

# Predicting the Solubility of the Sparingly Soluble Solids 1,2,4,5-Tetramethylbenzene, Phenanthrene, and Fluorene in Various Organic Solvents by Molecular Simulation

Andrew S. Paluch, Dan D. Cryan, III, and Edward J. Maginn\*

Department of Chemical and Biomolecular Engineering, University of Notre Dame, 182 Fitzpatrick Hall, Notre Dame, Indiana 46556-5637, United States

**S** Supporting Information

**ABSTRACT:** We present a simple method to estimate the solubility of weakly soluble solids in different solvents. The method involves the calculation of the residual chemical potential of a single solute molecule in the solvent of interest using an appropriate atomistic free energy simulation technique. In the present method, an expanded ensemble calculation is used, along with a combined Wang–Landau/Bennett's acceptance ratio method. To avoid the simulation of the solid phase and the use of analytic reference states, a single experimental solubility data point for the solute in a single reference solvent is also required. The method has advantages over more empirical descriptor-based methods in that the simulations enable insight into the underlying molecular driving forces responsible for solubility trends. Results are presented for the solubility of 1,2,4,5-tetramethylbenzene, phenanthrene, and fluorene in the solvents hexane, octane, ethanol, 1-octanol, cyclohexane, benzene, and 1,4-dioxane. Overall, agreement between the results of the present study and available experimental data is good. In general, the predictions have a lower total absolute error when compared to experimental solubility data than those obtained using the Abraham general solvation model.

## INTRODUCTION

Knowledge of the solubility behavior of organic solids in different solvents is of substantial interest. For instance, in the formulation of pharmaceuticals, a molecule that has a promising activity but is poorly soluble in water may end up being abandoned or may require substantial modification. Advances in high throughput synthesis and testing, combined with the desire to have drugs be more sensitive in their binding affinity, have resulted in an increase in the molecular weight and lipophilicity of drug molecule candidates, with a resulting decrease in aqueous solubility and hence bioavailability.<sup>1,2</sup> Often times, multiple cosolvents may be mixed to satisfy pharmaceutical requirements of solubility, toxicity, and price. It follows that the ability to predict which solvents will be most effective in dissolving the solute would be exceedingly useful.

As a result of the vast possibility of solvents and chemical composition of the candidate pharmaceutical, along with the time and expense associated with experimental solubility measurements, it is desirable to have methods that can predict the ability of a given solvent to dissolve a particular solute. Because solvation is a complex phenomenon in which many different competing forces interact to determine the behavior of a given solute–solvent system, the development of such methods is extremely challenging. An astounding amount of research has been done in this area, but it is still generally the case that the prediction of solvation behavior relies upon semiempirical correlations and group contribution methods.<sup>2–9</sup> These methods often require a substantial amount of experimental data to make meaningful predictions. Given this situation, it is desirable to have atomistic-based molecular dynamics and Monte Carlo models that can predict solubility with little or no experimental

data as input. Not only would such methods provide guidance in solvent selection, but atomistic models give insight into the molecular level details governing solubility behavior.

Only a limited number of molecular modeling approaches have been used to predict the solubility limit of solids.<sup>10–13</sup> In some cases, atomistic simulations have been performed to compute parameters for empirical correlations,<sup>10</sup> while in others, simulations have been used to directly compute the solubility limit of solutes in solvents.<sup>11–13</sup> In the latter case, the phase equilibria condition requires equality of the temperature, pressure, and chemical potential of each species between the phases in equilibrium. Computing the chemical potential in such systems is a nontrivial task. In addition, these methods require knowledge of the experimental crystal structure of the solute and the ability to accurately model the solute in both the solid and the condensed phases. While many force fields are optimized for performance in the liquid phase,<sup>14</sup> extending these force fields to model crystalline solids often leads to unsatisfactory results.<sup>15,16</sup> It follows that the ability to avoid modeling the solid phase altogether would prove advantageous.

In this work, a procedure for estimating the solubility limit of large, sparingly soluble solutes in a variety of solvents is presented. A single experimental reference data point is used for each solute in conjunction with an atomistic free energy simulation of the solute in the solvents of interest. In this manner,

**Special Issue:** John M. Prausnitz Festschrift

**Received:** November 19, 2010

**Accepted:** January 6, 2011

**Published:** March 02, 2011

simulations of the crystalline solid are avoided altogether. Solubility predictions may be made in any solvent of interest, given that a reasonable force field for the solvent exists. Furthermore, any combination of cosolvents may be studied without increasing the computational difficulty of the prediction. If the solubility limit is known in a pure solvent or at a given cosolvent concentration, the method allows one to readily predict the influence of adding or removing additional cosolvent on the solubility limit.

## METHODOLOGY

**Solid–Liquid Equilibrium.** The solubility of solid substances in pure and mixed solvents is described by the classical equations of solid–liquid equilibrium.<sup>3–5</sup> Consider the case of a pure solid solute “A” dissolved in a solvent. The chemical potential of species A in the liquid phase (“L”) at a temperature  $T$  and pressure  $P$  is

$$\mu_A^L(T, P, x_A) = \mu_A^0(T, P) + RT \ln \frac{f_A^L(T, P, x_A)}{f_A^0(T, P)} \quad (1)$$

where  $\mu_A^0$  and  $f_A^0$  are the reference state chemical potential and fugacity, respectively, and  $f_A^L(T, P, x_A)$  is the fugacity of the solute in the liquid solution at the solubility limit mole fraction  $x_A$ . The chemical potential of the pure solid phase (“S”) of A in equilibrium with the solution is

$$\mu_A^S(T, P) = \mu_A^0(T, P) + RT \ln \frac{f_A^S(T, P)}{f_A^0(T, P)} \quad (2)$$

where  $f_A^S(T, P)$  is the fugacity of the pure solid at  $T$  and  $P$ , and it is assumed that no solvent dissolves into the crystalline solute.

The phase equilibrium condition is

$$f_A^S(T, P) = f_A^L(T, P, x_A) \quad (3)$$

The liquid phase fugacity can be written in terms of the reference fugacity and the appropriate activity coefficient  $\gamma_A$  as

$$f_A^L(T, P, x_A) = \gamma_A(T, P, x_A) x_A f_A^0(T, P) \quad (4)$$

Substituting eq 4 into eq 3, the limiting solubility is given by

$$x_A = \frac{f_A^S(T, P)}{\gamma_A(T, P, x_A) f_A^0(T, P)} \quad (5)$$

Equation 5 is exact given the assumption of a pure solid phase and requires knowledge of the pure solid fugacity and the activity coefficient of the solute in the solution phase relative to the (arbitrary) reference state. These quantities can be computed from a molecular simulation to estimate the limiting solubility, as has been shown in previous work.<sup>11–13</sup> To do so, however, requires that the free energy of the solid phase be computed, which besides being difficult requires knowledge of the solid phase structure. In addition, the composition dependence of the activity coefficient must be found by performing a series of calculations at varying solute compositions. For low solubility substances, we make the assumption that the activity coefficient is relatively insensitive to the solute concentration up to the (low) solubility limit. Specifically, we assume that

$$\gamma_A(T, P, x_A) \approx \gamma_A(T, P, x_A^*) \quad (6)$$

where  $x_A^*$  corresponds to the solute concentration associated with a single solute molecule in a simulation box of solvent molecules.

This assumption is analogous to that made in similar studies.<sup>7,8</sup> From the definition of the residual chemical potential and activity coefficient

$$RT \ln \gamma_A(T, P, x_A^*) = \mu_A^{\text{res}}(T, P, x_A^*) - \mu_{A, \text{pure}}^{\text{res}}(T, P) \quad (7)$$

the solubility limit can be written as

$$\ln x_A = -\frac{1}{RT} \mu_A^{\text{res}}(T, P, x_A^*) + \frac{1}{RT} \mu_{A, \text{pure}}^{\text{res}}(T, P) + \ln \left( \frac{f_A^S(T, P)}{f_A^0(T, P)} \right) \quad (8)$$

Recognizing that the last two terms are pure component properties of A, the limiting solubility can be expressed as

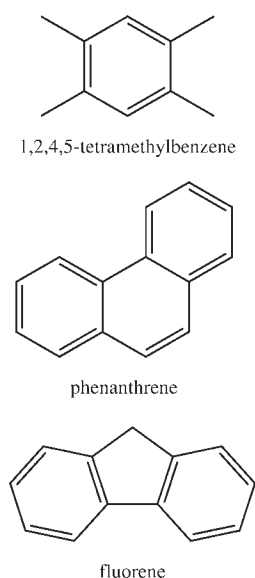
$$\ln x_A = -\frac{1}{RT} \mu_A^{\text{res}}(T, P, x_A^*) + \zeta_A(T, P) \quad (9)$$

where  $\zeta_A$  is a constant parameter for a given solute (regardless of the solvent) and is set by a single experimental solubility data point. The term  $\mu_A^{\text{res}}$  is readily obtained from a single molecular simulation at concentration  $x_A^*$ , as discussed below. Equation 9 is the main result and will be used to estimate the solubilities of the three solutes in a range of different solvents.

## COMPUTATIONAL DETAILS

**Computing the Residual Chemical Potential.** An expanded ensemble (EE) procedure was used to compute  $\mu_A^{\text{res}}$ . Details of this method are provided elsewhere,<sup>17</sup> so only a brief summary of the essential concepts is provided here. The basic idea behind the EE method<sup>18–21</sup> is to construct an augmented ensemble as a sum of  $M + 1$  subensembles. This series of subensembles connects two systems of interest by gradually performing transitions between the two systems. In the present work, the systems of interest are a noninteracting solute molecule in a pure solvent and a single fully interacting solute in the solvent, with both states at the same temperature and pressure. A single solute in the solvent corresponds to a solute mole fraction  $x_A^*$ . The free energy difference between these two systems gives  $\mu_A^{\text{res}}$ . The intermediate subensembles between the noninteracting and the fully interacting solute subensembles serve to scale the intermolecular interaction potential of the solute. A specific subensemble is designated by index  $m$ , while the intermolecular Lennard–Jones (LJ) and electrostatic interactions are regulated by the subensemble dependent coupling parameters  $\lambda_m^{\text{LJ}}$  and  $\lambda_m^{\text{elec}}$ , respectively. These coupling parameters vary from  $0 \leq \lambda_m^{\text{LJ}} \leq 1$  and  $0 \leq \lambda_m^{\text{elec}} \leq 1$ .

While within a given subensemble, the configurational phase space is sampled by molecular dynamics within the isothermal–isobaric (NPT) ensemble. Periodically, a stochastic transition to an adjacent subensemble is attempted. These transitions are accepted using an appropriate acceptance rule.<sup>17</sup> As a result of the bulky size of the solute molecules, the free energy difference between adjacent subensembles tends to be large, and hence the probability of accepting transitions is small. To increase the acceptance probability of these moves, a biasing scheme that utilizes a combined Wang–Landau (WL)<sup>22–24</sup> and Bennett’s acceptance ratio (BAR)<sup>25,26</sup> method is used. The difference in free energy between the two end states at  $m = 0$  and  $m = M$  (and hence  $\mu_A^{\text{res}}$  in eq 9) is determined using the BAR procedure. Complete details can be found elsewhere.<sup>17</sup>



**Figure 1.** Chemical structures of the studied solutes.

**Force Fields.** For the solvents and solutes, nonbonded intermolecular interactions were treated using a combined LJ and fixed point charge model of the form

$$U_{\text{nb}}(r_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (10)$$

where  $r_{ij}$ ,  $\epsilon_{ij}$ ,  $\sigma_{ij}$ ,  $q_i$ , and  $q_j$  are the site separation distance between atoms  $i$  and  $j$ , well-depth of the LJ interaction, distance at which the LJ interaction is zero, and partial charge values, respectively. For interactions between unlike LJ sites, Lorentz–Berthelot<sup>27</sup> combining rules were employed, in which the LJ size parameter was obtained as an arithmetic mean

$$\sigma_{ij} = \frac{1}{2} (\sigma_{ii} + \sigma_{jj}) \quad (11)$$

and the LJ well-depth was obtained as a geometric mean

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad (12)$$

LJ parameters and partial charges for hexane, octane, ethanol, 1-octanol, and cyclohexane were taken from the united-atom transferable potential for phase equilibria (TraPPE-UA) force field.<sup>28–30</sup> Parameters for benzene were taken from the explicit hydrogen analog (TraPPE-EH).<sup>31</sup> The model for 1,4-dioxane was taken from the work of Yazaydin and Thompson.<sup>32</sup>

The solutes studied were 1,2,4,5-tetramethylbenzene, phenanthrene, and fluorene. Structures of these solutes are given in Figure 1. These solutes were chosen because experimental solubility data exists for these species in the solvents studied here, they are large cyclic compounds representative of drug-like molecules, and they have relatively low solubilities in the studied solvents. LJ parameters for the solute molecules were taken from the general AMBER force field (GAFF).<sup>33,34</sup> GAFF has been shown to perform well in simulations of hydration free energies,<sup>35–37</sup> and it is used extensively to model pharmaceutical compounds.

To obtain partial charges for the solutes, geometries were first optimized at the B3LYP/cc-pVTZ level of theory, followed by single point energy calculations at the HF/6-31G\* level of

theory<sup>38</sup> as suggested previously.<sup>33,35,39,40</sup> All of the ab initio calculations were performed with Gaussian 09.<sup>41</sup> Partial charges were then obtained from the electrostatic potential using the restrained electrostatic potential (RESP)<sup>42,43</sup> method in ANTECHAMBER (part of the AMBER 11 simulation suite).<sup>44,45</sup>

To prevent instabilities in the trajectory when the solute was nearly decoupled from the system, that is when  $\lambda_m^{\text{LJ}} \approx 0$ , solute–solvent intermolecular nonbonded LJ interactions were modeled with a modified, “soft-core” potential,  $U_{\text{LJ}}^{\text{sc}}$  of the form<sup>46–48</sup>

$$U_{\text{LJ}}^{\text{sc}}(r_{ij}; m) = 4\lambda_m^{\text{LJ}} \epsilon_{ij} \left\{ \frac{\sigma_{ij}^{12}}{[(1 - \lambda_m^{\text{LJ}})\alpha_{\text{LJ}}\sigma_{ij}^6 + r_{ij}^6]^2} - \frac{\sigma_{ij}^6}{[(1 - \lambda_m^{\text{LJ}})\alpha_{\text{LJ}}\sigma_{ij}^6 + r_{ij}^6]} \right\} \quad (13)$$

where  $r_{ij}$ ,  $\epsilon_{ij}$ , and  $\sigma_{ij}$  are the same LJ parameters as in eq 10,  $\lambda_m^{\text{LJ}}$  is the subensemble dependent coupling strength of the LJ potential, and  $\alpha_{\text{LJ}}$  is a constant, taken in this study to be 1/2. When the solute is fully coupled to the system,  $\lambda_m^{\text{LJ}} = 1$ , and eq 13 reduces to the normal LJ potential given by eq 10. When the solute is nearly decoupled,  $\lambda_m^{\text{LJ}}$  approaches 0, and eq 13 becomes a smooth interaction function that allows solvent molecules to overlap the solute with finite energy. When the solute is decoupled from the system,  $\lambda_m^{\text{LJ}} = 0$  and the solute has no interaction with the solvent (but it maintains its intramolecular potential). Thus, the potential form in eq 13 correctly represents the limiting behavior of the solute–solvent interactions, while eliminating instabilities when  $\lambda_m^{\text{LJ}} \rightarrow 0$ . Electrostatic intermolecular interactions are decoupled in a linear fashion via the coupling parameter  $\lambda_m^{\text{elec}}$ ; a detailed description regarding the decoupling of intermolecular interactions with Ewald summation may be found elsewhere.<sup>17</sup>

The same standard LJ interaction potential (eq 10) and combining rules were used for all intramolecular nonbonded interactions by all pairs of atoms separated by four or more bonds. For the case in which the intramolecular sites are separated by exactly three bonds, the LJ and electrostatic interactions were scaled by factors of 1/2 and 5/6, respectively, for the solutes and benzene. This was done to be consistent with the AMBER formalism. For 1,4-dioxane, the scaling factors for LJ and electrostatic interactions were both 1/2.<sup>32</sup> For the TraPPE-UA solvents, no scaling was used.

All of the molecules were modeled with flexible bonds, angles, and dihedral angles. The bond stretching and angle bending intramolecular interaction between sites separated by one and two bonds respectively were modeled by simple harmonic potentials of the form

$$U_{\text{bond}}(r_{ij}) = k_{ij}(r_{ij} - r_{ij}^0)^2 \quad (14)$$

and

$$U_{\text{angle}}(\theta_{ijk}) = k_{ijk}(\theta_{ijk} - \theta_{ijk}^0)^2 \quad (15)$$

where  $k_{ij}$ ,  $r_{ij}$ , and  $r_{ij}^0$  are the force constant, distance between sites  $i$  and  $j$ , and the corresponding nominal bond length, respectively. Likewise,  $k_{ijk}$ ,  $\theta_{ijk}$ , and  $\theta_{ijk}^0$  are the force constant, angle between sites  $i$ ,  $j$ , and  $k$ , and the corresponding nominal bond angle, respectively. The torsional potential describing the intramolecular interaction between sites separated by three bonds was



modeled by a potential of the form

$$U_{\text{tors}}(\phi_{ijkl}) = \sum_{n=0}^5 K_n \cos^n(\phi_{ijkl} - 180^\circ) \quad (16)$$

where  $\phi_{ijkl}$  is the dihedral angle between sites  $i, j, k$ , and  $l$ , and the  $K_n$  coefficients are constants. The same torsional potential was used to describe improper dihedral angles, meant to keep planar groups planar. The TraPPE-UA models have rigid bond lengths, and TraPPE-EH benzene is completely rigid; to avoid the use of constraints during the molecular dynamics simulations, missing harmonic bond, harmonic angle, and torsional potential parameters were taken from the AMBER Parm99 force field.<sup>39,40</sup> For all of the other molecules, except for 1,4-dioxane (where the parameters came directly from the literature),<sup>32</sup> intramolecular parameters were taken from GAFF. All of the force field files used in the present study are provided in the Supporting Information.

**Simulation Parameters.** All simulations were performed with a modified version of the molecular dynamics simulation package M.DynaMix 5.2.<sup>49,50</sup> For all systems studied in this work, LJ interactions were truncated at a distance of  $r_{\text{cut}} = 16 \text{ \AA}$ , and standard uniform fluid tail corrections were applied to both the energy and pressure, assuming  $g(r) = 1$  beyond the cutoff.<sup>27</sup> Electrostatic interactions were evaluated with an Ewald summation with tin foil boundary conditions,<sup>27,51</sup> with real space interactions truncated at  $r_{\text{cut}}$ . A damping parameter of  $\alpha r_{\text{cut}} = 3.14$  was used, and the maximum number of reciprocal space lattice vectors was set by  $K_{\text{max}} = 7.55$ . The integration of the equations of motion was performed with the multiple-time step method of Tuckerman and coworkers<sup>52</sup> in Cartesian coordinates. A short time step of 0.2 fs was used for fast intramolecular degrees of freedom and nonbonded interactions within a cutoff of  $r_{\text{short}} = 5 \text{ \AA}$ , and a time step of 2 fs was used for all other interactions. An Andersen thermostat<sup>53</sup> and Andersen–Hoover barostat<sup>53–55</sup> were used with the collision time for the thermostat set to 0.5 ps, and the time constant for the barostat set to 2 ps. Modifications to M.DynaMix include implementation of the Andersen thermostat, the “soft-core” potential (eq 13), separate decoupling of LJ and electrostatic interactions for EE calculations, WL-BAR, modification of the Ewald summation with EE fractional particles, and other minor additions.

The systems were set up by first randomly packing a cubic box<sup>56</sup> with pure solvent molecules. In all cases, the number of solvent molecules was chosen such that for a box length of approximately 40  $\text{\AA}$ , the system was near the experimental density at ambient conditions.<sup>57</sup> This led to systems having  $N_{\text{solvent}}$  solvent molecules of 312 hexane, 252 octane, 696 ethanol, 228 1-octanol, 384 cyclohexane, 456 benzene, and 444 1,4-dioxane molecules. Thus,  $x_A^* = 1/(N_{\text{solvent}} + 1)$  varied from 0.001 to 0.004. Short microcanonical (NVE) runs of 2 ps were then performed on each system to remove any intramolecular strain. Next, the systems were equilibrated in an NPT ensemble at elevated temperatures below the normal boiling point<sup>57</sup> and at a pressure of 1 bar. The NPT runs were for 2.25 ns, with configurations saved every 0.25 ns over the last 1.25 ns. For each solvent, these five configurations were saved, the velocities were reinitialized from a Maxwell–Boltzmann distribution with a unique seed to the random number generator and then further equilibrated for 1 ns at 298 K and 1 bar. This yielded five independent solvent boxes. For every combination of independent solvent box and solute, the gas phase minimized solute molecule was randomly

placed inside the solvent box, and the velocities of the entire system were again reinitialized with a unique random seed.

Production runs were carried out in an EE-NPT ensemble at 298 K and 1 bar for a total of 20 ns. Each of the five independent systems for each solvent was initialized with a unique random seed for the random number generator used by the thermostat and for the EE random walk. The system began in the subensemble with a noninteracting solute molecule, and attempts to change subensembles were made every 20 fs. Over the first 0.5 ns, the random walk was carried out with WL biasing. During the entire course of the simulation, transition energies (in both directions) were computed each time a transition between subensembles was attempted/proposed, and new subensemble weights were computed from BAR every 0.5 ns.<sup>17</sup> The solute was taken from noninteracting to fully interacting by first bringing the intermolecular LJ interaction to full strength and then adding in any intermolecular electrostatic interactions. The addition of the LJ and electrostatic interactions were performed separately in  $M_{\text{LJ}} = 15$  and  $M_{\text{elec}} = 5$  steps, respectively, for a total of  $M = 20$  steps (where in the reference subensemble  $m = 0$ , the solute molecule is decoupled from the system). For the first  $M_{\text{LJ}}$  steps, the intermolecular electrostatic interactions were turned off, and the intermolecular LJ interactions were strengthened as  $\lambda_m^{\text{LJ}} = \{0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, \text{ and } 1.0\}$  over the range  $1 \leq m \leq 15$ . The unequally spaced  $\lambda_m^{\text{LJ}}$  values were chosen to agree with previous work of Dill and coworkers.<sup>35,36,58</sup> If the solvent had partial charges, the intermolecular electrostatic interactions were then strengthened as  $\lambda_m^{\text{elec}} = \{0.2, 0.4, 0.6, 0.8, \text{ and } 1.0\}$  over the range  $16 \leq m \leq 20$ . Note that, for hexane, octane, and cyclohexane, electrostatic steps were not necessary. A detailed description regarding the decoupling of intermolecular interactions with Ewald summation may be found in our previous work.<sup>17</sup>

The reported residual chemical potentials were taken as the mean value of the five independent production runs for each solute–solvent combination, and the uncertainty was taken as the bootstrap standard error.<sup>59–61</sup> To compute the bootstrap standard error for each solute–solvent combination, the estimate of the chemical potential from each of the five production runs was taken to be an independent data point. Next, 1000 sets containing five data points each were created by randomly selecting five of our independent data points, with replacement. The mean of each set was computed, creating a bootstrap sample of 1000 estimates of the residual chemical potential. The bootstrap standard error was then found as the standard error of the bootstrap sample relative to the mean of the five independent production runs for each solute–solvent combination.

## RESULTS AND DISCUSSION

By application of the Markov principle,<sup>62</sup> ensemble averages may be computed within each EE subensemble. Therefore, the partial molar residual enthalpy of the solute,  $h_A^{\text{res}}$ , was determined from ensemble averages of the enthalpy of the solution in the fully coupled ( $H_M$ ) and decoupled ( $H_0$ ) states as

$$\frac{1}{RT} h_A^{\text{res}}(T, P, x_A^*) = \frac{1}{RT} H_M(T, P, x_A^*) - \frac{1}{RT} H_0(T, P, x_A^*) \quad (17)$$

Likewise, the partial molar residual entropy of the solute,  $s_A^{\text{res}}$ , was obtained from the fundamental thermodynamic

**Table 1. Summary of the Computed Residual Chemical Potential and Each of Its Components for 1,2,4,5-Tetramethylbenzene**

solvent	$(1/RT)\mu^{\text{res}}$	$(1/RT)h^{\text{res}}$	$(1/R)s^{\text{res}}$
hexane	$-12.19 \pm 0.02$	$-0.052 \pm 0.003$	$12.14 \pm 0.02$
octane	$-12.17 \pm 0.05$	$-0.083 \pm 0.010$	$12.09 \pm 0.05$
ethanol	$-10.09 \pm 0.04$	$-0.024 \pm 0.002$	$10.07 \pm 0.04$
1-octanol	$-11.27 \pm 0.11$	$-0.122 \pm 0.009$	$11.15 \pm 0.11$
cyclohexane	$-12.43 \pm 0.03$	$-0.046 \pm 0.001$	$12.38 \pm 0.03$
benzene	$-11.68 \pm 0.04$	$-0.035 \pm 0.003$	$11.65 \pm 0.04$
1,4-dioxane	$-11.06 \pm 0.03$	$-0.033 \pm 0.002$	$11.03 \pm 0.03$

**Table 2. Summary of the Computed Residual Chemical Potential and Each of Its Components for Phenanthrene**

solvent	$(1/RT)\mu^{\text{res}}$	$(1/RT)h^{\text{res}}$	$(1/R)s^{\text{res}}$
hexane	$-16.44 \pm 0.02$	$-0.065 \pm 0.014$	$16.38 \pm 0.02$
octane	$-16.74 \pm 0.04$	$-0.093 \pm 0.007$	$16.65 \pm 0.04$
ethanol	$-15.00 \pm 0.06$	$-0.029 \pm 0.001$	$14.97 \pm 0.06$
1-octanol	$-16.63 \pm 0.18$	$-0.157 \pm 0.026$	$16.47 \pm 0.18$
cyclohexane	$-16.79 \pm 0.02$	$-0.062 \pm 0.001$	$16.73 \pm 0.02$
benzene	$-17.19 \pm 0.04$	$-0.050 \pm 0.005$	$17.14 \pm 0.04$
1,4-dioxane	$-16.94 \pm 0.07$	$-0.056 \pm 0.003$	$16.88 \pm 0.07$

**Table 3. Summary of the Computed Residual Chemical Potential and Each of Its Components for Fluorene**

solvent	$(1/RT)\mu^{\text{res}}$	$(1/RT)h^{\text{res}}$	$(1/R)s^{\text{res}}$
hexane	$-14.95 \pm 0.05$	$-0.073 \pm 0.004$	$14.88 \pm 0.05$
octane	$-15.00 \pm 0.04$	$-0.092 \pm 0.006$	$14.91 \pm 0.04$
ethanol	$-13.54 \pm 0.05$	$-0.024 \pm 0.004$	$13.52 \pm 0.05$
1-octanol	$-14.54 \pm 0.20$	$-0.185 \pm 0.018$	$14.36 \pm 0.20$
cyclohexane	$-15.20 \pm 0.04$	$-0.053 \pm 0.001$	$15.15 \pm 0.04$
benzene	$-15.51 \pm 0.05$	$-0.045 \pm 0.003$	$15.47 \pm 0.05$
1,4-dioxane	$-15.34 \pm 0.07$	$-0.047 \pm 0.004$	$15.29 \pm 0.07$

relation<sup>5</sup>

$$\frac{1}{R}s^{\text{res}}(T, P, x_A^*) = \frac{1}{RT}h^{\text{res}}(T, P, x_A^*) - \frac{1}{RT}\mu^{\text{res}}(T, P, x_A^*) \quad (18)$$

A summary of the computed residual chemical potential, partial molar residual enthalpy, and partial molar residual entropy for each solute in each solvent is given in Tables 1 to 3. In all cases, the computed partial molar residual enthalpy is the same order of magnitude as the uncertainty of the computed residual chemical potential. The results indicate that the residual entropy is the dominant component in determining the residual chemical potential. In other words, the solvation process for these solutes is nearly entirely entropically driven, as might be expected given their low solubility.

To compute the solubility limit using eq 9, a reference experimental solubility is needed to determine the parameter  $\zeta_A$  for each solute. The results in Tables 1 to 3 suggest that the calculated chemical potential for the solutes in hexane are found with the greatest precision. Therefore, the required solute-dependent constant  $\zeta_A$  was found by using the simulation results for solute solubility in hexane. By forcing the hexane solubility in eq 9 to match the experimental data,  $\zeta_A$  was set

**Table 4. Comparison of the Experimental Solubility,<sup>65</sup> the Solubility Predicted in This Study, and the Solubility Predicted by the Water to Solvent ( $W \rightarrow S$ ) and Gas to Solvent ( $G \rightarrow S$ ) Forms of the Abraham General Solvation Model<sup>64,65,68</sup> for 1,2,4,5-Tetramethylbenzene<sup>a</sup>**

solvent	ln X (mole fraction)			
	experiment	this work	Abraham $W \rightarrow S$	Abraham $G \rightarrow S$
hexane	-1.545	$-1.54 \pm 0.03$	-2.434	-2.071
octane	-1.552	$-1.56 \pm 0.05$	-1.931	-1.987
ethanol	-3.709	$-3.64 \pm 0.04$	-3.530	-3.344
1-octanol	-2.123	$-2.46 \pm 0.11$	-2.524	-2.281
cyclohexane	-1.587	$-1.30 \pm 0.04$	-2.013	-2.024
benzene		$-2.05 \pm 0.04$	-1.276	-0.865
1,4-dioxane		$-2.67 \pm 0.04$	-2.035	-1.842
AE total		0.448	1.144	0.903
AE alkane		0.295	1.694	1.398
AE alcohol		0.344	0.439	0.398

<sup>a</sup>The predictions of the current study use eq 9 with hexane as a reference, giving  $\zeta_A = -13.73 \pm 0.02$ . The reported solubility is in units of mole fraction, and AE corresponds to the absolute error.

**Table 5. Comparison of the Experimental Solubility,<sup>6</sup> the Solubility Predicted in This Study, and the Solubility Predicted by the Water to Solvent ( $W \rightarrow S$ ) and Gas to Solvent ( $G \rightarrow S$ ) Forms of the Abraham General Solvation Model<sup>6,64,68</sup> for Phenanthrene<sup>a</sup>**

solvent	ln X (mole fraction)			
	experiment	this work	Abraham $W \rightarrow S$	Abraham $G \rightarrow S$
hexane	-3.445	$-3.45 \pm 0.03$	-3.793	-3.305
octane	-3.114	$-3.15 \pm 0.04$	-2.939	-2.908
ethanol	-4.497	$-4.89 \pm 0.06$	-4.187	-3.992
1-octanol	-2.915	$-3.26 \pm 0.18$	-3.032	-2.832
cyclohexane	-3.311	$-3.10 \pm 0.03$	-2.629	-2.890
benzene	-1.710	$-2.70 \pm 0.04$	-1.046	-0.465
1,4-dioxane		$-2.95 \pm 0.07$	-1.480	-1.228
AE total		0.565	0.929	0.708
AE alkane		0.247	1.205	0.767
AE alcohol		0.523	0.497	0.512

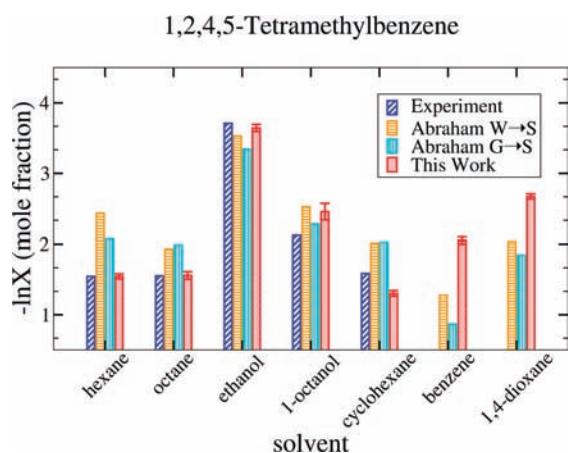
<sup>a</sup>The predictions of the current study use eq 9 with hexane as a reference, giving  $\zeta_A = -19.89 \pm 0.02$ . The reported solubility is in units of mole fraction, and AE corresponds to the absolute error.

for each solute and used to estimate the solubility in other solvents. Experimental solubility data are given in Tables 4 to 6. Using eq 9, the solute constants at 298 K and 1 bar are found to be  $\zeta_A = -13.73 \pm 0.02$ ,  $\zeta_A = -19.89 \pm 0.02$ , and  $\zeta_A = -18.44 \pm 0.05$  for 1,2,4,5-tetramethylbenzene, phenanthrene, and fluorene, respectively. Using this constant, the solubility limit of the solute in any other solvent can be estimated from an EE simulation. Note that the constant  $\zeta_A$  could have been determined using data for any solvent; results were found to be essentially the same if the reference solvent were chosen to be octane, ethanol, or cyclohexane. Differences were observed if 1-octanol was chosen as the reference solvent, which is attributed to the rather larger uncertainties in the computed residual chemical potential for 1-octanol, presumably due to sampling

**Table 6.** Comparison of the Experimental Solubility,<sup>66,67</sup> the Solubility Predicted in This Study, and the Solubility Predicted by the Water to Solvent ( $W \rightarrow S$ ) and Gas to Solvent ( $G \rightarrow S$ ) Forms of the Abraham General Solvation Model<sup>64,66–68</sup> for Fluorene<sup>a</sup>

solvent	ln X (mole fraction)			
	experiment	this work	Abraham $W \rightarrow S$	Abraham $G \rightarrow S$
hexane	-3.493	-3.49 ± 0.07	-3.974	-3.532
octane	-3.283	-3.44 ± 0.06	-3.268	-3.242
ethanol	-4.902	-4.90 ± 0.07	-4.529	-4.423
1-octanol	-3.248	-3.90 ± 0.21	-3.464	-3.268
cyclohexane	-3.279	-3.24 ± 0.06	-3.118	-3.238
benzene		-2.93 ± 0.07	-1.883	-1.397
1,4-dioxane		-3.10 ± 0.09	-2.352	-2.177
AE total		0.672	0.666	0.484
AE alkane		0.196	0.657	0.121
AE alcohol		0.652	0.431	0.479

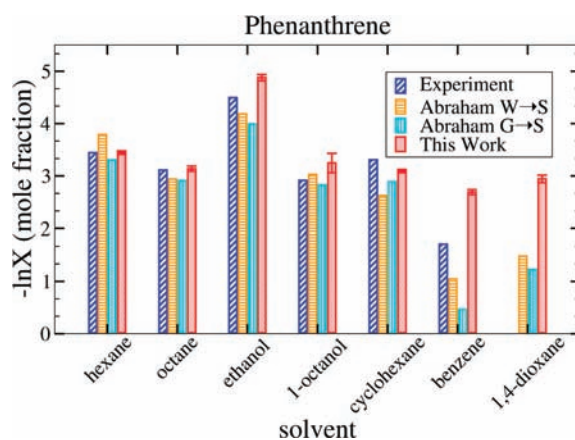
<sup>a</sup>The predictions of the current study use eq 9 with hexane as a reference, giving  $\zeta_A = -18.44 \pm 0.05$ . The reported solubility is in units of mole fraction, and AE corresponds to the absolute error.



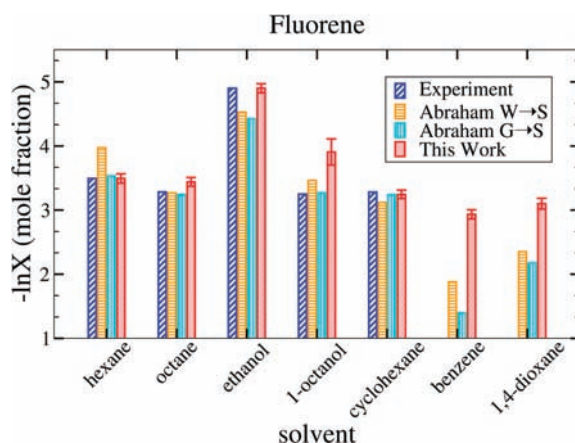
**Figure 2.** Comparison of the experimental solubility,<sup>65</sup> the solubility predicted by the water to solvent ( $W \rightarrow S$ ) and gas to solvent ( $G \rightarrow S$ ) forms of the Abraham general solvation model,<sup>64,65,68</sup> and the solubility predicted in this study for 1,2,4,5-tetramethylbenzene. The reported solubility is in units of mole fraction.

difficulties. It has been well-documented<sup>63</sup> that 1-octanol is difficult to sample in atomistic simulations due to the formation of aggregates and other inhomogeneities. To minimize these difficulties, five independent 20 ns simulations were conducted.

A summary of the solubility limit predictions obtained in this study, along with a comparison with experiment and the predictions obtained from the Abraham general solvation model,<sup>64–68</sup> are presented graphically in Figures 2 to 4. Results are tabulated in Tables 4 to 6. In general, the simulations perform extremely well compared to the Abraham model. This is encouraging, since the Abraham model is generally thought to perform at least as well as the mobile order theory and the universal functional activity coefficient (UNIFAC) and modified UNIFAC (Dortmund) models currently used in the pharmaceutical and chemical industries.<sup>6</sup> To compare the simulation predictions and the predictions of the Abraham general solvation model



**Figure 3.** Comparison of the experimental solubility,<sup>6</sup> the solubility predicted by the water to solvent ( $W \rightarrow S$ ) and gas to solvent ( $G \rightarrow S$ ) forms of the Abraham general solvation model,<sup>64,68</sup> and the solubility predicted in this study for phenanthrene. The reported solubility is in units of mole fraction.



**Figure 4.** Comparison of the experimental solubility,<sup>66,67</sup> the solubility predicted by the water to solvent ( $W \rightarrow S$ ) and gas to solvent ( $G \rightarrow S$ ) forms of the Abraham general solvation model,<sup>64,66–68</sup> and the solubility predicted in this study for fluorene. The reported solubility is in units of mole fraction.

against experimental data quantitatively, the absolute error, AE, defined as

$$AE = \sum \sqrt{(\ln x_{\text{model}} - \ln x_{\text{experiment}})^2} \quad (19)$$

was computed for each solute. In eq 19,  $x_{\text{model}}$  and  $x_{\text{experiment}}$  are the predicted solubility of the model and the experimental solubility, respectively, in units of mole fraction. The summation is over all systems except the hexane system against which  $\zeta_A$  was regressed. Note that the solubilities predicted by the Abraham general solvation model are in units of molarity. These were converted to mole fractions using the solute molar volumes published with the Abraham general solvation model,<sup>65–67</sup> reference solvent molar volumes,<sup>57</sup> and the corresponding conversion used in the Abraham general solvation model to convert experimental mole fractions to molarities.<sup>65–67</sup>

For 1,2,4,5-tetramethylbenzene, the estimated solubility obtained from the present work is in much better agreement with the experiment than the predictions of the Abraham solvation



model for both alkanes and alcohols. In the case of phenanthrene, the present study exhibits superior performance for the alkanes, but not for alcohols. The AE for alcohols is enclosed by the AE of the two predictions of the Abraham model. The overall AE for the alkanes and alcohols, however, is smaller for the present study than for that obtained with the Abraham model. We note that, for the solubility of phenanthrene in benzene, reference data are presented that are extrapolated from solubility data at other conditions.<sup>6</sup> For this case, the AE for the model of the present study is in between the AE for the two predictions of the Abraham general solvation model, with the present study under predicting the solubility, and the Abraham general solvation model over predicting the solubility. Lastly, for fluorene the AE for the model of the present study is in between the AE for the two predictions of the Abraham general solvation model for alkanes and is inferior to the Abraham general solvation model for alcohols. However, the overall deviation is in close agreement with the water to solvent prediction of the Abraham general solvation model. For the cases in which experimental data are not available for benzene and 1,4-dioxane, the present study deviates from the Abraham general solvation model. Unfortunately, the lack of experimental data prevents a determination of which model is more accurate. It is important to note that the predictions from the present study show the same general trends as the Abraham solvation model, but that only a single experimental data point is required. This is in contrast to the Abraham model, which requires extensive parametrization against experimental data.

This suggests that, in the absence of large amounts of experimental data, the approach presented here can give relatively accurate estimates of the solubility limit of low solubility solutes and most certainly is capable of capturing solubility trends.

## SUMMARY AND CONCLUSIONS

A method has been proposed to predict the solubility limit of low solubility solids. A single experimental reference solubility is required for each solute, as well as a single free energy simulation of the solute–solvent system. For the case of three solutes (1,2,4,5-tetramethylbenzene, phenanthrene, and fluorene), the method yields accurate solubility predictions for alkane and alcohol solvents. The predictions are found to be comparable to the Abraham general solvation model, but the method requires far fewer experimental data points to parameterize the model. Predictions for the solubility of these three solutes in solvents for which experimental data are not available are in qualitative agreement with the predictions of the Abraham general solvation model. While the results are not perfect and only a limited range of solvents have been explored, the results are extremely promising, suggesting that the proposed method may be used to help guide experiment with regards to selecting a solvent to dissolve a particular solid. Furthermore, now that the solute-dependent constants have been found for the studied systems, the solubility of these solutes in any pure or mixed solvent may be predicted by performing a single free energy calculation in the solvent of interest. These promising results warrant further studies to probe the application to a wider range of solvents and solutes of pharmaceutical interest.

Lastly, the method enables one to obtain physical insight into the reasons for the observed solubility trends, something that is often difficult to obtain with more empirical models. In the

present study, it was found that enthalpic interactions play a small role in the solubility; instead, the solubility of these solute–solvent systems is dominated by entropic effects. Using a single experimental data point obviates the need to compute the free energy of the crystalline phase, something that can be done but with great difficulty.<sup>13</sup> It should be relatively straightforward to extend this method to cosolvent systems.

We are happy to submit this paper as part of the John M. Prausnitz Festschrift. Much of the underlying molecular thermodynamic theory for the method can be found in the seminal monograph “Molecular Thermodynamics of Fluid-Phase Equilibria”<sup>5</sup> by Prausnitz, Lichtenthaler, and de Azevedo. This text and the writings of Prof. Prausnitz have continued to inspire us to think about ways of linking the very fundamental topic of molecular interactions with applications that are at the heart of chemical engineering. The present paper attempts to follow this model by using advanced molecular simulation methods and a small amount of experimental data to make solubility predictions.

## ASSOCIATED CONTENT

**S Supporting Information.** M.DynaMix 5.2 force field files for all of the solutes and solvents studied in this work. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [ed@nd.edu](mailto:ed@nd.edu). Phone: (574) 631-5687. Fax: (574) 631-8366.

## ACKNOWLEDGMENT

The authors thank Dr. Jindal Shah and Dr. Neeraj Rai for stimulating discussions. A.S.P. gratefully acknowledges a fellowship from the Arthur J. Schmitt Foundation. Funding was also provided by the Notre Dame Sustainable Energy Initiative. Computing support was provided by Notre Dame's Center for Research Computing.

## DEDICATION

On a personal note, I (E.J.M.) would like to thank John Prausnitz for his role in helping shape me as a scholar. I had the privilege of learning molecular thermodynamics from John as a graduate student. Needless to say, after his class I was hooked. I served as his teaching assistant and learned how a great teacher operates. Like many of my peers, I also had the pleasure of receiving one of his coveted “letters of encouragement” that he frequently sends to young faculty to let them know he had read (and liked) a paper they had written. Receiving such encouragement from so eminent a scholar does wonders for an overworked assistant professor. We wish John many more years of productive scholarship, health, and happiness.

## REFERENCES

- (1) Llinas, A.; Glen, R. C.; Goodman, J. M. Solubility Challenge: Can You Predict Solubilities of 32 Molecules Using a Database of 100 Reliable Measurements? *J. Chem. Inf. Model.* **2008**, *48*, 1289–1303.
- (2) Liu, R., Ed. *Water-Insoluble Drug Formulation*, 2nd ed.; CRC Press: Boca Raton, FL, 2008.

- (3) Hildebrand, J. H.; Scott, R. L. *Regular Solutions*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1962.
- (4) Acree, W. E. *Thermodynamic Properties of Nonelectrolyte Solutions*; Academic Press, Inc.: Orlando, FL, 1984.
- (5) Prausnitz, J. M.; Lichtenthaler, R. N.; de Azevedo, E. G. *Molecular Thermodynamics of Fluid-Phase Equilibria*, 3rd ed.; Prentice-Hall PTR: Upper Saddle River, NJ, 1999.
- (6) Acree, W. E.; Abraham, M. H. Solubility predictions for crystalline nonelectrolyte solutes dissolved in organic solvents based upon the Abraham general solvation model. *Can. J. Chem.* **2001**, *79*, 1466–1476.
- (7) Ruckenstein, E.; Shulgin, I. Solubility of drugs in aqueous solutions Part 1. Ideal mixed solvent approximation. *Int. J. Pharm.* **2003**, *258*, 193–201.
- (8) Abildskov, J.; O'Connell, J. P. Predicting the Solubilities of Complex Chemicals I. Solutes in Different Solvents. *Ind. Eng. Chem. Res.* **2003**, *42*, 5622–5634.
- (9) Nordstrom, F.; Rasmuson, A. C. Prediction of solubility curves and melting properties of organic and pharmaceutical compounds. *Eur. J. Pharm. Sci.* **2009**, *36*, 330–344.
- (10) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from Monte Carlo simulations. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1155–1158.
- (11) Ferrario, M.; Ciccotti, G.; Spohr, E.; Cartailier, T.; Turq, P. Solubility of KF in water by molecular dynamics using the Kirkwood integration method. *J. Chem. Phys.* **2002**, *117*, 4947–4953.
- (12) Sanz, E.; Vega, C. Solubility of KF and NaCl in water by molecular simulation. *J. Chem. Phys.* **2007**, *126*, 014507.
- (13) Paluch, A. S.; Jayaraman, S.; Shah, J. K.; Maginn, E. J. A method for computing the solubility limit of solids: Application to sodium chloride in water and alcohols. *J. Chem. Phys.* **2010**, *133*, 124504.
- (14) Kukol, A., Ed. *Molecular Modeling of Proteins*; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2008; Vol. 443.
- (15) Svard, M.; Rasmuson, A. C. Force Fields and Point Charges for Crystal Structure Modeling. *Ind. Eng. Chem. Res.* **2009**, *48*, 2899–2912.
- (16) Jayaraman, S.; Thompson, A. P.; von Lilienfeld, A.; Maginn, E. J. Molecular Simulation of the Thermal and Transport Properties of Three Alkali Nitrate Salts. *Ind. Eng. Chem. Res.* **2010**, *49*, 559–571.
- (17) Paluch, A. S.; Shah, J. K.; Maginn, E. J. Efficient Solvation Free Energy Calculations of Amino Acid Analogs by Molecular Simulation. *J. Chem. Theory Comput.* **2011**, submitted.
- (18) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.* **1992**, *96*, 1776–1783.
- (19) Lyubartsev, A. P.; Laaksonen, A.; Vorontsov-Velyaminov, P. N. Free energy calculations for Lennard-Jones systems and water using the expanded ensemble method: A Monte Carlo and molecular dynamics simulation study. *Mol. Phys.* **1994**, *82*, 455–471.
- (20) Lyubartsev, A. P.; Laaksonen, A.; Vorontsov-Velyaminov, P. N. Determination of Free Energy from Chemical Potentials. Application of the Expanded Ensemble Method. *Mol. Simul.* **1996**, *18*, 43–58.
- (21) Lyubartsev, A. P.; Forrisdahl, O. K.; Laaksonen, A. Solvation free energies of methane and alkali halide ion pairs: An expanded ensemble molecular dynamics simulation study. *J. Chem. Phys.* **1998**, *108*, 227–233.
- (22) Wang, F.; Landau, D. P. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.
- (23) Yan, Q.; Faller, R.; de Pablo, J. J. Density-of-states Monte Carlo method for simulation of fluids. *J. Chem. Phys.* **2002**, *116*, 8745–8749.
- (24) Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. Generalization of the Wang-Landau method for off-lattice simulations. *Phys. Rev. E* **2002**, *66*, 056703.
- (25) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (26) Fenwick, M. K.; Escobedo, F. A. On the use of Bennett's acceptance ratio method in multi-canonical-type simulations. *J. Chem. Phys.* **2004**, *120*, 3066–3074.
- (27) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press Inc.: New York, NY, 1987.
- (28) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569–2577.
- (29) Chen, B.; Potoff, J. J.; Siepmann, J. I. Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Tertiary Alcohols. *J. Phys. Chem. B* **2001**, *105*, 3093–3104.
- (30) Lee, J.; Wick, C. D.; Stubbs, J. M.; Siepmann, J. I. Simulating the vapour-liquid equilibria of large cyclic alkanes. *Mol. Phys.* **2005**, *103*, 99–104.
- (31) Rai, N.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 9. Explicit Hydrogen Description of Benzene and Five-Membered and Six-Membered Heterocyclic Aromatic Compounds. *J. Phys. Chem. B* **2007**, *111*, 10790–10799.
- (32) Yazaydin, A. O.; Thompson, R. W. Simulating the vapour-liquid equilibria of 1,4-dioxane. *Mol. Simul.* **2006**, *32*, 657–662.
- (33) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (34) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (35) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* **2007**, *111*, 2242–2254.
- (36) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.
- (37) Klimovich, P. V.; Mobley, D. A. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *J. Comput. Aided Mol. Des.* **2010**, *24*, 307–316.
- (38) Cramer, C. J. *Essentials of Computational Chemistry*, 2nd ed.; John Wiley and Sons Ltd.: West Sussex, U.K., 2004.
- (39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T. F.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (40) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (41) Frisch, M. J.; et al. *Gaussian 09*, Revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
- (42) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (43) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357–1377.
- (44) Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (45) Case, D. A.; et al. *AMBER 11*; University of California: San Francisco, CA, 2010.
- (46) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.



- (47) Shirts, M. R.; Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **2005**, *122*, 134508.
- (48) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.* **2007**, *127*, 214108.
- (49) Lyubartsev, A. P.; Laaksonen, A. MDynaMix—a scalable portable parallel MD simulation package for arbitrary molecular mixtures. *Comput. Phys. Commun.* **2000**, *128*, 565–589.
- (50) Lyubartsev, A. P.; Laaksonen, A. *MDynaMix: a Molecular Dynamics Program*; 2010; <http://www.fos.su.se/~sasha/mdynamix/>.
- (51) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2002.
- (52) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.
- (53) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- (54) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.
- (55) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87*, 1117–1157.
- (56) Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (57) Lide, D. R., Ed. *CRC Handbook of Chemistry and Physics*, 80th ed.; CRC Press: Boca Raton, FL, 1999.
- (58) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125*, 084902.
- (59) Efron, B. Computers and the Theory of Statistics: Thinking the Unthinkable. *SIAM Review* **1979**, *21*, 460–480.
- (60) Efron, B. Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods. *Biometrika* **1981**, *68*, 589–599.
- (61) Moore, D. S.; McCabe, G. P.; Craig, B. *Introduction to the Practice of Statistics*, 6th ed.; W. H. Freeman and Company: New York, NY, 2009.
- (62) Lawler, G. *Introduction to Stochastic Processes*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, 2006.
- (63) Chen, B.; Siepmann, J. I. Microscopic Structure and Solvation in Dry and Wet Octanol. *J. Phys. Chem. B* **2006**, *110*, 3555–3563.
- (64) Abraham, M. H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* **1993**, *22*, 73–83.
- (65) Flanagan, K. B.; Hoover, K. R.; Acree, W. E.; Abraham, M. H. *Phys. Chem. Liq.* **2006**, *44*, 173–182.
- (66) Monarrez, C. I.; Acree, W. E.; Abraham, M. H. *Phys. Chem. Liq.* **2002**, *40*, 581–591.
- (67) Stovall, D. M.; Hoover, K. R.; Acree, W. E.; Abraham, M. H. *Polycyclic Aromat. Compd.* **2005**, *25*, 313–326.
- (68) Acree, W. E.; Abraham, M. H. Solubility predictions for crystalline polycyclic aromatic hydrocarbons (PAHs) dissolved in organic solvents based upon the Abraham general solvation model. *Fluid Phase Equilib.* **2002**, *201*, 245–258.