce the molecular polarizability tensors of numerous small molecules with reasonable accuracy using only element-based isotropic atomic polarizabilities and a single value for the damping factor.[62] In our water study, it was discovered that the dependence of molecular polarizability on the damping coefficient is weak, but the polarization energy is much more sensitive to the strength of damping. After fitting the interaction energies of a series of small water clusters, we have chosen a universal damping factor of $a = 0.39$, rather than the value of 0.572 suggested by Thole. We adopt the atomic polarizabilities ($Å^3$) as originally given by Thole, i.e., 1.334 for carbon, 0.496 for hydrogen, 1.073 for nitrogen, and 0.837 for oxygen. The only exception is for aromatic carbon and hydrogen atoms, where we found that the use of larger values greatly improves the molecular polarizability tensor of benzene and polycyclic aromatics. The AMOEBA values for atomic polarizability are given in Table 1. In addition, for metal dications, we have found it necessary to use stronger damping ($a < 0.39$) to better represent the electric field around the ions.[21,61,64]

Intramolecular Polarization. For a large molecule such as a multifunctional organic or a biopolymer, polarization arises not only from the electric field of other molecules but also from distal portions of the same molecule. It is crucial to describe the intra- and the intermolecular response in a consistent manner. In prior work, we investigated the effect of intramolecular polarization on the conformational dependence of the electrostatic potential surrounding a dipeptide.[15] As observed by others,[65] the electrostatic parameters derived for alanine dipeptide vary significantly depending upon the conformation used to derive the values. A simple average over multipole moments obtained from a set of conformers does not transfer well between conformers, i.e., gives poor electrostatic potentials on different conformers. Furthermore, when short-range polarization between bonded atoms is ignored, use of intramolecular polarization yields only marginal improvement over current nonpolarizable potentials. To overcome this problem, a group based intramolecular polarization scheme has been devised.[15] The "groups" are typically functional groups with limited conformational degrees of freedom, such as an amide group or phenyl ring. *In this scheme, the permanent atomic multipoles (PAM) polarize between, and not within, groups. For a small molecule consisting of a single polarization group, such as water or ammonia, permanent atomic multipoles do not polarize sites within the same molecule, while "mutual" induction occurs among all polarizable sites as described above. This design offers a clear connection between the treatment of small molecules and that of the analogous fragments inside a larger*

Table 1. vdW Parameters and Atomic Polarizabilities for AMOEBA Atom Classes

atom	description	R^0 (Å)	ε (kcal/mol)	polarizability (Å^3)
C	alkane (CH$_3$− or −CH$_2$−)	3.820	0.101	1.334
H	alkane (CH$_3$−)	2.960	0.024 (0.92)	0.496
H	alkane (−CH$_2$−)	2.980	0.024 (0.94)	0.496
C	alkane (−CH<)	3.650	0.101	1.334
H	alkane (−CH<)	2.980	0.024 (0.94)	0.496
O	hydroxyl (water, alcohol)	3.405	0.110	0.837
H	hydroxyl (water, alcohol)	2.665	0.0135 (0.91)	0.496
O	carbonyl (aldehyde, amide, acid)	3.300	0.112	0.837
H	acid (HO)	2.665	0.0150 (0.91)	0.496
C	carbonyl (aldehyde, amide, acid)	3.820	0.106	1.334
C	aromatic carbon	3.800	0.091	1.750
H	aromatic (HC)	2.980	0.026 (0.92)	0.696
N	amine nitrogen (ammonia, amine)	3.710	0.105	1.073
H	amine (HN)	2.700	0.020 (0.91)	0.496
N	amide nitrogen	3.710	0.110	1.073
H	amide (HN)	2.590	0.022 (0.90)	0.496
S	sulfur	4.005	0.355	2.800
H	sulfhydryl (HS)	2.770	0.024 (0.96)	0.496
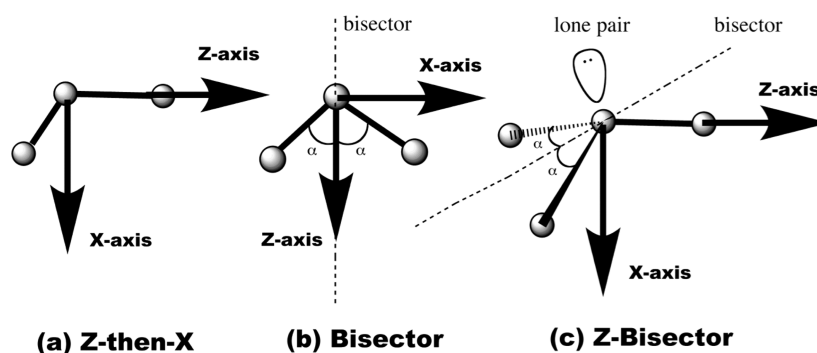
Figure 1. Local coordinate frame definitions for atomic multipole sites. (a) The Z-then-X frame is used for general sites and, with the addition of a third orthogonal *y* axis, can treat chiral centers. The majority of AMOEBA multipole sites are defined using this local frame. (b) The Bisector frame is useful for molecules with 2-fold local symmetry or pseudosymmetry, such as water and aliphatic methylene carbon atoms. (c) The Z-Bisector frame is used for sites such as the sulfur atom of dimethylsulfoxide, which have a distinct primary ("Z") axis and symmetry or pseudosymmetry along a secondary direction.

molecule, thereby facilitating the transfer of PAM values from small model compounds to larger species such as polypeptides and nucleic acids.

We use Distributed Multipole Analysis[66] (DMA) to extract atomic multipoles from *ab initio* calculations. Starting from the DMA atomic multipoles for an arbitrary conformer of a model compound, M_i^{DMA}, one can derive the intrinsic "permanent" atomic multipole moments, M_i, that satisfy

$$M_i^{\text{DMA}} = M_i + \mu_i \tag{3}$$

where μ_i is the dipole induced by intramolecular polarization by M_i The M_i are obtained by substituting μ_i from eq 4 below into the preceding equation. This approach allows derivation of conformation-independent atomic multipole parameters for larger organic compounds with multiple polarization groups. A local coordinate frame is defined at each atomic site and used to rotate atomic multipole moments as neighboring atoms move during structure manipulation. As shown in Figure 1, three types

of local frames are sufficient to handle essentially all situations arising in organic chemistry.

Polarization Energy. Formally, the induced dipole vector on any polarizable site *i* can be expressed as

$$\mu_i^{\text{ind}} = \alpha_i \left(\sum_{j \neq i} T_{ij}^1 M_j + \sum_{k \neq i} T_{ik}^{11} M_k \right) \tag{4}$$

and the associated energy is

$$U_{\text{ele}}^{\text{ind}} = -\frac{1}{2} \sum_i (\mu_i^{\text{ind}})^T E_i \tag{5}$$

where $T_{ij}^1[\nabla^1, \nabla^2, \nabla^3,$ etc.] is a 3 × 13 matrix with $\nabla^{l+m+n} = \partial^l/\partial x^l \; \partial^m/\partial y^m \; \partial^n/\partial z^n$ representing the second through fourth rows of the multipole—multipole interaction matrix T_{ij} (see Appendix, eq A2). $T_{ij}^{11} = \nabla_{ik}^2$ is a 3 × 3 submatrix consisting of elements in T_{ij}^1 corresponding to the dipole moments. As discussed above, the atomic polarizability is isotropic. Therefore, the off-diagonal elements of the tensor, α_i, are all zero, and the

three diagonal elements take the same scalar value. The factor of 1/2 is a result of the induction cost for the formation of induced dipoles.

In eq 4, the first term inside the parentheses on the right-hand side is the "direct" electric field, E, due to permanent multipoles outside the polarization group of atom i (index j). The second term corresponds to "mutual" induction by all induced dipoles (index k). *Thus, direct induction due to permanent multipoles only occurs between groups, while mutual polarization between induced dipoles involves every atom pair.* When computing energies, as opposed to induced dipoles, the scaling of $1-2$, $1-3$, and other local interactions is applied to permanent and polarized electrostatic terms as with other molecular mechanics models. In addition, dipole induction is damped at short-range to avoid the "polarization catastrophe", and damping is applied consistently to the induced field, energy, and force. For convenience, scaling and damping is assumed to be implicitly included with the **T** matrix elements in the present discussion.

The set of induced dipole equations is solved iteratively to obtain the final dipole values. The convergence is accelerated via a successive over-relaxation (SOR) procedure.[67]

$$\mu_i(n+1) = (1-\omega)\,\mu_i(n) + \omega\Big[\mu_i(0) + \alpha_i \sum_{\{k\}} T^{11}_{ik}\mu_k(n)\Big] \qquad (6)$$

where $\mu_i(0) = \alpha_i \sum T^1_{ij}M_j$ is the "direct" induced dipole moment generated by the permanent field. The default ω value is 0.7, while for the case $\omega = 1$, eq 6 reduces to eq 4.

Energy Gradient and Ewald Summation. The energy gradient due to permanent multipole moments, including force and torque components, was derived by Smith for a standard Ewald summation.[68] We have previously reported the AMOEBA Ewald force, torque, and virial arising from dipole induction in water systems.[18] Note that the pairwise direct (non-Ewald) formula can be obtained by replacing the real-space screening factor $B(r)$ with the corresponding function of $1/r^{18,68}$ and vice versa. The torque components are converted to atomic forces on the relevant frame-defining atoms in our implementation. It is also possible to derive the analytical forces corresponding to the torques directly via an infinitesimal rotation,[69] or by taking the derivative of the rotation matrix.[70] When evaluating the energy derivative directly, the additional chain rule terms due to the local frame rotation matrices are equivalent to the forces converted by means of the torque implementation. In the Appendix, we provide a derivation of the polarization energy gradient, with a focus on terms arising from intramolecular polarization.

The Ewald real-space interactions need to be modified to accommodate short-range scaling of electrostatics and damping of dipole induction as mentioned above. To scale the interaction between an atom pair, a term $(f_{scale} - 1)\,U'$ is added to the total Ewald energy, where U' is the full (non-Ewald) interaction between the pair, and the scaling factor, f_{scale}, ranges from 0 to 1. Analogous approaches are used in computing forces, fields, and torques.

Particle-mesh Ewald (PME) for point multipoles[69] has been implemented in the TINKER and AMBER/PMEMD software packages. PME significantly improves the computational efficiency as its cost scales as $N \log N$, where N is the number of particles. The addition of dipole and quadrupole moments to the PME method roughly doubles the computational expense versus point charge only models. Calculation of induced dipoles can be

time-consuming with the simple iterative solution method, depending upon the level of SCF convergence required. Alternative fast predictive induced dipole schemes have been suggested.[71,72] Acceleration via extended Lagrangian methods has been reported for induced dipole polarization[73−75] and is under investigation for the AMOEBA model.

A standard Ewald summation implies the use of "tinfoil" boundary conditions, corresponding to a system immersed in a conducting dielectric environment (i.e., $\varepsilon = \infty$). It is possible to include a boundary correction to the Ewald energy if other environments, such as insulating boundary conditions, are desired. For a cubic box, the correction term is a function of the total cell dipole moment, while for other system shapes the analytical form is difficult to derive.[76,77] Note that the energy obtained via Ewald summation is equivalent to the energy obtained using an infinitely long atom-based cutoff for the same periodic system. However, group-based cutoffs are often applied to preserve local charge neutrality. When using group cutoffs, the energy asymptotically approaches a different value from atom cutoffs as the cutoff length increases. The difference between the two energies is exactly equal to the above boundary correction term. This suggests that care must be taken if cutoff methods are applied to a system containing multipoles since the dipole and higher order moments are intrinsically group based.

Parameterization. The atomic polarizabilities are listed for each AMOEBA atom type in Table 1. The values are the same as those derived by Thole[62] except for aromatic carbon and hydrogen atoms, which have been systematically refined using a series of aromatic systems, including a small carbon nanotube (see Table 3). The molecular polarizabilities computed using the current model are compared to experimental values for selected compounds in Table 2. Reducing the damping factor from Thole's original value of 0.567 to AMOEBA's 0.39 is critical to correctly reproducing water cluster energetics.[18] On the other hand, AMOEBA's greater damping leads to a slight systematic underestimation of molecular polarizabilities. However, given the simplicity of the model, the agreement is generally satisfactory for both average polarizabilities and their anisotropies. As described above, polarization groups are defined for purposes of treating intramolecular polarization. Typically, a functional group is treated as a single polarization group. For example, methylamine is a group by itself, while ethylamine has two groups: $-CH_2NH_2$ and CH_3-. The groups are specified in AMOEBA parameter files in the following format: "polarize A α 0.39 B C", where α is the polarizability for atom type A, 0.39 is the damping coefficient in eq 2, and B and C are possible bonded atom types that belong to the same polarization group as atom type A.

The permanent atomic multipoles were derived for each molecule from *ab initio* QM calculations. *Ab initio* geometry optimization and a subsequent single-point energy evaluation were performed at the MP2/6-311G(1d,1p) level using Gaussian 03.[78] For small molecules with less than six heavy atoms, Distributed Multipole Analysis (DMA v1.2[79]) was used to compute the atomic multipole moments in the global frame using the density matrix from the QM calculation. Next, the TINKER POLEDIT program rotates the atomic multipoles into a local frame and extracts Thole-based intramolecular polarization to produce permanent atomic multipole (PAM) parameters. Thus, when the AMOEBA polarization model is applied to the permanent atomic moments, the original *ab initio* derived DMA is recovered. Finally, the POTENTIAL program from the TINKER package is used to optimize the permanent atomic multipole
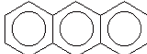
Table 2. Comparison of Experimental and Computed Molecular Polarizabilities ($\mathring{A}^3$)[a]

		α_{avg}	α_x	α_y	α_z
methane	AMOEBA	2.48	2.48	2.48	2.48
	Thole	2.55	2.55	2.55	2.55
	exptl	2.62	2.62	2.62	2.62
ethane		4.25	4.66	4.05	4.05
		4.46	4.93	4.24	4.24
		4.48	4.99	4.22	4.22
propane		6.01	6.75	5.78	5.51
		6.29	7.18	5.98	5.68
		6.38	7.66	5.74	5.74
formaldehyde		2.44	2.77	2.55	2.01
		2.54	3.07	2.70	1.86
		2.45	2.76	2.76	1.83
formamide		3.65	4.32	3.87	2.74
		3.79	4.86	4.04	2.50
		4.08 (4.22)	5.24	($\alpha_y + \alpha_z = 7.01$)	
acetamide		5.43	6.26	5.72	4.30
		5.71	6.70	6.30	4.13
		5.67	6.70	($\alpha_y + \alpha_z = 10.3$)	
methanol		3.19	3.61	3.02	2.93
		3.35	3.92	3.13	2.99
		3.32 (3.26)	4.09	3.23	2.65
ethanol		4.94	5.44	4.84	4.54
		5.08	5.76	4.98	4.50
		5.26 (5.13)	6.39	4.82	4.55
propanol		6.73	7.63	6.53	6.03
		7.21	8.42	6.89	6.30
		6.97 (6.96)			
NH_3		1.92	2.07	2.07	1.62
		1.95	2.17	2.17	1.52
		2.22			
dimethylether		4.99	5.92	4.55	4.52
		5.24	6.55	4.58	4.57
		5.24	6.38	4.94	4.39
benzene		9.68	11.42	11.42	6.20
		9.71	11.70	11.70	5.72
		9.01 (10.44)	11.03	11.03	4.97

[a] Experimental data are taken from Tables 5 and 6 of Applequist et al.[146] Where available, more recent experimental values for α_{avg} from Bosque and Sales[147] are reported in parentheses.

Table 3. Molecular Polarizability ($\mathring{A}^3$) of Aromatic Systems[a]

		α_x	α_y	α_z
Benzene	Expt	11.70	11.70	5.72
	Expt	12.26	12.26	6.66
	AMOEBA	12.30	12.30	6.64
Naphthalene	Expt	20.20	18.80	10.70
	Expt	22.20	18.20	7.30
	AMOEBA	21.78	18.51	9.77
Anthracene	Expt	35.20	25.60	15.20
	Expt	44.70	25.80	9.80
	DFT(B3LYP/6-31G*)	38.65	21.65	6.51
	AMOEBA	32.85	24.67	12.63
Nanotube Armchair (3,3)	DFT(B3LYP/6-31G*)	59.65	39.06	39.06
	DFT(B3LYP/6-31+G*)	64.45	47.33	47.33
	AMOEBA	61.68	41.20	41.20

[a] Experimental data are taken from Table 8 of Applequist.[148]

simple averaging would lead to a large error in the molecular dipole moment. The potential-optimized PAM, where methyl hydrogens are constrained to adopt equivalent values, will reproduce almost exactly both the *ab initio* potential and the molecular multipole moments. Our standard procedure is to use a molecular potential grid consisting of a 2 Å shell beginning 1 Å out from the vdW surface. The DMA monopole values are generally fixed during the potential fitting procedure.

This electrostatic parametrization protocol is particularly important for larger molecules and for molecules with high symmetry. It is known that the original DMA approach tends to give "unphysical" multipole values for large molecules when diffuse functions are included in the basis set even though the resulting electrostatic potential is correct. A recent modification of DMA[80] has been put forward to address this issue. However, in our hands, the multipoles from the modified scheme seem less transferable between conformations. The above protocol allows derivation of PAM corresponding to larger basis sets than would be practical with the original DMA method. Note that this procedure is different from restrained potential fits commonly used to fit fixed atomic charge models, as the starting DMA values are already quite reasonable and the fitting can be considered as a small perturbation biased toward the larger basis set potential. The overall procedure has been extensively tested in a small molecule hydration study[81] and will be used in future AMOEBA parameterization efforts.

Empirical vdW parameters were determined by fitting to both gas and liquid phase properties. The gas phase properties include homodimer binding energy (BSSE corrected) and structure from *ab initio* calculations at the MP2/aug-cc-pVTZ level or above. Liquid properties include experimental density and the heat of vaporization of neat liquids. The vdW parameters were first estimated by comparing the structure and energy of the AMOEBA-optimized dimer with *ab initio* results and then fine-tuned to

parameters by fitting to the electrostatic potential on a grid of points outside the vdW envelope of the molecule. The reference potential for the fitting step is typically derived from a single point calculation at the MP2/aug-cc-pVTZ level. Only a partial optimization to the potential grid is used to keep the atomic moments close to their DMA-derived values while still providing an improved molecular potential. The fitting approach is also useful for molecules containing symmetry-averaged atoms of the same atomic multipole type. In this case, simple arithmetic averaging would degrade the quality of the PAM. For example, in dimethyl- or trimethylamine, all of the methyl hydrogen atoms are indistinguishable and adopt the same atom type. The DMA multipole values for these atoms are somewhat different due their nonequivalence in any single conformation, and PAM derived by

reproduce the experimental liquid density and heat of vaporization via molecular dynamics simulation. Additional homodimers at alternative configurations, heterodimers with water, and liquid properties were computed *post facto* for the purpose of validation. A more generic force field atom classification for vdW parameters was enforced to ensure transferability. Table 1 lists the common vdW atom classes used by AMOEBA, together with the corresponding vdW parameters and polarizabilities. The vdW atom classes are also used to define parameters for all of the valence potential energy terms. The parameters for bonded terms, initially transferred from MM3, are optimized to reproduce *ab initio* geometries and vibration frequencies. In the final parametrization step, after all other parameters are fixed, torsional parameters are obtained by fitting to *ab initio* conformational energy profiles at the MP2/6-311++(2d,2p) level of theory.

■ COMPUTATION DETAILS

All force field calculations were carried out using the TINKER molecular modeling package.[82] The *ab initio* molecular orbital calculations were performed using Gaussian 03.[78] AMOEBA energy minimization of gas phase dimers was performed to achieve a RMS gradient of 0.01 kcal/mol/Å per atom. For bulk phase simulations, Particle Mesh Ewald was applied to treat the long-range electrostatic interactions, with a 9 Å real-space cutoff. A 12 Å switched cutoff is used for vdW interactions. Crystal minimizations were terminated when the gradient fell below 0.1 kcal/mol/Å per degree of freedom (atomic coordinates and cell parameters). The periodic box for neat liquid simulation systems was a cube approximately 20 Å on a side. A 2 ns NVT simulation was performed for each neat liquid, with an integration time step of 1 fs and the density set to the experimental value. A Berendsen thermostat was used to control the temperature.[83] The average pressure, heat of vaporization, and diffusion constant were calculated from the trajectories using the same formula reported for water.[18] Starting from the final configuration of each NVT run, NPT simulations of 2 ns were performed and used to compute average densities. NPT simulations of up to 6 ns were used to estimate the dielectric constants for selected compounds. NPT simulations used the Berendsen barostat with a relaxation time τ of 5 ps to control pressure.[83] Solvation free energies were computed using the same free energy perturbation procedure reported previously.[84] The hydration free energy of each small molecule was calculated by summing up free energies for three thermodynamic cycle steps: solute discharging in a vacuum, solute vdW coupling with a solvent (water), and solute recharging in water. Charging steps were performed over seven windows, and vdW coupling steps were performed in 16 windows. A softcore modifcation of the buffered-14−7 function was used in the vdW coupling.[84] Samples of solutes in a vacuum were collected every 0.5 ps from 10 ns stochastic dynamics simulations with an integration time step of 0.1 fs. Condensed phase simulations were run for 1 ns under NVT in 850 water molecules as the solvent, with the system density fixed at 1.000 g cm^{-3}. Snapshots were saved every 0.5 ps for free energy evaluation. Induced dipoles are converged to a RMS change of 0.00001 D per step for simulations in a vacuum and 0.01 D in bulk simulations. Free energy was calculated by re-evaluating the energy from the saved MD snapshots with the induce dipoles converged to 0.00001 D RMS. The Bennett Acceptance Ratio (BAR) method[85] was used to estimate the free energy change between neighboring steps.
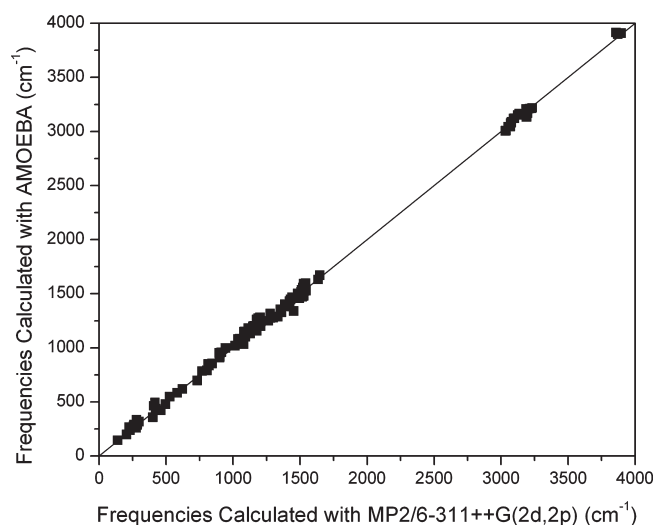

Figure 2. The comparison of gas-phase vibration frequencies calculated via MP2/6-311++G(2d,2p) and the AMOEBA force field. The signed average error is −9.59 cm^{-1}; the unsigned average error is 27.17 cm^{-1}. The RMSE is 33.77 cm^{-1}. Molecules included are methanol, ethanol, propanol, dimethylether, and phenol.

The TINKER VIBRATE program was used for normal mode calculations, implemented via diagonalization of the mass-weighted Hessian matrix of Cartesian second derivatives.

■ RESULTS AND DISCUSSION

Gas Phase Calculations. In conventional fixed charge force fields, atomic partial charges are often "pre-polarized" to match those in the liquid state by empirical scaling of *ab initio* charges from the gas phase. In contrast, the electrostatics parameters in AMOEBA are derived from high-level *ab initio* calculations in the gas phase, and electronic polarization by the environment is accounted for explicitly. Homodimer association energies and equilibrium geometries in the gas phase were used in conjunction with the condensed-phase properties to obtain vdW parameters for alkanes, aromatics, amines, alcohols, amides, and sulfides. Additional local minima corresponding to heterodimers with a water molecule, as hydrogen bond donor and acceptor, were utilized to further validate the parameters.

Intramolecular Interactions. The intramolecular valence parameters for AMOEBA were initially transferred from MM3.[1] These values were already known to perform satisfactorily in terms of producing reasonable equilibrium molecular geometries. We have further optimized the bond, angle, and other valence parameters against the *ab initio* equilibrium structures and vibration frequencies. As an illustration, the vibration frequencies of methanol, ethanol, propanol, dimethylether, and phenol (total of 117 data points) calculated using the final AMOEBA parameters are compared with MP2/6-311++G(2d,2p) data in Figure 2, where excellent agreement between AMOEBA and QM is seen to result from the parameter optimization. The average absolute vibration frequency error is 27.17 cm^{-1}, with a root-mean-square error (RMSE) of 33.77 cm^{-1}. The inclusion of a stretch−bend cross term is critical to the quality of the vibration frequencies. Examples of conformational energy are given in Figure 3 for butane, methanol, phenol, ethylamine, and ethylsulfide. The AMOEBA intramolecular nonbonded interactions, plus a standard torsional energy contribution computed via a
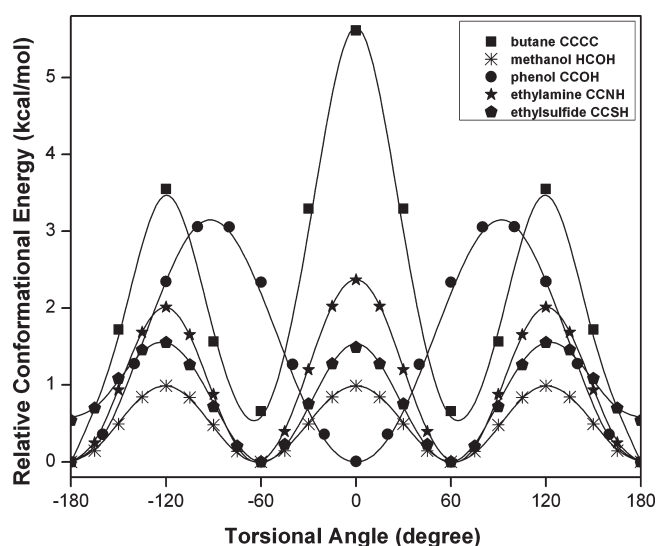
Figure 3. Relative conformational energies with respect to specific torsions. Solid line, MP2/6-311++G(2d,2p); symbols, AMOEBA. RMSEs between AMOEBA and QM results: butane CCCC, 0.14 kcal/mol; methanol HCOH, 0.016 kcal/mol; phenol CCOH, 0.069 kcal/mol; ethylamine CCNH, 0.014 kcal/mol; ethylsulfide CCSH, 0.031 kcal/mol.

Table 4. Relative Conformational Energies (kcal/mol) of *n*-Butane

	AMOEBA	*ab initio*[a]	experimental[b]
anti	0.00	0.00	0.00
syn	5.61	5.50	3.95
gauche	0.51	0.62	0.67
120°	3.55	3.31	3.62

[a] Ref 149. [b] Ref 150.

traditional three-term Fourier expansion, reproduce the *ab initio* conformational energy very well for these and other organic compounds. Table 4 gives the relative conformational energies (kcal/mol) of *n*-butane.

Intermolecular Interactions. The equilibrium geometry and association energy of 32 homo- and heterodimers of alkanes, amides, alcohols, amines, sulfides, and aromatics have been computed. Results are reported in Table 5. For homodimers, the global minima have been utilized in the parametrization process, with vdW parameters adjusted further on the basis of liquid properties (see below). The results reported here are computed using the final AMOEBA parameter values. In addition, heterodimers with a water molecule serve as a test of the transferability of the model. The parameters of the water model were as reported previously.[18] In Table 5, the dimer equilibrium geometries and association energies from AMOEBA and MP2/aug-cc-pVTZ, with and without basis set superimposition error (BSSE) correction, are compared. The excellent correlation ($R^2 = 0.99$) of the AMOEBA energy with BSSE corrected *ab initio* QM association energy is shown in Figure 4. Across the 32 dimer pairs, the average unsigned error in AMOEBA association energy is 0.31 kcal/mol with a RMSE of 0.38 kcal/mol. The AMOEBA association energy is calculated from the AMOBEA-optimized dimer structures, which are also in good agreement with *ab initio* structures, as indicated by the hydrogen bond distance and angle

comparison in Table 5. In comparison to the aug-cc-pVTZ dimer geometry, AMOEBA gives a 0.056 Å RMSE in hydrogen bond lengths and 9.53° in angle values. Recently, Faver et al.[86] compared AMOEBA results for a series of dimer interaction energies. They found AMOEBA energies to be in better agreement with high-level QM results than the GAFF and MMFF force fields as well as many lower-level QM protocols. The CHARMM fluctuating charge force field, another systematically derived polarizable force field, reported a 0.19 Å RMSE in hydrogen bond distance and 0.98 kcal/mol in dimerization energy for a series of solute–water complexes when compared to DFT results.[29] Harder et al. reported good agreement on NMA–water dimer energies with MP2/6-311+G(3df,2p) using a Drude oscillator base polarizable force field.[34]

In addition to the global minimum structures discussed above, several additional local minima for the dimers of formamide, DMF, NMA–water, ammonia, and benzene have been investigated using AMOEBA. These calculations provide an additional check of the potential energy surface beyond the global energy basin.

Hydrogen Bond Directionality. One of the most important interactions in organic molecules, and one that molecular mechanics methods should model accurately, is the hydrogen bond. Classical molecular orbital arguments describe the directional dependence of hydrogen bonding as a balance between electrostatics and charge transfer.[87] Since the mid-1980s,[88] most potentials for biological simulation have used simple Coulombic interactions to describe hydrogen bonds, while some organic force fields have incorporated special directionally dependent terms.[89] Recently, a new energy decomposition analysis of the water dimer based on absolutely localized molecular orbitals (ALMOs) by the Head-Gordon group suggested that the water–water interaction is largely due to electrostatics and polarization, with only minimal formal charge transfer.[90] In Figure 5, we compare AMOEBA results for the directionality of the formaldehyde–water hydrogen bond with those from MP2/aug-cc-pVTZ calculations and the fixed partial atomic charge-based OPLS-AA force field. For the planar hydrogen bonded structures, both AMOEBA and *ab initio* calculations yield minima at an acceptor angle near 120° with the linear structure lying about 1.5 kcal/mol higher in energy. This is in rough agreement with statistical distributions compiled from small molecule X-ray structures.[91] For OPLS-AA and other fixed partial charge models such as Amber and CHARMM (data not shown), there is very little angular dependence of the energy for a broad range of values centered at 180°. In AMOEBA, the quadrupole value on the carbonyl oxygen plays a major role in favoring the nonlinear configuration. A recently reported NEMO potential for formaldehyde provides independent evidence for the importance of local quadrupole moments.[92] Models based entirely on atomic charges (or atomic dipoles) are unable to break the symmetry along the C=O axis in order to provide a more favorable electrostatic potential at the "lone pair" angles. While the directional dependence of atomic charge models can be improved by including additional charges at lone pair[93] or π-cloud sites,[94] this is not a general solution and may not adequately address nonstandard hydrogen bonds.[95,96]

Amides. Ab initio studies of a series of formamide and dimethylformamide dimers have been reported previously[97,98] and are compared with our own quantum calculations and AMOEBA results in Table 6. In earlier work, we reported the inter- and intramolecular electrostatic models for NMA and alanine dipeptide,[15]

3149

dx.doi.org/10.1021/ct200304d |*J. Chem. Theory Comput.* 2011, 7, 3143–3161

Table 5. Gas Phase Dimer Equilibrium Structure and Binding Energy from QM and AMOEBA

	bond dist (Å)[a]/angle (degree)[b]		binding energy (kcal/mol)		
dimer	MP2/aug-cc-pVDZ	AMOEBA	MP2/aug-cc-pVTZ[c]	BSSE corrected	AMOEBA
methane—water	3.49/86.90	3.48/84.90	−1.18	−0.92	−1.21
methane—methane	4.01/179.96	3.93/165.00	−0.52	−0.36	−0.53
methanol—WD[d]	2.84/165.73	2.83/174.14	−6.10	−5.51	−5.85
methanol—WA[e]	2.90/177.05	2.99/178.30	−5.30	−4.78	−4.77
methanol—methanol	2.85/167.78,179.80[f]	2.88/179.42	−6.33	−5.26	−5.66
Ethanol-WD	2.84/161.12	2.87/175.48	−6.32	−5.70	−5.65
ethanol—WA	2.91/177.12	2.93/179.81	−5.27	−4.71	−4.67
ethanol—ethanol	2.86/166.46	2.89/167.63	−6.43	−5.62	−5.80
isopropanol—WD	2.84/161.68	2.89/168.37	−6.85	−6.11	−5.87
isopropanol—WA	2.92/177.72	2.89/176.15	−5.47	−4.85	−5.28
dimethylether—WD	2.81/159.96	2.84/174.51	−6.68	−5.93	−6.28
phenol—WD	3.01/163.55	2.94/167.73	−4.78	−3.98	−4.58
phenol—WA	2.84/176.64	2.88/176.95	−7.45	−6.71	−6.32
p-cresol—WD	3.00/162.90	2.92/167.62	−4.91	−4.14	−4.79
p-cresol—WA	2.85/176.76	2.88/176.85	−7.28	−6.55	−6.46
H₂S—WD	3.49/164.81	3.36/170.65	−3.35	−2.89	−3.48
H₂S—-WA	3.54/176.38	3.60/173.92	−3.04	−2.70	−2.78
H₂S—H₂S	4.09/172.62	4.04/167.90	−2.18	−1.81	−2.09
methylsulfide—WD	3.33/150.58	3.28/164.20	−4.94	−4.34	−4.84
methylsulfide—WA	3.57/170.47	3.62/176.74	−2.73	−2.38	−2.52
dimethylsulfide—WD	3.25/150.17	3.23/166.87	−6.07	−5.36	−5.19
methylamine—WD	2.86/162.25	2.86/175.87	−8.09	−7.45	−8.46
methylamine—methylamine	3.16/153.85	3.20/157.57	−4.74	−4.14	−4.09
ethylamine—WD	2.87/162.09	2.91/173.27	−8.17	−7.49	−7.57
imidazole—WA	2.87/160.21	2.93/177.67	−8.34	−7.05	−7.68
indole—WA	2.96/179.99	3.04/171.55	−6.52	−5.78	−5.58
ethylsulfide—WD	3.32/151.03	3.23/165.79	−5.52	−4.84	−5.46
ethylsulfide—WA	3.57/168.35	3.66/172.76	−2.69	−2.33	−2.10
methylethylsulfide—WD	3.24/151.66	3.20/168.66	−6.46	−5.68	−5.78
formamide—WD	1.91/99.52	1.86/115.00	−7.32	−6.75	−6.94
formamide—formamide	1.84/174.28	1,87/176.92	−16.86	−15.62	−16.00
NMA—WD	1.85/165.32	1.82/172.52	−8.71	−7.98	−8.38

[a] Heavy atom distance in the hydrogen bond, link O—O or O—N. [b] Hydrogen bond angle N(O)—H···N(O) except for methane. [c] Single point with MP2/aug-cc-pVTZ after structural optimization with MP2/aug-cc-pVDZ. [d] WD denotes water as the hydrogen bond donor in dimer structure. [e] WA denotes water as the hydrogen bond acceptor in the dimer structure. [f] MP2/6-31+G* optimization result.

as well as free energy calculations of ion solvation in liquid formamide.[99]

The six formamide dimer configurations investigated here are similar to those found by Vargas et al.[97] The global minimum is a cyclic configuration where two O···H hydrogen bonds constrain the dimer to form an 8-membered ring. The hydrogen bonding structure and distance in this dimer were used to adjust vdW parameters for carbonyl oxygen and amide hydrogen, while liquid simulations of a series of amides were utilized to fine-tune vdW parameters for the amide C, O, N, and H atoms simultaneously. As shown in Table 6, the six formamide dimer configurations optimized using MP2/aug-cc-pVDZ and AMOEBA are in excellent agreement, with an average RMSE of ∼0.1 Å in atomic coordinates.

For comparison, association energies were computed using AMOEBA and at the MP2/aug-cc-pVQZ level (for MP2/aug-cc-pVDZ optimized geometries) for each configuration. At this level of ab initio theory, the BSSE correction is about 0.6 kcal/mol for

dimer A. Sponer and Hobza[98] have shown that the association energy is well converged between the MP2/aug-cc-pVQZ and MP2/cc-pV5Z levels. At the MP2/aug-cc-pVTZ level, the association energy of dimer A is about 0.5 kcal/mol less negative than that obtained with the aug-cc-pVQZ basis set. Agreement between the ab initio and AMOEBA association energies for all six dimer configurations is 0.32 ± 0.22 kcal/mol.

These same dimers, especially the cyclic structure A, have been widely investigated using various ab initio methods, including CCSD(T)/CBS.[98,100,101] Both association and binding energies have been reported. The association energy is defined as the energy to separate the dimer without relaxing the monomer geometry, while the binding energy refers to the energy of the dimer relative to those of fully relaxed monomers. The difference between the two, i.e., the deformation energy upon binding, is reasonably small except for the cyclic dimer A. Our calculations suggest a deformation energy per molecule in dimer A of 0.8 kcal/mol at the MP2/aug-cc-pVQZ level. Moving from the

3150

dx.doi.org/10.1021/ct200304d |J. Chem. Theory Comput. 2011, 7, 3143–3161
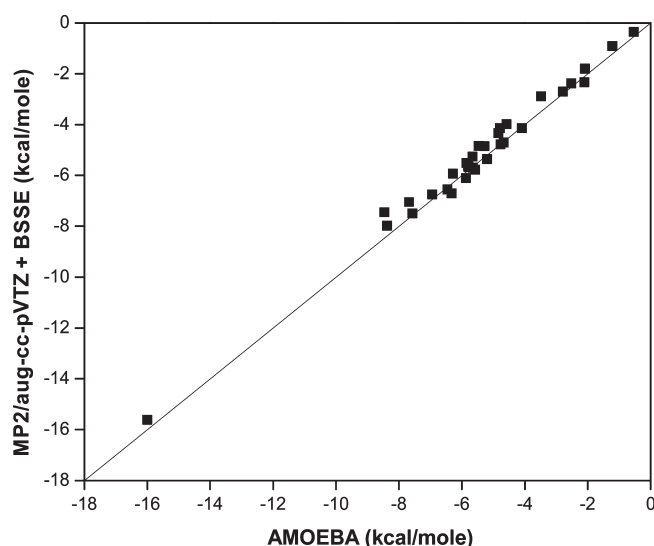
Figure 4. Comparison of dimer binding energies given by BSSE-corrected QM results and AMOEBA calculations. A total of 32 dimer structures are included, with a signed error of 0.22 kcal/mol, unsigned error of 0.31 kcal/mol, and RMSE of 0.38 kcal/mol.
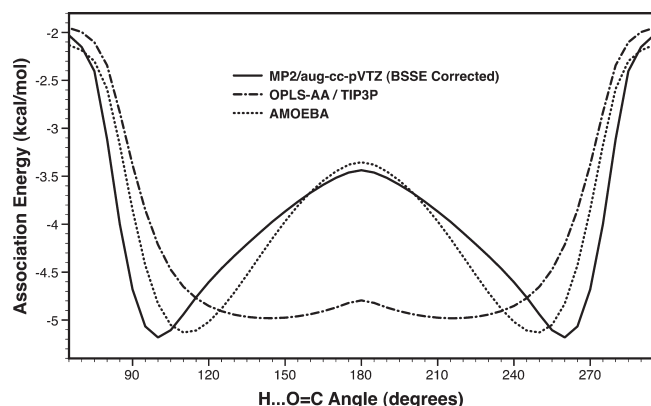
Figure 5. Association energy for the hydrogen bonded formaldehyde—water dimer as a function of the H···O=C angle. OPLS-AA/TIP3P is an OPLS-AA model for formaldehyde with a TIP3P water molecule. All energies are for structures fully optimized with a constrained H···O=C angle and with all atoms lying in a plane. The curves shown are interpolated from discrete calculations performed with each of the three methods at angle intervals of 5−15°.

equilibrium geometry of monomeric formamide to that in dimer A, the C=O bond elongates and the amide C−N bond becomes shorter, similar to the changes observed for amides upon moving from the gas to the liquid phase. The strong orbital delocalization favored by dimer A leads to changes in bond order, and the large deformation energy observed is not correctly described by the classical description of bond stretching and angle bending used by AMOEBA. In the force field calculations, the changes in the bond lengths are in the correct direction; however, the magnitudes are less than those observed in *ab initio* results. Thus, the AMOEBA deformation energy for dimer A is only 0.2 kcal/mol per molecule. Similar problems exist in the treatment of general conjugated, π-bonded systems. One possible solution, as implemented for the MM3 model,[102] involves using a simple VESCF

Table 6. Gas Phase Dimer Association Energy (kcal/mol) and Structure (Å) from *ab Initio* and AMOEBA Calculations of Multiple Configurations

	ab initio QM		AMOEBA	struct. RMSE[a]
formamide[b]				
A (cyc)	−16.1, −16.1,[c]	−15.96[d]	−16.0	0.03
B (s1)	−10.6		−10.3	0.05
C (np3)	−8.2		−8.9	0.07
D (np1)	−7.2		−7.5	0.23
E (s2)	−6.9		−7.3	0.04
F (HT)	−5.4		−5.5	0.08
DMF[e]	w/o BSSE	w/BSSE		
A	−6.95	−5.35	−5.01	0.08
B	−5.82	−4.14	−5.62	0.08
C	−11.41	−8.34	−7.39	0.26
D	−12.11	−8.90	−8.25	0.15
NMA−water[f]				
A		−8.07	−8.34	0.14
B		−8.01	−8.13	0.05
C		−5.18	−5.22	0.13
NH$_3$[b]				
linear		3.03	3.20	0.14
asymmetrical		3.07	3.21	0.17
benzene[g]				
T	−2.57[h]	−2.74[i]	−2.16	0.10
TT	−2.66	NA	−2.61	0.15
PD	−2.49	−2.78	−2.80	0.04
S	−1.51	−1.81	−2.05	0.13

[a] This study. RMS deviations between the AMOBEA and MP2/aug-cc-pVDZ optimized dimer structures; methyl hydrogen atoms excluded. [b] Data from this study. MP2/aug-cc-pVQZ with BSSE correction. [c] Ref 97. Geometry optimized at the MP2/aug-cc-pVTZ level. The binding energy was reported, which was adjusted to an association energy based on a deformation energy of 1.6 kcal/mol total. [d] Ref 101. CCSD(T)/CBS results for association energy. [e] Ref 97. MP2/aug-cc-pVTZ. [f] Data from this study. MP2/aug-cc-pVTZ with BSSE correction. [g] Benzene dimer structures: T-shaped (T), T-shaped tilted (TT), parallel displaced (PD), and parallel sandwich (S). [h] Ref 109. DFT-D structures and association energy from CCSD(T)$_2$|T 70% results. [i] Ref 108. Estimated CCSD(T)/CBS.

molecular orbital calculation to reassign bond orders on the fly. In the current work, we omit such MO-based corrections and attempt to compromise between agreement with the gas phase dissociation energies and reproduction of liquid phase thermodynamics, structure, and dynamics.

The dimethylformamide dimer provides an additional validation of the amide vdW parameters. *Ab initio* results, including binding energy and dimer structures at equilibrium, were reported by Vargas et al.[95] A comparison between AMOEBA and the *ab initio* results is made in Table 6. It should be noted that the BSSE corrections at the MP2/aug-cc-pVTZ level are almost 2 kcal/mol for dimers A and B, and more than 3 kcal/mol for dimers C and D. Overall, the AMOEBA results closely follow the BSSE corrected *ab initio* values.

Furthermore, three configurations of *N*-methylacetamide (NMA) complexed with a single water molecule have been identified as minima by MP2/6-31+G* energy minimization, similar to those

Table 7. Formic Acid Gas Phase Dimer Energy (kcal/mol) and Structure (Å) from *ab Initio* and AMOEBA Calculations

	ab initio QM[a]		AMOEBA	struct
	E_{assoc}	E_{bind}	E_{assoc}	RMSE
A	−18.6	−15.8	15.9	0.05
B	−10.3	−9.6	−9.7	0.04
C	−6.7	−6.1	−6.5	0.03
D	−3.4	−3.2	−3.5	0.06
E	−2.4	−2.2	−2.5	0.03
F	−4.4	−4.2	−4.4	0.05

[a] MP2/aug-cc-pVQZ energy at the MP2/aug-cc-pVDZ optimized geometry. Configurations A−F are analogous to those in Table 6 for the formamide dimer.

Figure 6. Comparison of heat of vaporizations (kcal/mol) from experimental results and from liquid simulations with AMOEBA.

reported previously.[29] The association energies were evaluated at the MP2/aug-cc-pVTZ level. In two of the minima, water is the hydrogen donor, while in the third the water oxygen atom is hydrogen bonded to the NMA amide hydrogen. In Table 6, the association energy and equilibrium structure given by AMOEBA are shown to be in excellent agreement with *ab initio* results for all three configurations.

Amines, Alcohols, and Acids. Among the stationary point structures of the ammonia dimer explored in the literature,[103,104] the so-called "linear" configuration is found to be a true minimum by both AMOEBA and *ab initio* optimization. A comparison of energies and geometries is given in Table 6. Analogous to AMOEBA water,[18] the quadrupole components of the ammonia N and H atoms are scaled downward by 40%. The scaling of quadrupole moments has no effect on the molecular dipole moment but is necessary for producing correct dimer equilibrium structures (i.e., "flap angles") in the case of both water and ammonia, possibly due to the limited basis set used in the PAM derivation or the lack of quadrupole polarization in the AMOEBA model. For ammonia, we have chosen a quadrupole scaling factor of 0.6, which yields a reduced molecular quadrupole moment (Qzz = 2.46 au) in close agreement with available experimental values of 2.42 ± 0.04 au[105] and 2.45 ± 0.30 au.[106,107] For consistency, we have similarly scaled the atomic quadrupole moments of the −NH and −OH groups in amines and alcohols so they are comparable to those of water and ammonia. AMOEBA-predicted methylamine and methanol dimer structures and energies are compared to corresponding *ab initio* results in Table 6. Results from liquid simulations are also satisfactory, as shown in the next section on condensed-phase simulations.

For both aldehydes and acids, the vdW parameters of the carbonyl group are transferred from amides, as given in Table 1. A number of dimer configurations for formic acid have been investigated using both AMOEBA and *ab initio* methods. In Table 7, we note that the deformation energy upon binding is 1.4 kcal/mol for formic acid configuration A (even higher than that of formamide), while in other configurations it is only a fraction of a kilocalorie per mole. This suggests a very strong electron delocalization in the cyclic configuration of dimer A. The BSSE from *ab initio* binding calculation using MP2/aug-cc-pVQZ at the MP2/aug-cc-pVTZ minimum geometry is 0.8 kcal/mol for dimer A. Deformation energies under the AMOEBA model are close to 0.1 kcal/mol, as expected.

Benzene and Other Aromatics. Benzene dimers have been widely studied in recent years using very high-level *ab initio* quantum mechanics.[108−111] A comparison between AMOEBA
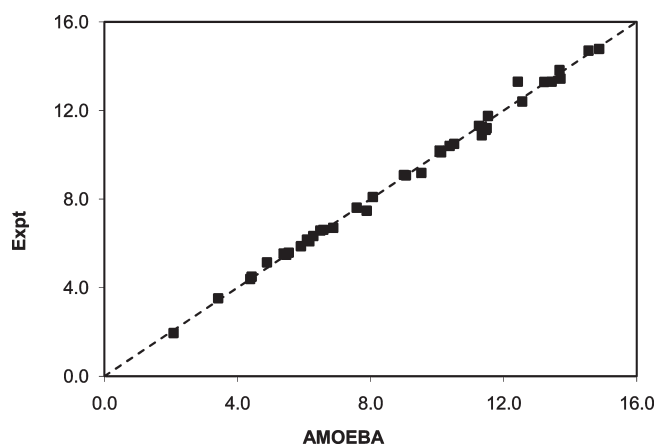
and QM results on four dominant stationary benzene dimer configurations is shown in Table 6. Overall agreement for the structure and energy is satisfactory, although AMOEBA selects the PD (parallel-displaced) structure as the global minimum while QM favors the TT (T-shaped tilted) configuration. As has been noted, the potential energy surface of the benzene dimer is extremely shallow.[109] According to AMOEBA, the two T-shaped dimers (T and TT) are transition state structures rather than true minima, as indicated by negative eigenvalues of the Hessian matrix. Many QM calculations impose symmetry and/or do not minimize completely, so it is not yet clear which structures are true minima on high-level *ab initio* surfaces.

The polarizability of various aromatics and a carbon nanotube section were calculated using AMOEBA with an atomic polarizability of 1.75 Å^3 for C and 0.696 Å^3 for H. The carbon nanotube studied is an "armchair" configuration made of three units of (3,3) structure. DFT calculations of molecular polarizability were performed using geometries optimized at the HF/6-31G* level. Many combinations of carbon and hydrogen atomic polarizabilities are able to reproduce the benzene molecular polarizability quite reasonably. However, upon comparing the molecular polarizabilities of a series of aromatic compounds, it is found that aromatic atomic polarizabilities for carbon and hydrogen must be increased from aliphatic values (see Table 3).

Condensed Phase Simulations. *Density and Heat of Vaporization.* The experimental density and heat of vaporization of a series of neat organic liquids were used to optimize vdW parameters. To enforce transferability, sets of compounds sharing the same atom types (Table 1) were parametrized together with compounds from different functional group families. This simultaneous parametrization helps to maintain the chemical consistency among different elements and functional groups. Liquid phase MD simulations were used to sample virial-based pressure values from NVT ensembles, and at the same time the heat of vaporization was computed from these trajectories. For liquids with low compressibility, such as water (5.1×10^{-5} per bar at 273 K), a pressure of 200 atm corresponds roughly to a 1% change in density. Thus, a reasonable target for molecules under AMOEBA was taken as an average pressure within the range 1 ± 200 atm, while keeping the heat of vaporization within ±0.5 kcal/mol of the experimental result. See Figure 6 for a comparison of heat of vaporizations (kcal/mol) from experimental results and from liquid simulations with AMOEBA.

Table 8. Heat of Vaporization (kcal/mol) and Pressure from NVT Simulations of Neat Liquids[a]

	E_{liq}	E_{gas}	ΔH_{sim}	ΔH_{exptl}	P_{sim}	T	ρ_{exptl}
water	−9.02	0.90	10.51	10.49[b]	−61	298.2	0.997[b]
MeOH	−3.57	4.90	9.06	8.95[c]	−40	298.2	0.786[c]
EtOH	−3.73	5.80	10.12	10.11[c]	−40	298.2	0.785[c]
n-PrOH	−1.41	9.26	11.26	11.31[c]	−42	298.2	0.800[c]
i-PrOH	−2.68	8.07	11.34	10.88[c]	−21	298.2	0.781[c]
NH$_3$	−2.89	2.17	5.54	5.58[e]	146	239.8	0.682[e]
MeNH$_2$	−2.89	2.67	6.09	6.17[f]	177	266.9	0.694[g]
EtNH$_2$	1.94	8.24	6.88	6.70[h]	−46	289.7	0.687[i]
PrNH$_2$	4.24	11.53	7.89	7.47[b]	−184	298.2	0.711[j]
di-Me amine	3.73	9.45	6.28	6.33[k]	−97	280.0	0.671[i]
tri-Me amine	15.88	20.95	5.62	5.48[l]	−231	276.0	0.653[i]
formic acid	−16.04	−5.17	11.46	11.13[m]	163	298.2	1.214[c]
acetic acid	−22.88	−10.91	12.56	12.49[d]	8	298.2	1.044[n]
formaldehyde	−3.47	1.41	5.38	5.54[b,o]	−65	254.0	0.812[b,o]
acetaldehyde	−6.73	−1.15	6.16	6.09[b,o]	184	293.2	0.783[b,o]
di-Me ether	3.00	7.39	4.89	5.14[b,o]	−62	248.3	0.736[b,o]
H$_2$S	−3.28	0.67	4.38	4.39[p]	−12	220.2	0.934[p]
MeSH	−2.95	2.40	5.91	5.87[q]	180	280.0	0.891[r]
EtSH	2.00	7.90	6.49	6.58[d]	109	298.1	0.833[s]
di-MeS	0.15	6.14	6.58	6.61[c]	28	298.2	0.842[t]
di-MeS$_2$	−7.39	1.54	9.53	9.18[d]	−15	298.2	1.057[s]
MeEtS	0.98	7.97	7.58	7.61[c]	20	298.2	0.837[s]
benzene	10.90	18.38	8.07	8.09[u,v]	96	298.0	0.874[u,v]
toluene	3.79	12.20	9.01	9.09[b]	142	298.2	0.865[b]
phenol	−5.87	7.22	13.68	13.82[c]	−52	298.2	1.058[c]
phenol	−4.44	8.14	13.22	13.36[b,o]	270	323.0	1.050[b,o]
ethylbenzene	10.68	20.17	10.08	10.10[b]	93	294.0	0.863[b]
cresol	−2.98	11.27	14.87	14.77[w]	173	313.2	1.019[b]
formamide	−17.96	−4.00	14.55	14.70[x]	−18	298.2	1.129[y]
N-MeForm	−12.57	0.54	13.71	13.43[o]	52	298.2	1.005[o]
di-MeForm	−5.86	5.04	11.49	11.21[b,o]	−129	298.2	0.944[b,o]
di-MeForm	−2.36	7.28	10.38	10.40[z]	107	373.2	0.873[aa,bb,cc]
acetamide	−18.88	−4.92	14.70	14.23[w]	6	373.0	0.984[w]
acetamide	−13.87	−2.42	12.43	13.30[w]	198	494.2	0.867[w]
N-MeAcet	−13.33	−0.62	13.44	13.30[dd]	30	373.2	0.894[dd]
di-MeAcet	−8.78	2.16	11.53	11.75[c]	349	298.2	0.936[c]
methane	−0.86	1.00	2.08	1.95[u,v]	293	111.0	0.424[u,v]
ethane	1.81	4.86	3.42	3.52[u,v]	98	184.0	0.546[u,v]
propane	5.24	9.21	4.43	4.49[u,v]	19	230.0	0.581[u,v]

[a] E is the potential energy in kcal/mol. ΔH is the heat of vaporization in kcal/mol, calculated as $\Delta H = E_{gas} - E_{liq} + RT$. P is pressure in atmospheres. T is temperature in Kelvin. ρ is density in g cm^{-3}. [b] Ref 151. [c] Ref 152. [d] Ref 153. [e] Ref 154. [f] Ref 155. [g] Ref 156. [h] Ref 157. [i] Ref 158. [j] Ref 159. [k] Ref 160. [l] Ref 161. [m] Ref 162. [n] Ref 163. [o] Ref 164. [p] Ref 165. [q] Ref 166. [r] Ref 167. [s] Ref 168. [t] Ref 169. [u] Ref 170. [v] Ref 171. [w] Ref 172. [x] Ref 173. [y] Ref 174. [z] Ref 175. [aa] Ref 176. [bb] Ref 177. [cc] Ref 178. [dd] Ref 179.

In Table 8, the results from liquid simulations are compared with experimental results. The overall RMSE in the heat of vaporization is 0.23 kcal/mol for the 37 compounds listed. The largest error of 0.9 kcal/mol was observed for acetamide at 494 K. The average pressure from the NVT simulations of all 37 compounds is 39 ± 124 atm, and the RMSE from the experimental

**Table 9. Comparison of Hydrogen Sulfide Liquid Properties from Experimental Results[165] and AMOEBA Simulation Results**

T (K)	E_{liq}	E_{gas}	ΔH_{expt}	ΔH_{sim}	P_{expt}	P_{sim}	ρ_{expt}
220.2	−3.28	0.67	4.39	4.38	1.441	−12	0.934
239.7	−3.00	0.73	4.19	4.20	3.326	54	0.902
252.4	−2.82	0.76	4.03	4.08	5.321	83	0.878
281.2	−2.39	0.85	3.63	3.80	12.969	163	0.816

value (1 atm) is 131 atm. This pressure deviation corresponds to a less than 1% change in density for most of these liquids. Selected NPT simulations were performed to confirm the density estimates. For liquid ammonia, we obtained an average pressure of 146 atm from the NVT simulation at the experimental density 0.682 g cm^{-3}, while a corresponding NPT trajectory at 1 atm gave an average density of 0.676 g cm^{-3} (relative error 0.8%). For formic acid, the liquid expanded slightly from 1.218 to 1.200 g cm^{-3} when the pressure changed from 163 to 1 atm (1.5%). For methanol, the density increased by 0.3% from the 0.786 g cm^{-3} experimental value in the NPT simulation, while the NVT simulation produced an average pressure of −40 atm. The largest error is for the NVT pressure of 349 atm for dimethylacetamide (DMA), resulting in a density decrease of 2.2% from 0.936 to 0.915 in the NPT simulation. The statistical error in the pressure, estimated using a block average approach, is on the order of 50 atm, while the statistical error in the heat of vaporization is negligible. For liquid hydrogen sulfide, the heat of vaporization values calculated at four different temperatures are in excellent agreement with experimental values, as shown in Table 9.

The overall performance on neat liquid properties seems slightly better than for fixed charge potentials such as OPLS-AA[4] and COMPASS.[112,113] Density and heat of vaporization are explicit targets in the AMOEBA force field optimization and development, as they are in nearly all force field models intended for bulk simulation. However, it should be kept in mind that gas-phase cluster properties computed using the same AMOEBA parameter set are also in good agreement with *ab initio* MP2 results. The error in heat of vaporization for 14 organic liquids given by the CHARMM fluctuating charge force field was about 1 kcal/mol, which is somewhat greater than the fixed charge CHARMM force field.[29] Other polarizable force fields, including Drude-oscillator[34,35] and PIPF models,[22] have reported an accuracy comparable to AMOEBA for selected molecules.

Dielectric and Diffusion Constants. Dielectric constants and diffusion constants were computed for selected compounds and are compared with available experimental measurements in Figures 7 and 8 (see also Table 10). For both static dielectric and self-diffusion constants, the overall agreement between AMOEBA and experiment values is satisfactory. Static dielectric constants were computed numerically from the cell dipole moment fluctuations and generally required multiple nanoseconds of simulation time to achieve convergence. Dielectric constants of individual liquids have been reported previously from fixed charge and polarizable force field simulations. Among the common fixed-charge water models, TIP5P[54] reproduces the experimental static dielectric constant accurately, while other TIPxP,[53,114] SPC,[115] and SPC/E[116] models give values that are either far too low or much too high.[54,117] There seems to be no obvious correlation between the molecular electric moments and the

3153

dx.doi.org/10.1021/ct200304d |*J. Chem. Theory Comput.* 2011, 7, 3143–3161
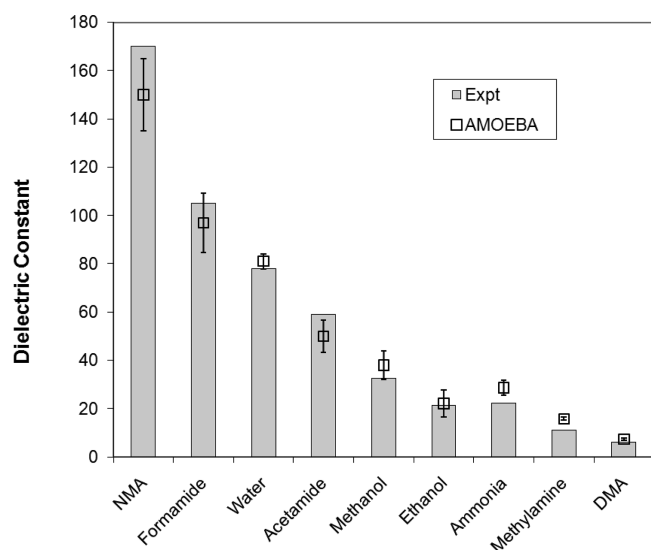
Figure 7. Dielectric constants from AMOEBA liquid MD simulations. The squares with vertical error bars are values computed from MD simulations. The filled bars are experimental values.
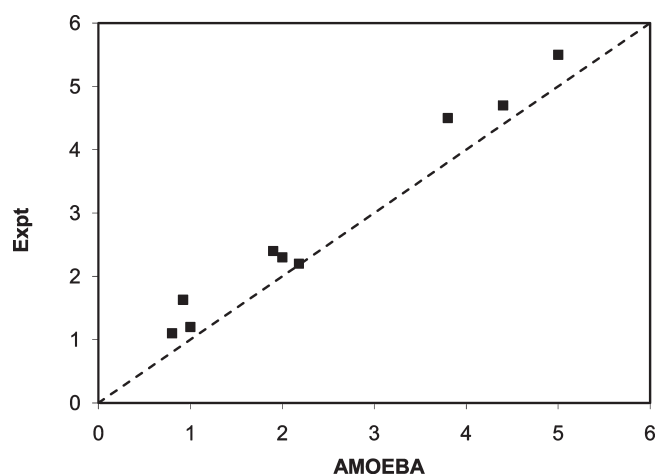
Figure 8. Comparison of diffusion constants from experimental measurement and MD simulations using AMOEBA. Diffusion constant data ($\times 10^{-9}$ m/s): dimethylformamide, $D_{sim} = 0.92$, $D_{exptl} = 1.63$, ref 186; ammonia, $D_{sim} = 5.0$, $D_{exptl} = 5.5$, ref 188; trimethylamine, $D_{sim} = 4.4$, $D_{exptl} = 4.7$ (273K), ref 189; methylamine, $D_{sim} = 3.8$, $D_{exptl} = 4.5$ interpolated from ref 189; methanol, $D_{sim} = 1.9$, $D_{exptl} = 2.4$, ref 190.

Table 10. Static Dielectric Constant and Self-Diffusion Coefficient $(\times 10^{-9}$ m^2 s$^{-1})^a$

| | T (K) | dielectric constant | | self-diffusion | |
		exptl.	AMOEBA	exptl.	AMOEBA
water	298.2	78.4[b]	81.0 (3.1)	2.3[h]	2.1
formamide	298.2	105.0[c] 84 (293 K)[d]	97.8 (12.3)		
DMF	298.2			1.6[i]	0.9
NMA	308.2	170.0[d]	153 (15.0)	1.2[j]	1.0
ammonia	240.0	22.0[d]	28.6 (3.0)	5.5[k]	5.0
methylamine	266.9	10.5[e] 16.7 (215 K)[b]	15.8 (0.6)	4.5[l]	3.8
dimethylamine	280.0	6.0[d]	7.3 (0.5)		
trimethyamine	276.0	2.4[b]	1.9 (0.4)	4.7[l]	4.4
methanol	298.2	33.0[b]	38.0 (5.8)	2.4[m]	1.9
ethanol	298.2	24.3[f,g]	22.1 (5.6)	1.1[h]	0.8
acetamide	373.2	59 (355 K)[d]	52.4 (6.6)		
benzene	298.2	2.3[b]	1.1 (0.5)	2.2[n]	2.2

^{a} The uncertainty of the calculated static dielectric constant is given in the parentheses. The uncertainty in self-diffusion constants is less than 0.1×10^{-9} m^2 s^{-1}. ^{b} Ref 151. ^{c} Ref 180. ^{d} Ref 181. ^{e} Ref 182. ^{f} Ref 183. ^{g} Ref 184. ^{h} Ref 185. ^{i} Ref 186. ^{j} Ref 187. ^{k} Refs 186 and 188. ^{l} Ref 189. ^{m} Ref 190. ^{n} Ref 191.
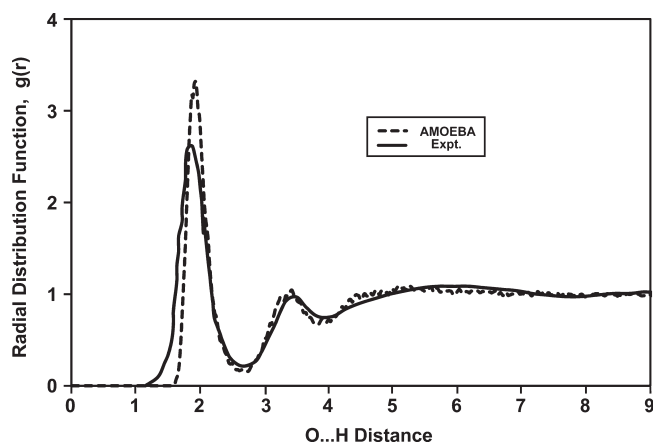
dielectric constant of NMA has a strong dependence on its average dipole moment, and a 0.2 D drop in the dipole moment lowered the dielectric constant by 30.[34] However, the induced dipole-based PIPF model overestimated the dielectric constant for NMA by 50% even though its liquid molecular dipole moment (5.0 D) is lower than that of a Drude model (5.7 D). The average NMA dipole according to AMOEBA is 5.5 D. Therefore, the dependence of the dielectric constant on the molecular dipole moment may only hold for a given, specific model. Dielectric constants of small alcohols are generally in the 20–30 range. The dielectric constant of methanol was reproduced accurately by a polarizable force field,[49] whereas a fixed charge potential underestimated the ethanol dielectric constant.[119] These results suggest that it is difficult for the classical models to capture the static dielectric constant without explicit incorporation of polarization effects.

For self-diffusion coefficients, there seems to be a systematic underestimation by the AMOEBA model, although the errors for water, ethanol, NMA, TMA, and benzene are insignificant. The polarizable Drude oscillator model also reported reasonable diffusion coefficients for benzene and toluene.[35] It was noticed by the early developers of polarizable force fields[43] that polarization slowed diffusion in neat liquids significantly compared to the fixed charge counterparts. For water, most of the fixed charge models overestimate the diffusion coefficient by as much as a factor of 2,[120] with TIP5P and SPC/E[116] being notable exceptions. On the hand other hand, the diffusion coefficients given by the CHARMM FQ model are very similar to fixed charge CHARMM22, with random errors in both directions.[29] Radial distribution functions for methanol and ammonia are compared to those derived from neutron scattering experiments in Figure 9–11. AMOEBA gives a more dominant first peak in the O–H RDF for liquid methanol, corresponding to the hydrogen bonding of the hydroxyl group, than that inferred from the experiment.[121] A previous CPMD study has also suggested a similar peak height at about 3.3.[122] For weakly hydrogen-bonded ammonia, the "experimental" radial distribution function[123] given

ability to reproduce the dielectric constant in the series of TIPxP models. Nonetheless, the adoption of five sites in TIP5P clearly has an effect on electrostatic interactions, as reflected in the water dimer energy surface and an increased tendency to form tetrahedral structure in the bulk.[54] Both the polarizable AMOEBA and the Drude oscillator model[33] can accurately reproduce the dielectric constant for water. Formamide has a high dielectric constant of 105. An early calculation using OPLS reported a value of 59 for formamide, while for DMF the computed value of 32 was in reasonable agreement with the experiment value of 37.[118] In the current study, the dielectric constant of formamide is slightly underestimated by AMOEBA at 98. The dielectric constant given by the recent CHARMM Drude oscillator model was somewhat too low as well. It was suggested that the static

Figure 9. Radial distribution, $g(r)$, for oxygen−oxygen atom pairs in liquid methanol at 298 K.

Figure 10. Radial distribution, $g(r)$, for oxygen−hydrogen atom pairs in liquid methanol at 298 K.
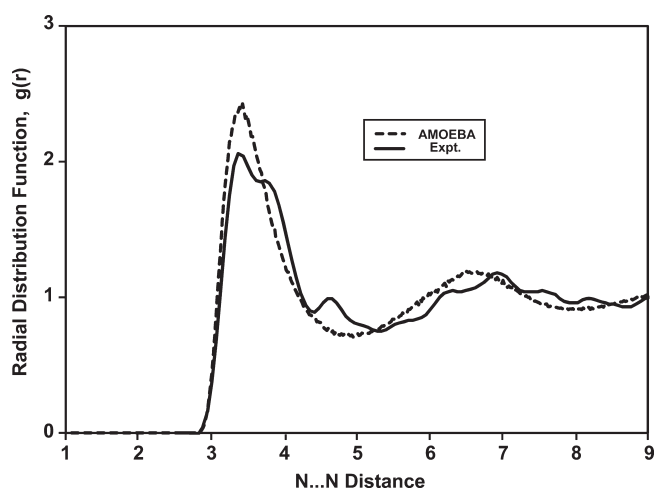

Figure 11. Radial distribution, $g(r)$, for nitrogen−nitrogen atom pairs in liquid ammonia at 277 K.

in Figure 11 was derived for molecular centers from X-ray scattering assuming spherical symmetry. A more recent neutron diffraction experiment[124] and first principle calculations[125] have argued that the shoulder at 3.7 A in the early X-ray results may be artifactual, in agreement with our simulation.

Crystal Structures. The crystal structures of formamide, acetamide, acetic acid, imidazole, and 1H-indole-3-carboxaldehyde have been examined using the AMOEBA potential. Crystal models were constructed from experimental fractional coordinates and subjected to full geometry optimization of the system energy by varying both atomic coordinates and cell parameters (i.e., all cell lengths and angles). Since the unit cells of these crystals are fairly small, replicated supercells were computed to allow use of particle mesh Ewald for long-range electrostatics. After full optimization, the atomic coordinates deviated from the experimental crystal by at most 0.3 Å in all cases (Table 11). In general, the overall cell volume shrank slightly, as expected for energy minimization. Molecular dynamics simulations of these and other crystals at experimental temperatures are underway and will be reported in due course.

Hydration Free Energy. Solvation plays a critical role in many chemical and biological processes. Accurate knowledge of solvation energetics is needed as part of the calculation of absolute association energies, for example, the binding of ligands to proteins. There is an extensive history of estimating the solvation free energy for small organics, protein side chain analogs, etc. using various force fields and water models. Vialla and Mark calculated the hydration free energies of 18 small molecules using a GROMOS96 force field[126] and SPC water model.[127] The average error was 2.8 kcal/mol using the original GROMOS partial charges, and it was suggested that the error might be reduced to 1 kcal/mol if the charge values were increased by 10%. In similar work by Maccallum and Tieleman using the OPLS-AA force field for solutes, an average unsigned error of 1.1 kcal/mol was reported for OPLS-AA in TIP4P water, 1.2 for OPLS-AA in SPC water, and 2.1 kcal/mol for GROMOS96 in SPC water.[128] Later, the GROMOS force field was optimized to reproduce the solvation free energies in water and cyclohexane, resulting in a much smaller error (0.2 kcal/mol).[129] However, this last study required the solutes to adopt different atomic charges in water and cyclohexane. Recently, effort has been devoted to increasing precision in hydration free energy calculations and optimizing the force field charges to capture solvation free energy more accurately. Shirts and co-workers[130] showed that it is possible to reduce the statistical uncertainty in the calculated hydration free energy to below 0.05 kcal/mol, and exhaustive sampling of various parameters was achieved using the folding@home computing resource.[131] Subsequently, the hydration free energy of 15 amino acid side chain analogs was determined using OPLS-AA and the above-mentioned water models plus SPC/E, TIP4P-EW,[132] and TIP3P-MOD.[133,134] The TIP3P-MOD, a modified TIP3P model to improve the solvation free energy of methane, gave the most accurate hydration free energy values with a RMSE of 0.51 kcal/mol. It is interesting to note that the TIP4P-EW model that yields the best overall pure water properties led to the worst hydration energies among all water models tested. Further modification of TIP3P vdW parameters in the spirit of TIP3P-MOD has been able to optimize the solvation energy of all 15 compounds to a RMSE of 0.39 kcal/mol. However, it should be cautioned that changes in the vdW parameters have profound effects on bulk water properties. While the heat of vaporization and density may remain reasonable, the structure of water (e.g., radial distribution function for O···O and O···H pair distances) is very sensitive to the vdW parameters.

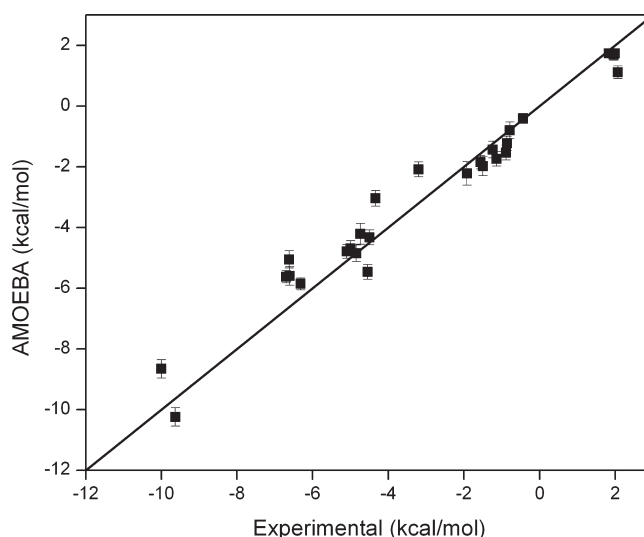**Table 11. Comparison of Experimental and AMOEBA-Optimized Crystal Structures and Cell Parameters of Organic Molecules[a]**

		struct RMSE	cell	a	b	c	α	β	γ	ref
formamide	exptl (90K)		$3 \times 1 \times 2$	10.812	9.041	13.988	90	100.5	90	192
	calcd	0.3		10.643	9.340	13.497	90	104.3	90	
acetamide	exptl (23K)		$1 \times 1 \times 1$	11.513	11.513	12.883	90	90	120	193
	calcd	0.1		11.564	11.564	12.289	90	90	120	
acetic acid	exptl (83K)		$1 \times 3 \times 2$	13.214	11.772	11.532	90	90	90	194and195
	calcd	0.1		13.424	11.573	11.352	90	90	90	
imidazole	exptl (293K)		$2 \times 3 \times 2$	15.464	16.374	19.558	90	117.3	90	196
	calcd	0.3		14.756	15.718	20.100	90	117.2	90	
1H-indole 3-carbox-aldyhyde	exptl (295K)		$1 \times 2 \times 2$	14.145	11.664	17.428	90	90	90	197
	calcd	0.3		14.458	11.964	16.683	90	90	90	

[a] The cell lengths are in Ångstroms, and angles are in degrees.

Table 12. Hydration Free Energies of Small Molecules (kcal/mol)[a]

molecule	AMOEBA	exptl
methane	1.73 (0.13)	1.99[b]
ethane	1.73 (0.15)	1.83[b]
propane	1.69 (0.17)	1.96[b]
n-butane	1.11 (0.21)	2.08[b]
methanol	−4.79 (0.23)	−5.11[b]
ethanol	−4.69 (0.25)	−5.00[b]
propanol	−4.85 (0.27)	−4.83[b]
isopropanol	−4.21 (0.34)	−4.76[b]
phenol	−5.05 (0.28)	−6.62[b]
p-cresol	−5.60 (0.31)	−6.14[b]
methylether	−2.22 (0.38)	−1.90[b]
benzene	−1.23 (0.23)	−0.87[b]
toluene	−1.53 (0.25)	−0.89[b]
ethylbenzene	−0.80 (0.28)	−0.80[b]
methylamine	−5.46 (0.25)	−4.56[b]
ethylamine	−4.33 (0.24)	−4.50[b]
dimethylamine	−3.04 (0.26)	−4.29[b]
trimethylamine	−2.09 (0.24)	−3.24[b]
imidazole	−10.25 (0.30)	−9.63[c]
N-methylacetamide	−8.66 (0.30)	−10.00[d]
acetic acid	−5.63 (0.20)	−6.70[b]
hydrogen sulfide	−0.41 (0.17)	−0.44[b]
methylsulfide	−1.43 (0.27)	−1.24[b]
ethylsulfide	−1.74 (0.24)	−1.30[b]
dimethylsulfide	−1.85 (0.22)	−1.54[b]
methylethylsulfide	−1.98 (0.32)	−1.50[e]
water	−5.86 (0.19)	−6.32[e]

[a] Statistical uncertainties of AMOEBA calculations are given in parentheses. [b] Ref 198. [c] Ref 199. [d] Ref 200. [e] Ref 201.

Figure 12. Comparison of solvation free energies of 27 small molecules calculated with the AMOEBA force field with the experimental values. Signed average error = −0.11 kcal/mol; unsigned average error = 0.56 kcal/mol; RMSE = 0.69 kcal/mol.

molecules reported a RMSE of 1.24 kcal/mol and a correlation of 0.89 between simulation and experimental values.[138]

In the current study, we have computed the hydration free energy of 27 compounds as a validation of the AMOEBA force field. None of the hydration free energy data was incorporated into the parametrization process. Results are listed in Table 12, and the correlation with experimental data is plotted in Figure 12. For this small set of compounds, the RMSE between AMOEBA and experimental results is 0.69 kcal/mol. The average signed error is 0.11, and the average unsigned error is 0.56 kcal/mol. The correlation (R^2) between the calculated and experimental value is 0.96 with a slope of 1.09. The largest error was observed for phenol, at 1.57 kcal/mol. Five out of 27 compounds had a deviation from experimental results greater than 1 kcal/mol. There is no obvious correlation between the error in hydration free energy and errors in gas-phase dimer energy or neat liquid heat of vaporization. The errors for methyl, dimethyl, and trimethyl amine are all about 1 kcal/mol. The densities of both dimethyl- and trimethylamine are higher than experimental results, while the density of methylamine is underestimated, as

More recently Mobley et al. took a different approach, investigating the effect of solute charges on hydration free energy.[135] Among the protocols they tested, RESP charges from HF/6-31G* performed the best, with a RMSE of 1.04 kcal/mol in hydration free energy of 44 compounds, closely followed (RMSE = 1.10 kcal/mol) by semiempirical charges from an AM1-BCC method tuned to reproduce HF/6-31G* charges.[136,137] Recent calculations using Amber GAFF parameters for 504 neutral

3156

dx.doi.org/10.1021/ct200304d |J. Chem. Theory Comput. 2011, 7, 3143–3161

indicated by the average pressure from NVT simulations reported in Table 8. In retrospect, it seems likely that the methyl vdW parameters, shared by all three amines, are not fully optimized for the condensed phase. The solvation energy RMSE for the other 24 compounds is 0.45 kcal/mol, indicating that there remains room for improvement in the amine solvation energies.

We note that AMOEBA gives a RMSE of 0.23 kcal/mol for liquid heat of vaporization and 0.38 kcal/mol for dimer energy in the gas phase, two properties that were actively utilized during parameter optimization. Thus, the ideal RMSE target of the calculated hydration free energy should probably lie below 0.5 kcal/mol. The error may come from various sources. The three main contributions to intermolecular interaction in the AMOEBA model arise from permanent electrostatics by atomic multipoles, polarization via atomic polarizability, and vdW interactions. It is possible that the level of *ab initio* theory and basis set used to derive the atomic multipoles is not sufficient. The buffered 14−7 vdW potential, and particularly the effect of a combining rule on heteroatomic vdW interactions in differing environments, is a likely impediment to improved accuracy.[139]

■ CONCLUSIONS

A polarizable point multipole potential has been developed for a range of common small organic molecules. Molecular electrostatics is represented by atomic multipole moments through the quadrupole, while polarization effects are treated via classical induced dipole interactions. The permanent atomic multipoles are derived from *ab initio* theoretical calculations at the MP2/6-311G(1d,1p) and MP2/aug-cc-pVTZ levels. Dipole polarization is modeled empirically with a damped, interactive atomic dipole scheme and using a small set of highly transferable atomic polarizabilities. The vdW parameters are derived via simultaneous fitting to gas-phase dimer calculations and liquid thermodynamic properties such as the density and heat of vaporization. Sharing atom types within and across families of organic molecules ensures transferability of parameters. A number of other gas phase and condensed phase properties were computed for use in validation, including stable homo- and heterodimer energies and structures, liquid diffusion and dielectric constants, radial distribution functions, molecular crystal structures, and solvation free energies in water. Overall, satisfying agreement between the polarizable potential, *ab initio*, and experimental results has been achieved in both the gas and condensed phases. The improvement in energy and density of homogeneous liquids is modest compared to a well-tuned fixed charge force field such as OPLS-AA, but the introduction of polarization and atomic multipoles significantly improves the ability of AMOEBA to describe details of molecular interactions across different environments. Current and prior results using this polarizable force field[34] suggest that the inclusion of induction effects is crucial for capturing diffusion and dielectric properties. The hydration free energy of 27 organic compounds computed using the current parameters verifies the general transferability of the model (RMSE = 0.69 kcal/mol), but further improvement is likely still possible.

A realistic physical model is critical for an accurate and transferable empirical force field. Obtaining consistent parameters for such a model presents an immense challenge. On the basis of the lessons learned in the current study and recent work by others on polarizable force fields, we expect that the overall performance of the AMOEBA model can be further refined.

Nowadays, accurate QM calculations can be performed routinely on small- to moderate-sized organics. The DMA procedure combined with potential fitting allows us to utilize high level *ab initio* calculations directly in AMOEBA parametrization. While our current feeling is that the AMOEBA permanent electrostatics are sufficient to construct a highly accurate force field, there are some indications that the polarization model can be improved in comparison to rigorous quantum results.[140] Given the increasing availability of computing resources, it should be possible to systematically optimize vdW parameters to reproduce neat liquid properties and transfer free energies simultaneously. The inclusion of Tang−Tonnies damping of dispersion interactions at short-range[141] is formally analogous to Thole damping of polarization effects and may be important for applications such as crystal structure prediction.[142] An important omission in most current force fields, including AMOEBA, is explicit coupling of electrostatics to the valence parameters. Simple schemes have been proposed to include this coupling,[143] and it is known to play a role in bond angle deformation in liquid water,[144] pyramidalization at amide nitrogen atoms,[145] and other important structural features. Future studies should also move beyond calculation of hydration free energy to include an examination of free energies of transfer, solvation structure around the solutes, and additional dynamic properties using AMOEBA and alternative polarizable force fields.

■ APPENDIX

Polarization Energy Gradient. A derivation of the gradient of the AMOEBA polarization energy is provided below. It is convenient to express the system energy via super-matrices:

$$U_{ele}^{perm} = \frac{1}{2} M^T T M$$

with $M^T = [M_1 ... M_i ... M_N]$

$$\text{and } T = \begin{bmatrix} 0 & T_{12} & \cdots & T_{1n} \\ T_{21} & 0 & \cdots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1} & T_{n2} & \cdots & 0 \end{bmatrix} \quad \text{(A1)}$$

where M_i is the transposed permanent multipole vector at site i, and T_{ij} is the interaction matrix between site i and j:

$$T_{ij} = \begin{bmatrix} 1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & \cdots \\ \frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i x_j} & \frac{\partial^2}{\partial x_i y_j} & \frac{\partial^2}{\partial x_i z_j} & \cdots \\ \frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \cdots \\ \frac{\partial}{\partial zi} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \left(\frac{1}{r_{ji}} \right) \quad \text{(A2)}$$

Rewriting eq 4 in terms of a super-matrix yields

$$(\alpha^{-1} - T^{11}) \mu^{ind} = T^1 M = E \quad \text{(A3)}$$

Here, μ^{ind} is a vector of length $3N$, where N is the number of polarizable sites, $\mu^{ind} = [\mu_{1x}, \mu_{1y}, \mu_{1z}, ..., \mu_{Nz}]^T$. α^{-1} is a $3N \times 3N$ matrix with α_{1x}^{-1}, α_{1y}^{-1}, etc. as diagonal components and all

off-diagonal components equal to zero. T^1 is a super matrix with elements corresponding to the field tensor T^1_{ij} in eq 4 (*i.e.*, $3N \times 13N$).

Now, we define $C = \alpha^{-1} - T^{11}$ and note that it is a symmetric matrix such that $C^T = C$. The induction energy is then defined by the product of the induced dipole with the permanent field

$$U^{ind}_{ele} = -\frac{1}{2}(\mu^{ind})^T E = -\frac{1}{2}E^T C^{-1} E \tag{A4}$$

Subsequently, the energy gradient on site k is given by

$$\frac{\partial U^{ind}_{ele}}{\partial x_k} = -\frac{1}{2}\left(\frac{\partial E^T}{\partial x_k}C^{-1}E + E^T\frac{\partial C^{-1}}{\partial x_k}E + E^T C^{-1}\frac{\partial E}{\partial x_k}\right),$$
$$k = 1, 2, 3, \ldots 3N \tag{A5}$$

The gradient on the left is a $3N$ vector in the above equation. Given $C^{-1}C = I$, i.e., $(\partial C^{-1}/\partial x_k)C + C^{-1}(\partial C/\partial x_k) = 0$, we can simplify to obtain

$$\frac{\partial U^{ind}_{ele}}{\partial x_k} = -\frac{1}{2}\frac{\partial E^T}{\partial x_k}\mu^{ind} + \frac{1}{2}E^T C^{-1}\frac{\partial C}{\partial x_k}C^{-1}E - \frac{1}{2}(\mu^{ind})^T\frac{\partial E}{\partial x_k}$$
$$= -(\mu^{ind})^T\frac{\partial T^1 M}{\partial x_k} - \frac{1}{2}(\mu^{ind})^T\frac{\partial T^{11}}{\partial x_k}\mu^{ind} \tag{A6}$$

Given the total multipoles at each site as $M^t = M + M^{ind}$, the net force at the site becomes

$$\frac{\partial U^t_{ele}}{\partial x_k} = -\frac{1}{2}(M^t)^T\frac{\partial T^1}{\partial x_k}M^t - (M^t)^T T^1\frac{\partial \mathbf{R}}{\partial x_k}M \tag{A7}$$

where the factor $1/2$ takes care of the redundancy due to the inclusion of both ij and ji in the above summation. While permanent multipoles in the local frame are invariant parameters, the permanent multipole moments, M, in the global frame are a function of the local frame rotation matrix, leading to additional chain rule terms related to the rotational force, *i.e.*, a torque. The exact formula for the force and torque components can be easily obtained by comparing the above to the permanent–permanent terms[18] and keeping in mind the additional factor of $1/2$. Note that the torque term does not include an induced–induced contribution since induced dipoles are always defined in the global frame.

When the field "E" in the induction energy is different from the "E" that produces the induced dipoles, the gradient formula requires further modification. The induction energy is then given by

$$U^{ind}_{ele} = -\frac{1}{2}(\mu^{ind}_d)^T E_P = -\frac{1}{2}E^t_d C^{-1}E_P \tag{A8}$$

where E_p is the field acutally used in the polarization energy calculation, and E_d is the "direct" field due to permanent multipoles responsible for polarization. The difference between the two subscripts, d and p, results from differing local interaction scaling. In traditional molecular mechanics, short range nonbonded interactions between bonded atoms are generally neglected. In the current model, the interaction energy between E_p and induced dipoles is ignored for 1–2, 1–3, etc. bonded pairs, as these effects are implicitly included in bond and angle terms, while we recall that intramolecular "direct" polarization occurs between polarization groups.

The gradient of the above energy becomes

$$\frac{\partial U^{ind}}{\partial x_k} = -\frac{1}{2}E^T_d C^{-1}E_p \tag{A9}$$

$$\frac{\partial U^{ind}}{\partial x_k} = -\frac{1}{2}\left(\frac{\partial E^T_d}{\partial x_k}C^{-1}E_p + E^T_d\frac{\partial C^{-1}}{\partial x_k}E_p + E^T_d C^{-1}\frac{\partial E_p}{\partial x_k}\right) \tag{A10}$$

Now define an intermediate quantity, $\mu^{ind}_p = C^{-1}E_p$. Recall that C is invariant with respect to local interaction scaling since "mutual" induction always occurs between every atom pair.

$$\frac{\partial U^{ind}}{\partial x_k} = -\frac{1}{2}\left(\frac{\partial E^T_d}{\partial x_k}\mu^{ind}_p - E^T_d C^{-1}\frac{\partial C}{\partial k}C^{-1}E_p + (\mu^{ind}_d)^T\frac{\partial E_p}{\partial xk}\right)$$
$$= -\frac{1}{2}\left[\frac{\partial E^T_d}{\partial x_k}\mu^{ind}_p + (\mu^{ind}_d)^T\frac{\partial E_p}{\partial x_k}\right] - \frac{1}{2}(\mu^{ind}_d)^T\frac{\partial T^{11}}{\partial x_k}\mu^{ind}_p$$
$$= -\frac{1}{2}\left[\frac{\partial T^1_d}{\partial x_k}\mu^{ind}_p + (\mu^{ind}_d)^T\frac{\partial T^1_P M}{\partial x_k}\right] - \frac{1}{2}(\mu^{ind}_d)^T\frac{\partial T^{11}}{\partial x_k}\mu^{ind}_p \tag{A11}$$

In cases where there is no intermolecular polarization (e.g., water molecule), μ^{ind}_p equals μ^{ind}_d, and the above equation reduces to eq A6. In practice, the two sets of μ are converged simultaneously, as the only difference between the two is the scaling of real-space local interactions.

Upon comparing eqs A6 and A11, it can be seen that within induced–permanent terms the induced dipole on site i (where force is computed) is replaced by $1/2(\mu^{ind}_d + \mu^{ind}_p)$ and the induced–induced term for a given pair interaction, $(\mu^{ind}_i)^T(\partial T^{11}/\partial x_k)\mu^{ind}_j$, is replaced by $1/2((\mu^{ind}_{id})^T(\partial T^{11}/\partial x_k)\mu^{ind}_{jp} + (\mu^{ind}_{jd})^T(\partial T^{11}/\partial x_k)\mu^{ind}_{ip})$. Throughout the energy, force, torque, and virial terms, similar substitutions can be made for induced-permanent terms and induced–induced terms in the Ewald formulation. The algorithms for computing the permanent and induced force, torque, and virial, both for pairwise non-periodic systems and when using PME, have been implemented and verified numerically in the TINKER and AMBER software packages.

■ ASSOCIATED CONTENT

ⓢ Supporting Information. Detailed protocol outlining a step-by-step parametrization procedure for determination of AMOEBA values for new organic molecules is provided. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: ponder@dasher.wustl.edu (J.W.P.); pren@mail.utexas.edu (P.R.).

■ ACKNOWLEDGMENT

TINKER modeling package, which can be obtained from http://dasher.wustl.edu/tinker/.

■ REFERENCES

(1) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

(2) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(3) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(4) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(5) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(6) Silberstein, L. *Philos. Mag. Ser. 6* **1917**, *33*, 92–128.

(7) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

(8) Vesely, F. J. *J. Comput. Phys.* **1977**, *24*, 361–371.

(9) Sprik, M. *J. Phys. Chem.* **1991**, *95*, 2283–2291.

(10) Dang, L. X.; Chang, T. M. *J. Chem. Phys.* **1997**, *106*, 8149–8159.

(11) Brdarski, S.; Astrand, P. O.; Karlstrom, G. *Theor. Chem. Acc.* **2000**, *105*, 7–14.

(12) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730–4737.

(13) Burnham, C. J.; Li, J. C.; Xantheas, S. S.; Leslie, M. *J. Chem. Phys.* **1999**, *110*, 4566–4581.

(14) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 2237–2251.

(15) Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497–1506.

(16) Burnham, C. J.; Xantheas, S. S. *J. Chem. Phys.* **2002**, *116*, 1479–1492.

(17) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515–1531.

(18) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(19) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621–627.

(20) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2004**, *108*, 13427–13437.

(21) Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P. Y. *J. Phys. Chem. B* **2006**, *110*, 18553–18559.

(22) Xie, W. S.; Pu, J. Z.; MacKerell, A. D.; Gao, J. L. *J. Chem. Theory Comput.* **2007**, *3*, 1878–1889.

(23) Rappé, A. K.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358–3363.

(24) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(25) Rick, S. W.; Stuart, S. J.; Bader, J. S.; Berne, B. J. *J. Mol. Liq.* **1995**, *65–6*, 31–40.

(26) Banks, J. L.; Kaminski, G. A.; Zhou, R. H.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741–754.

(27) Ando, K. *J. Chem. Phys.* **2001**, *115*, 5228–5237.

(28) Yoshii, N.; Miyauchi, R.; Miura, S.; Okazaki, S. *Chem. Phys. Lett.* **2000**, *317*, 414–420.

(29) Patel, S.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1–15.

(30) Patel, S.; MacKerell, A. D.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1504–1514.

(31) van Maaren, P. J.; van der Spoel, D. *J. Phys. Chem. B* **2001**, *105*, 2618–2626.

(32) Yu, H. B.; Hansson, T.; van Gunsteren, W. F. *J. Chem. Phys.* **2003**, *118*, 221–234.

(33) Lamoureux, G.; MacKerell, A. D.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185–5197.

(34) Harder, E.; Anisimov, V. M.; Whitfield, T. W.; MacKerell, A. D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3509–3521.

(35) Lopes, P. E. M.; Lamoureux, G.; Roux, B.; MacKerell, A. D. *J. Phys. Chem. B* **2007**, *111*, 2873–2885.

(36) Rick, S. W.; Stuart, S. J. *Rev. Comp. Ch.* **2002**, *18*, 89–146.

(37) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(38) Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J. M. *J. Phys., Condens. Mater.* **2009**, *21*, 333101.

(39) Illingworth, C. J.; Domene, C. *Proc. R. Soc. London, Ser. A* **2009**, *465*, 1701–1716.

(40) Lopes, P. E. M.; Lamoureux, G.; Mackerell, A. D. *J. Comput. Chem.* **2009**, *30*, 1821–1838.

(41) Masia, M.; Probst, M.; Rey, R. *J. Chem. Phys.* **2004**, *121*, 7362–7378.

(42) Masia, M.; Probst, M.; Rey, R. *J. Chem. Phys.* **2005**, *123*, 164505.

(43) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208–6219.

(44) Gao, J. L.; Pavelites, J. J.; Habibollazadeh, D. *J. Phys. Chem.* **1996**, *100*, 2689–2697.

(45) Cabaleiro-Lago, E. M.; Rios, M. A. *J. Chem. Phys.* **1998**, *108*, 3598–3607.

(46) Hermida-Ramon, J. M.; Rios, M. A. *J. Phys. Chem. A* **1998**, *102*, 10818–10827.

(47) Qian, W. L.; Krimm, S. *J. Phys. Chem. A* **2001**, *105*, 5046–5053.

(48) Mannfors, B.; Mirkin, N. G.; Palmo, K.; Krimm, S. *J. Comput. Chem.* **2001**, *22*, 1933–1943.

(49) Yu, H. B.; Geerke, D. P.; Liu, H. Y.; van Gunsteren, W. F. *J. Comput. Chem.* **2006**, *27*, 1494–1504.

(50) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1587–1597.

(51) Williams, D. E. *J. Comput. Chem.* **1988**, *9*, 745–763.

(52) Dykstra, C. E. *Chem. Rev.* **1993**, *93*, 2339–2353.

(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(54) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910–8922.

(55) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U. K., 1996.

(56) Poplelier, P. L. A.; Joubert, L.; Kosov, D. S. *J. Phys. Chem. A* **2001**, *105*, 8254–8261.

(57) Buckingham, A. D.; Fowler, P. W. *J. Chem. Phys.* **1983**, *79*, 6426–6428.

(58) Buckingham, A. D.; Fowler, P. W. *Can. J. Chem.* **1985**, *63*, 2018–2025.

(59) Golubkov, P. A.; Ren, P. *J. Chem. Phys.* **2006**, *125*, 064103.

(60) Golubkov, P. A.; Wu, J. C.; Ren, P. Y. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2050–2057.

(61) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(62) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341–350.

(63) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399–2407.

(64) Piquemal, J.-P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 054511.

(65) Price, S. L.; Faerman, C. H.; Murray, C. W. *J. Comput. Chem.* **1991**, *12*, 1187–1197.

(66) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233–239.

(67) Young, D. M. *Iterative Solution of Large Linear Systems*; Academic Press: New York, 1971.

(68) Smith, W. *CCP5 Newsletter* **1998**, *46*, 18–30.

(69) Sagui, C.; Darden, T.; Pedersen, L. G. *J. Chem. Phys.* **2004**, *120*, 73–87.

(70) Kong, Y. Ph.D. thesis, Molecular Biophysics, Washington University Medical School, St. Louis, Missouri, 1997.

(71) Kolafa, J. *J. Chem. Phys.* **2005**, *122*, 164105.

(72) Sala, J.; Guardia, E.; Masia, M. *J. Chem. Phys.* **2010**, *133*, 234101.

(73) van Belle, D.; Wodak, S. J. *Comput. Phys. Commun.* **1995**, *91*, 253–262.

(74) Harder, E.; Kim, B.; Friesner, R. A.; Berne, B. J. *J. Chem. Theory Comput.* **2005**, *1*, 169–180.

(75) Souaille, M.; Loirat, H.; Borgis, D.; Gaigeot, M. P. *Comput. Phys. Commun.* **2009**, *180*, 276–301.

(76) Darden, T. A.; Toukmaji, A.; Pedersen, L. G. *J. Chim. Phys. PCB* **1997**, *94*, 1346–1364.

(77) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol.* **1999**, *28*, 155–179.

(78) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian Inc.: Wallingford, CT, 2003.

(79) Stone, A. J. *GDMA*; Cambridge University Technical Services: Cambridge, England, 1998.

(80) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.

(81) Shi, Y.; Wu, C.; Ponder, J. W.; Ren, P. *J. Comput. Chem.* **2010**, *32*, 967–977.

(82) Ponder, J. W. *TINKER Molecular Modeling Package*, V5.1; Washington University Medical School: St. Louis, MO, 2010.

(83) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(84) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6290−6295.

(85) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(86) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M. *J. Chem. Theory Comput.* **2011**, *7*, 790–797.

(87) Kollman, P. A. *J. Am. Chem. Soc.* **1971**, *94*, 1837–1842.

(88) Reiher, W. E. Ph.D. Thesis, Dept. of Chemistry, Harvard University, Cambridge, MA, 1985.

(89) Lii, J.-H.; Allinger, N. L. *J. Comput. Chem.* **1998**, *19*, 1001–1016.

(90) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. *Chem.—Eur. J.* **2009**, *15*, 851–855.

(91) Steiner, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 48–76.

(92) Holt, A.; Bostrom, J.; Karlstrom, G.; Smith, R. *J. Comput. Chem.* **2010**, *31*, 1583–1591.

(93) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048–1057.

(94) Baker, C. M.; Grant, G. H. *J. Chem. Theory Comput.* **2006**, *2*, 947–955.

(95) Vargas, R.; Garza, J.; Dixon, D. A.; Hay, B. P. *J. Am. Chem. Soc.* **2000**, *122*, 4750–4755.

(96) Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. *Nat. Chem. Biol.* **2010**, *6*, 615–620.

(97) Vargas, R.; Garza, J.; Friesner, R. A.; Stern, H.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105*, 4963–4968.

(98) Sponer, J.; Hobza, P. *J. Phys. Chem. A* **2000**, *104*, 4592–4597.

(99) Grossfield, A.; Ren, P. Y.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.

(100) Jurecka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89–94.

(101) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(102) Allinger, N. L.; Li, F.; Yan, L.; Tai, J. C. *J. Comput. Chem.* **1990**, *11*, 868–895.

(103) Boese, A. D.; Chandra, A.; Martin, J. M. L.; Marx, D. *J. Chem. Phys.* **2003**, *119*, 5965–5980.

(104) Janeiro-Barral, P. E.; Mella, M. *J. Phys. Chem. A* **2006**, *110*, 11244–11251.

(105) van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *THEOCHEM* **1982**, *89*, 185–201.

(106) Kukolich, S. G. *Chem. Phys. Lett.* **1970**, *5*, 401–404.

(107) Kukolich, S. G.; Casleton, K. H. *Chem. Phys. Lett.* **1973**, *18*, 408–410.

(108) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656–10668.

(109) Pitonak, M.; Neogrady, P.; Rezac, J.; Jurecka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829–1834.

(110) Dinadayalane, T. C.; Leszczynski, J. *Struct. Chem.* **2009**, *20*, 11–20.

(111) Dinadayalane, T. C.; Leszczynski, J. *J. Chem. Phys.* **2009**, *130*, 081101.

(112) Sun, H. *J. Phys. Chem. B* **1998**, *102*, 7338–7364.

(113) Sun, H.; Ren, P.; Fried, J. R. *Comput. Theor. Polym. Sci.* **1998**, *8*, 229–246.

(114) Jorgensen, W. L.; Jenson, C. *J. Comput. Chem.* **1998**, *19*, 1179–1186.

(115) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullmann, B., Ed.; D. Reidel Pub. Co.: Dordrecht, The Netherlands, 1981; pp 331−342.

(116) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(117) Hohtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. *J. Chem. Phys.* **1998**, *109*, 4927–4937.

(118) Essex, J. W.; Jorgensen, W. L. *J. Phys. Chem.* **1995**, *99*, 17956–17962.

(119) Saiz, L.; Guardia, E.; Padro, J. A. *J. Chem. Phys.* **2000**, *113*, 2814–2822.

(120) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2001**, *114*, 363–366.

(121) Yamaguchi, T.; Hidaka, K.; Soper, A. K. *Mol. Phys.* **1999**, *97*, 603–605.

(122) Pagliai, M.; Cardini, G.; Righini, R.; Schettino, V. *J. Chem. Phys.* **2003**, *119*, 6655–6662.

(123) Narten, A. H. *J. Chem. Phys.* **1976**, *66*, 3117–3120.

(124) Ricci, M. A.; Nardone, M.; Ricci, F. P.; Andreani, C.; Soper, A. K. *J. Chem. Phys.* **1995**, *102*, 7650–7655.

(125) Hannongbua, S. *J. Chem. Phys.* **2000**, *113*, 4707–4712.

(126) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–1218.

(127) Villa, A.; Mark, A. E. *J. Comput. Chem.* **2002**, *23*, 548–553.

(128) Maccallum, J. L.; Tieleman, D. P. *J. Comput. Chem.* **2003**, *24*, 1930–1935.

(129) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(130) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740–5761.

(131) Shirts, M.; Pande, V. S. *Science* **2000**, *290*, 1903–1904.

(132) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(133) Sun, Y.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1164–1169.

(134) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(135) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242–2254.

(136) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.

(137) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(138) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.

(139) Al-Matar, A. K.; Rockstraw, D. A. *J. Comput. Chem.* **2004**, *25*, 660–668.

(140) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2004**, *120*, 9903–9906.

(141) Tang, K. T.; Tonnies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726–3741.

(142) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872–9882.

(143) Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S. *Chem. Phys. Lett.* **2006**, *429*, 628–632.

(144) Fanourgakis, G. S.; Xantheas, S. S. *J. Chem. Phys.* **2006**, *124*, 174504.

(145) Mannfors, B.; Mirkin, N. G.; Palmo, K.; Krimm, S. *J. Phys. Chem. A* **2003**, *107*, 1825–1832.

(146) Applequist, J.; Carl, J. R.; Fung, K.-K. *J. Am. Chem. Soc.* **1972**, *94*, 2952–2960.

(147) Bosque, R.; Sales, J. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1154–1163.

(148) Applequist, J. *J. Phys. Chem.* **1993**, *97*, 6016–6023.

(149) Allinger, N. L.; Fermann, J. T.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* **1997**, *106*, 5143–5150.

(150) Murphy, W. F.; Fernandezsanchez, J. M.; Raghavachari, K. *J. Phys. Chem.* **1991**, *95*, 1124–1139.

(151) *CRC Handbook of Chemistry and Physics*, 82nd ed.; Lide, D. R., Ed.; CRC Press LLC: Boca Raton, FL, 2001.

(152) Riddick, J. A.; Bunger, W. B.; Sakano, T.; Weissberger, A. *Organic Solvents: Physical Properties and Methods of Purification*, 4th ed.; Wiley: New York, 1986.

(153) Wagman, D. D.; Evans, W. H.; Parker, V. B.; Schumm, R. H.; Halow, I. *J. Phys. Chem. Ref. Data* **1982**, *11 (Suppl 2)*, 1–405.

(154) Haar, L.; Gallagher, J. S. *J. Phys. Chem. Ref. Data* **1978**, *7*, 635–793.

(155) Aston, J. G.; Siller, C. W.; Messerly, G. H. *J. Am. Chem. Soc.* **1937**, *59*, 1743–1751.

(156) Felsing, W. *Ind. Eng. Chem.* **1929**, *21*, 1269–1272.

(157) Reid, R. C.; Prausnitz, J. M.; Sherwood, T. K. *The Properties of Gases and Liquids*, 3d ed.; McGraw-Hill: New York, 1977.

(158) Swift, E. *J. Am. Chem. Soc.* **1942**, *64*, 115–116.

(159) Letcher, T. M. *J. Chem. Thermodyn.* **1972**, *5*, 159–173.

(160) Aston, J. G.; Eidinoff, M. L.; Forster, W. W. *J. Am. Chem. Soc.* **1939**, *61*, 1539–1543.

(161) Aston, J. G.; Sagenkahn, M. L.; Szasz, G. J.; Moessen, G. W.; Zuhr, H. F. *J. Am. Chem. Soc.* **1944**, *66*, 1171–1177.

(162) Majer, V.; Svoboda, V. *Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation*; Blackwell Scientific Publications: Oxford, U. K., 1985.

(163) Hales, J. L.; Gundry, H. A.; Ellender, J. H. *J. Chem. Thermodyn.* **1983**, *15*, 211–215.

(164) Beaton, C. F.; Hewitt, G. F.; Liley, P. E. *Physical Property Data for the Design Engineer*; Hemisphere Pub. Corp.: New York, 1989.

(165) Goodwin, R. D. *Hydrogen Sulfide Provisional Thermophysical Properties From 188 to 700 K at Pressures to 75 MPa*; Report, NBSIR-83-1694; NTIS No. PB84-122704; 1983

(166) Russell, H., Jr.; Osborne, D. W.; Yost, D. M. *J. Am. Chem. Soc.* **1942**, *64*, 165.

(167) Berthoud, A.; Brun, R. *J. Chim. Phys. PCB* **1924**, *21*, 143–160.

(168) Haines, W. E.; Helm, R. V.; Bailey, C. W.; Ball, J. S. *J. Phys. Chem.* **1954**, *58*, 270–278.

(169) Haines, W. E.; Helm, R. V.; Cook, G. L.; Ball, J. S. *J. Phys. Chem.* **1956**, *60*, 549–555.

(170) *Selected Values of Physical and Thermodyanmic Properties of Hydrocarbons and Related Compounds American Petroleum Institute Research Project 44*; Carnegie Press, Pittsburgh, PA, 1953.

(171) *Physical Constants of Hydrocarbons*; ASTM Technical Publication No. 109A, American Society for Testing and Materials: Philadelphia, PA, 1963.

(172) Yaws, C. L. *Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds (online book)*; Knovel: Norwich, NY, 2003.

(173) Somsen, G.; Coops, J. *Recl. Trav. Chim. Pay-B.* **1965**, *84*, 985–1002.

(174) Covington, A. K.; Dickinson, T. *Physical Chemistry of Organic Solvent Systems*; Plenum Press: London, 1973.

(175) *DMF Product Bulletin*; E. 1. duPont, Inc: Wilmington, DE, 1971.

(176) Gopal, R.; Rigzi, S. A. *J. Indian Chem. Soc.* **1966**, *43*, 179.

(177) Geller, B. E. *Zh. Fiz. Khim.* **1961**, *35*, 2210.

(178) Zegers, H. C.; Somsen, G. *J. Chem. Thermodyn.* **1984**, *16*, 225–235.

(179) Lemire, R. J.; Sears, P. G. *Top. Curr. Chem.* **1978**, *74*, 45–91.

(180) Wohlfarth, C. *Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures*; Springer-Verlag: Berlin, 1991; Vol. 6.

(181) Speight, J. G. *Perry's Standard Tables and Formulas for Chemical Engineers*; McGraw-Hill: New York, 2002.

(182) Schlundt, H. Ph.D. Thesis, Dept. of Chemistry, University of Wisconsin, Madison, WI, 1901.

(183) Barthel, J.; Backhuber, K.; Buchner, R.; Hetzenauer, H. *Chem. Phys. Lett.* **1990**, *165*, 369–373.

(184) Kindt, J. T.; Schmuttenmaer, C. A. *J. Phys. Chem.* **1996**, *100*, 10373–10379.

(185) Tofts, P. S.; Lloyd, D.; Clark, C. A.; Barker, G. J.; Parker, G. J. M.; McConville, P.; Baldock, C.; Pope, J. M. *Magn. Reson. Med.* **2000**, *43*, 368–374.

(186) Holz, M.; Mao, X. A.; Seiferling, D.; Sacco, A. *J. Chem. Phys.* **1996**, *104*, 669–679.

(187) Williams, W. D.; Ellard, J. A.; Dawson, L. R. *J. Am. Chem. Soc.* **1957**, *79*, 4652–4654.

(188) O'Reilly, D. E.; Peterson, E. M.; Scheie, C. E. *J. Chem. Phys.* **1973**, *58*, 4072–4075.

(189) Chen, L. P.; Gross, T.; Ludemann, H. D. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3503–3508.

(190) Hurle, R. L.; Woolf, L. A. *Aust. J. Chem.* **1980**, *33*, 1947–1952.

(191) Kamei, Y.; Oishi, Y. *B. Chem. Soc. Jpn.* **1972**, *45*, 2437–2439.

(192) Stevens, E. D. *Acta Crystallogr., Sect. B* **1978**, *34*, 544–551.

(193) Jeffrey, G. A.; Ruble, J. R.; Mcmullan, R. K.; Defrees, D. J.; Binkley, J. S.; Pople, J. A. *Acta Crystallogr., Sect. B* **1980**, *36*, 2292–2299.

(194) Nahringbauer, I. *Acta Chem. Scand.* **1970**, *24*, 453–462.

(195) Boese, R.; Blaser, D.; Latz, R.; Baumen, A. *Acta Crystallogr., Sect. C* **1999**, *55*, IUC9900001.

(196) Craven, B. M.; Mcmullan, R. K.; Bell, J. D.; Freeman, H. C. *Acta Crystallogr., Sect. B* **1977**, *33*, 2585–2589.

(197) Golubev, S. N.; Kondrashev, Y. D. *Zh. Strukt. Khim.* **1984**, *25*, 147–150.

(198) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563–595.

(199) Wolfenden, R.; Liang, Y.; Matthews, M.; Williams, R. *J. Am. Chem. Soc.* **1987**, *109*, 463–466.

(200) Wolfenden, R. *Biochemistry* **1978**, *17*, 201–204.

(201) Abraham, M. H.; Whiting, G. S. *J. Chem. Soc., Perkin Trans. 2* **1990**, 291–300.