Supporting Information for:

Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning

Daniel J. Cole,[1,2] Jonah Z. Vilseck,[1] Julian Tirado-Rives,[1] Mike C. Payne,[2]

William L. Jorgensen[1,]*

[1] Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, United States

[2] TCM Group, Cavendish Laboratory, 19 JJ Thomson Avenue, Cambridge CB3 0HE,

United Kingdom

* To whom correspondence should be addressed. E-mail: william.jorgensen@yale.edu

S1. SUPPORTING METHODS

Free Energies of Hydration. Free energies of hydration for 41 neutral, organic molecules were calculated using combined Metropolis Monte Carlo[1] and Free Energy Perturbation[2] (MC/FEP) simulations.[3-7] All small molecule calculations were performed using the BOSS software package.[8] For each molecule, eleven λ-windows of simple overlap sampling (SOS)[5,9] were employed to perturb OPLS-AA Lennard-Jones non-bonded parameters[10,11] and 1.14*CM1A partial atomic charges[4,7,12] (OPLS/CM1A) into non-bonded parameters derived from AIM electron density partitioning. Within a thermodynamic cycle, 5 million configurations of equilibration and averaging were performed in gas (5 M/5 M) and 10 M/15M configurations were performed in water. The TIP4P water model was used to represent the aqueous environment.[13] All simulations were run in the isothermal-isobaric ensemble at 25 °C and 1 atm.

For aqueous simulations, molecules were placed within boxes of 500 TIP4P water molecules with periodic boundary conditions. Solvent-solvent and solute-solvent cutoff distances were set to 10 Å for all non-hydrogen atom pairs and non-bonded interaction energies were quadratically smoothed to zero over the last 0.5 Å. Solute and volume moves were attempted every 100 and 3125 configurations, respectively. Solvent molecules were constrained to be rigid, with no intra-molecular degrees of freedom sampled, and were subject to translations and rotations within ranges of ±0.15 Å and ±15.0°. Solutes were translated within ranges of ±0.06 Å and ±6.0°. Solute bonded interactions were represented by the OPLS-AA force field, and were fully sampled. For select solutes, extra off-center charges were added to account for anisotropy in the electron density (Section S3). The positions of the extra sites were constrained to be rigid in the MC/FEP simulations. Introducing flexibility for these sites did not change free energy results above the statistical noise. Final uncertainties were calculated by the batch means

method with batch sizes of 1M configurations and were less than 0.25 kcal/mol.[14] Finally, a long range correction was added *post hoc* to computed free energies to account for long-range solute-solvent van der Waals interactions neglected beyond the cutoff.[15,16]

Pure Liquid Simulations. To further examine the accuracy of AIM non-bonded parameters, Metropolis Monte Carlo[1] (MC) pure liquid simulations were performed for a set of 36 neutral organic molecules for 20 M/20 M configurations. Each simulation was run in the isothermal-isobaric ensemble at 1 atm and 25 ºC or at the solvent's boiling point, if the boiling point is below 25 ºC; *N*-methylacetamide was run at 100 °C. Pure liquids consisted of 267 solvent molecules arranged in cubic cells with periodic boundary conditions. AIM parameters were used for Coulombic and Lennard-Jones non-bonded interactions, while bonded interactions were described by the OPLS-AA force field. All intra- and inter-molecular degrees of freedom were fully sampled, excepting extra point sites which were constrained in their optimized positions. Translations and rotations were performed within ranges of ±0.12 Å and ±12.0º, respectively. Non-bonded solvent-solvent cutoffs ranged between 11–15 Å depending on molecular sizes. In general, aromatic compounds employed 15 Å cutoffs, carbonyl compounds used 13 Å cutoffs, and all other molecules used 11 Å cutoffs. Quadratic smoothing was again performed over a final cutoff distance of 0.5 Å and solvent-solvent long range corrections were added.[15,16] Corresponding gas phase MC simulations of a single solvent molecule were performed with 4 M/4 M configurations. Averaged gas phase and condensed phase energies were subsequently used in eq S1 to compute heats of vaporization for each pure liquid.[10] Average liquid volumes were used to compute pure liquid densities.

$$\Delta H_{vap} = E_{intra}(g) - \left(E_{intra}(l) + E_{inter}(l)\right) + RT \tag{S1}$$

Protein-Ligand Simulations. The initial coordinates of the complex between indole and the L99A mutant of T4 lysozyme were constructed from the 185L PDB file[17] using the MCPRO software.[8] The 105 amino acids closest to the ligand were retained, and solvated in a 20 Å water cap using the JAWS algorithm.[18] No water molecules were placed in the binding pocket by JAWS, as expected. The protein energetics were described using the OPLS-AA force field, the ligands with OPLS/CM1A, and water with TIP4P. Following the JAWS equilibration stage, the water molecules and ligand were removed from the structure, and the protein coordinates were input to the ONETEP software for force field derivation as described in the main text.

Free energy calculations were performed using a modified version of MCPRO version 2.3, which allows the atomic charges and vdW parameters for the protein to be read from a file at the start of the calculation. The indole to benzofuran transformation was achieved by simultaneously mutating -NH to -CH and -CH to O, thus avoiding sampling errors caused by the ligand rotating in the binding pocket. FEP calculations were performed using 11 λ-windows of simple overlap sampling,[5] with each window comprising 10 M configurations of equilibration and 40 M configurations of averaging. Protein backbone sampling was achieved using the concerted rotation algorithm in MCPRO.[19] Unbound simulations were run in a 20 Å water cap for 20 M / 80 M configurations. All condensed phase simulations were run in triplicate with different starting coordinates.

S2. OPTIMUM DIELECTRIC FOR CONDENSED PHASE MODELING

Consider the transfer of a benzene molecule from the gas phase to water as depicted in Figure S1(a). Interaction between the benzene molecule and the polar water environment causes distortion of the ground state wave function (ψ^0) to its polarized state (ψ') and an associated enhanced benzene-water electrostatic interaction. Modeling this induction effect with a fixed charge force field is generally not possible.

Instead, it is necessary to account for induction in an effective manner. Consider now the transfer shown in Figure S1(b). The start and end points are the same as before, but two extra intermediate states are depicted in which the benzene molecule is "half-polarized". That is, the new wave function (ψ) is distorted from the ground state in the gas phase (ψ^0) but not as much as in water (ψ'). The free energy ΔG_{calc} represents the quantity computed in an FEP calculation using a fixed charge force field, that is, one in which the polarization of the molecule is unchanged upon moving from gas to water. Following the arguments of Chipot,[20] the energy change required to distort the electronic wave function from its gas phase ground state to ψ is given by:

$$\varepsilon_{distort} = \langle \psi | \hat{H}^0 | \psi \rangle - \langle \psi^0 | \hat{H}^0 | \psi^0 \rangle > 0 \tag{S2}$$

where $\hat{H}^0$ represents the DFT Hamiltonian in the gas phase and ψ is the wave function of the molecule computed in a dielectric medium with $1 < \varepsilon < 80$. Similarly, the energy gained by polarizing the wave function from ψ to ψ' in water is:

$$\varepsilon_{pol} = \langle \psi' | \hat{H}' | \psi' \rangle - \langle \psi | \hat{H}' | \psi \rangle < 0 \tag{S3}$$

where $\hat{H}'$ is the Hamiltonian representing the DFT implicit solvent model described in the main text. Equating the two transfer processes shown in Figure S1(a) and (b), the best estimate of the experimental free energy of hydration is:

$$\Delta G_{exp} = \varepsilon_{distort} + \Delta G_{calc} + \varepsilon_{pol} \tag{S4}$$

which becomes equal to the free energy of hydration computed using FEP and a fixed charge model when:

$$\varepsilon_{distort} + \varepsilon_{pol} = 0 \tag{S5}$$

Figure S2 shows this quantity plotted as the dielectric medium of the intermediate states (that is, as the extent of polarization of ψ) is varied. The first point to note is that for very small or very large values of ε, the magnitude of the sum of the distortion and polarization energies quickly becomes greater than 1 kcal/mol. Thus methods that compute charges in the gas phase are likely to strongly under-estimate polarization effects and, conversely, methods that compute charges in the water phase are likely to strongly under-estimate distortion effects. Fortunately, Figure S2 reveals that for a range of functional groups and molecular sizes, the distortion and polarization energies approximately cancel out when ε = 4. Therefore, charges computed in a dielectric medium with ε = 4 should provide a reasonable estimate of the experimental free energy of hydration that accounts in an effective manner for induction effects.

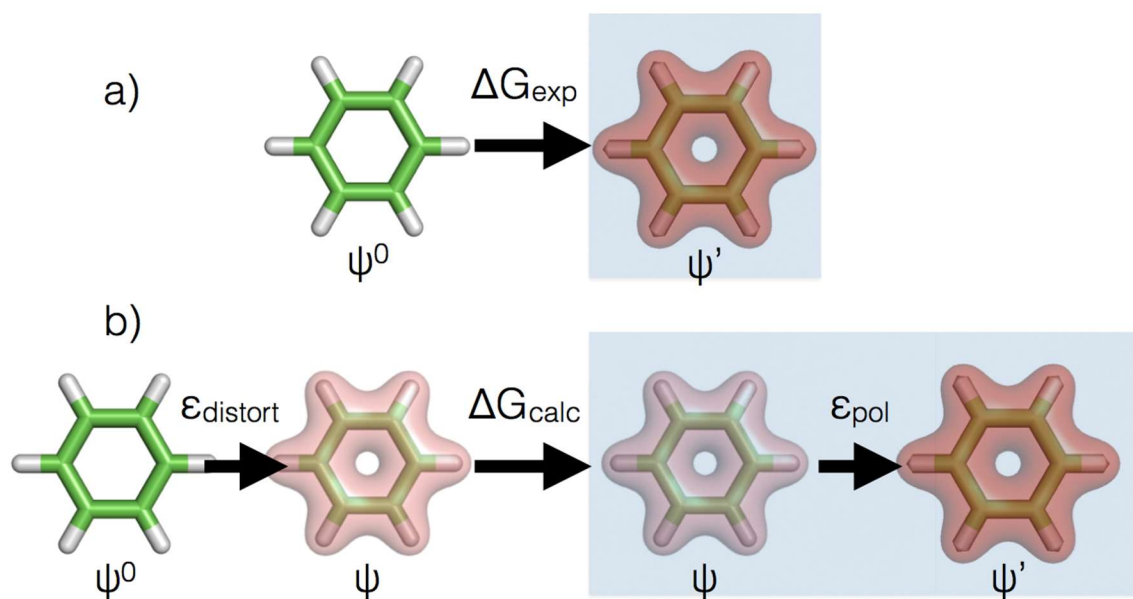**Figure S1.** (a) Pictorial representation of the transfer of a benzene molecule from the gas phase to water. Polarization of the electronic ground state density is denoted by red surfaces. (b) The same process but now the molecule is transferred through two hypothetical intermediate states in which the ground state electron density is first "half-polarized" in vacuum and then transferred to water.
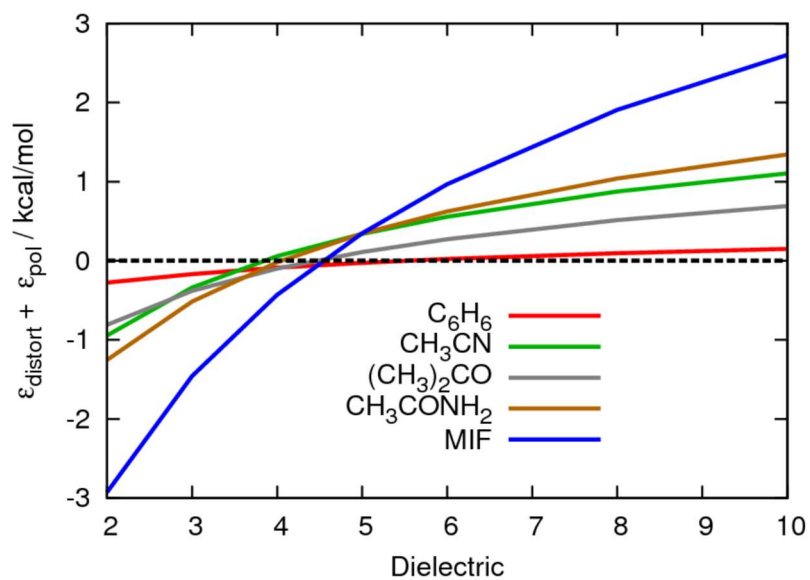
Figure S2. Variation of the summed distortion and polarization energies as computed using the ONETEP implicit solvent model with the dielectric medium of the intermediate states shown in Figure S1(a). Four small molecules have been selected from the present study and, in addition, a larger (34 atom) model of a drug-like aryl-1,2,3-triazole has been selected from Ref. 21.

Table S1. Free Atom Reference Data. V^{free} was computed using the MP4(SDQ)/aug-cc-pVQZ method in Gaussian09[22] and the chargemol code.[23] B^{free} are obtained from the database of Chu and Dalgarno.[24] R^{free} are fit to reproduce experimental liquid densities.

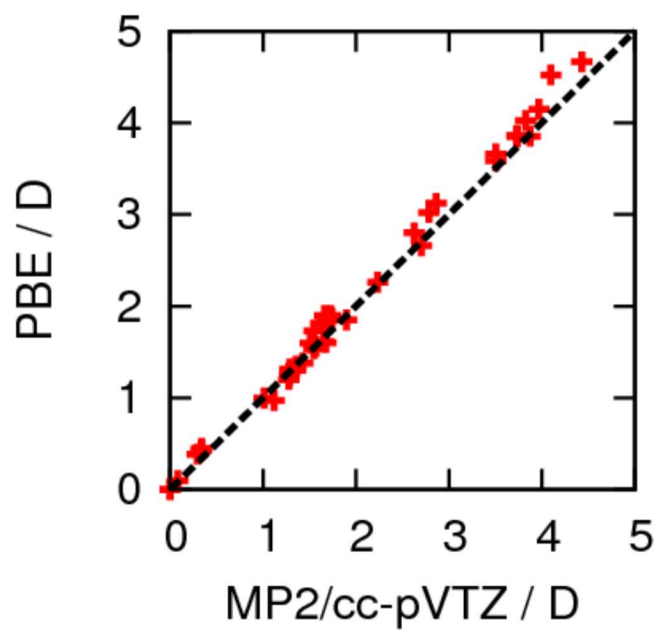| Element	V^{free} (Å^3)	B^{free} (Ha.Bohr^6)	R^{free} (Å)
H	7.6	6.5	1.64
C	34.4	46.6	2.08
N	25.9	24.2	1.72
O	22.1	15.6	1.60
F	18.2	9.5	1.58
S	75.2	134	2.00
Cl	65.1	94.6	1.88

Figure S3. Comparison between DFT, using the PBE exchange-correlation functional and an NGWF basis set, and MP2/cc-pVTZ gas phase dipole moments for the small molecule benchmark set.

S3. EXTRA POINT CHARGE SITE DERIVATION

In cases where the atomic electron density is close to spherically-symmetric, the electrostatic potential (ESP) outside the charge distribution is described well by an equivalent point charge placed at the center of the charge distribution. In other well-documented cases, such as atoms that possess lone pairs or σ-holes, the atomic electron density is anisotropic and an atom-centered point charge model will fail to reproduce the true ESP.[25]

In order to identify cases where there is a strong anisotropy in the electron density, we begin by defining for each atom a simple error function:

$$F(\boldsymbol{\mu}, \boldsymbol{Q}) = \Delta\mu_x^2 + \Delta\mu_y^2 + \Delta\mu_z^2 + \frac{1}{2\text{Å}^2}\left(\Delta Q_{xy}^2 + \Delta Q_{xz}^2 + \Delta Q_{yz}^2 + \Delta Q_{x^2-y^2}^2 + \Delta Q_{3z^2-r^2}^2\right) \qquad \text{(S6)}$$

where units of e and Å have been used. $\Delta\mu_x$ is the difference between the x-component of the dipole moment of the partitioned atomic electron density and the dipole moment of the point charges assigned to the atom of interest, using the nucleus as the center of the coordinate system. ΔQ_{xy} is the corresponding difference in quadrupole moment. For a standard atom-centered point charge model, all of the point charge multipole moments are zero by definition. With a threshold of $F = 0.015$ $e^2\text{Å}^2$, we identified 15 atoms in the test set with an anisotropic electron density (Table S3).

A commonly used technique for modeling anisotropy in the electron density is through the use of extra point sites, with extra site charge and positional parameters fit to reproduce *ab initio* or empirical data.[26-28] In keeping with our aims of reducing the number of adjustable parameters, we have sought to use the properties of the AIM atomic electron densities to derive the optimal positions of the extra point sites for the 15 molecules. Specifically, we have used a simplex minimization routine to place off-center extra point sites on the atom of interest to minimize the function $F(\boldsymbol{\mu},\boldsymbol{Q})$ subject to the constraint that the extra sites must lie within 1 Å of

the nucleus (2 Å for the Cl nucleus[27]). Extra point sites are added one at a time until little improvement in $F(\mu,Q)$ is observed. Convergence of $F(\mu,Q)$ becomes slow as the number of point sites is increased, so we have limited the number of extra points per site to a maximum of three. Preliminary tests showed that the addition of extra sites at just one atom can cause large errors in the total dipole moment of the molecule. However, adding a single extra site at neighboring heavy atoms is usually sufficient to correct this error. Table S2 compares the molecular dipole moment of 12 molecules, before and after the addition of extra sites, with the QM dipole moment.

Table S3 shows the differences between the multipole moments of the partitioned QM electron density and the corresponding multipole moments of the off-center point charges following optimization. In general, the residual multipole moments are significantly reduced ($F(\mu,Q) < 0.001$ $e^2Å^2$) compared to the simple atom-centered point charge model. For some sulfur atoms, there is a high quadrupole moment that cannot be modeled using two or three extra point sites. Following extensive minimization for S in dimethyl sulfide, the quadrupole moment was successfully reproduced using four extra sites, but the free energy of hydration and liquid properties were largely unchanged, so we limit ourselves to a maximum of three extra sites. Benzamines showed no change in properties with extra point sites, so they were excluded for simplicity.

Table S2. Molecular dipole moments (D) of the AIM charge model before and after the addition of extra point sites. Comparison is with the QM dipole moment computed using a dielectric of 4.

	AIM Before	AIM After	QM
C_5H_5N	2.41	2.81	2.74
CH_3NH_2	1.69	1.87	1.89
$(CH_3)_2NH$	0.90	1.16	1.17
CH_3Cl	2.55	2.14	2.17
C_6H_5Cl	2.28	1.86	1.92
CH_3SH	2.10	1.84	1.82
CH_3SCH_3	1.98	2.04	1.89
CH_3SOCH_3	5.14	4.59	4.61
C_6H_5SH	1.49	1.17	1.17
$C_6H_5SCH_3$	1.63	1.62	1.58
CH_3OH	2.01	1.82	1.84
CH_3OCH_3	1.37	1.40	1.40

Table S3. Errors in the atomic multipole moments (a.u.) of the AIM charge model before and after the addition of extra sites. The second column gives the atom name and the number of extra sites added.

			$\Delta\mu_x$	$\Delta\mu_y$	$\Delta\mu_z$	ΔQ_{xy}	ΔQ_{xz}	ΔQ_{xz}	ΔQ_{x2-y2}	ΔQ_{3z2-r2}
CH_3OH	O(3)	before	0.062	**-0.134**	0.000	0.066	0.000	0.000	0.099	**0.583**
		after	0.002	0.007	-0.008	-0.006	-0.007	-0.005	0.000	0.035
C_5H_5N	N(1)	before	0.000	0.000	**0.195**	-0.001	0.000	0.000	**-0.110**	**0.629**
		after	0.000	0.001	0.001	-0.001	0.000	0.001	**-0.110**	0.000
CH_3NH_2	N(2)	before	**-0.190**	-0.001	-0.085	0.002	**0.146**	-0.001	0.034	**1.008**
		after	-0.007	0.006	0.000	-0.002	0.016	-0.013	-0.040	-0.001
$(CH_3)_2NH$	N(2)	before	0.000	**0.200**	**-0.121**	0.000	0.000	**-0.170**	**-0.165**	**0.550**
		after	-0.002	0.003	0.000	-0.001	0.000	0.016	-0.008	-0.001
CH_3OCH_3	O(2)	before	-0.001	**0.172**	-0.001	0.000	0.000	0.001	**-0.229**	**0.415**
		after	-0.002	0.000	0.001	-0.009	-0.007	0.000	-0.005	-0.002
CH_3SH	S(3)	before	-0.068	**0.146**	0.000	**0.171**	0.000	0.001	**-0.273**	**2.544**
		after	0.021	0.006	-0.001	**0.120**	-0.049	0.010	-0.062	-0.008
CH_3SCH_3	S(2)	before	0.000	**0.111**	0.000	-0.001	0.001	-0.001	**-0.448**	**2.391**
		after	-0.002	-0.050	-0.004	-0.002	0.015	0.023	**-0.389**	0.020
CH_3SCH_3	S(4)	before	0.000	**0.111**	0.000	-0.001	0.001	-0.001	**-0.448**	**2.391**
		after	0.043	-0.001	-0.011	0.041	0.017	-0.033	-0.021	-0.037
CH_3SOCH_3	S(2)	before	-0.001	-0.064	**-0.360**	0.001	-0.001	0.016	-0.082	**1.506**
		after	0.000	0.004	0.000	-0.004	0.013	0.023	-0.041	0.016
C_6H_5SH	S(3)	before	**0.150**	-0.044	0.000	**-0.182**	0.000	0.000	**0.206**	**1.941**
		after	0.004	-0.010	0.004	-0.008	0.043	0.031	0.027	0.004
$C_6H_5SCH_3$	S(2)	before	-0.051	-0.078	0.000	**-0.156**	-0.001	0.000	**-0.206**	**1.757**
		after	-0.010	0.010	-0.001	**-0.166**	0.010	-0.022	**-0.189**	-0.011
CH_3Cl	Cl(2)	before	0.021	0.000	0.000	0.000	0.000	0.000	**-0.892**	**0.893**
		after	-0.004	0.000	0.000	0.000	0.000	0.000	-0.010	0.011
C_6H_5Cl	Cl(2)	before	0.000	**0.100**	0.000	0.000	0.000	0.000	**0.741**	**0.659**
		after	-0.004	-0.002	-0.003	-0.003	0.000	0.000	0.044	-0.038
Lysozyme (Met102)	S(2)	before	-0.016	**-0.132**	0.002	**0.302**	**-0.442**	-0.028	**0.311**	**-0.301**
		after	0.007	-0.002	0.007	0.019	**-0.106**	**0.143**	-0.041	-0.003

Table S4. Computed Free Energies of Hydration with DDEC-derived Non-bonded Parameters (kcal/mol)

	ΔG_{hyd}	Expt.[b]
ethane	2.61 ± 0.09	1.83
acetonitrile	-4.66 ± 0.10	-3.89
propane	3.00 ± 0.09	1.96
propene	2.56 ± 0.12	1.32
acetone	-4.59 ± 0.11	-3.81
acetaldehyde	-3.53 ± 0.11	-3.50
acetic acid	-6.89 ± 0.11	-6.70
methyl acetate	-4.30 ± 0.13	-3.32
methyl formate	-3.55 ± 0.13	-2.78
acetamide	-10.37 ± 0.14	-9.71[c]
E-*N*-methylacetamide	-10.04 ± 0.17	-10.08[c]
Z-*N*-methylacetamide	-11.07 ± 0.16	-10.08[c]
dimethylacetamide	-9.33 ± 0.20	-8.55[c]
nitroethane	-3.99 ± 0.14	-3.71
benzene	0.01 ± 0.17	-0.86
benzonitrile	-4.23 ± 0.18	-4.21
toluene	0.49 ± 0.18	-0.89
nitrobenzene	-4.83 ± 0.23	-4.12
acetophenone	-5.80 ± 0.19	-4.58
benzamide	-12.36 ± 0.22	-11.01
α-methylstyrene	0.79 ± 0.20	-1.24
methyl benzoate	-4.44 ± 0.22	-3.93
pyridine[a]	-5.27 ± 0.19	-4.70
methylamine[a]	-6.01 ± 0.09	-4.56
dimethylamine[a]	-3.38 ± 0.10	-4.30
anisole	-0.99 ± 0.19	-2.46
phenol	-4.71 ± 0.18	-6.62
dimethyl ether[a]	-0.98 ± 0.10	-1.91
aniline	-3.62 ± 0.18	-5.49
chlorobenzene[a]	0.28 ± 0.17	-1.12
chloromethane[a]	-0.28 ± 0.09	-0.55
methanol[a]	-4.01 ± 0.08	-5.10
N-methylaniline	-2.04 ± 0.20	-4.69
N,*N*-dimethylaniline	-0.98 ± 0.21	-3.45

dimethyl sulfide[a]	-2.51 ± 0.12	-1.61		
dmso[a]	-9.06 ± 0.13	-10.11		
methanethiol[a]	-1.54 ± 0.11	-1.24		
thiophenol[a]	-0.77 ± 0.18	-2.55		
thioanisole[a]	-1.02 ± 0.20	-2.73		
fluorobenzene	0.68 ± 0.17	-0.80		
trifluorobenzene	0.47 ± 0.18	-0.25		

[a] extra point sites included. [b] Ref. (29). [c] Ref. (30).

Table S5. Computed Pure Liquid Heats of Vaporization (kcal/mol) and Densities (g/cm^3) with DDEC-Derived Non-bonded Parameters

	ΔH_{vap}	Expt.	density	Expt.
ethane	3.53 ± 0.019	3.52[c]	0.552 ± 0.001	0.546[c]
acetonitrile	8.79 ± 0.019	8.01[e]	0.790 ± 0.001	0.776[e]
propane	4.41 ± 0.022	4.49[c]	0.577 ± 0.001	0.581[c]
propene	4.14 ± 0.018	4.40[c]	N/A	N/A
acetone	7.57 ± 0.022	7.48[c]	0.779 ± 0.001	0.784[c]
acetaldehyde	6.48 ± 0.019	6.24[c]	0.775 ± 0.002	0.772[c]
acetic acid	13.59 ± 0.024	12.49[c]	1.109 ± 0.002	1.044[c]
methyl acetate	7.81 ± 0.025	7.72[e]	0.925 ± 0.001	0.928[c]
methyl formate	6.92 ± 0.021	6.78[j]	0.974 ± 0.002	0.971[b,k]
acetamide		N/A		N/A
E-N-methylacetamide		N/A		N/A
Z-N-methylacetamide	15.54 ± 0.049	13.30[c]	0.897 ± 0.002	0.894[c]
dimethylacetamide	11.93 ± 0.041	11.75[c]	0.924 ± 0.001	0.936[c]
nitroethane	9.69 ± 0.021	9.94[e]	1.019 ± 0.001	1.040[e]
benzene	7.61 ± 0.031	8.09[g]	0.883 ± 0.001	0.874[g]
benzonitrile	11.65 ± 0.050	12.54[e]	0.998 ± 0.001	1.001[e]
toluene	8.43 ± 0.032	9.08[f]	0.859 ± 0.001	0.862[f]
nitrobenzene	12.68 ± 0.034	13.15[e]	1.192 ± 0.001	1.200[e]
acetophenone	13.04 ± 0.033	13.24[j]	1.025 ± 0.001	1.028[b,k]
benzamide		N/A		N/A
α-methylstyrene		N/A		N/A
methyl benzoate	12.83 ± 0.055	13.28[e]	1.084 ± 0.001	1.089[e]
pyridine[a]	8.88 ± 0.026	9.61[i]	0.971 ± 0.001	0.978[i]

methylamine[a]	7.56 ± 0.021	6.17^{d}	0.684 ± 0.002	0.694^{d}
dimethylamine[a]	6.78 ± 0.030	6.33^{d}	0.667 ± 0.001	0.671^{d}
anisole	10.31 ± 0.042	11.21^{f}	0.973 ± 0.001	0.989^{f}
phenol	13.32 ± 0.040	13.82^{c}	1.057 ± 0.001	1.058^{c}
dimethyl ether[a]	4.82 ± 0.020	5.14^{c}	0.699 ± 0.002	0.735^{c}
aniline	11.07 ± 0.047	12.60^{d}	0.982 ± 0.001	1.017^{d}
chlorobenzene[a]	8.65 ± 0.030	9.79^{h}	1.095 ± 0.001	1.101^{h}
chloromethane[a]		N/A		N/A
methanol[a]	8.36 ± 0.014	8.95^{c}	0.741 ± 0.002	0.786^{c}
N-methylaniline	11.94 ± 0.035	12.70^{d}	0.962 ± 0.001	0.984^{d}
N,N-dimethylaniline	12.22 ± 0.057	11.90^{d}	0.946 ± 0.001	0.953^{d}
dimethyl sulfide[a]	6.16 ± 0.021	6.61^{c}	0.822 ± 0.002	0.842^{c}
dmso[a]	11.87 ± 0.032	12.64^{n}	1.094 ± 0.001	1.096^{n}
methanethiol[a]	4.92 ± 0.016	5.87^{c}	0.850 ± 0.002	0.888^{c}
thiophenol[a]	9.54 ± 0.026	11.37^{j}	1.057 ± 0.001	$1.078^{b,k}$
thioanisole[a]	11.66 ± 0.029	12.52^{p}	1.045 ± 0.001	$1.057^{b,m}$
fluorobenzene	7.24 ± 0.037	8.26^{h}	0.993 ± 0.001	1.019^{h}
trifluorobenzene	8.70 ± 0.046	8.99^{j}	1.195 ± 0.001	1.178^{o}

[a] extra point sites included. [b] Expt. data at 20 °C. [c] Ref. (10). [d] Ref. (31). [e] Ref. (32). [f] Ref (33). [g] Ref. (34). [h] Ref. (28). [i] Ref. (35). [j] Ref. (36). [k] Ref. (37). [l] Ref. (38). [m] Ref. (39). [n] Ref. (40). [o] Ref. (41). [p] Ref. (42). N/A = Not available

S4. SUPPORTING REFERENCES

(1) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087-1092.

(2) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420-1426.

(3) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J. Comput. Chem.* **1993**, *14,* 1240-1249.

(4) Udier-Blagovic, M.; Morales De Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 1322-1332.

(5) Jorgensen, W. L.; Thomas, L. L. *J. Chem. Theory Comput.* **2008**, *4*, 869-876.

(6) Kaminski, G.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787-1796.

(7) Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2014**, *10*, 2802-2812.

(8) (a) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689-1700. (b) Jorgensen, W. L. *BOSS,* Version 4.9; Yale University, New Haven, CT, 2011. (c) Jorgensen, W. L. *MCPRO,* Version 2.3; Yale University, New Haven, CT, 2012.

(9) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Comput. Chem.* **2004**, *25*, 28-39.

(10) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

(11) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Nat. Acad. Sci. USA* **2005***, 102*, 6665-6670.

(12) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput-Aided Mol. Des.* **1995**, *9*, 87-110.

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

(14) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Clarendon: Oxford, 1987; pp 191-198.

(15) Leach, A. R. *Molecular Modelling: Principles and Applications*, 2nd ed.; Pearson Education Limited: Essex, England, 2001, pp. 325-327.

(16) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638-6646.

(17) Morton, A.; Matthews, B. W. *Biochemistry* **1995**, *34*, 8576-8588.

(18) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2009**, *113*, 13337-13346.

(19) Ulmschneider, J. P.; Jorgensen, W. L. *J. Chem. Phys.* **2003**, *118*, 4261-4271.

(20) Chipot, C. *J. Comput. Chem.* **2003**, *24*, 409-415.

(21) Jorgensen, W. L.; Gandavadi, S.; Du, X.; Hare, A. A.; Trofimov, A.; Leng, L.; Bucala, R. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 7033-7036.

(22) Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

(23) Manz, T. A.; Gabaldon Limas, N. Chargemol program for performing DDEC analysis. Available from http://ddec.sourceforge.net (accessed September 2014).

(24) Chu, X.; Dalgarno, A. *J. Chem. Phys.* **2004**, *121*, 4083-4088.

(25) Kramer, C.; Spinn, A.; Liedl, K. R. *J. Chem. Theory Comput.* **2014**, *10*, 4488-4496.

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

(27) Dixon, R. W.; Kollman, P. A. *J. Comput. Chem.* **1997**, *18*, 1632-1646.

(28) Jorgensen, W. L.; Schyman, P. *J. Chem. Theory Comput.* **2012**, *8*, 3895-3901.

(29) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. *J. Chem. Soc. Perkin Trans. 2* **1994**, 1777-1791.

(30) (a) Wolfenden, R. *Biochemistry* **1978**, *17*, 201-204. (b) Radzicka, A.; Pedersen, L.; Wolfenden, R. *Biochemistry* **1988**, *27*, 4538-4541.

(31) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827-4836.

(32) Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. *J. Comput. Chem.* **2001**, *22*, 1340-1352.

(33) Jorgensen, W. L.; Laird, E. R.; Nguyen, T. B.; Tirado-Rives, J. *J. Comput. Chem.* **1992**, *14*, 206-215.

(34) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768-4774.

(35) Jorgensen, W. L.; McDonald, N. A. *J. Mol. Struct. (THEOCHEM)* **1998**, *424*, 145-155.

(36) Enthalpies of Vaporization. *CRC Handbook of Chemistry and Physics*, 95th ed.; Boca Raton, FL, 2014-2015; pp 6-128 - 6-145.

(37) Laboratory Solvents and Other Liquid Reagents. *CRC Handbook of Chemistry and Physics*, 95th ed.; Boca Raton, FL, 2014-2015; pp 15-13 - 15-22.

(38) Good, W.D. *J. Chem. Eng. Data* **1972**, *17*, 158-162.

(39) Sigma-Aldrich. http://www.sigmaaldrich.com/catalog/ (Accessed Jan. 1, 2015).

(40) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. *J. Chem. Theory Comput*. **2012**, *8*, 61-74.

(41) Yaws, C. L. *Thermophysical Properties of Chemicals and Hydrocarbons*; William Andrew Inc.: Beaumont, Texas, 2008.

(42) NIST Chemistry Webbook. http://webbook.nist.gov/chemistry/ (Accessed January 1, 2015)