Strathprints Institutional Repository

This version is available at http://strathprints.strath.ac.uk/42566/

Predicting Hydration Gibbs Energies of Alkyl-aromatics Using Molecular Simulation: A Comparison of Current Force Fields and the Development of a New Parameter Set for Accurate Solvation Data

Nuno M. Garrido[1], Miguel Jorge[1], António J. Queimada[1], José R. B. Gomes[2],

Ioannis G. Economou[3] and Eugénia A. Macedo[1,*]

1. LSRE Laboratory of Separation and Reaction Engineering, Departamento de Engenharia Química, Faculdade de Engenharia, Universidade do Porto, Rua do Dr. Roberto Frias, 4200 - 465 Porto, Portugal

2. CICECO, Department of Chemistry, University of Aveiro, Campus Universitário de Santiago, Aveiro 3810-193, Portugal

3. The Petroleum Institute, Department of Chemical Engineering, PO Box 2533, Abu Dhabi, United Arab Emirates

[*]Author to whom all correspondence should be addressed at: eamacedo@fe.up.pt

1

ABSTRACT

The Gibbs energy of hydration is an important quantity to understand the molecular behavior in aqueous systems at constant temperature and pressure. In this work we review the performance of some popular force fields, namely TraPPE, OPLS-AA and Gromos, in reproducing the experimental Gibbs energies of hydration of several alkyl-aromatic compounds – benzene, mono-, di- and tri-substituted alkylbenzenes – using molecular simulation techniques. In the second part of the paper, we report a new model that is able to improve such hydration energy predictions, based on LJ parameters from the recent TraPPE-EH force field and atomic partial charges obtained from natural population analysis of density functional theory calculations. We apply a scaling factor determined by fitting the experimental hydration energy of only two solutes, and then present a simple rule to generate atomic partial charges for different substituted alkyl-aromatics. This rule has the added advantages of eliminating the unnecessary assumption of fixed charge on every substituted carbon atom and providing a simple guideline for extrapolating the charge assignment to any multi-substituted alkyl-aromatic molecule. The point charges derived here yield excellent predictions of experimental Gibbs energies of hydration, with an overall absolute average deviation of less than 0.6 kJ/mol. This new parameter set can also give good predictive performance for other thermodynamic properties and liquid structural information.

1. Introduction

Over the last decades, free energy calculations have continuously become more important in the chemical and biochemical fields, as noted in several literature reviews[1-4]. Free energy studies allow scientists and engineers to achieve microscopic-level insights into the behavior of complex systems that may be helpful in chemical process design, e.g. to estimate important properties such as solubility[5,6] or partition coefficients[7,8]. Accurate predictions of the Gibbs energy of solvation, also called solvation free energy, have the capacity to revolutionize several fields, predominantly in the pharmaceutical industry and/or in biological chemistry[9,10]. In this context, molecular simulation has emerged as a highly promising method to estimate free energies of drug-like molecules[10,11]. The importance and at the same time the difficulty of connecting classical atomistic models to experimental results in order to predict the Gibbs energy of solvation is also illustrated in a recent *blind challenge*[12].

One of the major drawbacks of free energy predictions from molecular simulations is that they tend to be quite sensitive to the force field and in some cases to the simulation parameters (cutoff radii, treatment of long-range electrostatics, etc.).[13] For this reason, several studies have focused on comparing existing force fields[14-17] or on fine-tuning existing ones[18-20] for solvation Gibbs energy predictions. Some of those studies have shown that the electrostatic contribution to the calculated Gibbs energy, commonly dependent on an adequate assignment of partial charges to each atom or group of atoms, is particularly prone to uncertainties, and sometimes hinders the transferability of the force field parameters[21,22]. In fact, an important shortcoming of fixed-charge force fields is that they do not normally take into account the different polarization environment when a molecule is transferred from the liquid phase to the gas phase (or vice-versa)[23,24].

Recently, our group has also tested the performance of three force-fields in predicting the Gibbs energy of hydration for the *n*-alkane series[7] and some representative polar compounds[25]. One of the conclusions was that "generic" force fields were able to provide reasonable predictions of Gibbs energy of solvation of simple and monofunctional compounds, but they generally failed for more complex polyfunctional molecules, where the deviation from experimental data can be larger than 10 kJ/mol. This justifies the need for a more detailed study of the effects of force field parameters on the Gibbs energy of solvation, most notably of partial atomic charges, since the electrostatic contribution usually represents a very large percentage of

the total hydration energy for polar solutes[21,22]. Our strategy is to systematically study a homologous series of compounds, rather than to apply the usual approach of testing the force fields against a large database of solutes with many different functional moieties. In this way, we hope to identify more easily the reasons for the success or failure of certain parameter sets, and to establish a well-defined set of rules for force field development.

In the present work, we focus on the aqueous behavior of benzene and several alkyl-substituted aromatics: mono-, di- and tri-substituted alkylbenzenes. Such aromatic fragments are widely present in chemical and pharmaceutical compounds, biological assemblies and petrochemical systems. Xylenes are a particular family of di-substituted alkylbenzene isomers that are used as large scale industrial solvents and intermediates for many derivatives, e.g, *p*-xylene is used in the polyester industry. Besides their importance as industrial chemicals, aromatic compounds were also found to play a key role in biochemistry[26-28]. This class of compounds is also particularly interesting from a theoretical point of view, as aromatic interactions are usually difficult to quantify experimentally[29]. To the best of our knowledge there is no published systematic study about the prediction of the Gibbs energy of hydration of alkyl-aromatic compounds from molecular simulation. Thus, in the first part of this work we predict the Gibbs energies of hydration of benzene and five mono-substituted alkyl-benzenes using five current force field parameter sets from the TraPPE, Gromos and OPLS families.

When modeling polyatomic molecules, two approaches are generally followed to represent the dispersion and repulsive forces: i) all-atom (AA) models; ii) united-atom (UA) models. In AA models, a separate force center is assigned to each atom, located on its nucleus. In UA models specific groups of atoms, such as CH, CH2 or CH3, are treated as single force centers. The advantage of AA models is that they generally give a better account of molecular geometry and structure, as well as dynamic properties[30,31]. The counterpart is that they require a great deal of computer time because of the larger number of force centers, bending angles and torsion angles involved. UA methods generally neglect the hydrogen atoms connected to carbon atoms, and their influence is considered implicitly in the parameterisation of potential parameters for the pseudo-atoms. A "hybrid" approach is also possible, where aromatic hydrogens are treated explicitly, while aliphatic hydrogens are treated within a UA approximation. In this paper, we will call this an explicit-hydrogen (EH) model.

The TraPPE force field was originally developed under a UA philosophy, where both aromatic CH and aliphatic CH_x groups were treated as pseudo-atoms whose molecular weights are equivalent to those of the whole fragment (henceforth denoted by TraPPE-UA[31-34]). More recently, an EH version of the TraPPE force-field was proposed for benzene[35] and we have also decided to test this version of TraPPE (henceforth denoted by TraPPE-EH). However, this force field has not yet been extended to alkyl-aromatic compounds. Likewise, the Gromos biomolecular force field was originally developed as a united atom force field[36] (Gromos-UA). However, recent developments have provided parameters for an EH description of the aromatic CH groups (Gromos-EH, version 53A5[37]), which were parameterized to reproduce, among other properties, the Gibbs energy of hydration of amino acids containing aromatic groups[37]. Finally, although starting with a UA description, the Optimized Potential for Liquid Simulations (OPLS)[30] force field quickly became popular with its AA description of the solute molecules. Moreover, for aromatic molecules the OPLS-UA model was not parameterized. One of the purposes of this paper is to understand how accurate the current parameter sets are in reproducing experimental hydration energies, and to evaluate the need for using optimized parameters to correctly achieve hydration data predictions.

In the second part of this paper, we attempt to improve the prediction of hydration energies, by generating a new charge set for an EH model for alkylbenzenes, based on the recent TraPPE-EH force field[35], using information from natural population analysis (NPA) of density functional theory (DFT) calculations. NPA is a method for deriving charges and orbital populations of atoms in a molecular system, based on a natural bond orbital analysis[38]. The new charge set is obtained by scaling the calculated gas-phase charges to fit the experimental Gibbs energy of solvation of benzene and of 1,3,5-trimethylbenzene. A new rule for assigning partial atomic charges for multisubstituted alkyl-aromatic compounds is also presented. The ability of this new charge set to predict solvation energies of several other alkyl-aromatic compounds is examined.

The remainder of this paper is organized as follows: in Section 2 the methodology to calculate hydration energies and thermodynamic properties of the pure components is explained together with the main computational details. In Section 3.1 we test the ability of current force field parameter sets to predict the Gibbs energy of hydration. Section

3.2 is dedicated to improving force field parameters for better reproducing experimental hydration energies. Finally, our main conclusions are summarized in Section 4.

2. Theoretical and Computational Details
2.1. Calculation of the Gibbs energy of hydration

The Gibbs energy of hydration can be seen as the total reversible work required to transfer a solute molecule from the ideal gas phase to water, at constant pressure and temperature, representing the infinite dilution chemical potential. Free energies can be calculated for an appropriate molecular model by carrying out molecular dynamics (MD) simulations and a thermodynamic integration (TI) procedure. Based on a thermodynamic cycle, the Gibbs energy of hydration, $\Delta_{hyd}G$, is estimated from simulations in which solute-solvent interactions are progressively turned off. This has normally been accomplished by carrying out two separate simulations, one in solution (where both intra- and inter-molecular contributions are turned-off) and another one in vacuum (to subtract the intra-molecular contribution)[39]. Details of this methodology can be found in a previous publication[13]. However, a new feature in the MD GROMACS 4.0 suite[40] allows for these intermolecular solute-solvent interactions to be quantified by solely running a liquid phase simulation, thus avoiding the need for a vacuum step. The latter approach was adopted here.

The TI algorithm, free of hysteresis, has been described in detail in previous publications[41-43]. Briefly, if we consider any two generic well-defined states, the total Hamiltonian of the system can be made a function of a coupling parameter, λ, and used to describe the transition between those two states – in this case, an initial fully interacting system and a final "dummy" state where all solute-solvent interactions are turned off. Considering several discrete and independent λ values, equilibrium averages can be used to evaluate derivatives of the Gibbs energy with respect to λ. One then integrates these derivatives along a continuous path connecting the initial and final states to obtain the Gibbs energy between them.

It is both physically insightful and methodologically convenient to separate the Gibbs energy into a sum of two separate contributions: a non-polar Lennard-Jones (LJ) contribution (that includes the cavity formation work and favorable solute-solvent dispersion interactions) and an electrostatic (Coulombic) term. In our protocol, in order

to avoid charge-fusion effects[14,44] we have first turned off the electrostatic interactions and then the LJ interactions. The total Gibbs energy of hydration can then be estimated from:

$$\Delta_{hyd}G = -\int_0^1 \left\langle \frac{\partial \mathcal{H}(\lambda)}{\partial \lambda} \right\rangle_\lambda^{Total} d\lambda = -\left[\int_0^1 \left\langle \frac{\partial \mathcal{H}(\lambda)}{\partial \lambda} \right\rangle_\lambda^{Elec} d\lambda + \int_0^1 \left\langle \frac{\partial \mathcal{H}(\lambda)}{\partial \lambda} \right\rangle_\lambda^{LJ} d\lambda \right] \tag{1}$$

A MD simulation was performed for each discrete λ value, ranging from $\lambda = 0$ (fully interacting solute) to $\lambda = 1$ (non-interacting solute). We have used 5 λ values for the electrostatic decoupling and 15 λ points to evaluate the LJ contribution. As demonstrated in a previous work, this choice of intermediate states ensures sufficient accuracy in the free energy calculations[45]. The integration of the LJ Hamiltonian derivatives was carried out by fitting the data to a physically-based approximation to the cavity formation and dispersion interaction terms and then integrating the curve analytically[45]. The fitting function is:

$$\left(\frac{\partial \mathcal{H}(\lambda)}{\partial \lambda} \right)_{LJ} = A_0\lambda^2 + A_1\lambda - \frac{A_2}{\lambda^2 - A_3\lambda + A_4} + \frac{A_2}{A_4} \tag{2}$$

where A_i are fitting parameters[45]. In all cases we have employed non-linear weighted least squares fittings of the simulation data. Correlation coefficients above 0.99 and a global RMS error of 0.3 kJ/mol were obtained. This integration procedure was shown to increase the precision of the calculated hydration energies while decreasing the number of necessary intermediate points[45]. Simpson's rule was used for the integration of the electrostatic component. During the decoupling process, the electrostatic interactions were linearly interpolated between neighboring states while the LJ interactions were interpolated via soft-core interactions[46], with parameters given in our previous publications[13]. This soft-core dependence eliminates singularities in the calculation as the LJ interactions are turned off [47].

2.2. Molecular Dynamics Simulations

We have carried out MD simulations both for the calculation of the Gibbs energy of hydration, as described in the previous section, and for computing pure-fluid thermodynamic properties. All simulations were performed with version 4.0.7 of the

GROMACS software[40] using the leap-frog Verlet integration algorithm[48] with a time step of 2 fs to integrate Newton's equations of motion. Simulations were performed using periodic boundary conditions in all directions. Covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm[49] while the water geometry was fixed with the SETTLE algorithm[50]. For efficiency reasons (see reference[51] for details), the reaction-field method[52], which approximates the medium beyond a cut-off distance of 1 nm by a dielectric continuum of uniform permittivity, was used to handle long-range electrostatics. The dielectric constant was adjusted to be the permittivity of each solvent or pure component. The Ewald summation method was applied in a few test cases and the resulting hydration energies were practically the same as with the reaction field method (difference of at most 0.2 kJ/mol). The remaining cut-off radii were 1 nm for the short-range neighbor list and a 0.8-0.9 nm switched cut-off for the LJ interactions. We have also applied long range dispersion corrections for energy and pressure[14].

In the case of free energy calculations, solvated systems consisted of one solute molecule and 500 water molecules at 298 K and 1 bar. As discussed in several previous studies, the choice of the water model may have implications in the final predicted Gibbs energy[15,17,21,53]. In the present work we have decided to use the Modified Extended Simple Point Charge (MSPC/E)[54] model for the simulation of water. MSPC/E is an accurate force field for pure water and aqueous phase equilibria thermodynamic properties, and includes a polarization correction expected to improve hydration energy predictions. A comparison of the free energy results predicted with the MSPC/E and SPC/E[55] water models was reported previously[51]. Nevertheless, we have also performed some tests with the TIP4P[56] water model in this paper, as discussed in Section 3.2.

Langevin stochastic dynamics[57] was used to control the temperature, with a frictional constant of 1 ps^{-1} while for constant pressure runs the Berendsen barostat[58], with a time constant of 0.5 ps and an isothermal compressibility of $4.5 \times 10^{-5}\,\mathrm{bar}^{-1}$, was used to enforce pressure coupling. For each simulation, we first run an energy minimization (the Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm[59] during 5000 steps followed by a steepest descent minimization during 500 steps were used) and afterwards a constant volume equilibration (100 ps), a constant pressure equilibration (500 ps), enough to obtain complete equilibration of the box volume, and

finally a 5 ns *NpT* production stage. It is worth noticing that 5 ns is considered to be enough to observe several transitions between stable configurations, indicating sufficient sampling of torsional degrees of freedom, even for complex solutes.[13] Nevertheless, this point was carefully confirmed in this work for the longer *n*-alkylbenzenes, where several transitions between different torsional configurations were verified. All simulations were performed in triplicate, or in the case of the larger molecules in quintuplicate, in order to obtain statistically meaningful results.

For the estimation of the pure component properties we have employed the Nosé-Hoover thermostat[60,61] for temperature coupling using a time constant of 1 ps, and the Parrinello-Rahman approach[62] with a time constant of 2 ps to enforce pressure coupling. The reference pressure was set to 1 bar and the compressibility was set according to the fluid under evaluation (data from Cibulka and Takagi[63]). Each simulation box contained 512 molecules and had equilibrated box dimensions ranging from 4.19 to 5.14 nm. Liquid densities were directly obtained from the GROMACS suite using the *g_energy*[64] tool, while heats of vaporization were estimated by taking the difference of enthalpy in the vapor and liquid phases:

$$\Delta_{vap}H = E_g - E_L + RT \tag{3}$$

where, E_g is the total energy in the gas phase and E_L is the total energy per mole in the liquid phase. Although other approaches for computing the heat of vaporization exist[65], equation (3) is sufficiently accurate for our purposes. Vacuum (gas phase) simulations were conducted without cutoffs or periodic boundary conditions for 50 ns.

2.3. Density Functional Theory Calculations

The gas-phase geometries of all benzene derivatives considered in this work were optimized, using very tight convergence criteria (*opt=verytight* and *int=ultrafine* keywords), with the B3LYP hybrid DFT approach[66,67] as included in the Gaussian 03 package[68]. The 6-31+G(d,p) basis set was used to describe the electronic densities of all atoms[69,70]. The optimized structures were characterized as true minima on the potential energy surface by calculation of the vibrational frequencies at the same level of theory (zero imaginary frequencies). The computational approach adopted above yielded accurate structures and thermochemical data for several different compounds derived

from benzene[71,72]. The atomic charges were calculated i) from a natural population analysis[73] (NPA) of the natural atomic orbitals (NAO) obtained by using the atomic blocks of the density matrix averaged over the spatial directions, e.g. p_x, p_y, p_z orbitals, and ii) from the application of the CHelpG scheme[74], based on the fitting of charges to the electrostatic potential calculated in a regularly spaced grid of points around the molecule under the constraint that the dipole moment is preserved. All charges thus calculated are provided in the Supporting Information (Tables S5 to S12).

3. Results and Discussion

3.1. Prediction of $\Delta_{hyd}G$ using current force field parameter sets

The Gibbs energy of hydration of benzene (BZ) and several mono-substituted alkylbenzenes, namely methylbenzene (MB), ethylbenzene (EB), propylbenzene (PB), butylbenzene (BB) and pentylbenzene (PeB), were initially estimated using the five parameter sets already mentioned: TraPPE-UA/EH, Gromos-UA/EH and OPLS-AA. Gibbs energy of hydration results using these force fields are presented in Tables S1-S4 of the Supporting Information and illustrated in Figure 1, while absolute average deviations (AAD) are summarized in Table 1. From these results one may generally conclude that with the exception of Gromos-EH version 53A6, which was specifically reparameterized to fit hydration energies, none of the current force field parameter sets are able to accurately predict the Gibbs energy of hydration of our molecular test set. Moreover, it is clear that OPLS-AA, in spite of its increased number of degrees of freedom arising from the all-atom description, is not able to accurately describe the Gibbs energy of hydration. Although it provides a reasonable prediction for BZ, it significantly overestimates the increase in free energy with the degree of substitution. Similar results were obtained for the hydration of alkanes[7] suggesting that the shortcoming may arise from a deficiency in the parameterization of the LJ component in this force field.

Both of the UA force fields yield Gibbs energies of hydration that are too positive by a large amount (AAD of 5.9 and 9.4 kJ/mol for Gromos-UA and TraPPE-UA, respectively). More precisely, all predictions for the Gibbs energies of hydration with UA models are positive, while the experimental values are all negative. This

Physical Chemistry Chemical Physics

shortcoming of the UA force fields is due to the fact that they neglect the electrostatic component of the interaction potential (all partial charges are set to zero), which is observed to be important in the solvation of aromatic solutes. Nevertheless, one may notice that the effect of increasing the alkyl chain length on going from toluene (methylbenzene) to pentylbenzene is correctly captured by both UA force fields, particularly in the case of TraPPE-UA. Because this series is generated by inserting neutral CH_x groups into the substituent chain, it is expected that the electrostatic component will be approximately the same for all solutes, and thus the free energy differences between them should be dominated by changes in the LJ component. Since the relative Gibbs energy between subsequent members of the series is adequately captured, it can be concluded that the LJ component of the Gibbs energy is well described by the UA force fields. The current TraPPE LJ force field parameters are thus a good starting point for a further refinement in order to better reproduce hydration energies.

Dramatic quantitative improvements (by about 5 kJ/mol, see Table 1) in the Gibbs energy predictions can be obtained when an explicit hydrogen approach (which includes point charges, as in the case of TraPPE-EH and Gromos-EH) is used to describe the aromatic ring, maintaining the UA description of the alkyl substituents. Nevertheless, the TraPPE-EH prediction for BZ still falls short of the experimental hydration energy. In the case of the Gromos-EH force field, the predictions are quantitatively much better for the entire series (on average 3.8 kJ/mol closer to experiment when compared to the united-atom version – Table 1), but the experimental trend from benzene to toluene is not correctly described. Looking at the experimental data ($\Delta_{hyd}G^{exp}$ is more negative for MB than for BZ), and taking into account that the LJ component monotonically increases from BZ to PeB for all force fields (see Tables S1-S4), one should expect a slightly more negative electrostatic term for MB. However, in the Gromos 53A6 parameter set, the point charges located on the first site of the alkyl substituent and on the substituted carbon are halved relative to the corresponding aromatic hydrogen and carbon atoms in benzene, which significantly reduces the magnitude of the electrostatic contribution. In the following, we will attempt to improve the description of the Gibbs energy of hydration of alkyl-aromatics by adopting an explicit hydrogen description of the aromatic ring, based on the TraPPE-EH force field, and adjusting the point charges based on insight obtained from DFT calculations.

3.2. A new parameter set for accurate $\Delta_{hyd}G$ predictions

Our starting point is to describe the benzene $\Delta_{hyd}G^{exp}$ by deriving new parameters modifying TraPPE-EH to obtain a better match to the experimental Gibbs energy of hydration. We have seen that the electrostatic term, which usually plays a major role in bio-molecular systems[22], seems to be incorrectly captured by the original charges in TraPPE-EH, as shown in Figure 1. This is not as unexpected as it may seem, if we consider that the experimental hydration energy includes additional contributions that arise from the change in polarization environment when moving from the gas phase to water, which are not captured by fixed-charge force fields. Indeed, the process described in section 2.1 computes the free energy to hydrate a molecular model of the solute in water, while holding its charge distribution fixed. However, in a real solvation process the electrostatic potential of the solute changes in response to the change in the polarization environment. Polarizable models can explicitly account for this change, but they are much more computationally demanding and are thus currently inappropriate for routine solvation free energy calculations. In the following, we will try to account for polarization effects in an approximate way using fixed-charge models. Swope *et al.*[75,76] have recently presented a detailed analysis of this subject, and we will mostly follow their reasoning here.

The entire solvation process can be represented using a thermodynamic cycle such as the one shown in Figure 2. The experimental solvation energy, described in the top line of the cycle, is the quantity we are interested in predicting. This corresponds to transferring the "experimental" solute *X* (i.e., with a fully accurate description of its electrostatic potential) from an unpolarized state in the gas phase to a polarized environment in the solvent phase. This process can be broken down into 3 steps, depicted in the remainder of the cycle. On the left-hand side we show the free energy contribution from polarizing the solute in the gas phase. In other words, a fixed-charge model is constructed, with parameters that provide a good representation of the solute in the liquid phase (we call this the "model" solute), and ΔG_{Pol}^{G} is the free energy difference between the experimental unpolarized solute and the polarized solute model in the gas phase. The next stage of the cycle is the hydration free energy computed in the simulations, i.e., the free energy of transferring the model solute from the gas phase to the solvent phase, keeping the charges constant. Most models are designed to represent the electrostatic potential of the solute in the pure-liquid phase. However, the

polarization environment in the actual solvent of interest (in this case, water) can be significantly different. Thus, in principle one must also account for the free energy cost of changing this pure-component description into one that adequately represents the "experimental" solute in the solvent phase (the branch on the right-hand side of the cycle). Finally, it is worth mentioning that we follow Swope *et al.*[75] in assuming that the model is able to sample the entire configuration space available to the "experimental" solute, so that all restructuring free energy contributions are zero.

Swope *et al.*[75,76] went on to develop accurate expressions for estimating ΔG_{Pol}^{G} for fixed-charge models. A simplified form of those expressions, equation (4), has been used previously, for example in the development of the SPC/E water model[55].

$$\Delta G_{Pol}^{G} = \frac{\left(|\mu_L| - |\mu_G|\right)^2}{2\alpha} \tag{4}$$

In the above expression, μ_G is the dipole moment of the solute in the gas phase, μ_L is the corresponding dipole moment in the liquid phase, and α is the isotropic polarizability. We have applied equation (4) to estimate this contribution for several alkyl-aromatic solutes, and found the energies to be lower than 0.1 kJ/mol, which is below the statistical error of the simulations (see below). This agrees with Swope *et al.*[75], who found that the polarization contribution was practically negligible for toluene. Based on these results, the ΔG_{Pol}^{G} contribution will be neglected in the remainder of this paper. It should be noted, however, that for more polar solutes the gas-phase polarization contribution is likely to be important and should be taken into account.

Swope *et al.*[75] also neglect ΔG_{Pol}^{L} by assuming that the model is able to provide an adequate description of the solute polarization in the solvent under consideration. Unfortunately, this is not always the case. Indeed, most fixed-charge force fields are parameterized to reproduce pure-component properties (in the case of TraPPE, these are vapor-liquid coexistence curves, critical and normal boiling points and heats of vaporization). Conversely, the prediction of Gibbs energy of hydration requires an accurate description of the solute interactions in water – which is a completely different environment than a pure organic liquid. When moving from the pure liquid to water, the polarization environment (included implicitly in the parameterization of the point charges) changes significantly. Therefore, one should expect that charges parameterized in the pure liquid lose some of their accuracy when transferred to a water environment. Indeed, this reason is behind the recent parameterization of the GROMOS 53A6 force

field, where different charges are fitted to reproduce Gibbs energy of solvation in solvents of different polarity[37]. This fact and the comparison we carried out in section 3.1 for standard "pure-component" force fields strongly suggest that ΔG_{Pol}^{L} is non-negligible. Unfortunately, it is not straightforward to estimate this contribution in a similar way as for ΔG_{Pol}^{G}, since in the former case the change in polarization of the solute will also change the solvent structure and the solute-solvent interaction energy in a non-trivial way[75].

An alternative approach to circumvent this limitation of standard force fields is to develop a fixed-charge model that is indeed able to provide an accurate representation of the polarization energy in the solvent phase. To achieve this, we have decided to evaluate different point charges in the aromatic carbons and hydrogens, based on insights from NPA analysis of hybrid DFT calculations. It is important to note that the development and comparison of methods to extract point charges from the full electrostatic potential is still a subject of active debate in the theoretical chemistry community[77]. A particularly sensitive point is related to the incorporation of charges derived from gas-phase DFT into MD simulations in the liquid phase, since the different environment will almost certainly induce changes into the electrostatic potential (and hence into the effective point charges) of the molecule. Although grid-based methods for point charge calculation (such as the CHelpG[74] or the ESP[78] methods) are generally chosen for direct use in classical potentials, some studies have shown that they yield physically inconsistent variations in the point charges of homologous compounds[79,80]. This effect was confirmed here for the substituted aromatics (Tables S5-S12). Charges obtained by dividing the occupancies of the natural orbitals between the constituent atoms, like NPA charges, tend to suffer much less from this problem[81]. Nevertheless, the latter are known to differ appreciably from the partial atomic charges derived from the quantum chemical electrostatic potential, often by being larger in magnitude[77]. Taking into account all of these factors, we have opted to use NPA charges for our study, mostly due to their chemical consistency across homologous series of compounds. To account for the change of environment when going from the gas to the aqueous phase, as well as to correct the magnitude of the NPA charges, we have multiplied all of the gas-phase charges by a constant scaling factor, which was determined as described below. The concept of rescaling point charges to describe a change in polarization environment is not new, and has been recently reviewed[82]. It is

important to reiterate that the new charges derived here include the effect of a change in the polarization environment only implicitly.

- *The Gibbs Energy of Hydration of Benzene*

Due to the symmetry of the benzene ring, each CH aromatic group must be neutral, and so the charges assigned to the carbon and to the hydrogen atoms by the NPA method are symmetric (see Table S5). The scaling factor was estimated by fitting simulation results to the experimental Gibbs energy of hydration of benzene. We have analyzed the sensitivity of the electrostatic contribution to the hydration energy by testing 8 different values of the scaling factor, implying a range of charges from $0.055 \leq q_H \leq 0.155$. From this analysis, we have verified that the system responds under a second order polynomial, going through the origin (see Figure S1 for details). Such a quadratic behavior of the electrostatic free energy is to be expected and has been reported before[83], but it is nevertheless useful for charge-fitting purposes. In order to match the experimental Gibbs energy of hydration, we propose the following parameters for benzene: $q_{C_{aro}} = -q_{H_{aro}} = -0.1225$, which yield an electrostatic contribution of $\Delta G^C_{benzene} = -9.13$ kJ/mol and a $\Delta_{hyd} G_{benzene} = -3.6$ kJ/mol. These charges were obtained by multiplying the NPA benzene partial charges by a scaling factor, $k_P = 0.5$. As described above, this scaling factor accounts for limitations of the NPA method to accurately reproduce the electrostatic potential of the molecule, but also takes into account the change of environment from the gas phase to the aqueous phase, which may have a strong effect on the values of the effective point charges. Given the importance of these effects in a highly polar solvent such as water, it is not surprising to find a scaling factor of 0.5.

Previous work has shown that the choice of water model may have a significant effect on Gibbs energy of hydration predictions[15]. Even though it is beyond the scope of the present work to perform an exhaustive analysis of these effects, we have tested the current benzene parameters in a different water model. Thus, for the TIP4P water model we have obtained $\Delta_{hyd} G^{Calc}_{TIP4P} = \Delta G^{LJ}_{TIP4P} + \Delta G^C_{TIP4P} = 5.6 - 9.2 = -3.6$ kJ/mol, which is to the same as the MSPC/E result.

- *The Gibbs Energy of Hydration of Substituted Aromatics*

Having determined the optimal charges for BZ, in a second stage we addressed the Gibbs energy of hydration of toluene, longer mono-substituted alkylbenzenes and multi-substituted aromatics. As observed both in the trend from MB to PeB, shown in Figure 1, and in a previous study for alkanes[7], the TraPPE-UA force field is able to correctly describe the non-polar component of the hydration energy for the series of alkanes or alkyl substituents, i.e., the TraPPE-UA force field reproduces the experimental trends in the relative Gibbs energy of solvation of the series MB to PeB. Thus, for all the alkyl substituents, we took the LJ parameters from this force field[32,33] and kept them unchanged throughout our subsequent analysis. With these LJ parameters we then focused on the electrostatic component of the hydration energy.

Substituted aromatics can be "built" by replacing one (or more) aromatic hydrogens with a given alkyl group. Such a substitution obviously changes the electrostatic potential of the substituted aromatic carbon(s) and of the corresponding alkyl substituent(s), but may also induce changes in the charge distribution of the rest of the aromatic ring. When inserting a substituent in the aromatic ring, it is expected that the charge symmetry in the ring will suffer a disruption. The information extracted from DFT analysis shows that, for example, the inclusion of a CH_3 group into the BZ ring to yield MB or the inclusion of a C_2H_5 into the BZ ring to yield EB, promotes a delocalization of the charges in the remaining aromatic sites. Concretely, in both cases the carbons located in *ortho-* and *meta-* positions relative to the substituted carbon become slightly more positive while the carbon in *para-* position becomes more negative (see Tables S6 and S7, respectively). As for hydrogens, those in *meta-* and *para-* positions remain practically unchanged while the *ortho-*hydrogen becomes less polar.

Although the NPA analysis provides interesting insight regarding the effect of substituents on the charge distribution within the aromatic ring, it is difficult to directly apply NPA charges to the substituted carbon and alkyl substituents. This is because the NPA/DFT calculations naturally consider an all-atom description of the aliphatic groups, while our classical model adopts a united-atom approach. Simply adding up the NPA point charges of each CH_x group present in the different substituent chains will yield practically neutral net charges for the pseudo-atoms and for the substituted carbon

(see Tables S6-S8), which will strongly underestimate the electrostatic contribution to the hydration energy.

In order to determine the charge assigned to the substituted carbon atoms for the different multisubstituted alkylbenzenes we have determined the optimal value of this charge by fitting the hydration energy of 1,3,5-trimethylbenzene (TMB). For the non-substituted carbons and hydrogens we use the scaled NPA charges ($k_P = 0.5$). We have thus gradually changed the charge of the (C)-CH$_x$ carbon (the charge of the CH$_x$ group is also adjusted, to keep the whole molecule neutral) until the experimental TMB Gibbs energy of hydration was reproduced. Once more, we obtained a quadratic dependence of the TMB Gibbs energy of hydration with the carbon charge (see Figure S2). The optimal charge for the substituted aromatic carbon was found to be $q_{C-(CH_x)} = -0.107$ (Table 2) and the corresponding TMB Gibbs energy of hydration was -3.8 kJ/mol.

If we simply transfer this charge on the substituted carbon to several other mono-, di- and tri-substituted aromatics, we still assume that the effect of each substitution on that charge is independent of other substitutions elsewhere in the ring (i.e., the charge is directly transferred from TMB to the other solutes). Instead, we have attempted to devise a general rule for determining the contribution of each substitution to the charge on neighboring atoms, based on an analysis of the NPA charges.

We begin by comparing the scaled NPA charges for the substituted aromatics with the analogous charges in benzene, thus obtaining the corresponding charge differences caused by the substitutions. More precisely, these differences were calculated for the charges on MB, *meta*-xylene (MX), and TMB. We have found that each substitution of an aromatic hydrogen by an alkyl fragment in those molecules induced charge increments (or decrements) on each of the neighboring atoms that were practically independent of the degree of substitution. For example, changing from MX to TMB induced practically the same relative differences as changing from MB to MX. This suggests that one can establish a simple procedure to compute the point charges on any alkyl-aromatic compound based simply on the benzene charges and on the charge increments caused by its respective substituent groups. Such a procedure satisfies the following equation:

$$q_i = k_P \times q_i^{BZ} + \sum_{j=0}^{3} \delta_j N_j \qquad\qquad (5)$$

where q_i is the charge on site i (with numbers given in Figure 3), q_i^{BZ} is the corresponding charge in the benzene molecule, δ_j is the charge increment caused by a substituent at position j, and N_j is the total number of substituents at position j (j = 1,2,3 for ortho, meta and para substitution, respectively, and j = 0 for the substituted C-CH$_x$ group). The charge increments for each substituent position, shown in Table 3, were adjusted to provide the best possible match to the scaled NPA charges for the 3 compounds mentioned above. Point charges estimated using equation 4 were in excellent agreement with the scaled NPA charges for all compounds of the test set (RMS error lower than 0.0001 a.u.). Remaining CH$_x$ groups (e.g terminal CH$_3$ groups in EB, C$_2$H$_5$ in PB, etc.) were considered to be neutral, as one might expect.

The above procedure involves simply determining, for each atom of a given alkyl-aromatic molecule, the corresponding values of N_j. Equation (5) is then applied to each atom, starting from the corresponding benzene charges and adding the contributions due to each substituent, to obtain the new charge set. For example, in TMB the charge on each non-substituted carbon atoms takes into account the contribution of one substituent in *para* position plus two substituents in *ortho* position. In the Supporting Information, we present an example (tutorial) of this calculation procedure for 1,2,4-trimethylbenzene. In Figure 4 we summarize the procedure followed to derive this new charge set and list the different quantities used as input data during its development. In Table 2 we give the detailed charge set of parameters proposed for the different compounds.

The results obtained by applying this rule are presented in Table 4 for a test set of 14 alkylaromatics and represented in Figure 5. The overall AAD for the mono-, di- and tri-substituted alkylaromatics was 0.4, 0.7 and 0.9 kJ/mol, respectively. For the monosubstituted aromatics, we have also included our prediction in Figure 1. As depicted in this figure, when comparing our new parameter set with the current force fields we can observe a remarkable improvement in the predicted free energies (See Table 1 for quantitative values). Also, in terms of comparison, for alkylaromatic compounds an AAD of more than 2 kJ/mol was previously reported in an extensive hydration test using the generic Amber force field[16]. This shows that the approach

proposed here for parameterizing point charges has a significant potential to improve the capacity of current force fields to predict Gibbs energy of hydration. Moreover, the rule presented by equation (5) has the added advantages of eliminating the unnecessary assumption of fixed charge on every substituted carbon atom and providing a simple guideline for extrapolating the charge assignment to any multi-substituted alkyl-aromatic molecule.

Although the main aim of our parameterization of the point charges for alkyl-aromatic solutes is to predict Gibbs energies of hydration, we have also made a short test to evaluate how the new parameters perform for thermodynamic properties of pure liquids, namely their ability to predict i) liquid density in a wide temperature range and ii) enthalpy of vaporization. Additionally, we have tested the impact of force field changes on the structure of pure liquid benzene by computing radial distribution functions (RDFs). The calculated solvent densities from *NpT* MD and the heats of vaporization at various temperatures are shown in Supporting Information in Tables S13 and S14, respectively. Reported results enable to observe that the new charge set also allows estimating liquid densities and vaporization enthalpies for the test set in good agreement with the experimental data, i.e., a global deviation of 2% for liquid densities in the overall temperature range and a global AAD of 1.4 kJ/mol for the enthalpies. The structure of liquid benzene can be observed in several RDFs, which were calculated and are presented in Figure S3 of the Supporting Information. For the case of the computed aromatic Carbon – Carbon interactions (C-C curve), the agreement with experimental data available from X-ray diffraction[84] is good. Moreover, both the experimental peaks in the 0.5-0.6 nm region and the shoulder around 0.7 nm are reproduced in the calculated data. The C-H and H-H functions are also similar to the ones obtained in previous calculations[85,86]. All three curves are practically indistinguishable from the corresponding RDFs obtained with the original TraPPE-EH model.

4. Conclusions

By evaluating the ability of existing molecular mechanics force fields to predict hydration energies of a set of alkyl-aromatic compounds, we have observed that a correct assignment of partial atomic charges is essential to obtain accurate results. Based on this observation, we have performed a detailed and systematic study of different force fields to predict hydration Gibbs energies.

A better performance was obtained when information from normal population analysis of DFT calculations was used to describe charge delocalization within the aromatic ring induced by alkyl substitutions. Although it is premature to recommend NPA charges as the method of choice for force field parameterization, since our study focuses only on a single class of solutes, the good performance observed here is rather encouraging. Other methods for computing point charges should be tested in the near future and the range of solute types should be expanded to include more polar molecules.

NPA charges were multiplied by a constant scaling factor to account for the change of environment on moving from the gas phase to the aqueous phase, and this scaling factor was optimized by fitting the benzene hydration energy. An optimization of the charge on the substituted carbon atom was also found to improve performance. At this point, we cannot be sure if the same scaling factor applies for the Gibbs energy of hydration of more polar molecules. Regarding solvation in different solvents, it is likely that a different factor will need to be employed to account for a different solvation environment. Further studies are needed to clarify these issues.

As a result of our study, we have proposed a general rule for assigning point charges to any alkyl-aromatic solute, based on charge increments caused by each substitution. This new charge set was able to predict experimental hydration energies of several aromatic compounds with remarkable accuracy. In principle, this rule could be extended to other types of substituents (e.g., halogen atoms, hydroxyl groups, etc.), with appropriate changes in the increment parameters. The new framework presented here for the development of point charges enables accurate predictions of Gibbs energies of hydration using molecular simulation and thermodynamic integration. In future publications, we intend to extend this framework to molecules containing other functional groups and possibly to other solvents.

Supporting Information Available:

Detailed results of predicted Gibbs energies of hydration for mono-substituted aromatics using the different force fields, NPA and CHelpG charges for molecules of the training and test sets, correlations obtained for the variation of the electrostatic component of the Gibbs energy with the values of the point charges for BZ and TMB are given as Supporting Information. A tutorial example of calculation of point charges using equation 5, numerical results on the prediction of liquid densities and enthalpies

of vaporization using the new data set are also included. Finally, computed RDFs for liquid benzene are also shown.

7. Cited References

(1)   Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.

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

(3)   Beveridge, D. L.; Dicapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.

(4)   Gilson, M.; Zhou, H. *Ann. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21.

(5)   Westergren, J.; Lindfors, L.; Hoglund, T.; Luder, K.; Nordholm, S.; Kjellander, R. *J. Phys. Chem. B* **2007**, *111*, 1872.

(6)   Palmer, D. S.; Llinas, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. *Molecular Pharmaceutics* **2008**, *5*, 266.

(7)   Garrido, N. M.; Queimada, A. J.; Jorge, M.; Macedo, E. A.; Economou, I. G. *J. Chem. Theory Comput.* **2009**, *5*, 2436.

(8)   Garrido, N. M.; Jorge, M.; Queimada, A. J.; Macedo, E. A.; Economou, I. G. *Phys. Chem. Chem. Phys.* **2011**, *DOI: 10.1039/c1cp20110g.*

(9)   Shirts, M. R.; Mobley, D. L.; Chodera, J. D. *Ann Rep Comput Chem* **2007**, *3*, 41.

(10)   Jorgensen, W. L. *Science* **2004**, *303*, 1813.

(11)   Jorgensen, W. L.; Duffy, E. M. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1155.

(12)   Guthrie, J. P. *J. Phys. Chem. B* **2009**, *113*, 4501.

(13)   Garrido, N. M.; Jorge, M.; Queimada, A. J.; Economou, I. G.; Macedo, E. A. *Fluid Phase Equilib* **2010**, *289*, 148.

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

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

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

(17)   Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616.

(18)   Davis, J. E.; Patel, S. *Chem. Phys. Lett.* **2010**, *484*, 173.

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

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

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

(22)   van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 4064.

(23)   Swope, W. C.; Horn, H. W.; Rice, J. E. *J. Phys. Chem. B* **2010**, *114*, 8631.

(24)   Swope, W. C.; Horn, H. W.; Rice, J. E. *J. Phys. Chem. B* **2010**, *114*, 8621.

(25)   Garrido, N. M.; Queimada, A. J.; Jorge, M.; Economou, I. G.; Macedo, E. A. *Fluid Phase Equilib.* **2010**, *296*, 110.

(26)   Hunter, C. A. *Chem. Soc. Rev.* **1994**, *23*, 101.

(27)   Benzing, T.; Tjivikua, T.; Wolfe, J.; Rebek, J. *Science* **1988**, *242*, 266.

(28)   Muehldorf, A. V.; Vanengen, D.; Warner, J. C.; Hamilton, A. D. *Journal of the American Chemical Society* **1988**, *110*, 6561.

(29)   Hobza, P.; Selzle, H. L.; Schlag, E. W. *Chem. Rev.* **1994**, *94*, 1767.

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

(31)   Chen, B.; Siepmann, J. I. *J. Phys. Chem. B* **1999**, *103*, 5370.

(32)   Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, *102*, 2569.

(33)   Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1999**, *103*, 4508.

(34)   Chen, B.; Potoff, J. J.; Siepmann, J. I. *J. Phys. Chem. B* **2001**, *105*, 3093.

(35)   Rai, N.; Siepmann, J. I. *J. Phys. Chem. B* **2007**, *111*, 10790.

(36)   Daura, X.; Mark, A. E.; van Gunsteren, W. F. *J. Comp. Chem.* **1998**, *19*, 535.

(37)   Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comp. Chem.* **2004**, *25*, 1656.

(38)   Reed, A.; Weinstock, R.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.

(39)   Leach, A. *Molecular Modeling: Principles and Applications*; Prentice-Hall, 2001.

(40)   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(41)   Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.

(42)   C. Chipot; Pohorille, A. *Free Energy Calculations - Theory and Applications in Chemistry and Biology*; Springer: Berlin, 2007.

(43)   Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235.

(44)   Deng, Y.; Roux, B. *J. Phys. Chem. B* **2004**, *108*, 16567.

(45)   Jorge, M.; Garrido, N. M.; Queimada, A. J.; Economou, I. G.; Macedo, E. A. *J. Chem. Theory Comput.* **2010**, *6*, 1018.

(46)   Beuler, T. M., R; van Schaik, RC; Gerber, PR; van Gunsteren, WF. *Chem. Phys. Lett.* **1994**, *222*, 529.

(47)   Pitera, J. W.; Van Gunsteren, W. F. *Molec. Sim.* **2002**, *28*, 45.

(48)   van Gunsteren, W.; Berendsen, H. *Mol. Sim.* **1988**, *1*, 173.

(49)   Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. *J. Comp. Chem.* **1997**, *18*, 1463.

(50)   Miyamoto, S.; Kollman, P. A. *J. Comp. Chem.* **1992**, *13*, 952.

(51)   Garrido, N.; Jorge, M.; Queimada, A.; Economou, I.; Macedo, E. *Fluid Phase Equilib.* **2010**, *289*, 148.

(52)   Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100.

(53)   Shivakumar, D.; Williams, J.; Wu, Y. J.; Damm, W.; Shelley, J.; Sherman, W. *J. Chem. Theory Comput.* **2010**, *6*, 1509.

(54)   Boulougouris, G. C.; Economou, I. G.; Theodorou, D. N. *J. Phys. Chem. B* **1998**, *102*, 1029.

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

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

(57)   Van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1982**, *45*, 637.

(58)   Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(59)   Liu, D. C.; Nocedal, J. *Math. Programm.* **1989**, *45*, 503.

(60)   Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

(61)   Nose, S. *Molecular Physics* **1984**, *52*, 255.

(62)   Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.

(63)   Cibulka, I.; Takagi, T. *J. Chem. Eng. Data* **1999**, *44*, 411.

(64)   Spoel, D. L., E; Hess, B; Buuren, A; Apol, E; Meuulenhof, P; Tieleman, D; Sijbers, A; Feenstra, K; Drunen, R; Berendsen, H; *Gromacs User Manual - version 3.3*The Netherlands, 2006.

(65)   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.

(66)   Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(67)   Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(68)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., J. A.; 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, Revision C.01*; Gaussian Inc.: Wallingford CT, 2004.

(69)   Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(70)   Hariharan, P. C.; Pople, J. A. *Teor. Chim. Acta* **1973**, *28*, 13.

(71)   Gomes, J. R. B. *J. Phys. Chem. A* **2009**, *113*, 1628.

(72)   Gomes, J. R. B.; Ribeiro da Silva, M. A. V. *Int. J. Quant. Chem.* **2005**, *101*, 860.

(73)   Reed, A. E.; Weinstock, R. B.; Weinhold, F. A. *J. Chem. Phys.* **1985**, *83*, 735.

(74)   Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(75)   Swope, W. C.; Horn, H. W.; Rice, J. E. *J. Phys. Chem. B* **2010**, *114*, 8631.

(76)   Swope, W. C.; Horn, H. W.; Rice, J. E. *J. Phys. Chem. B* **2010**, *114*, 8621.

(77)   Sigfridsson, E.; Ryde, U. *J. Comput. Chem.* **1998**, *19*, 377.

(78)   Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(79)   Pacios, L. F.; Gomes, P. C. *J. Mol. Struct. Theochem* **2001**, *544*, 237.

(80)   Gomes, J. R. B.; Cordeiro, M. N. D. S.; Jorge, M. *Geochim. Cosmochin. Acta* **2008**, *72*, 4421.

(81)   Wiberg, K.; Rablen, P. *J. Comput. Chem.* **1993**, *14*, 1504.

(82)   Leontyev, I.; Stuchebrukhov, A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2613.

(83)   Hummer, G.; Pratt, L. R.; Garca, A. E. *J. Phys. Chem.* **1996**, *100*, 1206.

(84)   Narten, A. H. *J. Chem. Phys.* **1977**, *67*, 2102.

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

(86)   Bartelll, L. S.; Sharkey, L. R.; Shi, X. *J. Am. Chem. Soc.* **1988**, *110*, 7006.

Table 1: Performance of the different force fields in reproducing the Gibbs energy of hydration of the mono-substituted *n*-alkylbenzene series. The subscripts give the statistical accuracy of the last decimal point shown.

Force-Field	Average Absolute Deviation (kJ/mol)
TraPPE-UA	9.4_6
TraPPE-EH[a]	4.2
OPLS-AA	8.0_5
Gromos-UA	5.9_4
Gromos-EH	2.1_3

[a] This result refers only to benzene

Table 2: Suggested partial atomic charges for improved hydration free energies, for benzene, TMB, mono-substituted *n*-alkylbenzenes[a] and di- and tri-substituted aromatics

#[b]	Benzene		TMB		n-alkylbenzenes		OX		MX	
	Site	q_i	Site	q_i	Site	q_i	Site	q_i	Site	q_i
1			C_{aro}^{subs}	-0.107	C_{aro}^{subs}	-0.1138	C_{aro}^{subs}	-0.1116	C_{aro}^{subs}	-0.1104
2			C_{aro}	-0.1225	C_{aro}^{ortho}	-0.1203	C_{aro}^{subs}	-0.1116	C_{aro}	-0.1181
3			C_{aro}^{subs}	-0.107	C_{aro}^{meta}	-0.1191	C_{aro}	-0.1169	C_{aro}^{subs}	-0.1104
4	C_{aro}	-0.1225	C_{aro}	-0.1225	C_{aro}^{para}	-0.1269	C_{aro}	-0.1235	C_{aro}	-0.1247
5			C_{aro}^{subs}	-0.107	C_{aro}^{meta}	-0.1191	C_{aro}	-0.1235	C_{aro}	-0.1157
6			C_{aro}	-0.1225	C_{aro}^{ortho}	-0.1203	C_{aro}	-0.1169	C_{aro}	-0.1247
7			CH_x	0.1123	CH_x	0.1129	CH_x	0.1103	CH_x	0.1126
8			H_{aro}	0.1172	H_{aro}^{ortho}	0.1199	CH_x	0.1103	H_{aro}	0.1173
9			CH_x	0.1123	H_{aro}^{meta}	0.1222	H_{aro}	0.1196	CH_x	0.1126
10	H_{aro}	0.1225	H_{aro}	0.1172	H_{aro}^{para}	0.1224	H_{aro}	0.1221	H_{aro}	0.1198
11			CH_x	0.1123	H_{aro}^{meta}	0.1222	H_{aro}	0.1221	H_{aro}	0.1219
12			H_{aro}	0.1172	H_{aro}^{ortho}	0.1199	H_{aro}	0.1196	H_{aro}	0.1198

[a] Aliphatic pseudo-atoms beyond the first site attached to the aromatic ring always have zero charge. [b] Numbering as in Figure 3

Table 2 (Continuation): Suggested partial atomic charges for improved hydration free energies, for benzene, TMB, mono-substituted *n*-alkylbenzenes[a] and di- and tri-substituted aromatics

#[b]	PX Site	q_i	*p*-ethyltoluene Site	q_i	1,2,3-TMB Site	q_i	1,2,4-TMB Site	q_i
1	C_{aro}^{subs}	-0.1182	C_{aro}^{subs}	-0.1182	C_{aro}^{subs}	-0.1082	C_{aro}^{subs}	-0.1082
2	C_{aro}	-0.1169	C_{aro}	-0.1169	C_{aro}^{subs}	-0.1094	C_{aro}^{subs}	-0.1160
3	C_{aro}	-0.1169	C_{aro}	-0.1169	C_{aro}^{subs}	-0.1082	C_{aro}	-0.1135
4	C_{aro}^{subs}	-0.1182	C_{aro}^{subs}	-0.1182	C_{aro}	-0.1213	C_{aro}	-0.1213
5	C_{aro}	-0.1169	C_{aro}	-0.1169	C_{aro}	-0.1201	C_{aro}^{subs}	-0.1148
6	C_{aro}	-0.1169	C_{aro}	-0.1169	C_{aro}	-0.1213	C_{aro}	-0.1147
7	CH_x	0.1128	CH_x	0.1128	CH_x	0.1100	CH_x	0.1100
8	H_{aro}	0.1196	H_{aro}	0.1196	CH_x	0.1077	CH_x	0.1102
9	H_{aro}	0.1196	H_{aro}	0.1196	CH_x	0.1100	H_{aro}	0.1193
10	CH_x	0.1128	CH_x	0.1128	H_{aro}	0.1195	H_{aro}	0.1195
11	H_{aro}	0.1196	H_{aro}	0.1196	H_{aro}	0.1218	CH_x	0.1125
12	H_{aro}	0.1196	H_{aro}	0.1196	H_{aro}	0.1195	H_{aro}	0.1170

[a] Aliphatic pseudo-atoms beyond the first site attached to the aromatic ring always have zero charge. [b] Numbering as in Figure 3

Table 3: Charge increments on each position of the aromatic ring caused by an alkyl substitution.

Position	Index j	δ_{CARBON}	$\delta_{HYDROGEN}$[a]
Substituted	0	+0.0087	-0.0096
ortho	1	+0.0022	-0.0026
meta	2	+0.0034	-0.0003
para	3	-0.0044	-0.0001

[a] We use this notation for simplicity, but the charge increments apply also to CH_x substituent groups.

Table 4: LJ $\left(\Delta G^{LJ}\right)$ and electrostatic contributions (ΔG^{C}) to the predicted Gibbs energy of hydration $(\Delta_{hyd}G^{sim})$ of the full test and training sets. All data in kJ/mol. The subscripts give the statistical accuracy of the last decimal point shown.

Solute	$\overline{\Delta G^{LJ}}$	ΔG^{C}	$\Delta_{hyd}G^{sim}$	$\Delta_{hyd}G^{exp}$	AAD
BZ	5.5_5	-9.13_6	-3.6_5	-3.62	0.0
MB	5.9_4	-10.15_7	-4.2_4	-3.71	
EB	7.1_2	-10.32_7	-3.3_2	-3.33	
PB	8.1_4	-10.42_7	-2.3_4	-2.23	0.4_2
BB	9.6_9	-10.62_7	-1.1_9	-1.66	
PeB	10.2_8	-10.61_7	-0.4_8	-0.96	
OX	6.6_8	-10.79_7	-4.2_8	-3.77	
MX	7.1_4	-11.22_7	-4.1_4	-3.47	0.7_3
PX	6.6_3	-10.93_7	-4.3_3	-3.35	
1,3,5-TMB	8.3_8	-12.12_7	-3.8_8	-3.77	0.0
p-ethyltoluene	7.9_3	-11.48_6	-3.6_3	-4.0	
1,2,3-TMB	7.4_2	-12.29_6	-4.9_2	-5.1	
1,2,4-TMB	6.8_1	-12.16_5	-5.5_1	-3.6	0.9_6
isopropylbenzene	8.0_3	-9.95_6	-1.9_3	-1.3	
p-isopropyltoluene	9.7_3	-11.47_6	-1.7_3	-2.8	
isobutylbenzene	9.3_3	-9.89_6	-0.6_3	0.7	
				Overall	**0.58**
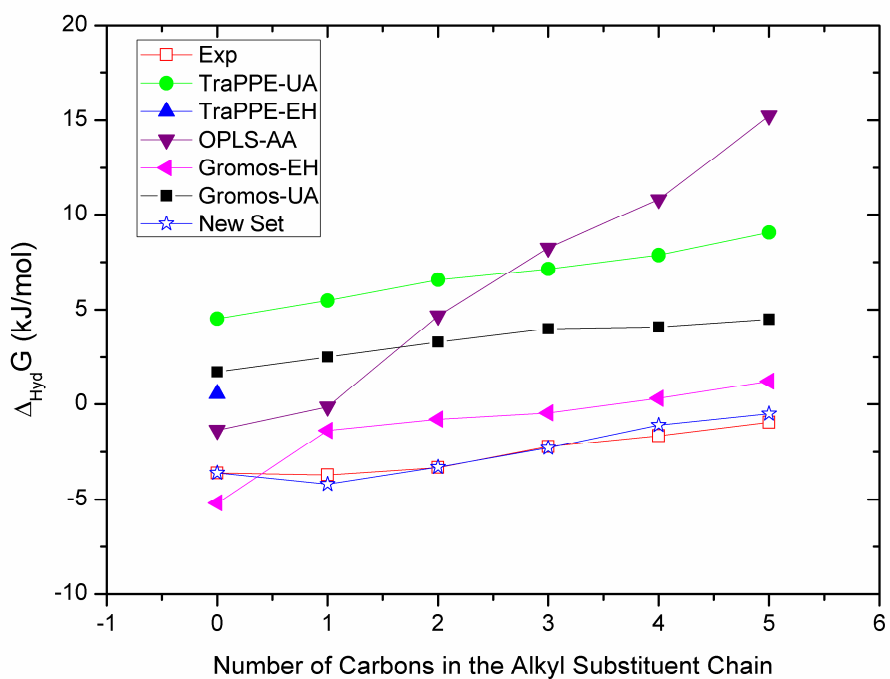
Figure 1: Comparison of the performance of existing force fields in predicting the Gibbs energy of hydration for different mono-substituted *n*-alkylbenzenes together with the prediction obtained by the new proposed charge set.

Solute $\xrightarrow{\Delta_{Hyd}G^{Exp}}$ Solute

("Experimental", Unpolarized, Gas)   ("Experimental", Polarized, Solvated)

$\downarrow \Delta G^{G}_{Sol}$ $\uparrow \Delta G^{L}_{Pol}$

Solute $\xrightarrow{\Delta_{Sim}G^{Exp}}$ Solute

("Model", Polarized, Gas)   ("Model", Polarized, Solvated)
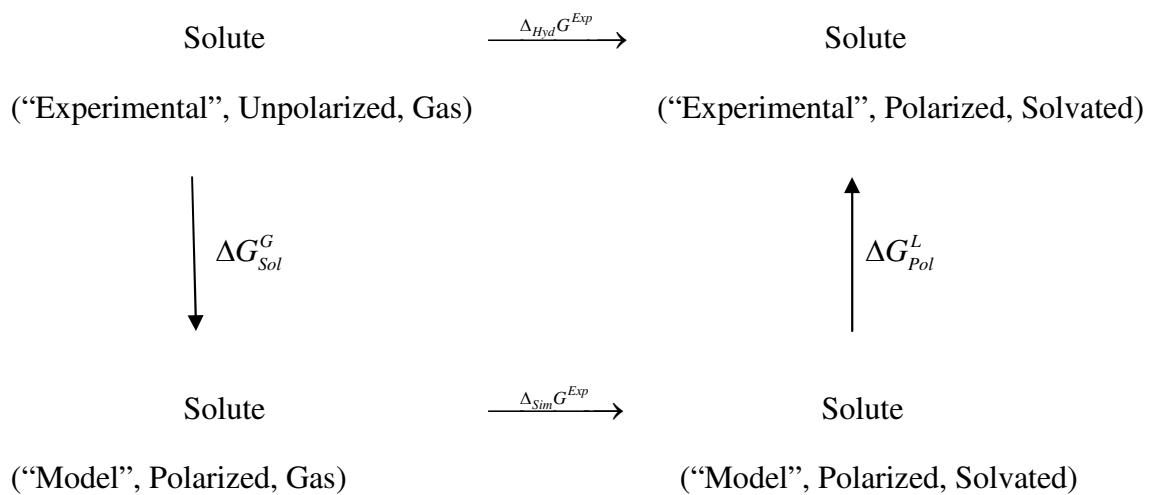
Figure 2: Thermodynamic Cycle
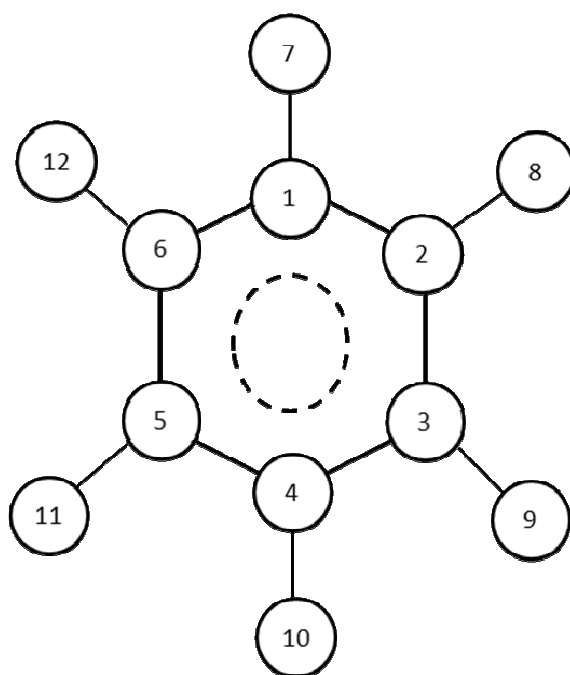
Figure 3: Atom numbering used for point charge assignment to the alkylaromatic solutes (Table 2). Notice that positions 7 to 12 may correspond to hydrogen atoms or to the first pseudo atoms of alkyl chains (the remaining pseudo-atoms are not shown because they are always neutral).
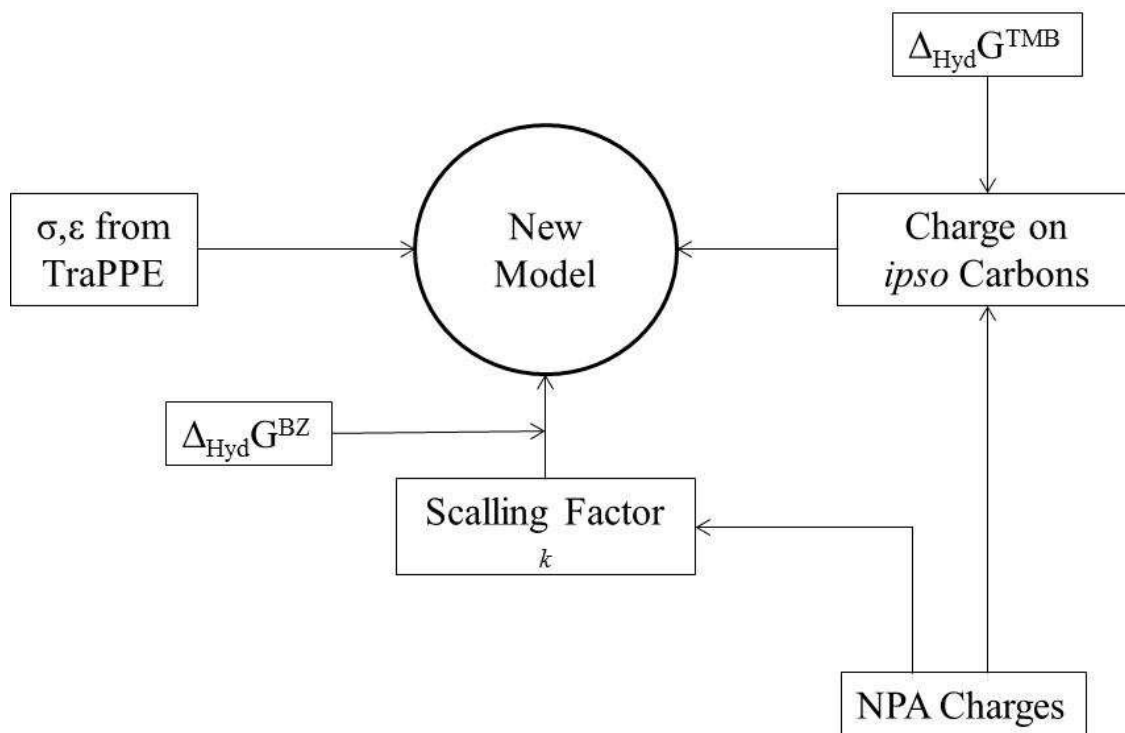
Figure 4: General procedure followed to derive the new charge set and different quantities used as input data to the development of the new model.
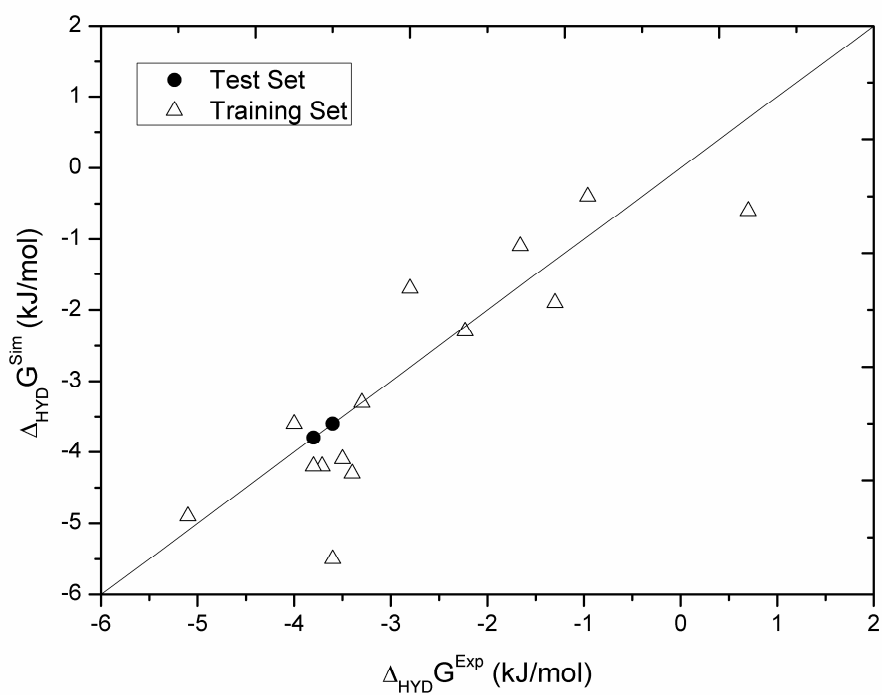
Figure 5: Comparison between experimental and predicted Gibbs energy of hydration for all the compounds under study using the proposed new force field. Test set refers to benzene and 1,3,5-trimethylbenzene and training set to the remaining solutes.

Supporting Information:

Predicting Hydration Gibbs Energies of Alkyl-aromatics Using Molecular Simulation: A Comparison of Current Force Fields and the Development of a New Parameter Set for Accurate Solvation Data

Nuno M. Garrido[1], Miguel Jorge[1], António J. Queimada[1], José R. B. Gomes[2],

Ioannis G. Economou[3] and Eugénia A. Macedo[1,*]

1. LSRE Laboratory of Separation and Reaction Engineering, Departamento de Engenharia Química, Faculdade de Engenharia, Universidade do Porto, Rua do Dr. Roberto Frias, 4200 - 465 Porto, Portugal

2. CICECO, Department of Chemistry, University of Aveiro, Campus Universitário de Santiago, Aveiro 3810-193, Portugal

3. The Petroleum Institute, Department of Chemical Engineering, PO Box 2533, Abu Dhabi, United Arab Emirates

[*]Author to whom all correspondence should be addressed at: eamacedo@fe.up.pt.

Table S1: LJ and electrostatic contributions to the predicted hydration Gibbs energies of benzene and linear alkylbenzenes using the TraPPE-UA force field. All data in kJ/mol. The subscripts give the statistical accuracy of the last decimal point shown. Abbreviations as detailed in the main manuscript.

Solute	ΔG^{LJ}	ΔG^{C}	$\Delta_{hyd} G^{sim}$	$\Delta_{hyd} G^{exp}$
BZ	4.5_5	0	4.5_5	-3.62
MB	5.5_6	0	5.5_6	-3.71
EB	6.6_6	0	6.6_6	-3.33
PB	7.1_{10}	0	7.1_{10}	-2.23
BB	7.9_9	0	7.9_9	-1.66
PeB	9.1_6	0	9.1_6	-0.96
			AAD	**9.4_6**

Solute	ΔG^{LJ}	ΔG^{C}	$\Delta_{hyd} G^{sim}$	$\Delta_{hyd} G^{exp}$
BZ*	5.5_5	-5.04_9	0.5_5	-3.62
			AAD	**4.2**

* Considering an explicit hydrogen description of the aromatic ring.

Table S2: LJ and electrostatic contributions to the predicted hydration Gibbs energies of benzene and linear alkylbenzenes using the OPLS-AA force field. All data in kJ/mol. The subscripts give the statistical accuracy of the last decimal point shown.

Solute	ΔG^{LJ}	ΔG^{C}	$\Delta_{hyd} G^{sim}$	$\Delta_{hyd} G^{exp}$
BZ	6.8_2	-8.2_1	-1.4_2	-3.62
MB	12.1_5	-12.24_8	-0.1_5	-3.71
EB	16.4_4	-11.70_8	4.7_4	-3.33
PB	19.4_4	-11.17_8	8.2_4	-2.23
BB	22.0_5	-11.18_8	10.8_5	-1.66
PeB	26.3_2	-11.09_8	15.2_2	-0.96
			AAD	**8.8 ± 4.9**

Table S3: LJ and electrostatic contributions to the predicted hydration Gibbs energies of benzene and linear alkylbenzenes using the Gromos-EH force field. All data in kJ/mol. The subscripts give the statistical accuracy of the last decimal point shown.

Solute	ΔG^{LJ}	ΔG^{C}	$\Delta_{hyd} G^{sim}$	$\Delta_{hyd} G^{exp}$
BZ	7.6_5	-12.66_8	-5.2_5	-3.62
MB	7.9_4	-9.23_7	-1.4_4	-3.71
EB	8.7_6	-9.53_7	-0.8_6	-3.33
PB	9.0_6	-9.44_7	-0.5_6	-2.23
BB	9.9_7	-9.54_7	0.3_7	-1.66
PeB	10.7_8	-9.49_7	1.2_8	-0.96
			AAD	**2.1 ± 0.3**

Table S4: LJ and electrostatic contributions to the predicted hydration Gibbs energies of benzene and linear alkylbenzenes using the Gromos-UA force field. All data in kJ/mol. The subscripts give the statistical accuracy of the last decimal point shown.

Solute	ΔG^{LJ}	ΔG^{C}	$\Delta_{hyd} G^{sim}$	$\Delta_{hyd} G^{exp}$
BZ	1.7_6	0	1.7_6	-3.62
MB	2.5_3	0	2.5_3	-3.71
EB	3.3_3	0	3.3_3	-3.33
PB	4.0_3	0	4.0_3	-2.23
BB	4.1_8	0	4.1_8	-1.66
PeB	4.5_9	0	4.5_9	-0.96
			AAD	**5.9** ± 0.4

Tables S5-S12: CHelpG and NPA charges for the different molecules studied

- **Table S5:** Benzene (BZ)

Atom	CHelpG	NPA
C_1	(4x) -0.062	-0.245
C_2	(2x) -0.103	
H_1	(4x) 0.072	0.245
H_2	(2x) 0.085	

- **Table S6:** Toluene (MB)
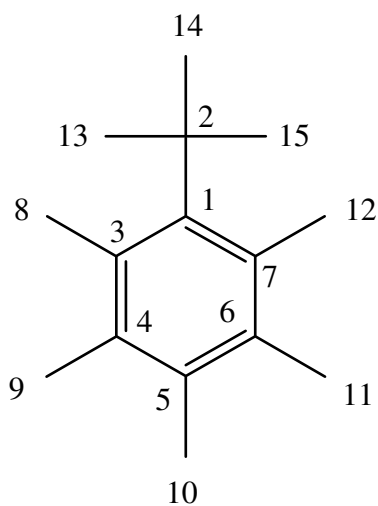
Atom#	CHelpG	NPA	Atom#	CHelpG	NPA	Atom#	CHelpG	NPA
1	0.212	-0.042	6	-0.026	-0.238	11	0.072	0.245
2	-0.223	-0.709	7	-0.203	-0.241	12	0.098	0.240
3	-0.212	-0.240	8	0.109	0.240	13	0.066	0.245
4	-0.020	-0.238	9	0.069	0.245	14	0.059	0.252
5	-0.155	-0.254	10	0.094	0.245	15	0.059	0.252

- **Table S7:** Ethylbenzene (EB)
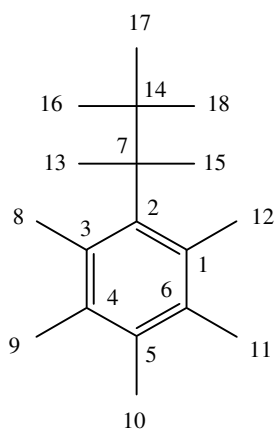
Atom#	CHelpG	NPA	Atom#	CHelpG	NPA	Atom#	CHelpG	NPA
1	-0.167	-0.239	7	0.033	-0.488	13	-0.016	0.247
2	0.093	-0.036	8	0.105	0.240	14	0.164	-0.690
3	-0.167	-0.239	9	0.075	0.245	15	-0.016	0.246
4	-0.063	-0.239	10	0.080	0.245	16	-0.035	0.239
5	-0.096	-0.253	11	0.075	0.245	17	-0.054	0.239
6	-0.063	-0.239	12	0.105	0.245	18	-0.054	0.239

- **Table S8:** Propylbenzene (PB)
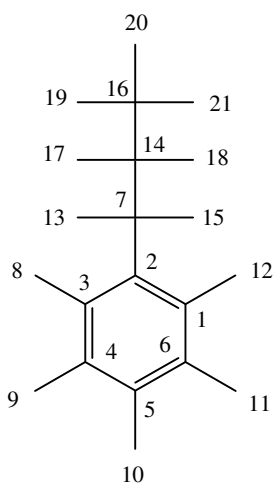
Atom#	CHelpG	NPA	Atom#	CHelpG	NPA	Atom#	CHelpG	NPA
1	-0.134	-0.239	8	0.099	0.240	15	0.022	0.245
2	0.096	-0.035	9	0.081	0.245	16	-0.137	-0.699
3	-0.134	-0.239	10	0.079	0.245	17	-0.144	0.239
4	-0.091	-0.239	11	0.080	0.245	18	-0.114	0.239
5	-0.079	-0.253	12	0.099	0.240	19	0.013	0.244
6	-0.091	-0.239	13	0.022	0.245	20	0.015	0.235
7	-0.155	-0.480	14	0.429	-0.472	21	0.015	0.235

- **Table S9:** Ortho-xylene (OX)
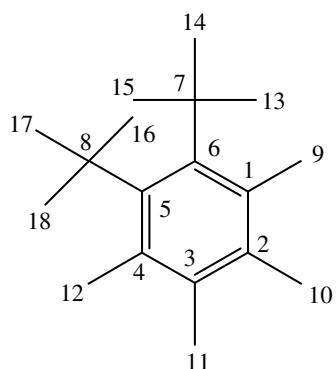
Atom#	CHelpG	NPA	Atom#	CHelpG	NPA	Atom#	CHelpG	NPA
1	-0.210	-0.233	7	-0.176	-0.708	13	0.051	0.247
2	-0.071	-0.247	8	-0.176	-0.708	14	0.054	0.249
3	-0.071	-0.247	9	0.114	0.239	15	0.054	0.249
4	-0.210	-0.233	10	0.077	0.244	16	0.051	0.247
5	0.106	-0.039	11	0.077	0.244	17	0.054	0.249
6	0.106	-0.039	12	0.114	0.239	18	0.054	0.249

- **Table S10:** Meta-xylene (MX)
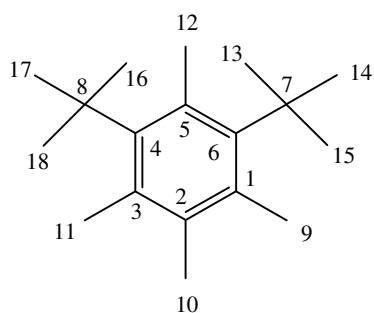
Atom#	CHelpG	NPA	Atom#	CHelpG	NPA	Atom#	CHelpG	NPA
1	-0.247	-0.250	7	-0.149	-0.708	13	0.042	0.248
2	-0.009	-0.231	8	-0.130	-0.708	14	0.041	0.246
3	-0.250	-0.250	9	0.115	0.240	15	0.044	0.253
4	0.215	-0.035	10	0.074	0.244	16	0.035	0.247
5	-0.327	-0.237	11	0.115	0.240	17	0.040	0.253
6	0.22	-0.035	12	0.134	0.235	18	0.037	0.247

- **Table S11:** Para-xylene (PX)
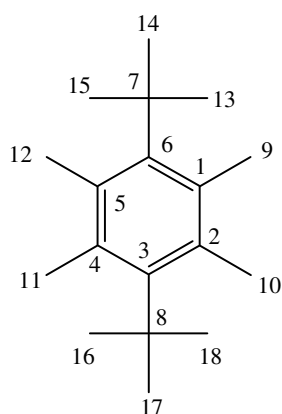
Atom#	CHelpG	NPA	Atom#	CHelpG	NPA	Atom#	CHelpG	NPA
1	-0.173	-0.233	7	-0.172	-0.707	13	0.048	0.247
2	-0.179	-0.233	8	-0.165	-0.707	14	0.050	0.252
3	0.166	-0.051	9	0.105	0.239	15	0.048	0.247
4	-0.179	-0.233	10	0.106	0.239	16	0.046	0.247
5	-0.171	-0.233	11	0.106	0.239	17	0.049	0.252
6	0.162	-0.051	12	0.105	0.239	18	0.046	0.247

- **Table S12:** 1,3,5-Trimethylbenzene (TMB)
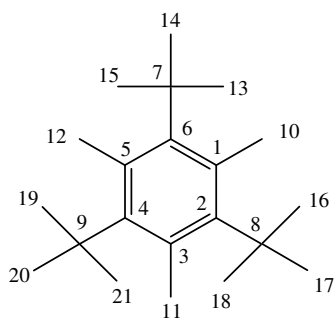
Atom#	CHelpG	NPA	Atom#	CHelpG	NPA	Atom#	CHelpG	NPA
1	-0.379	-0.245	8	-0.171	-0.707	15	0.042	0.250
2	0.263	-0.025	9	-0.184	-0.707	16	0.047	0.250
3	-0.379	-0.245	10	0.145	0.234	17	0.049	0.251
4	0.253	-0.025	11	0.146	0.234	18	0.046	0.244
5	-0.371	-0.245	12	0.142	0.234	19	0.050	0.244
6	0.256	-0.025	13	0.042	0.244	20	0.052	0.251
7	-0.154	-0.707	14	0.044	0.251	21	0.051	0.251

Table S13: Prediction of liquid densities (g/l) at P = 1 bar using the new parameters set.

Solute	T/K	ρ^{calc}	AAD (%)	T/K	ρ^{calc}	AAD (%)	T/K	ρ^{calc}	AAD (%)
BZ	313	900.89	5.0	294	920.81	5.0	347	864.59	5.1
MB	311	873.63	2.5	293	891.78	2.6	345	840.36	2.5
EB	311	862.77	1.2	292	879.66	1.1	344	832.81	1.4
PB	309	861.21	1.4	291	876.55	1.3	343	832.87	1.6
BB	309	861.71	1.5	290	876.61	1.5	342	834.94	1.7
PeB	308	858.97	1.4	289	873.41	1.4	341	833.22	1.3
OX	309	879.67	1.5	291	896.16	1.6	343	850.61	1.4
MX	309	864.05	1.5	291	880.09	1.6	342	835.42	1.6
PX	309	860.23	1.5	291	875.86	1.5	342	831.99	1.6
TMB	308	874.50	2.4	289	889.19	2.3	340	847.92	2.6
		Overall	2.0		*Overall*	2.0		*Overall*	2.1

Table S14: Prediction of vaporization enthalpies (kJ/mol) at P = 1 bar using the new

parameters set.

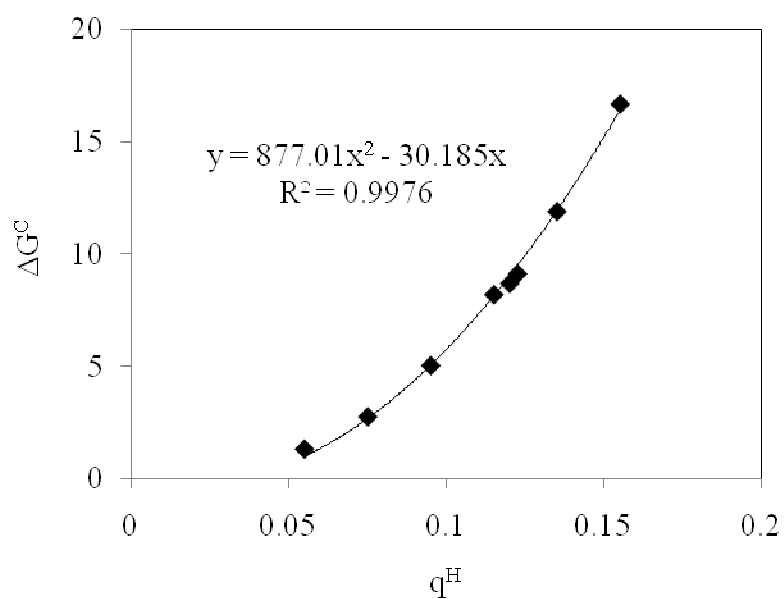| Solute	T/K	$\Delta_{vap}H^{exp}$	E^{Gas}	E^{Liq}	$\Delta_{vap}H^{sim}$	AAD (kJ/mol)
BZ	313	33.0	102.4	69.1	35.9	2.9
MB	311	37.4	100.0	64.9	37.7	0.3
EB	311	41.5	121.8	83.8	40.6	1.0
PB	309	46.6	129.5	87.2	44.9	1.7
BB	309	49.9	137.9	91.0	49.5	0.4
PeB	308	54.8	145.9	94.9	53.7	1.2
OX	309	42.9	97.2	58.6	41.2	1.7
MX	309	42.2	97.2	59.1	40.6	1.6
PX	309	41.7	97.2	59.4	40.4	1.3
TMB	308	46.9	94.6	52.6	44.6	2.3
					Overall	1.4

Figure S1: Correlation between ΔG^C (kJ/mol) and aromatic hydrogen point charges for BZ (the charge on the aromatic carbon is always symmetric to the hydrogen charge).
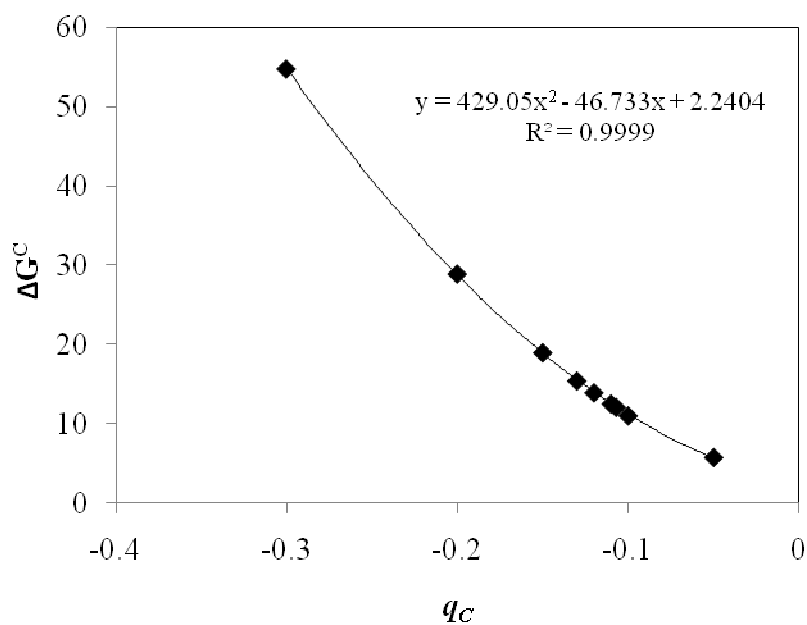

Figure S2: Correlations between ΔG^C (kJ/mol) for TMB and different point charges on the substituted aromatic carbon atom (remaining charges are kept equal to their corresponding values in BZ).
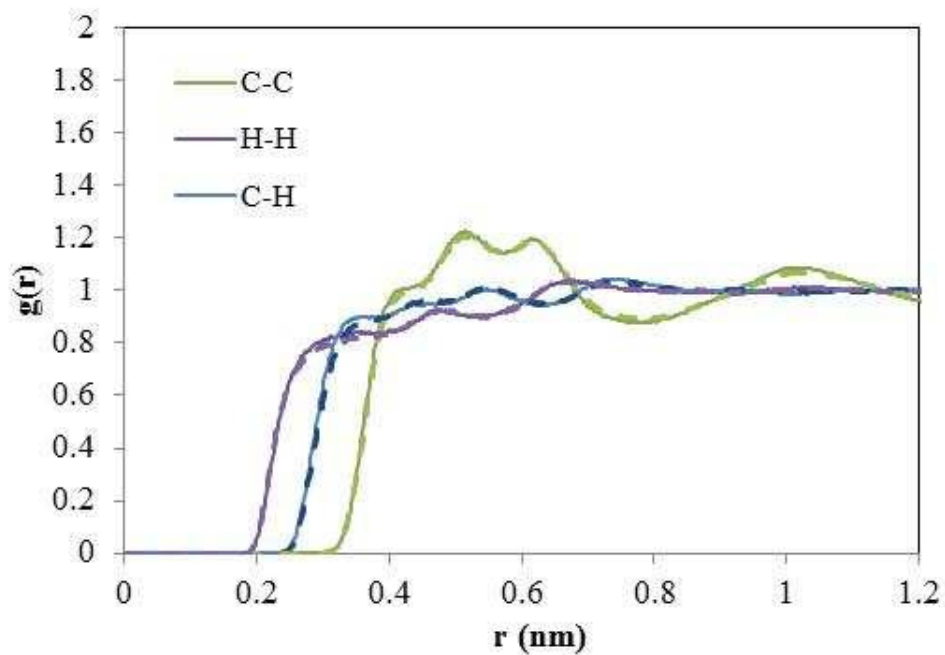
Figure S3: Benzene liquid structure: computed $C_{aro} - C_{aro}$ (aromatic carbon – aromatic

carbon), $H_{aro} - H_{aro}$ (aromatic hydrogen – aromatic hydrogen) and $C_{aro} - H_{aro}$

(aromatic carbon – aromatic hydrogen) radial distribution functions at 298 K using the

new charge set (solid lines) and the original TraPPE-EH parameters (dashed lines).

Tutorial: How to assign charges for **1,2,4-TMB** using the rule:

1) Assign each site (Carbon, Hydrogen or CH_x) to its corresponding position in the diagram of Figure 2.

2) For each site, determine the total number of substituents (N_j) on each position j (0 for the current C/H atoms, 1 for C/H atoms in ortho position, 2 for C/H atoms in meta position, and 3 for C/H atoms in para position). Recall that the maximum value of N is 1 for $j=0,3$ and 2 for $j=1,2$.

3) Compute the charge on each site by applying equation (4) and the charge increments of Table 4.

The table below shows the number of substituents and the total charge on each site of the 1,2,4-TMB molecule.
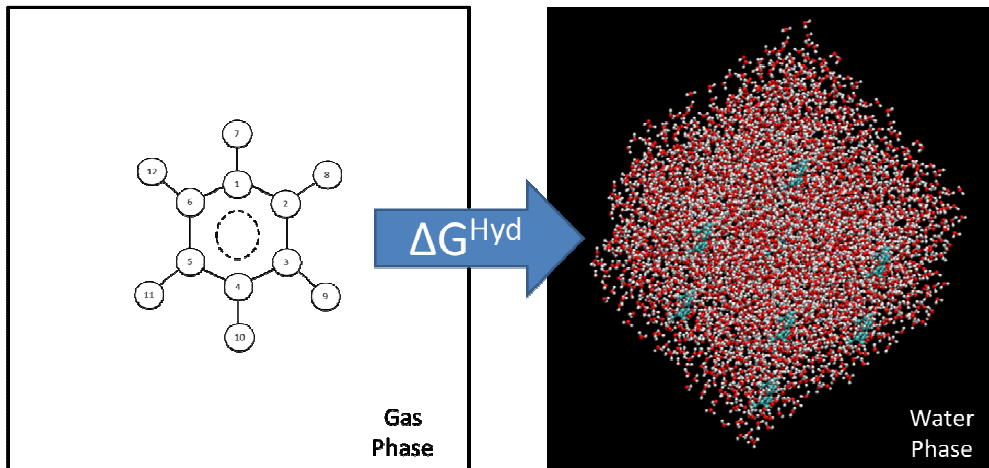
Atom	N_0	N_1	N_2	N_3	q
C_1	1	1	1	0	-0.1082
C_2	1	1	0	1	-0.1160
C_3	0	1	2	0	-0.1135
C_4	0	1	1	1	-0.1213
C_5	1	0	1	1	-0.1148
C_6	0	2	1	0	-0.1147
CH_7	1	1	1	0	0.1100
CH_8	1	1	0	1	0.1102

H_9	0	1	2	0	0.1193
H_{10}	0	1	1	1	0.1195
CH_{11}	1	0	1	1	0.1125
H_{12}	0	2	1	0	0.117
					$\sum = 0.0000$

Below are two examples of application of equation (4) to calculate the charge on the carbon atom at position 1 and on the CH_x pseudo-atom at position 7.

Example: $q_{C_1} = -0.1225 + 1*0.0087 + 1*0.0022 + 1*0.0034 - 0*0.0044 = -0.1082$

$q_{CH_7} = 0.1225 - 1*0.0096 - 1*0.0026 - 1*0.0003 - 0*0.0001 = 0.1100$

Caption: The Gibbs energy of hydration can be seen as the total reversible work required to transfer a solute molecule from the ideal gas phase to water