Title
Predicting the excess solubility of acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine in binary water/ethanol mixtures via molecular simulation

Journal
The Journal of Chemical Physics, 142(4)

ISSN
0021-9606

Authors
Paluch, Andrew S
Parameswaran, Sreeja
Liu, Shuai
et al.

Publication Date
2015-01-28

DOI
10.1063/1.4906491

Peer reviewed

Predicting the Excess Solubility of Acetanilide, Acetaminophen, Phenacetin, Benzocaine, and Caffeine in Binary Water/Ethanol Mixtures via Molecular Simulation

Andrew S. Paluch,[1] Sreeja Parameswaran,[2] Pavel V. Klimovich,[3]
Shuai Liu,[3] Anasuya Kolavennu,[2] and David L. Mobley[4]

[1]*Department of Chemical, Paper, and Biomedical Engineering, Miami University, Oxford, Ohio, 45056, USA*[*]

[2]*Department of Chemistry, University of New Orleans, New Orleans, Louisiana, 70148, USA*

[3]*Department of Pharmaceutical Sciences, University of California Irvine, Irvine, California, 92697, USA*[†]

[4]*Departments of Pharmaceutical Sciences and Chemistry, University of California Irvine, Irvine, California, 92697, USA*[‡]

(Dated: October 2, 2014)

Abstract

We present a general framework to predict the excess solubility of small molecular solids (such as pharmaceutical solids) in binary solvents via molecular simulation free energy calculations at infinite dilution with conventional molecular models. The present study used molecular dynamics with the General AMBER Force Field to predict the excess solubility of acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine in binary water/ethanol solvents. The simulations are able to predict the existence of solubility enhancement and the results are in good agreement with available experimental data. The accuracy of the predictions in addition to the generality of the method suggest that molecular simulations may be a valuable design tool for solvent selection in drug development processes.

[*]Previous Address: Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA; Electronic address: PaluchAS@MiamiOH.edu; To whom correspondence should be addressed.

[†]Previous Address: Department of Chemistry, University of New Orleans, New Orleans, Louisiana, 70148, USA

‡Previous Address: Department of Chemistry, University of New Orleans, New Orleans, Louisiana, 70148, USA; Electronic address: DMobley@uci.edu

I. INTRODUCTION AND MOTIVATION

Solubility drives a huge range of biological and physical processes, and so prediction and control of solubility is a key challenge for modeling in the physical and biological sciences. For example, crystallization is an important solid-liquid separation process capable of creating high purity products, but control of crystallization requires knowledge of the equilibrium solubility of the solid solute(s) of interest [1–4]. Here, we are particularly interested in solids of small molecules, such as pharmaceutical solids. Pharmaceutical solids constitute an important class of compounds subject to many design challenges. Knowledge of the solubility of pharmaceutical solids is important not just for purification, formulation, and production (e.g. crystallization) but additionally for drug synthesis and bio-availability [5–7]. Each of these processes can have different solubility requirements, complicating design. Additionally to satisfy pharmaceutical requirements of solubility and price, multiple co-solvents may be used during manufacturing and formulation [5, 7]. This is further complicated by the often complex chemistries involved.

Since solubility plays such a key role in molecular and process design, we want methods which can predict the ability of a given solvent to dissolve a particular solute. These methods could not only guide the design process by predicting solubility in advance of experiment, but could also provide new insight into mechanisms driving solubility. But, as a consequence of the many competing intermolecular forces, the development of such methods is extremely challenging. While considerable work has been done in this area, it is still generally the case that solubility predictions rely on empirical and semi-empirical correlations [8]. These correlations are often limited to a specific solvent, and require extensive experimental solubility data for a range of chemically different solute molecules to allow for their training. Recently, promising solubility predictions have been made with the theoretically based NRTL-SAC model [9], the 2005 revised MOSCED model [10], and the PC-SAFT equation of state [11–13]. Normally, however, solubility data is first required for the solute of interest to determine the necessary model parameters prior to making predictions in the solvent of interest. All of the PC-SAFT parameters may be computed using *ab initio* methods 11 . However, as common with equation of state methods [14], the results are sensitive to the employed combining rules, and are difficult to predict *a priori* [11].

Since many of these methods require extensive input data, then, they are often poorly suited for design problems like designing processes for dealing with a novel drug candidate, since this

extensive data may not yet be available. Thus, we seek methods which predict solubility from physical principles. Currently, there are very few available approaches which do this. Previous attempts have been made to use the predictive UNIFAC method to characterize the solution phase behavior of pharmaceutical solids along with limited pure component data to predict equilibrium solubility (see for example refs. 15, 16). But UNIFAC parameters are missing for many molecules of pharmaceutical interest, and predictions with existing parameters may show appreciable error [15, 16]. Efforts are currently being made to extend UNIFAC to pharmaceutical compounds [17] and to develop methods to estimate missing parameters [18, 19]. With use of *ab initio* calculations, COSMO based methods have shown great promise in predicting solubility in the absence of experimental data [20–22]. While COSMO models only solution-phase properties, the solid-phase properties necessary to make solubility predictions are estimated using an empirical correlation of experimental data [20]. COSMO is under active development, and has been successfully used in both academic and industrial settings [22]. A comparison of *a priori* predictions of infinite dilution activity coefficients, binary vapor-liquid equilibrium, and the solubility of pharmaceutical solids with the various COSMO and UNIFAC was conducted by Gmehling and co-workers [17, 23], and the interested reader is directed therein for a detailed comparison. Nonetheless, these methods are unable to probe the atomistic details of the solvation process. Such insight may be useful for rational design purposes.

Thus, we seek physical techniques which can predict solubilities in the absence of experimental data. Here, our focus is on atomistic molecular simulations, which can do exactly this. In this case, not only will such a method help guide solubility control, but also provide insight into the molecular level details governing solubility. However, only a limited number of molecular simulation studies have attempted to directly predict the solubility of solids [24–31]. Solubility calculation requires determining the concentration at which the chemical potential of the solute in solution is equal to that in a solid (crystalline) phase, at the same temperature and pressure. The solubility may be computed in the complete absence of experimental thermodynamic data so long as the solute solid crystal structure is known [24–26, 28, 29], or may alternatively be computed from experimental data [27, 31] or estimated using group-contribution methods [30] or other predictive means [31].

In the present study we describe a framework using molecular simulation to compute the excess solubility of small molecular solids in binary solvents via molecular simulation. By excess

solubility, we refer to the solubility in the actual solution relative to what would be observed if the solvents mixed to form an ideal solution [32, 33]. The study of excess solubilities is advantageous for a number of reasons, the most important of which is that framing these as relative or excess solubility calculations means it is not necessary to calculate the properties of the pure solid. Thus, this work can focus entirely on solution-phase properties. If absolute solubilities are desired, excess solubilities can be converted to absolute solubilities using knowledge of a single reference solubility or by modeling the solid phase directly (as done in refs 24–26, 28, 29).

We use our framework to study the excess solubility of acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine (see fig. 1) in binary water/ethanol mixtures. The solutes were all modeled using the General AMBER Force Field. Acetaminophen and phenacetin are both acetanilide derivatives, allowing for a systematic test of the models. Overall, the molecular simulations are in good qualitative agreement with experiment and are able to predict the existence of an extrema in the the excess solubility at intermediate binary solvent concentrations. The encouraging results highlight the potential use of molecular simulation as a design tool for solvent selection.

II. METHODOLOGY

A. Thermodynamic Approach to Solid-Liquid Equilibrium

III. METHODOLOGY

The solubility of a solid solute (component 1) in a binary solvent (components 2 and 3) may be described by the classical equations of phase-equilibrium thermodynamics. At equilibrium, the temperature, pressure, and chemical potential of component 1 in both phases are equivalent. Assuming that the solid phase is pure component 1 (i.e. that the solvent does not dissolve into the crystalline solid solute in equilibrium with the solution phase), the equality of chemical potential criteria takes the form

$$\beta\mu_1^{\mathrm{S}}\left(T,p\right) = \beta\mu_{1,m}^{\mathrm{L}}\left(T,p,N_1^{\mathrm{L}},N_2^{\mathrm{L}},N_3^{\mathrm{L}}\right) \tag{1}$$

where μ_1^{S} and $\mu_{1,m}^{\mathrm{L}}$ are the chemical potential of component 1 in a pure solid (S) and in the liquid

(L) phase, respectively, T is the temperature, p is the pressure, N_1^{L}, N_2^{L}, and N_3^{L} are the number of molecules of component 1, 2, and 3, respectively, in the liquid solution phase, and $\beta^{-1} = k_{\mathrm{B}}T$ where k_{B} is Boltzmann's constant. The chemical potential may be written as the sum of a residual (res) and ideal gas (ig) term. The residual chemical potential of component 1 in the liquid solution phase may be computed using conventional free energy calculations in the isothermal-isobaric (*NpT*) ensemble [34–36]. Using classical molecular models this leads to the following re-expression of eq. (1) [37, 38]

$$\begin{aligned}
\beta\mu_1^{\mathrm{S,res}}\left(T,p\right)+\ln\left(\frac{\Lambda_1^3}{v_1^{\mathrm{S}}\left(T,p\right)}\right)-\ln Z_1^{\mathrm{ig}} &= \beta\mu_{1,m}^{\mathrm{L,res}}\left(T,p,N_1^{\mathrm{L}},N_2^{\mathrm{L}},N_3^{\mathrm{L}}\right)+\ln\left(\frac{N_1^{\mathrm{L}}\Lambda_1^3}{\left\langle V\right\rangle_{T,p,N_1^{\mathrm{L}}-1,N_2^{\mathrm{L}},N_3^{\mathrm{L}}}}\right)-\ln Z_1^{\mathrm{ig}} \\
\beta\mu_1^{\mathrm{S,res}}\left(T,p\right)-\ln v_1^{\mathrm{S}}\left(T,p\right) &= \beta\mu_{1,m}^{\mathrm{L,res}}\left(T,p,N_1^{\mathrm{L}},N_2^{\mathrm{L}},N_3^{\mathrm{L}}\right)+\ln\left(\frac{N_1^{\mathrm{L}}}{\left\langle V\right\rangle_{T,p,N_1^{\mathrm{L}}-1,N_2^{\mathrm{L}},N_3^{\mathrm{L}}}}\right) \\
\beta\mu_1^{\mathrm{S,res}}\left(T,p\right)-\ln v_1^{\mathrm{S}}\left(T,p\right) &= \beta\mu_{1,m}^{\mathrm{L,res}}\left(T,p,N_1^{\mathrm{L}},N_2^{\mathrm{L}},N_3^{\mathrm{L}}\right)+\ln\left(\frac{x_{1,m}\left(N_1^{\mathrm{L}}+N_2^{\mathrm{L}}+N_3^{\mathrm{L}}\right)}{\left\langle V\right\rangle_{T,p,N_1^{\mathrm{L}}-1,N_2^{\mathrm{L}},N_3^{\mathrm{L}}}}\right) \quad (2)
\end{aligned}$$

where Λ_1 is the thermal de Broglie wavelength of component 1 (which is only a function of T), v_1^{S} is the molar volume of component 1 in the solid phase, Z_1^{ig} is the ideal gas configuration integral of a single molecule (which is also only a function of T), $\left\langle V\right\rangle_{T,p,N_1^{\mathrm{L}}-1,N_2^{\mathrm{L}},N_3^{\mathrm{L}}}$ is the ensemble average volume of the liquid solution phase system in the absence of the solute molecule that is being coupled/decoupled to the system (to compute $\beta\mu_{1,m}^{\mathrm{L,res}}$), and $x_{1,m}$ is the mole fraction of component 1 in the liquid solution phase. Since the two phases are at the same temperature at equilibrium, Λ_1 and Z_1^{ig} conveniently cancel out of the equality. Equation (2) may then be solved for the mole fraction of component 1 in the liquid solution phase at equilibrium

$$\begin{aligned}
\ln x_{1,m} = &\left[\beta\mu_1^{\mathrm{S,res}}\left(T,p\right)-\beta\mu_{1,m}^{\mathrm{L,res}}\left(T,p,N_1^{\mathrm{L}},N_2^{\mathrm{L}},N_3^{\mathrm{L}}\right)\right] \\
&-\ln v_1^{\mathrm{S}}\left(T,p\right)+\ln\left(\frac{\left\langle V\right\rangle_{T,p,N_1^{\mathrm{L}}-1,N_2^{\mathrm{L}},N_3^{\mathrm{L}}}}{N_1^{\mathrm{L}}+N_2^{\mathrm{L}}+N_3^{\mathrm{L}}}\right) \quad (3)
\end{aligned}$$

Equation (3) is rigorously correct and forms the basis of the present study.

Next, let us assume that the equilibrium solubility of component 1 is sufficiently small that it may be considered infinitely dilute. The infinite dilution limit is achieved in a molecular simulation free energy calculation by adding a single solute molecule ($N_1^{\mathrm{L}} = 1$) to the binary solvent. If $N_2^{\mathrm{L}} + N_3^{\mathrm{L}} \gg 1$, eq. (3) reduces to

$$\ln x_{1,m} = \left[\beta\mu_1^{\mathrm{S,res}}\left(T,p\right) - \beta\mu_{1,m}^{\mathrm{L,res},\infty}\left(T,p,x_2^{\mathrm{b}},x_3^{\mathrm{b}}\right)\right] + \ln\frac{v_m^{\mathrm{b}}\left(T,p,x_2^{\mathrm{b}},x_3^{\mathrm{b}}\right)}{v_1^{\mathrm{S}}\left(T,p\right)} \tag{4}$$

where the superscript ∞ is used to indicate the infinite dilution limit, v_m^{b} is the solute free molar volume of the binary solvent, and x_2^{b} and x_3^{b} are the solute free mole fractions of component 2 and 3 (the binary solvent), respectively. Since the solute is infinitely dilute, eq. (4) may be re-arranged to obtain the equilibrium solubility in molar concentrations ($c_{1,m}$ in moles/volume)

$$\ln c_{1,m} = \ln\left(\frac{x_{1,m}}{v^{\mathrm{b}}\left(T,p,x_2^{\mathrm{b}},x_3^{\mathrm{b}}\right)}\right) = \left[\beta\mu_1^{\mathrm{S,res}}\left(T,p\right) - \beta\mu_{1,m}^{\mathrm{L,res},\infty}\left(T,p,x_2^{\mathrm{b}},x_3^{\mathrm{b}}\right)\right] - \ln v_1^{\mathrm{S}}\left(T,p\right) \tag{5}$$

In the study of multicomponent solutions, it is often most interesting to study the excess properties of the solution [14]. Excess properties of solutions are the thermodynamic properties of a solution relative to those of an ideal solution at the same T, p, and composition. In an ideal solution, all excess properties are therefore zero. Following the work of O'Connell and co-workers [32, 33], the expression for the log excess solubility of component 1 (in mole fractions) in the binary solvent ($\ln x_{1,m}^{\mathrm{E}}$) is given by the expression

$$\ln x_{1,m}^{\mathrm{E}} = \ln x_{1,m} - \left[x_2^{\mathrm{b}}\ln x_{1,2} + x_3^{\mathrm{b}}\ln x_{1,3}\right] \tag{6}$$

where $x_{1,m}$, $x_{1,2}$, and $x_{1,3}$ are the equilibrium solubility of component 1 in the binary solvent, in pure component 2, and in pure component 3, respectively, and the terms in brackets correspond to the solubility if components 2 and 3 mixed to form an ideal solution. In the present study we will likewise compute the log excess solubility of component 1 (in moles/volume or mass/volume) in the binary solvent ($\ln c_{1,m}^{\mathrm{E}}$) as

$$\ln c_{1,m}^{\mathrm{E}} = \ln c_{1,m} - \left[x_2^{\mathrm{b}} \ln c_{1,2} + x_3^{\mathrm{b}} \ln c_{1,3}\right] \tag{7}$$

where $c_{1,m}$, $c_{1,2}$, and $c_{1,3}$ are the equilibrium solubility of component 1 in the binary solvent, in pure component 2, and in pure component 3, respectively, and the terms in brackets again correspond to the solubility if components 2 and 3 mixed to form an ideal solution. In eq. (7) the concentrations may either be expressed in units of moles/volume or mass/volume; the conversion factor of the molecular weight of the solute cancels out of the expression.

For an infinitely dilute solute, eqs. (4) to (7) lead to the following expressions

$$\begin{aligned}\ln x_{1,m}^{\mathrm{E}} = -\beta\mu_{1,m}^{\mathrm{L,res,\infty}}\left(T,p,x_2^{\mathrm{b}},x_3^{\mathrm{b}}\right) + \left[x_2^{\mathrm{b}}\beta\mu_{1,2}^{\mathrm{L,res,\infty}}\left(T,p\right) + x_3^{\mathrm{b}}\beta\mu_{1,3}^{\mathrm{L,res,\infty}}\left(T,p\right)\right] \\ + \ln v^{\mathrm{b}}\left(T,p,x_2^{\mathrm{b}},x_3^{\mathrm{b}}\right) - \left[x_2^{\mathrm{b}}\ln v_2\left(T,p\right) + x_3^{\mathrm{b}}\ln v_3\left(T,p\right)\right]\end{aligned} \tag{8}$$

and

$$\ln c_{1,m}^{\mathrm{E}} = -\beta\mu_{1,m}^{\mathrm{L,res,\infty}}\left(T,p,x_2^{\mathrm{b}},x_3^{\mathrm{b}}\right) + \left[x_2^{\mathrm{b}}\beta\mu_{1,2}^{\mathrm{L,res,\infty}}\left(T,p\right) + x_3^{\mathrm{b}}\beta\mu_{1,3}^{\mathrm{L,res,\infty}}\left(T,p\right)\right] \tag{9}$$

where $\beta\mu_{1,2}^{\mathrm{L,res,\infty}}$ and $\beta\mu_{1,3}^{\mathrm{L,res,\infty}}$ correspond to the dimensionless infinite dilution residual chemical potential of component 1 in pure component 2 and 3, respectively, and v_2 and v_3 correspond to the molar volume of pure component 2 and 3, respectively. Notice that the properties of the pure solid solute cancel out of eqs. (8) and (9).

It is interesting to point out that $\ln c_{1,m}^{\mathrm{E}}$ is equivalent to the change in the negative dimensionless infinite dilution residual chemical potential of component 1 upon mixing [14], $-\Delta\left(\beta\mu_{1,m}^{\mathrm{L,res,\infty}}\right)^{\mathrm{mix}}$. Furthermore, $\ln x_{1,m}^{\mathrm{E}}$ is equivalent to the sum of $-\Delta\left(\beta\mu_{1,m}^{\mathrm{L,res,\infty}}\right)^{\mathrm{mix}}$ and the change in the log molar volume of the solvent upon mixing, $\Delta\left(\ln v_m^{\mathrm{b}}\right)^{\mathrm{mix}}$. The decomposition of $\ln x_1^{\mathrm{E}}$ and $\ln c_{1,m}^{\mathrm{E}}$ is interesting because $\Delta\left(\ln v_m^{\mathrm{b}}\right)^{\mathrm{mix}}$ is a solvent dependent property (independent of the solute), whereas $-\Delta\left(\beta\mu_{1,m}^{\mathrm{L,res,\infty}}\right)^{\mathrm{mix}}$ results from intermolecular interactions (solute–solvent and solvent–solvent) and the size of the cavity formed by the solute in the solvent. Additionally, it is interesting to note that if components 2 and 3 form an ideal solution, $x_{1,m}$ and $c_{1,m}$ are exponential functions

of the solvent composition whereas $\beta\mu_{1,m}^{\mathrm{L,res},\infty}$ is a linear function of the solvent composition [32].

Using eqs. (8) and (9), $\ln x_1^{\mathrm{E}}$ and $\ln c_1^{\mathrm{E}}$ may readily be computed by performing a series of conventional molecular simulation free energy calculations for the solute (component 1) at infinite dilution in pure component 2 and 3, and at the binary solvent concentrations of interest. Equations (8) and (9) assume that component 1 is infinitely dilute, typically corresponding to values of $x_{1,m} \leq 0.01$ [33]. However, the work of Ellegaard et al. [33] showed that this assumption frequently works well when modeling $\ln x_1^{\mathrm{E}}$ of solid pharmaceutical solutes in binary mixtures for solute mole fractions in excess of 0.01. The application of eqs. (4) and (5) to compute the relative solubility of a solute in two pure solvents is discussed in the appendix to this manuscript.

IV. COMPUTATIONAL DETAILS

Force Fields

Acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine (the solutes, see fig. 1) were all modeled with the General AMBER Force Field (GAFF) [39, 40]. To obtain partial charges for GAFF, the geometry of each molecule was optimized followed by a single point energy calculation at the MP2/cc-pVTZ level of theory in a self-consistent reaction field (SCRF) using the polarizable continuum model (PCM) for water using the default parameters provided in Gaussian 09 [41]. Partial charges were then obtained from the electrostatic potential using the restrained electrostatic potential (RESP) [42, 43] method in ANTECHAMBER (part of the AMBER 11 simulation suite) [44, 45].

Since the ultimate goal of the present study is to predict $\ln x_1^{\mathrm{E}}$ and based on our previous experience [46], solvent force fields were chosen which accurately model the excess thermodynamic properties of the binary solvent. The force field for ethanol was constructed as an extension of the united-atom methanol model of Weerasinghe and Smith [47] which was optimized to reproduce the composition dependent Kirkwood-Buff integrals (and hence the excess Gibbs free energy) of aqueous methanol solutions at ambient conditions when used with the SPC/E water model [48]. The extension of the methanol force field was performed so as to remain consistent with the original methanol parameterization. The molecular geometry, angle bending force constants, and

the dihedral potential for ethanol were all adopted from the united-atom version of the Transferable Potential for Phase Equilibria (TraPPE-UA) force field [49]. Bond vibration constants were adopted from the united-atom GROMOS87 force field (gmx.ff) in GROMACS 4.5.5 [50, 51]. The united-atom Lennard-Jones parameter for the additional methyl site was taken from the united-atom GROMOS96 force field [52]. The partial charges of the original methanol model were retained. The SPC/E water model was adopted for water.

All of Gaussian 09 input files and GROMACS force field files used in the present study may be found in the Supplementary Material of this manuscript [53].

Simulation Details

System Preparation

System preparation began with structures of the solvents (water and ethanol) and solutes (acetaminophen, acetanilide, benzocaine, caffeine, and phenacetin) in PDB and .mol2 format. These were generated by exporting 2D structures of the molecules from MarvinSketch [54], then generating 3D structures via OpenEye's Python toolkits[55] and Omega[56, 57] and writing these out. GROMACS topology and coordinate files were then generated from the .mol2 files by parameterizing with ANTECHAMBER using GAFF and partial charges as described above, then using acpype [58] to convert to GROMACS format.

Solvent boxes were set as cubic, with a box edge length of 32 Å on each side. Solvated systems containing water/ethanol mixtures at specified mole fractions containing each solute were then generated. These were generated using Packmol [59] and a custom Python script, and (solute free) solvent mole fractions in increments of 10% ranging from pure water (0% ethanol) to zero water (100% ethanol). We used Packmol to fill a box of the target volume (32 Å on a side) with one solute molecule, plus a number of solvent molecules estimated by the ideal molar volume of the binary solvent mixture ($v^{\mathrm{b,id}} = x_2^{\mathrm{b}} v_2 + x_3^{\mathrm{b}} v_3$, where $v^{\mathrm{b,id}}$ is the estimated molar volume of the binary solvent). The number of solvent molecules was rounded up to the nearest integer, the Packmol tolerance for the minimum distance between atoms was set to 2.05 Å, and the molecules were initially placed randomly. Following placement of molecules via Packmol, output coordinates (in PDB format) were converted to GROMACS .gro format for simulation. All of the .gro files may

be found in the Supplementary Material of this manuscript [53].

Because the AMBER force fields use different intramolecular 1-4 scaling factors than TraPPE-UA, we used the AMBER intramolecular 1-4 scaling factors (1/2 for Lennard-Jones, 5/6 for Coulomb) were used for the studied solutes. Consistent with TraPPE-UA, no intramolecular 1-4 interactions were used for ethanol.

Simulations

Simulations were conducted in a developmental version of GROMACS 4.5.3 customized for free energy calculations [60], and consisted of a minimization, followed by a constant volume equilibration, constant pressure equilibration, an affine transformation to ensure the box volume corresponds to the average volume over the constant pressure equilibration, and a constant volume "production" simulation. Minimization consisted of up to 1500 steps of steepest descents minimization. For dynamics, the integrator was the GROMACS "stochastic dynamics" integrator corresponding to Langevin dynamics, with a 2 fs timestep, with 5000 steps for the initial equilibration period, followed by an additional 100,000 steps for the NpT equilibration. Production simulations were 5 ns for each alchemical λ value (and minimization and equilibration were also done separately for each λ). The first 100 ps of production was discarded as additional equilibration. Cut-offs, etc., were generally as described elsewhere [61], with PME used for electrostatics and long-range Lennard-Jones dispersion corrections applied to the energy and pressure [62].

Free energy simulations were used to compute $\beta\mu_{1,m}^{\mathrm{L,res},\infty}$ (a.k.a. solvation free energies) in each different binary mixture, in essentially an identical manner as applied previously to hydration free energy calculations [61]. The simulations used 20 λ values, first turning off the solute electrostatic interactions with the environment (Coulomb λ = 0.0, 0.25, 0.75, and 1.0), then turning solute Lennard-Jones parameters to zero (Lennard-Jones λ = 0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0). To close the thermodynamic cycle, the free energy of restoring these Lennard-Jones parameters in vacuum was also computed, thus computing the total transfer free energy from an ideal gas to solution, and thereby $\beta\mu_{1,m}^{\mathrm{L,res},\infty}$.

Analysis

Free energies were computed via the multistate Bennett acceptance ratio (MBAR) [63] method via our own Python analysis script which is distributed along with PyMBAR [64]. Error bars for the free energy difference between each pair of λ values were taken as the value provided by PyMBAR, and standard error propagation was used to estimate the error in the overall computed chemical potential.

The molar volume for each solvent mixture was computed as the average volume of the simulation box over the pure solvent simulation, and similarly for the excess enthalpy [65]. Five independent simulations using the same protocols as above were performed, and the computed values were the average across these five trials, while uncertainty estimates were computed as the standard error in the mean (the standard deviation divided by the square root of the number of trials).

V. RESULTS AND DISCUSSION

Binary Solvent

In the present study, we seek to understand the ability of molecular simulation to compute $\ln x_1^{\mathrm{E}}$ devoid of experimental solubility data. Recall from eq. (6) that $\ln x_1^{\mathrm{E}}$ is equal to the *actual* log solubility of the solute (component 1) in the binary solvent relative to the *ideal* log solubility of the solute in the binary solvent which would result if components 2 and 3 (which make up the binary solvent) mixed to form an ideal solution. Therefore, it is informative to first consider our ability to accurately model the binary solvent. Throughout this study, components 2 and 3 will refer to water and ethanol, respectively.

Figure 2 shows the solute-free composition dependence of the binary solvent molar volume (v^{b}), excess molar volume ($v^{\mathrm{b,E}}$), and excess molar enthalpy ($h^{\mathrm{b,E}}$) computed via molecular simulation in this study and from experiment [66, 67]. $v^{\mathrm{b,E}}$ and $h^{\mathrm{b,E}}$ are computed as the actual value of the solution relative that of an ideal solution, $v^{\mathrm{b,E}} = v^{\mathrm{b}} - \left[x_2^{\mathrm{b}} v_2 + x_3^{\mathrm{b}} v_3\right]$ and $h^{\mathrm{b,E}} = h^{\mathrm{b}} - \left[x_2^{\mathrm{b}} h_2 + x_3^{\mathrm{b}} h_3\right]$, where the terms in brackets correspond to v^{b} and h^{b} if water and ethanol mixed to form an ideal solution, and h_2 and h_3 are the molar enthalpy of pure water and

ethanol, respectively. We find that the computed pure component molar volume of ethanol is slightly too large by 2.8% (or $1.65\,\mathrm{cm}^3/\mathrm{mol}$), and as a consequence, v^{b} is slightly over-predicted for ethanol-rich concentrations. For $v^{\mathrm{b,E}}$, we find that the simulation and experimental results are in excellent agreement, with the simulations systematically predicting values too negative. However, qualitatively, the simulations capture the key experimental trends. Likewise, we find that $h^{\mathrm{b,E}}$ from the simulation results are in good qualitative agreement with experiment. The simulations are able to reasonably capture the the region of negative $h^{\mathrm{b,E}}$, suggesting that the models reasonably capture the formation of hydrogen bonds between water and ethanol [14]. The values of $h^{\mathrm{b,E}}$ for ethanol-rich concentrations are in excellent agreement with experiment while $h^{\mathrm{b,E}}$ for water-rich concentrations are underestimated. Note that $h^{\mathrm{b,E}}$ is plotted in units of J/mol, suggesting that the agreement is quite good considering typically hydrogen bond strengths are between 8–40 kJ/mol [14].

Overall, we find that the simulation results for v^{b}, $v^{\mathrm{b,E}}$, and $h^{\mathrm{b,E}}$ are in good agreement with experiment, suggesting that the molecular models reasonably accurately capture the thermodynamic properties of the binary solvent.

Excess Solubility

First consider $\ln x_1^{\mathrm{E}}$ of acetanilide, acetaminophen, and phenacetin as shown in fig. 3. As seen in fig. 1, the molecules are all chemically similar. All three molecules have a scaffold consisting of *N*-phenylethanamide. Acetanilide contains no other functional groups. On the other hand, acetaminophen and phenacetin contain a hydroxy and ethoxy group attached to the para (or 4) position, respectively.

We find that for acetanilide, the simulation results are in excellent quantitative agreement with experiment. The simulation and experimental results overlap, and the simulation results accurately predict the extrema at approximately $x_2^{\mathrm{b}} = 0.7$.

For acetaminophen, the simulation results and experiment are in good agreement. As compared to acetanilide, acetaminophen has an additional hydroxy group attached to the para position, which is capable of both accepting and donating hydrogen bonds. Comparing the experimental data for acetanilide and acetaminophen, we find that the hydroxy groups has only a minute effect on $\ln x_1^{\mathrm{E}}$;

the quantitative results are similar and the extrema remains at approximately $x_2^{\mathrm{b}} = 0.7$. On the other hand, the simulation results predict the location of the extrema to be near $x_2^{\mathrm{b}} = 0.6$, with the quantitative value of the extrema over-predicted by approximately one log unit. Given the excellent agreement of the simulation results for acetanilide with experiment, this would suggest that the hydroxy group be reparameterized (electrostatic and/or Lennard-Jones). This result is consistent with our previous work demonstrating the sensitivity of the hydration free energy to the charge parameterization scheme [68] and specifically for the case of acetaminophen [46], in addition to potential systematic errors with GAFF hydroxy groups [69, 70], and future work should test whether it is corrected by updated GAFF-DC hydroxyl parameters [71].

Likewise, for phenacetin, the simulation results and experiment are in good agreement. For this case, phenacetin has an additional ethoxy group attached to the para position as compared to acetanilide. Comparing the experimental data for acetanilide and phenacetin, we find that both have an extrema at at approximately $x_2^{\mathrm{b}} = 0.7$, with the addition of the ethoxy group increasing the value of $\ln x_1^{\mathrm{E}}$ at the extrema by roughly 0.5 log unit. The simulation result also predict the extrema at $x_2^{\mathrm{b}} = 0.7$, however the simulation results consistently over-predict $\ln x_1^{\mathrm{E}}$; at the extrema $\ln x_1^{\mathrm{E}}$ is over predicted by approximately 1.5 log units. Nonetheless, the results are in good agreement, and we find that the simulations correctly predict the trend of increasing $\ln x_1^{\mathrm{E}}$ in going from acetanilide to phenacetin. Just as with acetaminophen, the excellent agreement of the simulation results for acetanilide with experiment suggest that the ethoxy group be reparameterized (electrostatic and/or Lennard-Jones).

Next, consider $\ln x_1^{\mathrm{E}}$ of benzocaine and caffeine as shown in fig. 4. For both systems, we find that the simulation results are in good qualitative agreement with experiment. The simulation results are able to both predict an extrema in $\ln x_1^{\mathrm{E}}$ in addition to the corresponding value of x_2^{b}. However, we find that the simulation results over-predict $\ln x_1^{\mathrm{E}}$, with the extrema over-predicted by approximately 1.5–2 log units. This shortcoming with caffeine is consistent with our previous work [70] wherein caffeine was an outlier in a set of 23 small organic molecule hydration free energy calculations, where the simulation predicted hydration free energy was lower than experiment by approximately 5 kcal/mol (8.4 $k_{\mathrm{B}}T$).

To further allow for a quantitative comparison of the simulation results with experiment, the values of $\ln x_1^{\mathrm{E}}$ from experiment were fit to a third-order Redlich-Kister expansion of the form [14]

$$\ln x_{1,m}^{\mathrm{E}} = x_2^{\mathrm{b}} x_3^{\mathrm{b}} \left[A + B\left(x_2^{\mathrm{b}} - x_3^{\mathrm{b}}\right) + C\left(x_2^{\mathrm{b}} - x_3^{\mathrm{b}}\right)^2 \right] \tag{10}$$

where A, B, and C are model constants which were found for each system by least squares regression. The Redlich-Kister expansion is a series expansion of $\ln x_{1,m}^{\mathrm{E}} / \left(x_2^{\mathrm{b}} x_3^{\mathrm{b}}\right)$ in $x_2^{\mathrm{b}} - x_3^{\mathrm{b}}$, where the term $x_2^{\mathrm{b}} x_3^{\mathrm{b}}$ ensures that the equation exhibits correct limiting behavior (i.e. $\ln x_{1,m}^{\mathrm{E}} = 0$ when x_2^{b} or x_3^{b} are 1). Equation (10) is equivalent to the Jouyban-Acree model with solvent mole fractions replaced with solvent volume fractions; the Jouyban-Acree model is often used to correlate $\ln x_{1,m}$ in binary solvents [8]. The resulting fits are represented as solid lines in figs. 3 and 4. In all cases the third-order Redlich-Kister expansion fits the experimental data well with R^2 values greater than 0.9, and effectively smooths the data allowing for interpolation to the binary solvent concentrations at which molecular simulations were performed.

A plot of $\ln x_{1,m}^{\mathrm{E}}$ computed via molecular simulation $\left(\ln x_{1,m}^{\mathrm{E,\,sim}}\right)$ versus $\ln x_{1,m}^{\mathrm{E}}$ at the same binary solvent concentration from eq. (10) fit to experimental data $\left(\ln x_{1,m}^{\mathrm{E,\,exp}}\right)$ is shown in fig. 5. Figure 5 contains all of the data from figs. 3 and 4 at the binary solvent concentrations studied via molecular simulation except for $x_2^{\mathrm{b}} = 0$ and $x_2^{\mathrm{b}} = 1$ where by definition $\ln x_1^{\mathrm{E}} = 0$. Overall, we find that the data is reasonably correlated with a straight line through the origin with a slope of 1.44 ± 0.08 , yielding a value of $R^2 = 0.67$. The fact that the slope is greater than unity is in agreement with the observation in figs. 3 and 4 that the simulation results agree better with experimental results the lower the value of $\ln x_{1,m}^{\mathrm{E}}$. The linear relation could serve as a means of improving predictions for other solutes in water/ethanol mixtures.

Moreover, in the Methodology section, we derived expressions for the excess solubility in units of mole fractions ($\ln x_{1,m}^{\mathrm{E}}$, eq. (8)) and in concentration units of moles/volume or mass/volume ($\ln c_{1,m}^{\mathrm{E}}$, eq. (9)). We found that $\ln c_{1,m}^{\mathrm{E}}$ was only dependent on $-\Delta\left(\beta\mu_{1,m}^{\mathrm{L,res,}\infty}\right)^{\mathrm{mix}}$, which results from intermolecular interactions and the size of the cavity formed by the solute in the solvent. On the other hand, $\ln x_{1,m}^{\mathrm{E}}$ additionally contained the term $\Delta\left(\ln v_m^{\mathrm{b}}\right)^{\mathrm{mix}}$, which is a solvent dependent term (independent of the solute). To understand the contribution of $-\Delta\left(\beta\mu_{1,m}^{\mathrm{L,res,}\infty}\right)^{\mathrm{mix}}$ and $\Delta\left(\ln v_m^{\mathrm{b}}\right)^{\mathrm{mix}}$, $\ln x_{1,m}^{\mathrm{E}}$ and $\ln c_{1,m}^{\mathrm{E}}$ for acetaminophen are both shown in fig. 6. We find that $-\Delta\left(\beta\mu_{1,m}^{\mathrm{L,res,}\infty}\right)^{\mathrm{mix}}$ plays a dominant role, with $\Delta\left(\ln v_m^{\mathrm{b}}\right)^{\mathrm{mix}}$ only making a minute contribution. The results for acetaminophen are representative of the systems studied. This does not suggest that accurately computing the molar volume of the solvent is not important. The average inter-

molecular interaction distance between solvent molecules is related to to molar volume, which is important when computing the residual chemical potential at infinite dilution. However, the dominant contribution to $\ln x_{1,m}^{\mathrm{E}}$ is from the the solute residual chemical potential.

Lastly, in fig. 3 we additionally compare our simulations for acetaminophen to the state-of-the-art 2005 revision of the MOSCED model [10, 72, 73]. MOSCED is a semi-theoretical model used to estimate infinite dilution activity coefficients, and describes each molecule using five adjustable parameters. The interested reader is directed to the original publication for a description of its formulation and basis [72]. In the 2005 revision, the molecular descriptors of 132 organic solvents, water, and 5 permanent gases were fit to a set of 6,441 experimental infinite dilution activity coefficient data points [10, 74]. In addition, parameters for 26 organic solids were regressed. For the case of acetaminophen, the parameters were obtained by regressing to solubility data in 19 pure solvents, where the set of solvents included water and ethanol [10, 75]. MOSCED may therefore be seen as a correlation of experimental data in pure solvents, and is not truly a predictive method (as compared to our molecular simulation results). To compute $\ln x_{1,m}^{\mathrm{E}}$ in a binary water/ethanol solvent, MOSCED was used to compute the binary interaction parameters necessary for Wilson's equation [14, 46, 76, 77]. Wilson's equation is a semi-theoretical modification of the Flory-Huggins equation, where the point of departure is Wilson's use of local compositions. Wilson's equation contains two binary interaction parameters for each binary pair in a multicomponent mixture [14]. The results are shown graphically in fig. 3 as a dashed line. We find that MOSCED is in very good agreement with the experimental data.

This is a very interesting result. Given its parameterization, we would expect the infinite dilution activity coefficient of acetaminophen in pure water and in pure ethanol computed by MOSCED to be reasonably accurate. Given reasonable properties at infinite dilution in pure solvents, Wilson's equation is then able to accurately extrapolate to binary solvent compositions. Adopting a similar scheme using molecular simulation results would pose an intriguing possibility. However this is not without great challenge. Computing the activity coefficient of acetaminophen with respect to a Lewis-Randall standard state requires knowledge of the fugacity of pure (subcooled) liquid acetaminophen at the temperature of interest.

VI. SUMMARY AND CONCLUSION

We developed a framework to compute the excess solubility of small molecular solids (such as pharmaceutical solids) in binary solvents via molecular simulation. The use of molecular simulation is advantageous as it is a purely predictive method and may further be used to gain a molecular level understanding of the phenomenon. The present study considered the solid solutes acetanilide, acetaminophen, phenacetin, benzocaine and caffeine in binary water/ethanol mixtures. The solutes were all modeled using the General AMBER Force Field (GAFF) which is capable of modeling a wide range of solutes of various complexities devoid of experimental data. In all cases, the molecular simulations were in good qualitative agreement with experiment. The simulations were able to predict the existence of an extrema in the the excess solubility at intermediate binary solvent concentrations, captured the qualitative behavior (or relative solubility trend) of each solute as a function of the binary solvent composition, and was able to capture the trend of increasing excess solubility with changing functional groups in acetanilide and phenacetin.

While the simulation results were not always in excellent quantitative agreement with experiment, we found that the simulation results and experiment were linearly correlated, with the deviation increasing as the magnitude of the excess solubility increases. The resulting trend could be used to develop a calibration curve to improve predictions involving other solutes in the same water/ethanol binary solvent. We additionally found that while the excess solubility is composed of a term involving the change in the dimensionless infinite dilution residual chemical potential of the solute upon mixing and a term involving the molar volume of the solvent, the residual chemical potential dominates the numerical result.

Our results suggest that molecular simulation with conventional force fields may serve a valuable role for solvent selection in the drug development process. There is still room for improvement in the parameterization of the molecular models, and the present framework suggests a means by which this may be accomplished. In the present study we found that GAFF was able to accurately predict the excess solubility for acetanilide. However, noticeable deviations were observed for acetaminophen and phenacetin, which are both acetanilide deviates with a hydroxy and ethoxy group substituted in the 4 (or para) position, respectively. The results therefore suggest possible problems with GAFF hydroxy and ethoxy parameters, and a need for re-parameterization. Parameterizing force fields around excess solubility is particularly attractive because it may readily be

computed from experimental solubility data and is directly related to the residual chemical potential of the solute. Extracting the residual chemical potential (or solvation free energy) directly from experimental data may be challenging for solid solutes because one needs to know the pure (sub-cooled) liquid vapor pressure of the solute at the temperature of interest (see for example ref. 78).

Lastly, we used the MOSCED model with Wilson's equation to compute the excess solubility of acetaminophen in a binary water/ethanol mixture. Given reasonable predictions of infinite dilution properties in pure solvents, Wilson's equation was able to accurately extrapolate the results to binary mixture compositions. This suggests an intriguing possibility for efficiently using molecular simulation (which is a predictive method) to predict excess solubility which would require only a limited number of calculations in pure solvents. Before this can be accomplished, however, an accurate method to compute the fugacity (or vapor pressure or chemical potential) of pure (sub-cooled) liquid solutes at the temperature of interest is required.

Acknowledgments

A.S.P. gratefully acknowledges a fellowship from the Arthur J. Schmitt Foundation (Notre Dame) and start-up funding from the College of Engineering and Computing at Miami University. Computing support for all of the Gaussian 09 calculations was provided by Notre Dame's Center for Research Computing. DLM acknowledges support of the National Institutes of Health (1R15GM096257-01A1) and the National Science Foundation (CHE-1352608), as well as the Louisiana Board of Regents.

Appendix A: Relative Solubility Predictions

For the solubility of a solid solute (component 1) in pure solvent α, expressions analogous to eqs. (4) and (5) may be derived where we again assume that the solubility of component 1 is sufficiently small so as to be considered infinitely dilute

$$\ln x_1^{\alpha} = -\beta\mu_1^{\alpha,\mathrm{res},\infty}\left(T,p\right) + \ln v^{\alpha}\left(T,p\right) + \left[\beta\mu_1^{\mathrm{S,res}}\left(T,p\right) - \ln v_1^{\mathrm{S}}\left(T,p\right)\right] \tag{A1}$$

$$\ln c_1^{\alpha} = -\beta\mu_1^{\alpha,\mathrm{res},\infty}\left(T,p\right) + \left[\beta\mu_1^{\mathrm{S,res}}\left(T,p\right) - \ln v_1^{\mathrm{S}}\left(T,p\right)\right] \tag{A2}$$

where the superscript α is used to indicate properties of solvent α, $\mu_1^{\alpha,\mathrm{res},\infty}$ is the infinite dilution residual chemical potential of component 1 in solvent α computed via molecular simulation by adding a single solute molecule to solution, and the terms in brackets correspond to properties of the pure solid solute. It follows that the solubility of component 1 in solvent α *relative* to solvent γ may readily be computed via molecular simulation

$$\ln \frac{x_1^{\alpha}}{x_1^{\gamma}} = \beta\mu_1^{\gamma,\mathrm{res},\infty}\left(T,p\right) - \beta\mu_1^{\alpha,\mathrm{res},\infty}\left(T,p\right) + \ln \frac{v^{\alpha}\left(T,p\right)}{v^{\gamma}\left(T,p\right)} \tag{A3}$$

$$\ln \frac{c_1^{\alpha}}{c_1^{\gamma}} = \beta\mu_1^{\gamma,\mathrm{res},\infty}\left(T,p\right) - \beta\mu_1^{\alpha,\mathrm{res},\infty}\left(T,p\right) \tag{A4}$$

where the pure component solute properties conveniently cancel out of the expressions. This set of equations provides an efficient means to predict the relative solubility of a solute in pure solvents via molecular simulation devoid of experimental data and may prove useful for solvent selection processes. Equation (A4) is equivalent to the expression used by Essex et al. [79] to compute relative partition coefficients and was used previously by Paluch et al. [80] to predict relative solubilities. If a single experimental solubility is known, either eq. (A3) or eq. (A4) allows for the properties of the pure solid solute to be obtained by performing a single molecular simulation free energy calculation in the liquid phase [80].

[1] W. L. McCabe, J. C. Smith, and P. Harriott, *Unit Operations of Chemical Engineering* (McGraw-Hill, New York, NY, 2005), 7th ed.

[2] D. A. Dahlstrom, R. C. Bennett, R. C. Emmett, Jr., P. Harriott, T. Laros, W. Leung, C. McCleary, S. A. M. B. Morey, J. Y. Oldshue, G. Priday, et al., in *Perry's Chemical Engineers' Handbook*, edited by R. H. Perry and D. W. Green (McGraw-Hill Professional, New York, NY, 1997), 7th ed.

[3] C. Wibowo, Chem. Eng. Progress **110**, 37 (2014).

[4] S. Tanveer, Y. Hao, and C.-C. Chen, Chem. Eng. Progress **110**, 37 (2014).

[5] K. Grodowska and A. Parczewski, Acta Pol. Pharm. **67**, 3 (2010).

[6] D. J. C. Constable, C. Jimenez-Gonzalez, and R. K. Hendersen, Org. Process Res. Dev. **11**, 133 (2007).

[7] R. Liu, ed., *Water-Insoluble Drug Formulation* (CRC Press, Boca Raton, FL, 2008), 2nd ed.

[8] A. Jouyban, *Handbook of Solubility Data for Pharmaceuticals* (CRC Press, Boca Raton, FL, 2010).

[9] C. Chen and P. A. Crafts, Ind. Eng. Chem. Res. **45**, 4816 (2006).

[10] M. J. Lazzaroni, D. Bush, C. A. Eckert, T. C. Frank, S. Gupta, and J. D. Olson, Ind. Eng. Chem. Res. **44**, 4075 (2005).

[11] J. Cassens, F. Ruether, K. Leonhard, and G. Sadowski, Fluid Phase Equilib. **299**, 161 (2010).

[12] I. G. Economou, Ind. Eng. Chem. Res. **41**, 953 (2002).

[13] T. Spyriouni, X. Krokidis, and I. G. Economou, Fluid Phase Equilib. **302**, 331 (2011).

[14] J. M. Prausnitz, R. N. Lichtenthaler, and E. G. de Azevedo, *Molecular Thermodynamics of Fluid-Phase Equilibria* (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1986), 2nd ed.

[15] S. Gracin, T. Brinck, and A. C. Rasmuson, Ind. Eng. Chem. Res. **41**, 5114 (2002).

[16] H. Hojjati and S. Rohani, Org. Process Res. Dev. **10**, 1110 (2006).

[17] A. Diedrichs and J. Gmehling, Ind. Eng. Chem. Res. **50**, 1757 (2011).

[18] H. E. Gonzalez, J. Abildskov, R. Gani, P. Rousseaux, and B. Le Bert, AIChE J. **53**, 1620 (2007).

[19] A. A. Mustaffa, R. Gani, and G. M. Kontogeorgis, Fluid Phase Equilib. **366**, 24 (2014).

[20] A. Klamt, F. Eckert, M. Hornig, M. E. Beck, and T. Burger, J. Comput. Chem. **23**, 275 (2002).

[21] C. M. Hsieh, S. Wang, S. T. Lin, and S. I. Sandler, J. Chem. Eng. Data **56**, 936 (2011).

[22] A. Klamt, F. Eckert, and W. Arlt, Annu. Rev. Chem. Biomol. Eng. **1**, 101 (2010).

[23] Z. Xue, T. Mu, and J. Gmehling, Ind. Eng. Chem. Res. **51**, 11809 (2012).

[24] M. Ferrario, G. Ciccotti, E. Spohr, T. Cartailler, and P. Turq, J. Chem. Phys. **117**, 4947 (2002).

[25] E. Sanz and C. Vega, J. Chem. Phys. **126**, 014507 (2007).

[26] A. S. Paluch, S. Jayaraman, J. K. Shah, and E. J. Maginn, J. Chem. Phys. **133**, 124504 (2010).

[27] F. Moucka, M. Lisal, J. Skvor, J. Jirsak, I. Nezbeda, and W. R. Smith, J. Phys. Chem. B **115**, 7849 (2011).

[28] M. J. Schnieders, J. Baltrusaitis, Y. Shi, G. Chattree, L. Q. Zheng, W. Yang, and P. Y. Ren, J. Chem. Theory Comput. **8**, 1721 (2012).

[29] J. L. Argones, E. Sanz, and C. Vega, J. Chem. Phys. **136**, 244508 (2012).

[30] A. S. Paluch and E. J. Maginn, AIChE J. **59**, 2647 (2013).

[31] A. Ahmed and S. I. Sandler, J. Chem. Theory Comput. **9**, 2389 (2013).

[32] J. P. O'Connell and J. M. Prausnitz, Ind. Eng. Chem. Fundam. **3**, 347 (1964).

[33] M. D. Ellegaard, J. Abildskov, and J. P. O'Connell, Ind. Eng. Chem. Res. **49**, 11620 (2010).

[34] K. S. Shing and S. T. Chung, J. Phys. Chem. **91**, 1674 (1987).

[35] D. A. Kofke and P. T. Cummings, Mol. Phys. **92**, 973 (1997).

[36] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, J. Chem. Phys. **119**, 5740 (2003).

[37] T. L. Hill, *Statistical Mechanics: Principles and Selected Applications* (Dover Publications, Inc., Mineola, NY, 1987).

[38] M. D. Macedonia and E. J. Maginn, Mol. Phys. **96**, 1375 (1999).

[39] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).

[40] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, J. Mol. Graphics Modell. **25**, 247 (2006).

[41] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, et al., *Gaussian 09, Revision A.02* (2009).

[42] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269 (1993).

[43] P. Cieplak, W. D. Cornell, C. Bayly, and P. A. Kollman, J. Comput. Chem. **16**, 1357 (1995).

[44] D. A. Case, T. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. Woods, J. Comput. Chem. **26**, 1668 (2005).

[45] D. A. Case, T. A. Darden, T. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, et al., *Amber 11* (2010).

[46] A. S. Paluch and E. J. Maginn, Ind. Eng. Chem. Res. **52**, 13743 (2013).

[47] S. Weerasinghe and P. E. Smith, J. Phys. Chem. B **109**, 15080 (2005).

[48] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).

[49] B. Chen, J. J. Potoff, and J. I. Siepmann, J. Phys. Chem. B **105**, 3093 (2001).

[50] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindal, J. Chem. Theory Comput. **4**, 435 (2008).

[51] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, et al., Bioinformatics **29**, 845 (2013).

[52] L. D. Schuler, X. Daura, and W. F. van Gunsteren, J. Comput. Chem. **22**, 1205 (2001).

[53] All of the Gaussian 09 input files (type "com"), GROMACS force field files (type "itp" and "top") and GROMACS intial configuration files (type "gro") for each system studied may be found in the Supplementary Material documents. The Gaussian files may be found in "gaussian_com.tar.gz" and the GROMACS files may be found in "gromacs_top_gro.tar.gz". For information on Supplementary Material, see http://www.aip.org/pubservs/epaps.html.

[54] ChemAxon (2013), version 5.82, URL `http://www.chemaxon.com/products/marvin/marvinsketch`.

[55] (2011).

[56] P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson, and M. T. Stahl, J. Chem. Inf. Model **50**, 572 (2010).

[57] P. C. D. Hawkins and A. Nicholls, J. Chem. Inf. Model **52**, 2919 (2012).

[58] A. W. Sousa da Silva and W. F. Vranken, BMC Res. Notes **5**, 367 (2012).

[59] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, J. Comput. Chem **30**, 2157 (2009).

[60] VERSION 4.5.3-dev-20110324-722c2.

[61] D. L. Mobley, S. Liu, D. S. Cerutti, W. C. Swope, and J. E. Rice, J. Comput.-Aided Mol. Des. **26**, 551 (2012).

[62] M. R. Shirts, D. L. Mobley, J. D. Chodera, and V. S. Pande, J. Phys. Chem. B **111**, 13052 (2007).

[63] M. R. Shirts and J. D. Chodera, J. Chem. Phys. **129**, 124105 (2008).

[64] See code at: https://simtk.org/home/pymbar.

[65] Really, the simulation of the solute+solvent for the case where the solute has no interactions with the solvent, and so the simulation box volume is determined entirely by the solvent.

[66] J. P. E. Grolier and E. Wilhelm, Fluid Phase Equilib. **6**, 283 (1981).

[67] M. J. Costigan, L. J. Hodges, K. N. Marsh, R. H. Stokes, and C. W. Tuxford, Aust. J. Chem. **33**, 2103 (1980).

[68] D. L. Mobley, E. Dumont, J. D. Chodera, and K. A. Dill, J. Phys. Chem. B **111**, 2242 (2007).

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

[70] P. V. Klimovich and D. L. Mobley, J. Comput.-Aided Mol. Des. **24**, 307 (2010).

[71] C. J. Fennell, K. L. Wymer, and D. L. Mobley, J Phys Chem B **118**, 6438 (2014).

[72] E. R. Thomas and C. A. Eckert, Ind. Eng. Chem. Process Des. Dev. **23**, 194 (1984).

[73] L. C. Draucker, M. Janakat, M. J. Lazzaroni, D. Bush, C. A. Eckert, T. C. Frank, and J. D. Olson, Ind. Eng. Chem. Res. **46**, 2198 (2007).

[74] As pointed out to ASP by Wilfried Cordes and Bastian Schmid (DDBST GmbH), the 2005 revision contains a small typographical error. In the equation for ζ_1, the coefficient 3.24 should instead be 3.4 in agreement with the original MOSCED model from 1984. See also http://www.ddbst.com/prp-estimate.html.

[75] R. A. Granberg and A. C. Rasmuson, J. Chem. Eng. Data **44**, 1391 (1999).

[76] J. P. O'Connell, AIChE J. **17**, 658 (1971).

[77] See eqs. (7) and (8) of ref. [46] for the employed expression.

[78] M. T. Geballe, A. G. Skillman, A. Nicholls, J. P. Guthrie, and P. J. Taylor, J. Comput.-Aided Mol. Des. **24**, 259 (2010).

[79] J. W. Essex, C. A. Reynolds, and W. G. Richards, J. Chem. Soc., Chem. Commun. pp. 1152–1154 (1989).

[80] A. S. Paluch, D. D. Cryan, and E. J. Maginn, J. Chem. Eng. Data **56**, 1587 (2011).

[81] S. Romero, A. Reillo, B. Escalera, and P. Bustamante, Chem. Pharm. Bull. **44**, 1061 (1996).

[82] A. Jouyban, H. K. Chan, N. Y. K. Chew, M. Khoubnasabjafari, and W. E. Acree, Jr., Chem. Pharm. Bull. **54**, 428 (2006).

FIG. 1: The chemical structure of the studied solutes.

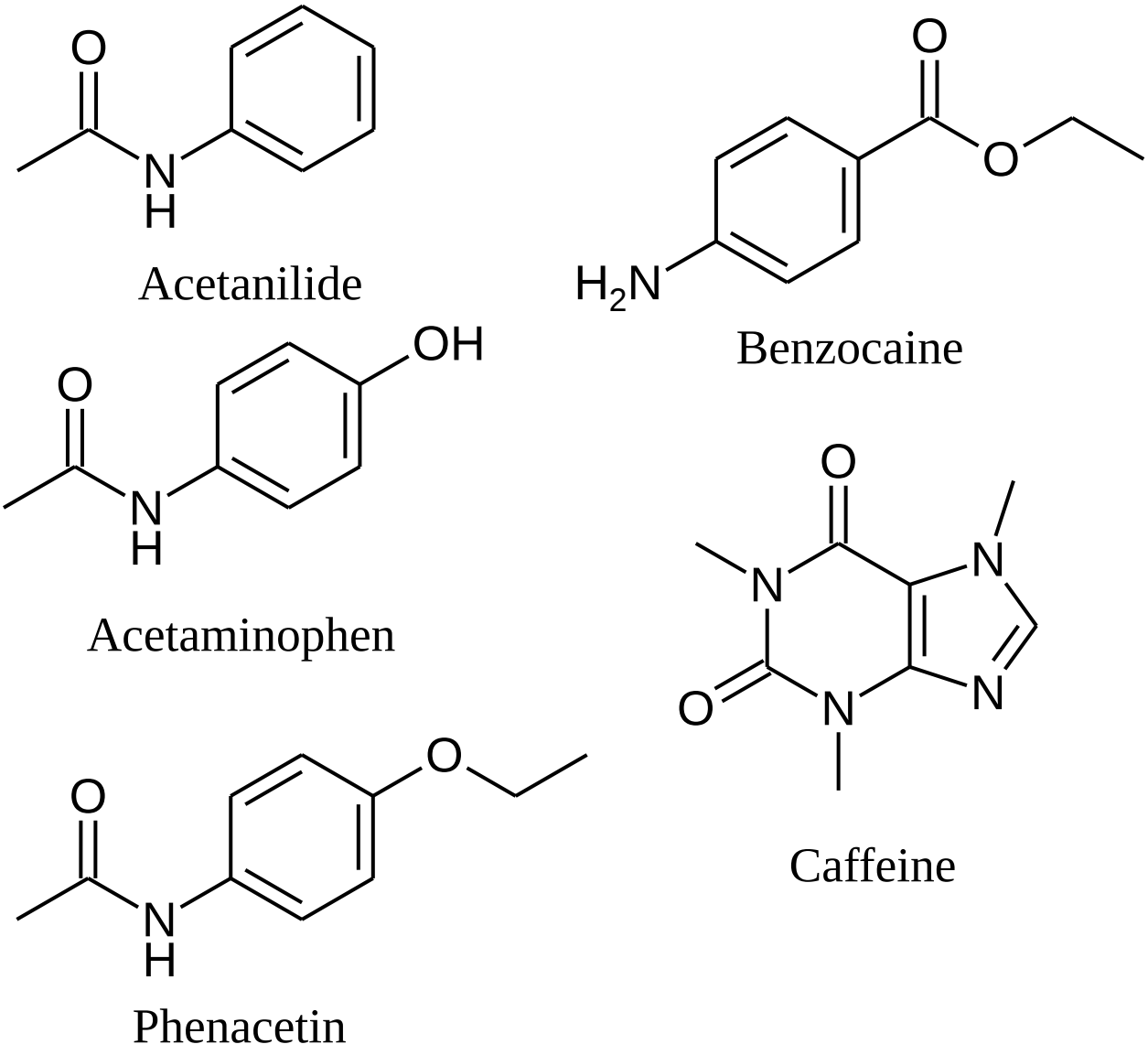

FIG. 2: The solute-free composition dependence of the molar volume (v^{b}), excess molar volume ($v^{\mathrm{b,E}}$), and excess molar enthalpy ($h^{\mathrm{b,E}}$) of the binary water(2)/ethanol(3) solvent. The solid lines are experimental data [66, 67] and the circles are simulation results. The uncertainties of the simulation results are smaller than the symbols.

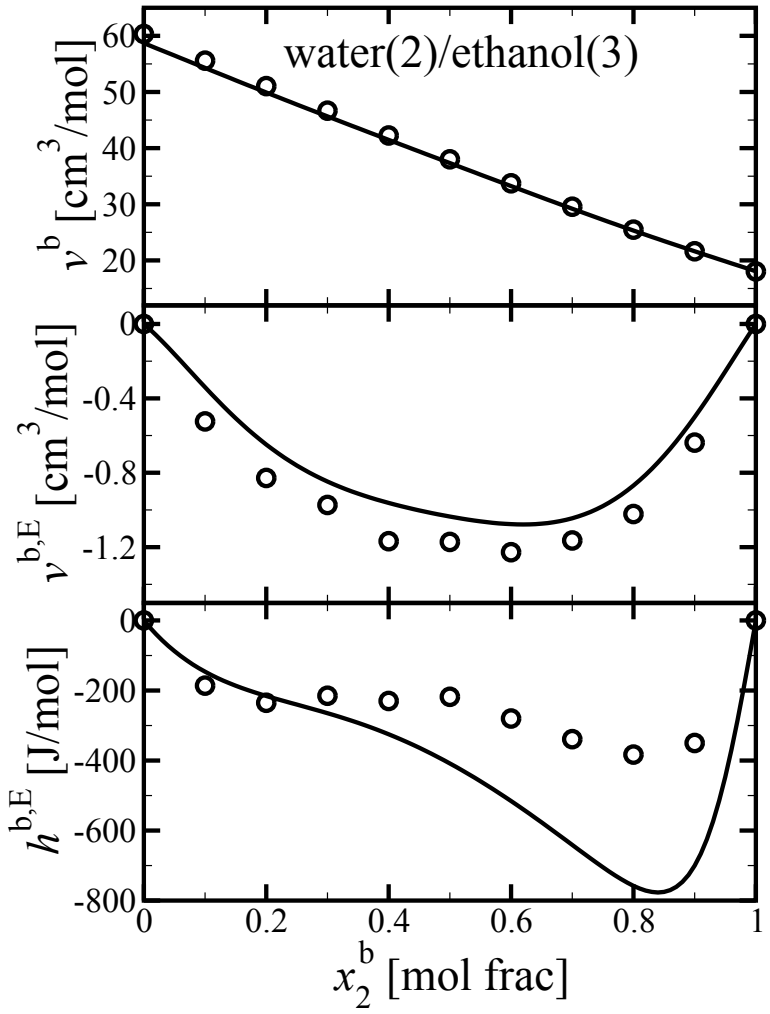

FIG. 3: The solute-free composition dependence of the the log excess solubility of component 1 (the solute, $\ln x_{1,m}^{\mathrm{E}}$) in the binary water(2)/ethanol(3) solvent, where component 1 is acetanilide, acetaminophen, or phenacetin, as indicated. Simulation results are shown as circles (◯) and experimental results are shown as: triangles up (△, ref. 8), squares (□, ref. 81), and diamonds (◊, ref. 82). For acetanilide, ref. 8 reports two independent sets of data which we differentiate here using open and filled symbols. The uncertainties of the simulation results are roughly the size of the symbols. The solid lines correspond to the third-order Redlich-Kister expansion (eq. (10)) fit to the experimental data [14]. The dashed line (for acetaminophen only) corresponds to MOSCED with Wilson's equation [10, 14, 72, 73, 76].

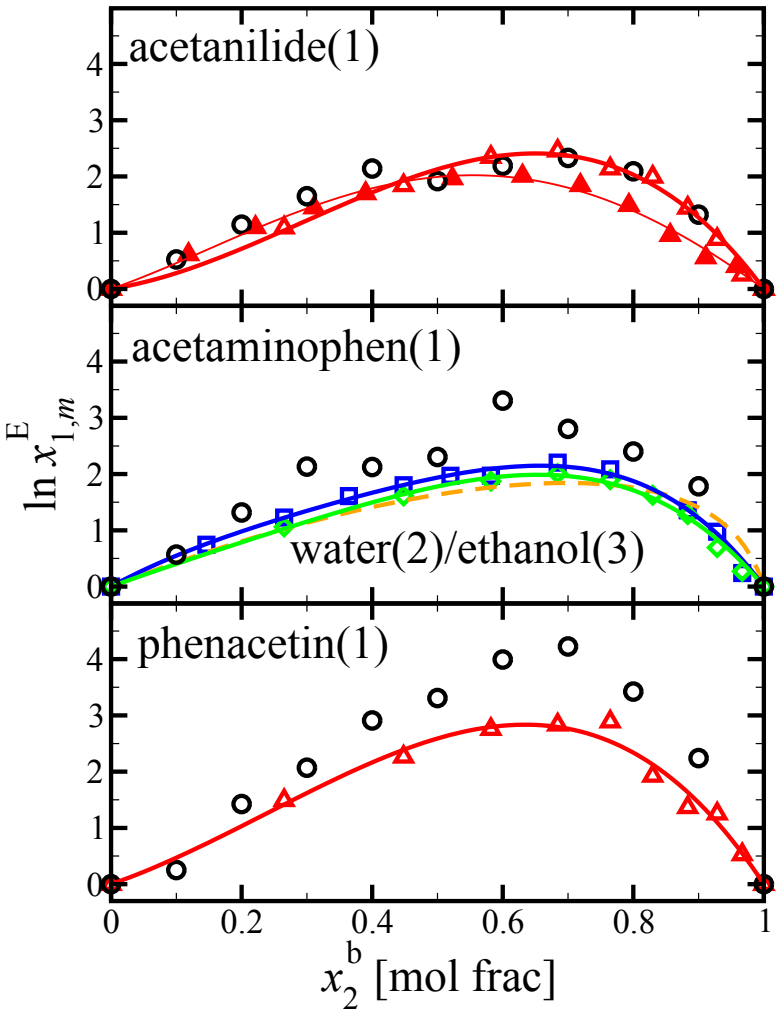

FIG. 4: The solute-free composition dependence of the the log excess solubility of component 1 (the solute, $\ln x_{1,m}^{\mathrm{E}}$) in the binary water(2)/ethanol(3) solvent, where component 1 is benzocaine or caffeine, as indicated. Simulation results are shown as circles (◯) and experimental results are shown as triangles up (△, ref. 8). The uncertainties of the simulation results are roughly the size of the symbols. The solid lines correspond to the third-order Redlich-Kister expansion (eq. (10)) fit to the experimental data [14].

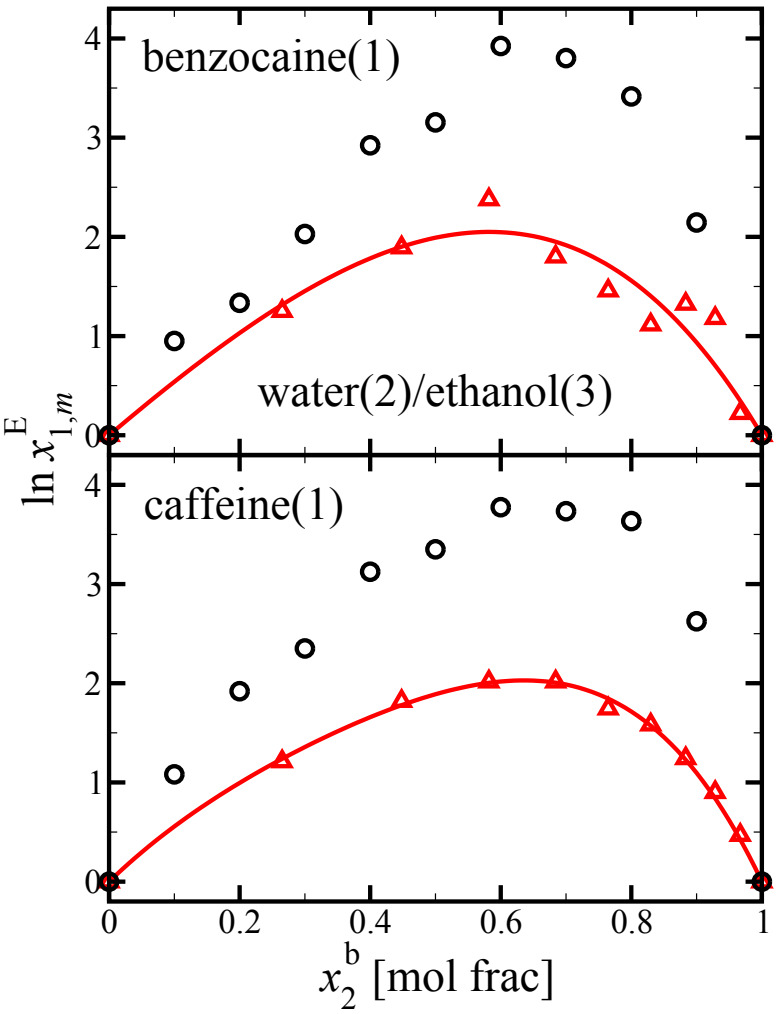

FIG. 5: A comparison of $\ln x_{1,m}^{\mathrm{E}}$ computed via molecular simulation $\left(\ln x_{1,m}^{\mathrm{E,\,sim}}\right)$ and from eq. (10) fit to experimental data $\left(\ln x_{1,m}^{\mathrm{E,\,exp}}\right)$ at the binary solvent concentrations studied via molecular simulation, except for $x_2^{\mathrm{b}} = 0$ and $x_2^{\mathrm{b}} = 1$ where by definition $\ln x_1^{\mathrm{E}} = 0$. The data is the same as in figs. 3 and 4.

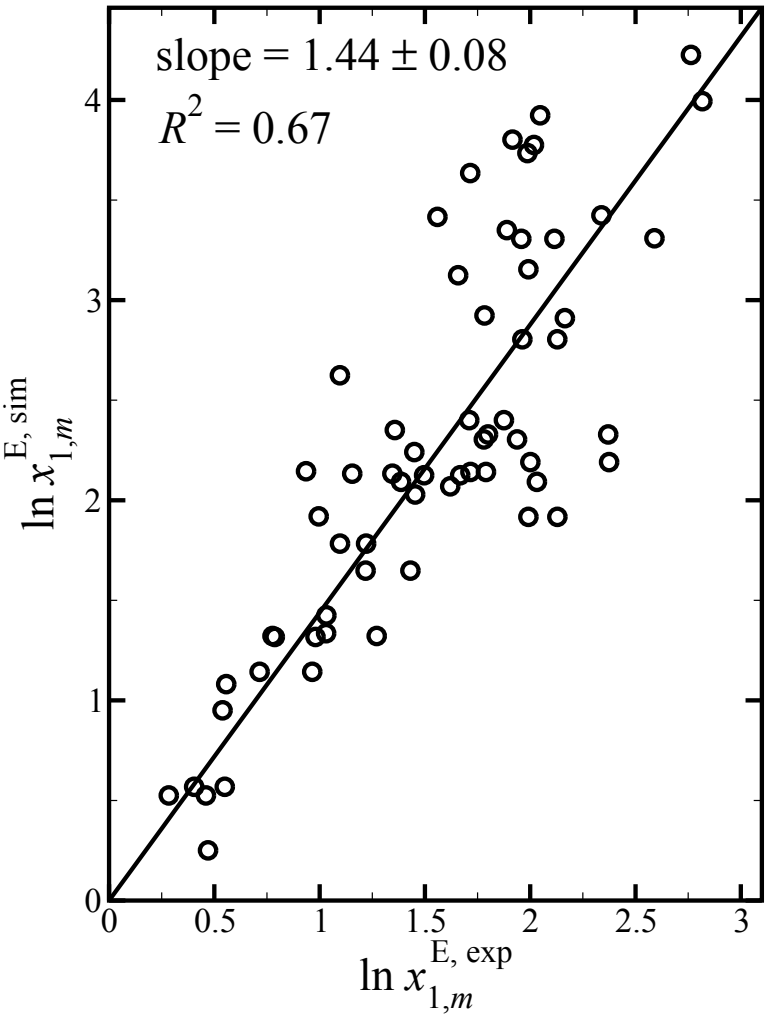

FIG. 6: The solute-free composition dependence of $\ln x^{\mathrm{E}}_{1,m}$ (circles, ◯) and $\ln c^{\mathrm{E}}_{1,m}$ (triangles left, ◁) of acetaminophen (component 1, the solute) in the binary water(2)/ethanol(3) solvent.