A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model

Shiang-Tai Lin and Stanley I. Sandler*

Center for Molecular and Engineering Thermodynamics, Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716

An activity coefficient model using molecular solvation based on the COSMO-RS method is proposed. In this model, quantum mechanical COSMO calculations are performed to obtain the screening charges for molecules in a perfect conductor. A statistical mechanical model that considers molecules to be a collection of surface segments is developed for the calculation of segment activity coefficients using these screening charges. Activity coefficients for molecules are then obtained by summing the contributions of the segments. This model requires only a single radius for each atom in the COSMO solvation calculations, one universal parameter to discern hydrogen-bond acceptors and donors, and two universal parameters to determine segment interactions. This is a significantly fewer number of parameters for phase equilibrium calculations than group contribution methods such as the UNIFAC (168 parameters) and modified UNIFAC (612 parameters) models. The resulting completely a priori prediction method results in absolute average deviations of 0.03 in vapor-phase mole fractions and 9% in total pressure for vapor−liquid equilibrium predictions of 243 binary mixtures and root-mean-square deviations of the octanol/water partition coefficient log K_{OW}, infinite dilution activity coefficients ln γ^{∞} in water, and in hexane for 64 compounds of 0.48, 1.65, and 0.50, respectively. This model is general and applicable for the a priori prediction of the phase behavior of most compounds, though admittedly it is less accurate than group contribution and other methods with many more parameters whose values have been obtained by regressing large amounts of data.

Introduction

In classical thermodynamics, the Lewis−Randall activity coefficient is frequently used to describe liquid-phase mixture nonidealities using activity coefficients. A variety of models have been proposed to describe the dependence of the activity coefficient on mixture composition, molecular structure, and intermolecular interactions. All of these models focus on the excess Gibbs free energy (G^{ex}), from which the activity coefficient is derived. At low and moderate pressures, the vapor phase can usually be assumed to be ideal so that only the G^{ex} model and pure component properties are needed to describe vapor−liquid-phase behavior, including the formation of azeotropes. In addition, G^{ex} models can be used to correlate or predict the mutual solubility of partially miscible binary mixtures, as well as the partitioning of a solute between two partially miscible solvents.

The most successful predictive G^{ex} models are UNIFAC[1] and modified UNIFAC[2] in which a molecule is described as a collection of independent functional groups and a mixture is formed from these functional groups. The activity coefficient of a molecule in a mixture is obtained from the sum of the activity coefficients of the constituent groups in the mixture. Such group contribution methods greatly reduce the number of parameters needed for predicting the phase behavior for a large variety of mixtures. However, as is true for all group contribution methods, UNIFAC and modified UNIFAC suffer from the inability to distinguish between isomers and inaccuracies when compounds with several strong nonalkyl functional groups are considered.[3] The functional group activity coefficients, which are dependent on the interactions between functional groups, cannot be measured; only the activity coefficient for the whole molecule can be obtained by fitting experimental phase equilibrium data. As a consequence, the quality of UNIFAC predictions depends on the similarity of the new system to the database used in its parametrization. There is no obvious way to improve these models other than empirical approaches, such as adding additional functional groups, introducing second- and higher-order functional groups (to account for the intramolecular interactions between functional groups), and increasing the database used to fit parameters.

Recently Klamt[4] proposed a completely new perspective in liquid-phase thermodynamics. In contrast to the G^{ex} models, Klamt started from the solvation of molecules in a conductor and developed a "conductor-like screening model for real solvent" (COSMO-RS) that, in principle, can be used to determine the chemical potential of any species in any mixture from quantum mechanical calculations. An interesting and unique feature of the COSMO-RS method is that it does not presuppose a specific concentration dependence of the G^{ex} function as is the case with UNIFAC, for example, which is based on the UNIQUAC equation. This provides greater flexibility in treating systems of very different chemical functionality but also suggests that care must be taking in formulating the model to ensure that all thermodynamic boundary conditions are met.

* To whom correspondence should be addressed. E-mail: sandler@udel.edu. Phone: 302-831-2945. Fax: 302-831-4466.

10.1021/ie001047w CCC: $22.00
Published on Web 12/22/2001

In COSMO-RS, molecules are treated as a collection of surface segments. An expression for the chemical potential of segments in the condensed phase was derived in which needed interaction energies between segments were determined from COSMO calculations.[5] The chemical potential of each molecule is then obtained by summing the contributions of the segments. This model was originally applied to the prediction of vapor pressures and partition coefficients, with an average accuracy of within a factor of 2. Recently, Tanpipat et al.,[6] Schäfer et al.,[7] and Klamt and Eckert[8] showed that COSMO-RS can also be used to make phase equilibrium predictions. Although the COSMO-RS method has provided promising results, the equation for the segment chemical potential used in COSMO-RS does not correctly converge to certain boundary conditions, and the final expression for the activity coefficient fails to satisfy thermodynamic consistency relations. Regardless of these problems, the need of scientists and engineers for predictive methods is so great that this model is being incorporated into process simulators.

Here, we present a derivation leading to an activity coefficient model in the COSMO-RS framework that resolves the difficulties that appear in the COSMO-RS model. The starting point is our previously developed group contribution solvation (GCS) model,[9] in which the activity coefficient is calculated from the solvation free energy of molecules in a solution. Originally the solvation free energy was obtained from ab initio solvation calculations.[10,11] In this work, this free energy is calculated in a two-step process: first, the dissolution of the solute into a perfect conductor and second, the restoration of the conductor to the actual solvent. Ideal solvation in the conductor is obtained from an ab initio COSMO calculation, and the restoring free energy was, similar to COSMO-RS, determined from consideration of segment interactions. An exact expression for the segment chemical potential that satisfies all boundary conditions is derived, and the resulting equation for the activity coefficient is thermodynamically consistent. An advantage of using ideal solvation from the COSMO model is that the dielectric properties of solvent, which are required in traditional continuum solvation models and generally not available especially for mixed solvents, are no longer necessary. Furthermore, only one atomic radius is needed for each atom in the ideal solvation calculation, and only two universal parameters are used for calculating segment interactions. This new model, based on surface segment contributions, is useful for making completely a priori phase equilibrium predictions, especially when traditional group contribution methods such as UNIFAC fail because of missing group parameters or the presence of more than one strong nonalkyl functional group on the same molecule.

Theory

Activity Coefficient from Solvation Free Energy. The solvation free energy, $\Delta G^{*sol}_{i/S}$, is defined as the free energy change when a solute molecule i is transferred from a fixed position in an ideal gas to a fixed position in a solution S at constant temperature and pressure. The activity coefficient $\gamma_{i/S}$ for the solute is then determined from[12]

$$\ln \gamma_{i/S} = \frac{\Delta G^{*sol}_{i/S} - \Delta G^{*sol}_{i/i}}{RT} + \ln \frac{c_S}{c_i} \qquad (1)$$

where c_j is the molar concentration of fluid j (which can be pure fluid i or solution S) and the superscript * is used to emphasize that the solute is at a fixed position. The solvation free energy is usually calculated through a two step process. First, the charges on the solute are turned off, and the remaining hard particle is inserted into the solvent. Then, the charges are turned on, and the electronic configuration of the solute is restored. The corresponding free-energy change in these two processes are referred to as the cavity formation free energy, ΔG^{*cav}, and the charging free energy, ΔG^{*chg}, respectively. Lin and Sandler[9,13] have shown that using the Staverman−Guggenheim combinatorial term[14,15] improved the calculation of cavity formation free energy from that used previously, and they proposed the following GCS model[9] for the activity coefficient

$$\ln \gamma_{i/S} = \frac{\Delta G^{*chg}_{i/S} - \Delta G^{*chg}_{i/i}}{RT} + \ln \gamma_{i/S}{}^{SG} \qquad (2)$$

where the Staverman−Guggenheim combinatorial term is

$$\ln \gamma_{i/S}{}^{SG} = \ln \frac{\phi_i}{x_i} + \frac{z}{2} q_i \ln \frac{\theta_i}{\phi_i} + l_i - \frac{\phi_i}{x_i} \sum_j x_j l_j \qquad (3)$$

with $\theta_i = (x_i q_i)/(\Sigma_j x_j q_j)$, $\phi_i = (x_i r_i)/(\Sigma_j x_j r_j)$, $l_i = (z/2)[(r_i - q_i) - (r_i - 1)]$, where x_i is the mole fraction of component i; r_i and q_i are the normalized volume and surface area parameters for i; z is the coordination number, usually taken to be 10; and the summation is over all of the species in the mixture.

Recently Klamt[4] proposed a new path for computing solvation energies. First, molecules are dissolved in a perfect conductor in which they are completely screened. Then, the screening charges are removed to restore the real fluid state. This process is equivalent to further decomposing the charging process used by Lin and Sandler into two steps: first, the dielectric properties of the solvent are increased to that of a perfect conductor, and the charges on the solute are turned on (ideal solvation), and then the solvent properties are restored to their normal values (removal of the screening charges). On the basis of this view, the charging free energy is the sum of the ideal solvation free energy, ΔG^{*is}, and the free energy required to remove the screening charges on the solute in the restoring process, ΔG^{*res}. Because the ideal solvation free energy is the same for dissolution of the solute in solvent S and in pure liquid i, with this new description, the GCS model reduces to

$$\ln \gamma_{i/S} = \frac{\Delta G^{*res}_{i/S} - \Delta G^{*res}_{i/i}}{RT} + \ln \gamma_{i/S}{}^{SG} \qquad (4)$$

The restoring free energy ΔG^{*res} can be obtained from the so-called "chemical potential of segment" in the COSMO-RS model. In the next section, a new expression is derived for this quantity in order to remove the inconsistencies in the original model.

Restoring Free Energy. Consider the formation of a liquid mixture from molecules in an ideal gas to occur using the thermodynamic cycle shown in Figure 1. The molecules are first dissolved in a perfect conductor (and consequently are perfectly screened), during which each molecule gains an electrostatic energy of $E_i^{COSMO} - E_i^{IG}$

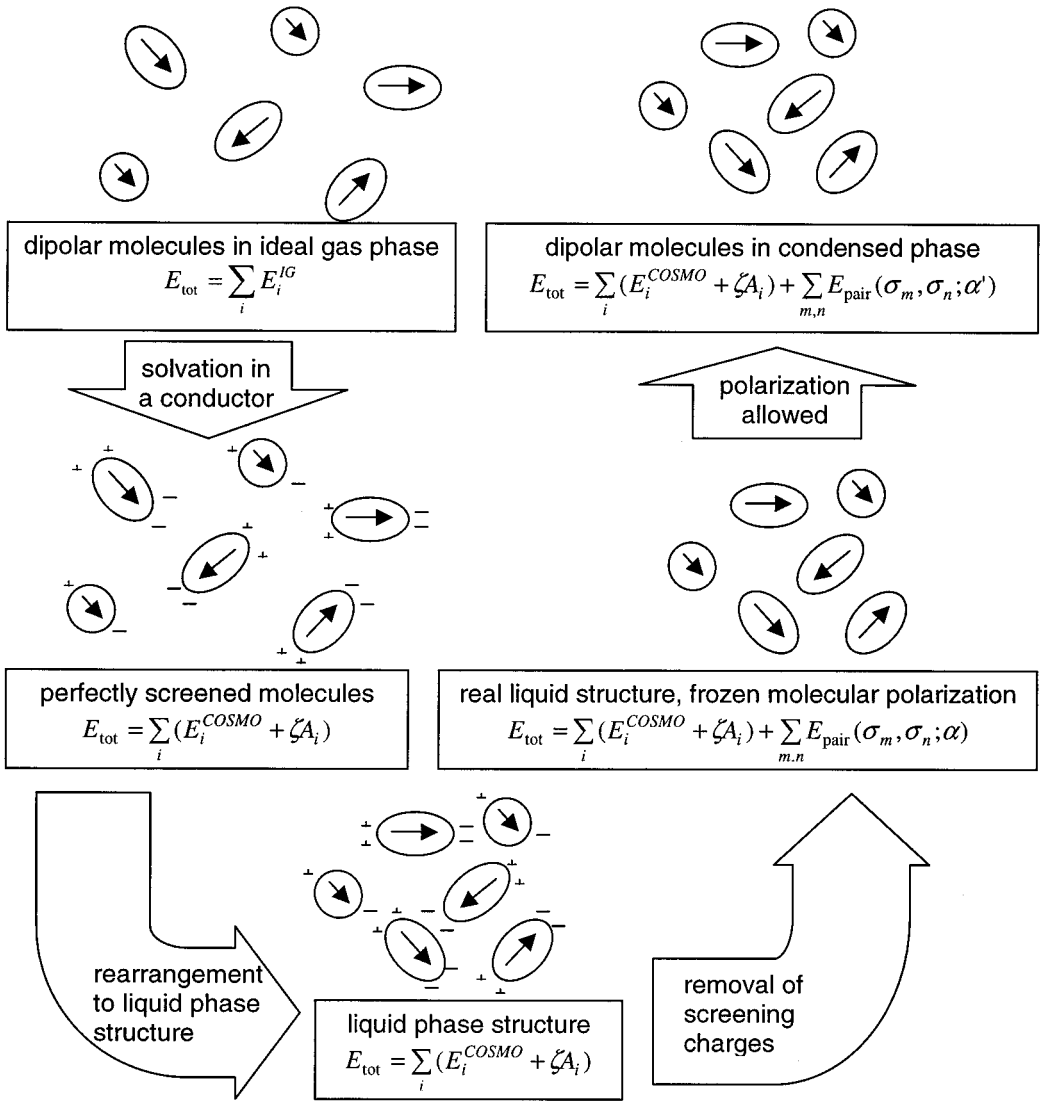

Figure 1. Schematic representation of condensation of dipolar molecules through an ideal solvation process. In the final step, molecules are allowed to polarize to lower the total free energy of the system ($\alpha' = 0.64\alpha$).

and a dispersion energy of ζA_i, where A_i is the surface area of molecule i and ζ is the dispersion constant. Because the molecules are ideally screened in the conductor, they can move freely without changing the total energy of the system.

Molecules in this ideally screened state are described by the Poisson equation. The induced screening charges appearing on the cavity surface of the conductor are determined from the requirement of zero total potential Φ_{tot} at the cavity boundary. That is,

$$\Phi_{tot} = \Phi_i + \Phi(q) = 0 \tag{5}$$

where Φ_i and $\Phi(q)$ are the electrostatic potentials resulting from molecule i and the ideal screening charge q. By defining a standard segment surface area a_{eff}, which is the same for all molecules and is one adjustable parameter in this method, each molecule contributes $n_i = A_i/a_{eff}$ surface segments to the mixture; the average screening charge q_{avg} characterizes the segment identity. The probability of finding a segment with a surface charge density $\sigma = q_{avg}/a_{eff}$ in pure liquid i is

$$p_i(\sigma) = n_i(\sigma)/n_i = A_i(\sigma)/A_i \tag{6}$$

where $n_i(\sigma)$ is the number of segments with charge density σ in a single molecule i and $A_i(\sigma) = a_{eff}n_i(\sigma)$ is the total surface area from all of these segments. The probability of finding a segment with charge density σ in a mixture is the weighted sum of the σ profiles of all of the components

$$p_S(\sigma) = \frac{\sum_i x_i n_i p_i(\sigma)}{\sum_i x_i n_i} = \frac{\sum_i x_i A_i p_i(\sigma)}{\sum_i x_i A_i} \tag{7}$$

The σ profile, that is $p_i(\sigma)$ as a function of σ, quantifies the electronic properties of a fluid and is the most important characteristic of each species in the COSMO-RS model. This quantity is obtained from quantum-chemical COSMO calculations, and the activity coefficient is determined solely on the basis of this information. In a condensed phase, molecules are in close contact with each other. Therefore, each segment is assumed to pair with another segment. For a system containing a total of N segments, $N/2$ pairs will form.

Because ideal screening charges do not exist in real solvents, these hypothetical charges must be removed

to return to the real fluid. This is done by adding additional paired segments to all surfaces, each of opposite charge to that in the ideally screened state, and these paired segments are assumed to have no interaction with each other. The total energy of the system is increased by the self-energies, E_{pair}, of the compensating paired surface segments

$$E_{tot} = \sum_i (E_i^{COSMO} + \zeta A_i) + \sum_{m,n} E_{pair}(\sigma_m, \sigma_n) \quad (8)$$

where σ_m and σ_n are charge densities of the paired segments m and n, the first summation is over all molecules, and the second summation is over all segment pairs.

The restoring free energy, ΔG^{*res}, is equivalent to the work required to add the additional segments on the surface of the solute to remove the screening charges. To simplify the problem, the segments of the molecules are decoupled; therefore, the physical picture of the liquid becomes that of an ensemble of paired surface segments, each independent of the other. An exact expression for the chemical potential of these segments based on a rigorous statistical mechanical argument is presented in Appendix 1; the result is

$$\mu_S(\sigma_m) = -kT \ln\left\{\sum_{\sigma_n} \exp\left[\frac{-E_{pair}(\sigma_m, \sigma_n) + \mu_S(\sigma_n)}{kT}\right]\right\} + kT \ln p_S(\sigma_m) \quad (9)$$

where $\mu_S(\sigma_m)$ is the chemical potential of a surface segment with charge density σ_m in a solution S. Defining the segment activity coefficient $\Gamma(\sigma)$ through the expression $\ln p_S(\sigma_m)\,\Gamma_S(\sigma_m) = (\mu_S(\sigma_m) - \mu^0(0))/kT$, with $\mu^0(0) = {}^1/_2 E_{pair}(0, 0)$ being the chemical potential of a segment in a fluid consisting of neutral segments, we have

$$\ln \Gamma_S(\sigma_m) = -\ln\left\{\sum_{\sigma_n} p_S(\sigma_n)\,\Gamma_S(\sigma_n) \exp\left[\frac{-\Delta W(\sigma_m, \sigma_n)}{kT}\right]\right\} \quad (10)$$

where $\Delta W(\sigma_m, \sigma_n) = E_{pair}(\sigma_m, \sigma_n) - E_{pair}(0, 0)$, called the exchange energy, is the energy required to obtain one (σ_m, σ_n) pair from a neutral pair. The physical meaning of the segment activity coefficient can be understood by relating this quantity to eq 2 (i.e., $RT \ln \Gamma(\sigma_m) = \Delta G^{*res}_{\sigma_{m/S}} - \Delta G^{*res}_{\sigma=0/\sigma=0} = \Delta G^{*res}_{\sigma_{m/S}}$), where the Staverman−Guggenheim combinatorial contribution is zero because the segments are equal in size. Therefore, $RT \ln \Gamma_S(\sigma_m)$ is the free energy required to add a segment with charge density σ_m at a fixed position in the solution (i.e., the restoring free energy of the segment). The restoring free energy of the solute can be obtained by summing the contributions from the segments

$$\frac{\Delta G^{*res}_{i/S}}{RT} = \sum_{\sigma_m}\left[n_i(\sigma_m)\frac{\Delta G^{*res}_{\sigma_{m/S}}}{RT}\right] = n_i \sum_{\sigma_m} p_i(\sigma_m) \ln \Gamma_S(\sigma_m) \quad (11)$$

so that the final equation for the activity coefficient is

$$\ln \gamma_{i/S} = n_i \sum_{\sigma_m} p_i(\sigma_m)[\ln \Gamma_S(\sigma_m) - \ln \Gamma_i(\sigma_m)] + \ln \gamma_{i/S}^{SG} \quad (12)$$

We will refer to this equation as the COSMO segment activity coefficient (COSMO-SAC) model.

The self-energy of a segment pair E_{pair} necessary for the calculation of the segment activity coefficient in eq 10 has been extensively discussed by the developers of COSMO-RS in their excellent papers.[4,8,16] This energy contains contributions from the electrostatic interactions, which they refer to as the misfit energy E_{mf}, the hydrogen-bonding interactions E_{hb}, and nonelectrostatic, mostly dispersion, interactions E_{ne}, so that

$$E_{pair}(\sigma_m, \sigma_n) = E_{mf}(\sigma_m, \sigma_n) + E_{hb}(\sigma_m, \sigma_n) + E_{ne}(\sigma_m, \sigma_n) = (\alpha'/2)(\sigma_m + \sigma_n)^2 + c_{hb} \max[0, \sigma_{acc} - \sigma_{hb}] \min[0, \sigma_{don} + \sigma_{hb}] + c_{ne} \quad (13)$$

where α' is a constant for the misfit energy that is calculated from the surface area of a standard segment as $(0.64 \times 0.3 \times a_{eff}^{3/2})/\epsilon_0$, with $\epsilon_0 = 2.395 \times 10^{-4}$ (e^2 mol)/(kcal Å) being the permittivity of free space; c_{hb} is a constant for the hydrogen-bonding interaction; σ_{hb} is a cutoff value for hydrogen-bonding interactions; σ_{acc} and σ_{don} are the larger and smaller values of σ_m and σ_n; max and min indicate that the larger and smaller values of their arguments are used, respectively; and the nonelectrostatic contribution E_{ne} is assumed to be a constant c_{ne}. Therefore, the exchange energy becomes

$$\Delta W(\sigma_m, \sigma_n) = (\alpha'/2)(\sigma_m + \sigma_n)^2 + c_{hb} \max[0, \sigma_{acc} - \sigma_{hb}] \min[0, \sigma_{don} + \sigma_{hb}] \quad (14)$$

It should be noted that nonelectrostatic interactions cancel out in $\Delta W(\sigma_m, \sigma_n)$ because of the assumption of constant E_{ne}.

Computational Details. In the COSMO-SAC model (eq 12), the activity coefficient of a molecule is determined from two separate contributions: the restoring free energy of the solute from the segment activity coefficient (eq 10) and the cavity formation free energy from the Staverman−Guggenheim model (eq 3). The former contribution is obtained from on the results from a quantum mechanical COSMO calculation, while the latter is determined from only the size and shape of the molecules. The procedure for the calculations is outlined as follows.

(1) When the procedure established for COSMO-RS as used in refs 4, 8, and 16 is followed, the first step is to establish the σ profile for each species in a mixture using the quantum chemistry package DMol3 implemented in Cerius2.[17−19] The equilibrium geometry of the molecules in the ideal gas phase is first obtained from molecular energy minimization using the density-functional theory with nonlocal VWN−BP functional at the DNP (double numeric with polarization functions, version 4.0.0, with a real space cutoff set to 5.50 Å) basis set level. Default values are used for all other parameters except that the numerical integration grid is set to fine. Next, the ideal screening charges on the molecular surface are determined from a solvation calculation for each molecule in the conductor (the COSMO calculation) at the same VWN−BP/DNP level with default atomic radii (Å) (H, 1.30 ; C, 2.00; N, 1.83; O, 1.72; Cl, 2.05) used to define the cavity. The screening charge densities σ^* from the COSMO output are averaged to give the "apparent" charge density σ on a

standard surface segment using the following expression:[16]

$$\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{\text{eff}}^2}{r_n^2 + r_{\text{eff}}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{\text{eff}}^2}\right)}{\sum_n \frac{r_n^2 r_{\text{eff}}^2}{r_n^2 + r_{\text{eff}}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{\text{eff}}^2}\right)} \quad (15)$$

where $r_{\text{eff}} = \sqrt{a_{\text{eff}}/\pi}$ ($a_{\text{eff}} = 7.50$ Å^2) is the radius of the standard surface segment, $r_n = \sqrt{a_n/\pi}$ is the radius of segment n, and d_{mn} is the distance between segments m and n. (It should be noted that Klamt used a smaller value of r_{eff} to achieve better results. As the use of a different r_{eff} for the averaging process appears arbitrary and was not justified, here, we use the natural value of r_{eff} based on the standard segment.) For most compounds, the ideal screening charge density falls within the range of -0.025 to 0.025 e/Å^2. This interval is partitioned to 50 parts, and the histogram (weighted by the area of each segment) of the averaged charge density is computed at each 0.001 e/Å^2 increment. Once the σ profile is obtained for the pure compounds, the σ profile for a mixture is calculated using eq 7.

(2) The second step is to determine the restoring free energy of the solute in the mixture $\Delta G^{*\text{res}}_{i/S}$ and in its neat liquid phase $\Delta G^{*\text{res}}_{i/i}$. The cutoff value σ_{hb} for hydrogen-bonding interactions is set to 0.0084 e/Å^2, the same as that in the COSMO-RS model.[8] For the calculation of the exchange energy ΔW, the surface area of a standard segment $a_{\text{eff}} = 7.50$ Å^2 and the hydrogen-bonding coefficient $c_{\text{hb}} = 85\,580$ (kcal/mol Å^4/e^2) are used. The values of these universal parameters were determined from fitting experimental vapor−liquid equilibrium (VLE) data (details are given later). The segment activity coefficient in the mixture and in the neat liquid, $\ln \Gamma_S(\sigma)$ and $\ln \Gamma_i(\sigma)$, are then determined from eq 10 by iteration using the σ profiles, $p_S(\sigma)$and $p_i(\sigma)$, respectively. The restoring free energies can then be obtained from eq 11.

(3) Next, the molecular volume and surface area obtained from the COSMO calculation of each compound are normalized to a standard volume (66.69 Å^3) and surface area (79.53 Å^2) of a functional group to yield the r and q parameters, which are then substituted into the Staverman−Guggenheim expression in eq 3. These normalization constants are different from the values originally used in the UNIQUAC model[20] because slightly larger atomic radii had been used to define the molecules in ref 8.

(4) Finally, the restoring free energies and the Staverman−Guggenheim contribution from the previous two steps are used in the COSMO-SAC model (eq 12) to obtain the activity coefficient.

Note that it is only to obtain the σ profiles of the molecules (step 1) that a quantum mechanics calculation must be done, and this single-molecule ideal solvation calculation needs to be done only once for each molecule, regardless of the mixture in which the molecule appears. Once these are known, only simple algebraic calculations are required for the computation of the activity coefficient as a function of solvent and composition for any mixture. Interestingly, the activity coefficient of a species in a mixture is obtained as a function of composition without specifying a priori the form of this composition dependence using a model.

Parameter Optimization. The COSMO-SAC model currently uses a total of eight parameters. These include the radii for the five atoms mentioned earlier for use in the COSMO calculation, one cutoff surface charge density σ_{hb} for distinguishing hydrogen-bond donors and acceptors, and two segment interaction-related variables (the surface area of a standard segment a_{eff} and the hydrogen-bonding coefficient c_{hb}). Only a_{eff} and c_{hb} are considered as adjustable parameters in this work. The other parameters are taken from previous work of Klamt, as stated earlier. Those parameters were optimized using a large data set with COSMO-RS, not the revised model we have developed here and, therefore, may not be optimal for this new model. The two remaining adjustable parameters were sequentially determined by minimizing the root-mean-square (rms) deviation of the calculated excess Gibbs free energy from that obtained from VLE data

$$\text{rms} = \left[\frac{1}{M}\sum_j^M \left(\frac{\underline{G}_{\text{calc},j}^{\text{ex}} - \underline{G}_{\text{expt},j}^{\text{ex}}}{RT}\right)^2\right]^{1/2} \quad (16)$$

with $\underline{G}_{\text{calc}}^{\text{ex}}/RT = \Sigma x_i \ln \gamma_{i/S}$ and $\underline{G}_{\text{expt}}^{\text{ex}}/RT = \Sigma x_i \ln((y_i^{\text{expt}} P^{\text{expt}})/(x_i P_i^{\text{vap}}))$. Here, M is the total number of data points, $\underline{G}^{\text{ex}}$ is the molar excess Gibbs free energy, and the vapor pressures P_i^{vap} of the pure compounds were calculated from the Antoine equation.[21] The a_{eff} parameter was first optimized using VLE data for 70 nonassociating binary mixtures (918 data points collected from the Dortmund Data Bank (DDB)[21]) formed from the 23 simple monofunctional compounds (butane, pentane, hexane, heptane, octane, 2,2-dimethylbutane, 2-methylpentane, 1-butene, 1-hexene, 1-heptene, diethyl ether, methyl *tert*-butyl ether, acetone, 2-butanone, 3-pentanone, propanal, butyraldehyde, methyl acetate, ethyl acetate, propoinitrile, butyronitrile, nitromethane, and nitroethane), covering a temperature range from 238.15 to 373.15 K. The physical constraint that the segment restoring free energy $RT \ln \Gamma(\sigma_m)$ from misfit energy be positive was imposed in the optimization. The standard segment was found to be 7.50 Å^2 with the rms deviation in $\underline{G}^{\text{ex}}/RT$ being 0.07. The absolute average deviations in the vapor-phase composition (AAD $= (1/M)\Sigma_j^M |y_j^{\text{exp}} - y_j^{\text{calc}}|$) and absolute average percentage deviation in total pressure (AAD% $= (1/M)\Sigma_j^M (|P_j^{\text{exp}} - P_j^{\text{calc}}|)/P_j^{\text{exp}}$) for these systems are 0.02 in mole fractions and 5%, respectively.

Next, 102 associating binary mixtures that consisted of two of the six hydrogen-bonding species (ethanol, 1-propanol, 1-butanol, 1-pentanol, 1-hexanol, and water) or one of the 24 compounds (previous 23 compounds and methylpropionate) and one of the six hydrogen-bonding species were retrieved from DDB. These systems (1526 data points from 273.15 to 383.15 K) were used to obtain the hydrogen-bonding coefficient $c_{\text{hb}} = 85\,580$ (kcal/mol Å^4/e^2), with a_{eff} fixed at 7.50 Å^2. The rms deviation in $\underline{G}^{\text{ex}}/RT$ for hydrogen-bonding mixtures was 0.08, slightly higher than for the nonassociating systems. The AAD in vapor-phase composition and AAD% in total pressures were 0.03 and 9%, respectively.

It is worth mentioning that because other atomic radii have not been determined, the COSMO-SAC model is currently limited to compounds containing atoms H, C, O, N, or Cl. However, this model is, in principle,

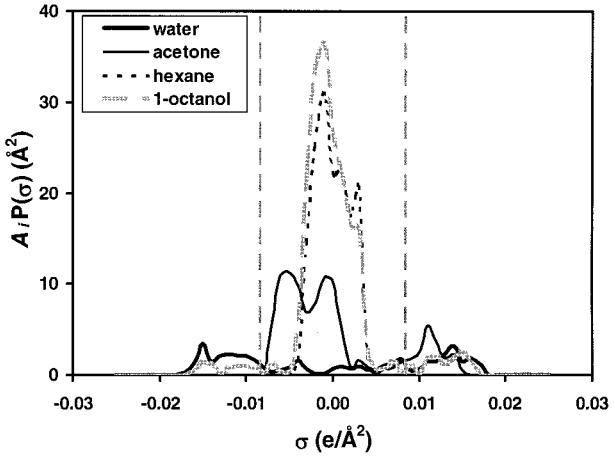

Figure 2. Surface area weighted σ profiles for water, acetone, hexane, and 1-octanol from COSMO calculations. The area under each curve corresponds to the molecular surface area of the compound.

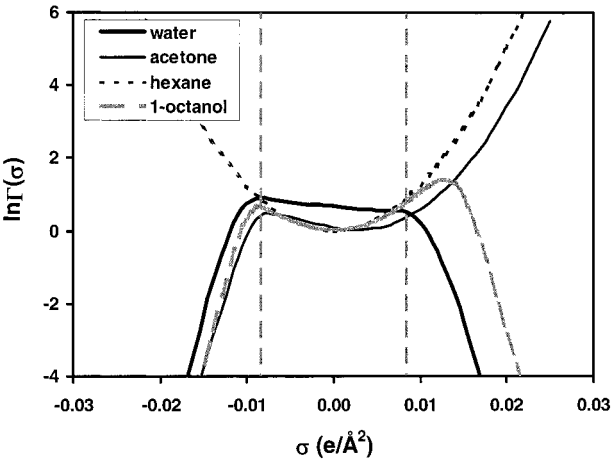

Figure 3. Segment activity coefficient ln $\Gamma(\sigma)$ as determined from eq 10 for water, acetone, hexane, and 1-octanol at 298.15 K.

applicable to all substances once all of the element-specific radii (F, Br, Si, etc.) are parametrized.

Results and Discussion

***σ* Profile and Restoring Free Energy.** Figure 2 shows the σ profiles for four common solvents, and the corresponding segment activity coefficients $\Gamma(\sigma)$ at 298.15 K are shown in Figure 3. The two vertical dashed lines in Figure 2 are the locations of the cutoff values for the hydrogen-bond donor ($\sigma < -0.0084$ e/Å^2) and acceptor ($\sigma > 0.0084$ e/Å^2). The σ profile for hexane lies within the hydrogen-bonding cutoff values, indicating its inability to form hydrogen bonds with other species. The resulting $\Gamma(\sigma)$ profile in Figure 3 for hexane is parabolic and positive, which means that work is required to remove the screening charges in the restoring process. For acetone, the peak in the σ profile at 0.011 e/Å^2 is a result of the oxygen atom in the carbonyl group and indicates that acetone is a hydrogen-bond acceptor. This peak leads to negative values of ln $\Gamma(\sigma)$ in Figure 3 for $\sigma < -0.01$ e/Å^2 because extra free energy is gained from forming hydrogen bonds when adding a hydrogen-bond donor segment to acetone. It is worth mentioning that there is a competing effect between the free energy of removing the screening charges (positive) and that of forming hydrogen bonds (negative). If a strong hydrogen bond is formed after the removal of the screening charges, the net restoring free energy would be negative. Water is both a hydrogen-bond donor and acceptor; therefore, its σ profile extends outside both H-bond cutoff values. The $\Gamma(\sigma)$ profile of water is negative at large positive and negative σ values because of the stabilizing effect of forming hydrogen bonds. 1-Octanol has a long alkane chain and a hydroxyl group that can hydrogen bond; therefore, its $\Gamma(\sigma)$ profile is similar to that of hexane at small σ values and to that of water at large σ values.

VLE Calculations. The utility of the proposed model is demonstrated by considering the VLE predictions for compounds not included in its parametrization. The DDB was used to search for consistent binary mixture data that contain one of the 30 compounds mentioned earlier and one of the new compounds listed in Table 1. The compounds chosen here include extensions of homologous series (1-octene and acetonitrile), isomers (2-butanol, *tert*-butyl alcohol, and isobutanol), new functionalities (cyclohexane, tetrahydrofuran, 1,4-dioxane, benzene, pyridine, phenol, butylamine, and trimethylamine), and one new atom (chlorine in chlorobenzene, 1,2-dichloroethane, chloroform, and tetrachloromethane). The accuracy of the predictions for these new compounds remains at the same level as that for those mixtures used in the parametrization, 0.03 in the deviation of vapor-phase mole fractions and 9% in total pressure. This indicates that the COSMO-SAC model is capable of making a priori predictions for new molecular species without any experimental data. The poor results for mixtures of chloroform are due to the overpolarization of the hydrogen atom in the COSMO calculation. The three electronegative chlorine atoms produce a significant positive charge on the hydrogen that exceeds the cutoff value for hydrogen-bond donor. Therefore, the interactions between chloroform and, for example, ketones and alcohols are incorrectly over predicted. This results in a large negative deviation in the calculated activity coefficient. One possible solution suggested in COSMO-RS is to turn off the hydrogen-bonding correction interactions (i.e., $c_{hb} = 0$) for chloroform with other species.[8] This solution is useful for chloroform/ketone systems; however, for chloroform/alcohol mixtures, it is impossible to distinguish between the hydrogen-bond donor from chloroform and from the alcohol molecule in the segment mixture. Therefore, it is necessary to refine the COSMO calculation to provide a better σ profile for chloroform, which is beyond the scope of our study here.

Although COSMO-SAC predictions are not as accurate as those from group contribution models such as UNIFAC and modified UNIFAC (Table 1), it should be noted that the COSMO-SAC model uses only two adjustable parameters, a_{eff} and c_{hb}. This is to be compared to the 168 and 612 parameters used in the UNIFAC and modified UNIFAC models for similar calculations. More importantly, the COSMO-SAC model is useful when no interaction parameter information is available for use in these group contribution methods. Figure 4 shows the VLE prediction of benzene/*n*-methylformamide binary mixture. Because *n*-methylformamide cannot be defined using current parameter sets in the UNIFAC and modified UNIFAC models, the COSMO-SAC model provides a valuable a priori prediction in such a case. Another advantage of the COSMO-SAC model is its consistent accuracy when extrapolated

Table 1. Predictions of VLE for Compounds Not Included in Parametrization

compound name	no. stm[a]	no. pts[b]	100 × AAD in y[c]			AAD% in P[d]			min T[e]	max T[f]
			COSMO-SAC	UNIFAC	modified UNIFAC	COSMO-SAC	UNIFAC	modified UNIFAC		
1-octene	1	23	0.30	1.81	1.60	2.24	2.54	1.96	348.15	348.15
2-butanol	8	129	2.55	1.75	1.15	5.91	3.63	1.84	298.15	358.15
tert-butyl alcohol	2	46	0.97	2.23	1.36	3.41	6.47	4.81	313.15	343.15
isobutanol	12	152	1.93	1.07	0.76	6.93	2.69	2.54	308.15	373.15
acetonitrile	13	171	2.21	1.77	0.73	4.04	4.07	1.72	313.15	351.22
cyclohexane	38	584	1.93	0.83	0.57	6.11	1.68	1.27	283.15	354.35
tetrahydrofuran	10	157	2.21	1.50	1.15	7.43	5.20	3.70	298.15	343.15
1,4-dioxane	12	126	2.44	5.35	1.19	5.28	12.08	2.31	308.15	358.15
benzene	62	865	2.69	0.64	0.52	6.89	1.15	1.11	293.15	348.15
pyridine	8	107	3.76	1.18	1.94	5.50	1.80	3.78	323.15	369.75
phenol	1	12	1.20	0.62	0.58	6.75	5.48	0.90	383.15	383.15
chlorobenzene	1	11	3.17	0.93	0.73	6.02	1.31	1.56	368.15	368.15
1,2-dichloroethane	13	207	2.86	1.06	0.72	6.05	1.98	1.55	298.15	343.15
chloroform	39	557	9.43	0.98	0.77	24.15	1.94	1.48	288.15	333.15
tetrachloromethane	17	274	2.35	0.77	0.67	7.02	1.62	1.26	298.15	353.15
butylamine	2	14	1.17	0.47	0.60	2.85	1.01	1.15	333.15	333.15
trimethylamine	4	52	0.57	0.24	0.22	1.15	0.23	0.26	333.15	353.15
overall	243	3487	3.48	1.13	0.75	9.04	2.43	1.64	283.15	383.15

[a] Number of binary systems. [b] Total number of data points. [c] Average absolute error in vapor-phase composition. [d] Average absolute percentage error in total pressure. [e] Minimum temperatures in the systems. [f] Maximum temperatures in the systems.

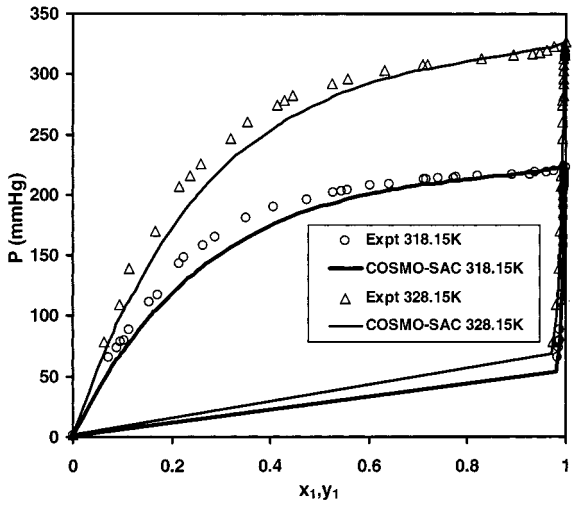

Figure 4. VLE prediction from COSMO-SAC for benzene (1)/*n*-methylformamide (2) at temperatures of 318.15 and 328.15 K.

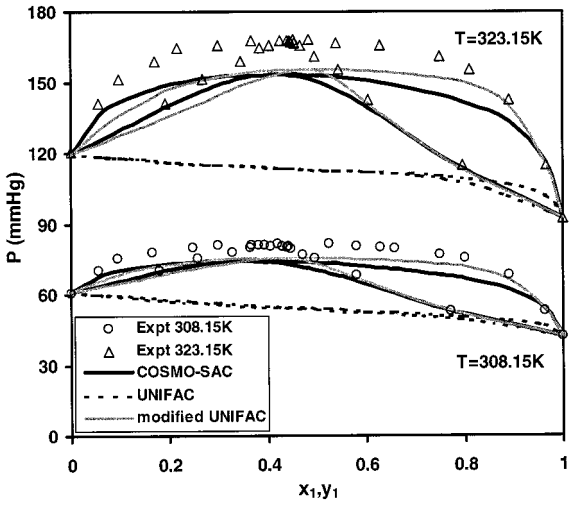

Figure 5. Comparison of VLE prediction from COSMO-SAC, UNIFAC, and modified UNIFAC models for water (1)/1,4-dioxane (2) at temperatures of 308.15 and 323.15 K.

to new systems. Water/1,4-dioxane binary mixtures, as shown in Figure 5, are known to be difficult to model using UNIFAC with one set of parameters for the ether group because of the strong hydrogen-bonding interactions between the two components[22] so that a new cyclic ether group was defined in the modified UNIFAC model. In contrast, the COSMO-SAC model is capable of describing these systems without introducing additional parameters.

A comparison between the COSMO-SAC model proposed here and COSMO-RS cannot be made here because we have not yet found an easy way to control the DMol3 program to calculate the activity coefficient for a specific solution composition for which experimental data are available. However, preliminary results show that COSMO-RS leads to better predictions than COSMO-SAC for some systems and poorer predictions for others. Two examples of VLE prediction from the two models are provided in Appendix 2.

Octanol/Water Partition Coefficients and Infinite Dilution Activity Coefficients. Calculated results for octanol/water partition coefficients and infinite dilution activity coefficients in water and hexane for 64 compounds are listed in Table 2 and shown in Figure 6. The predictions are satisfactory compared to the results from UNIFAC and modified UNIFAC model. The rms deviation in log K_{OW} from COSMO-SAC, UNIFAC, and modified UNIFAC are 0.48, 0.65, and 0.83, respectively, corresponding to an average error of 203%, 348%, and 575% in K_{OW}. The worst prediction of the COSMO-SAC model is for triethylamine. Amines, especially trialkylamines, are known to be difficult to model using classical methods.[16] The lone pair electrons tend to be hindered in the large cavity defined for the alkyl groups. Therefore, the screening charges may not be evaluated appropriately. The rms deviations of the natural log of the infinite dilution activity coefficient in water from the three models are 1.65, 1.92, 2.43, corresponding to average errors of 422%, 585%, and 1034% in γ^{∞}. There is a trend that the quality of the predictions from COSMO-SAC decreases for highly hydrophobic solutes. This is also reflected in the lower accuracy of K_{OW} predictions for large values of this parameter because the octanol/water partition coefficient is proportional to the infinite dilution activity coefficient in water.[13] For hydrophobic compounds, an accurate ln $\Gamma(\sigma)$ profile of water at small charge densities (i.e., $-\sigma_{hb} < \sigma < \sigma_{hb}$) is

Table 2. Calculations for Octanol/Water Partition Coefficient log $K_{OW,i}$, Infinite Dilution Activity Coefficient in Water ln $\gamma_{i/W}^{\infty}$, and Hexane ln $\gamma_{i/H}^{\infty}$ at 298.15 K

	log $K_{OW,i}$					ln $\gamma_{i/W}^{\infty}$					ln $\gamma_{i/H}^{\infty}$				
compound name	COSMO-SAC	UNIFAC	modified UNIFAC	COSMO-RS	expt	COSMO-SAC	UNIFAC	modified UNIFAC	COSMO-RS	expt	COSMO-SAC	UNIFAC	modified UNIFAC	COSMO-RS	expt
butane	2.45	1.85	1.48	2.64	2.89	8.34	6.85	6.33	9.09	9.99	−0.04	−0.05	0.00	0.03	0.00
pentane	2.89	2.28	1.96	3.23	3.39	9.53	8.08	7.56	10.60	11.55	−0.01	−0.01	0.00	0.02	−0.01
hexane	3.34	2.72	2.44	3.85	3.90	10.74	9.27	8.80	12.18	13.13	0.00	0.00	0.00	0.00	0.00
heptane	3.80	3.15	2.92	4.50	4.66	11.93	10.45	10.04	13.76	14.46	−0.01	−0.01	0.00	−0.01	−0.03
octane	4.25	3.58	3.41	5.14	5.15	13.12	11.61	11.28	15.35	16.08	−0.03	−0.03	0.00	−0.02	0.08
2,2-dimethyl-butane	3.05	2.69	2.44	3.49	3.82	9.96	9.21	8.79	12.18	12.47	0.03	0.00	0.00		
2-methyl-pentane	3.24	2.72	2.44	3.74		10.46	9.27	8.80	11.84	12.76	0.00	0.00	0.00	0.00	0.00
cyclohexane	2.87	2.52	1.43	2.75	3.44	9.48	8.73	6.36	10.35	11.31	0.00	0.03	0.09	0.02	0.13
1-butene	1.84	1.86	0.71	2.05	2.40	6.77	6.66	4.17	7.60	8.47	0.10	0.02	0.02		
1-hexene	2.72	2.72	1.67	3.28	3.39	9.10	9.08	6.67	10.64	11.38	0.12	0.07	0.07		
1-heptene	3.20	3.16	2.16	3.93	3.99	10.33	10.26	7.93	12.26	14.65	0.09	0.06	0.08		
1-octene	3.62	3.59	2.64	4.53	4.57	11.43	11.42	9.17	13.76	15.65	0.06	0.03	0.08		
1-hexyne	2.05	1.75	1.07	2.55	2.52	7.10	6.24	5.10	9.01	9.45	0.77	0.36	0.41		
1-heptyne	2.50	2.18	1.55	3.16		8.22	7.39	6.30	10.49	10.95	0.68	0.32	0.37		
diethyl ether	1.51	1.64	0.98	1.67	0.89	5.31	5.09	4.62	5.89	4.70	0.26	0.25	0.18		
methyl *tert*-butyl ether	1.70	1.91	1.22	1.96	0.94	5.59	5.43	5.12	6.40	4.72	0.34	0.45	0.25		
tetrahydrofuran	0.91	1.18	0.26	0.90	0.47	3.51	3.51	2.82	3.86	2.83	0.35	0.53	0.48		
1,4-dioxane	0.08	0.72	−0.23	0.07	−0.27	2.18	1.58	1.92	2.38	1.69	1.42	1.95	1.55	1.31	1.39
ethylamine	0.22	0.37	−0.58	0.19	−0.13	−0.70	0.02	−1.81	0.52	−0.99	1.96	0.94	0.94		
butylamine	1.13	1.24	0.38	1.43	0.88	1.42	2.27	0.65	3.28	1.39	1.77	0.82	0.93		
triethylamine	2.68	2.36	1.99	3.18	1.44	7.81	5.30	6.10	9.07	4.21	0.13	0.05	0.07		
acetone	−0.16	0.03	−0.08	−0.25	−0.24	1.53	2.44	2.08	1.80	1.95	1.40	1.84	1.84	1.44	1.90
2-butanone	0.40	0.47	0.39	0.48	0.35	2.80	3.47	3.18	3.38	3.24	1.11	1.68	1.71	1.08	1.51
2-pentanone	0.86	0.90	0.88	1.12	0.91	3.91	4.51	4.31	4.85	4.54	0.97	1.53	1.60		
3-pentanone	0.89	1.20	1.01	1.19	0.99	3.90	5.10	4.66	4.87	4.68	0.91	1.04	1.28		
propanal	0.17	0.43	0.67	0.11	0.59	2.65	2.98	2.69	2.96	2.57	1.17	1.53	1.44		
butyraldehyde	0.66	0.86	1.14	0.77	0.88	3.82	4.04	3.82	4.44	3.88	1.00	1.41	1.34		
n-methyl-formamide	−0.90			−1.05		0.21					7.25				
methylformate	−0.08	−0.28	−0.47	−0.13	0.03	2.69	1.62	1.59	3.01	2.74	1.87	1.74	1.81		
methyl acetate	0.22	0.44	0.22	0.31	0.18	3.04	3.44	3.20	3.58	3.12	1.46	1.29	1.43		
ethyl acetate	0.65	0.87	0.69	0.94	0.73	3.90	4.47	4.25	4.82	4.18	1.22	1.13	1.26	1.04	1.23
aceticanhydride	−0.40			−0.09		3.55					4.19				
dimethyl carbonate	−0.64			−0.49		2.01					3.16				
methyl-propionate	0.77	1.08	0.82	1.03		4.28	4.84	4.60	5.16	4.47	1.13	0.78	1.00		
nitromethane	−0.15	−0.16	−0.17	−0.15	−0.33	3.12	3.73	3.46	3.86	3.45	2.88	3.71	3.15	2.77	3.68
nitroethane	0.29	0.48	0.38	0.47	0.18	3.94	4.67	4.19	4.82	4.48	2.15	2.65	2.13	2.03	2.99
nitropropane	0.91	0.91	0.86	1.24	0.87	5.16	5.68	5.27	6.26	5.70	1.45	2.48	1.97	1.33	2.65
acetonitrile	−0.77	−0.16	−0.47	−0.97	−0.34	1.67	2.58	2.47	1.95	2.41	3.15	3.32	3.54	3.25	3.32
propoinitrile	−0.23	0.49	0.24	−0.25	0.16	2.41	3.63	3.58	2.93	3.63	2.18	2.22	2.45	2.24	2.92
butyronitrile	0.26	0.92	0.72	0.42	0.53	3.37	4.65	4.69	4.22	4.72	1.78	2.05	2.33		
cyanogen	0.77			0.81	0.07	4.29					0.83				
malonic acid dinitrile	−0.12			−0.53	−0.50	−1.11					7.22				
1,2-dicyano-ethane	−1.09	0.00	−0.04	−1.00	−0.99	2.43	4.14	5.73			8.09	5.85	7.34		
glutaronitrile	−1.39	0.44	0.45	−1.04	−0.72	2.60	4.90	6.48			7.25	5.43	6.85		
water	−1.36	−1.40	−1.59	−1.80	−1.40	0.00	0.00	0.00	0.00	0.00	11.43	7.25	4.91		
ethanol	0.03	0.03	−0.15	−0.08	−0.31	1.89	2.03	1.57	2.13	1.34	4.71	3.32	3.97	4.77	4.01
1-propanol	0.50	0.47	0.33	0.55	0.25	2.98	3.00	2.63	3.49	2.65	4.51	3.10	3.81	4.50	3.76
1-butanol	0.93	0.90	0.81	1.16	0.88	3.98	3.99	3.74	4.81	3.98	4.41	2.91	3.68	4.32	3.59
2butanol	1.07	0.90	0.75	1.25	0.61	4.37	3.99	3.59	5.06	3.27	3.36	2.91	3.41		
tert-butyl alcohol	0.80	0.87	0.60	0.98	0.35	3.53	3.93	3.30	4.30	2.48	3.57	2.92	3.06		
isobutanol	1.08	0.90	0.81	1.26	0.76	4.36	3.99	3.74	5.08	3.89	3.80	2.91	3.68		
1-pentanol	1.37	1.34	1.29	1.77	1.56	5.01	5.00	4.86	6.20	5.42	4.29	2.73	3.56		
1-hexanol	1.83	1.77	1.78	2.41	2.03	6.11	6.01	6.00	7.66	6.67	4.17	2.56	3.46		
1-octanol	2.73	2.64	2.75	3.66	3.15	8.22	8.06	8.31	10.58	9.36	3.93	2.24	3.28		
acetic acid	−0.21	−0.15	−0.48	−0.20	−0.17	0.86	1.25	1.19	1.32	1.06	5.53	2.39	2.49		
butyric acid	0.81	0.72	0.47	1.14	0.79	2.80	3.21	3.28	3.88	3.97	4.50	1.98	2.13		
benzene	1.47	2.15	2.10	1.78	2.13	6.17	7.79	7.75	7.41	7.82	0.71	0.45	0.56		
pyridine	0.48	1.47	0.32	0.64	0.65	2.61	1.95	1.56	3.49	2.99	1.55	1.33	1.77		
aniline	0.97	1.41	0.89	1.56	0.90	3.14	4.82	4.92	4.79	4.99	3.02	2.58	3.01		
phenol	0.93	1.72	1.09	1.38	1.47	3.00	2.40	3.78	4.46	6.31	4.20	2.92	3.46		
chlorobenzene	1.91	1.90	2.38	2.49	2.89	7.16	6.25	8.11	8.83	9.55	0.63	0.63	0.66		
chloroform	1.96	1.90	1.74	2.12	1.97	4.90	6.76	6.50	7.45	6.81	0.28	0.40	0.48	0.30	0.48
tetrachloro-methane	2.54	2.81	2.10	2.91	2.83	8.49	9.25	7.65	9.67	9.24	0.04	0.11	0.21	0.06	0.17
1,2-dichloro-ethane	1.15	1.54	1.35	1.63	1.47	5.40	6.50	6.29	6.90	6.46	1.12	0.89	1.17	0.93	1.15
max abs dev	1.24	1.57	2.01	1.74		4.32	4.47	6.73	4.86		1.20	0.70	0.86	1.31	
rms	0.48	0.65	0.83	0.43		1.65	1.92	2.43	1.14		0.50	0.33	0.29	0.53	

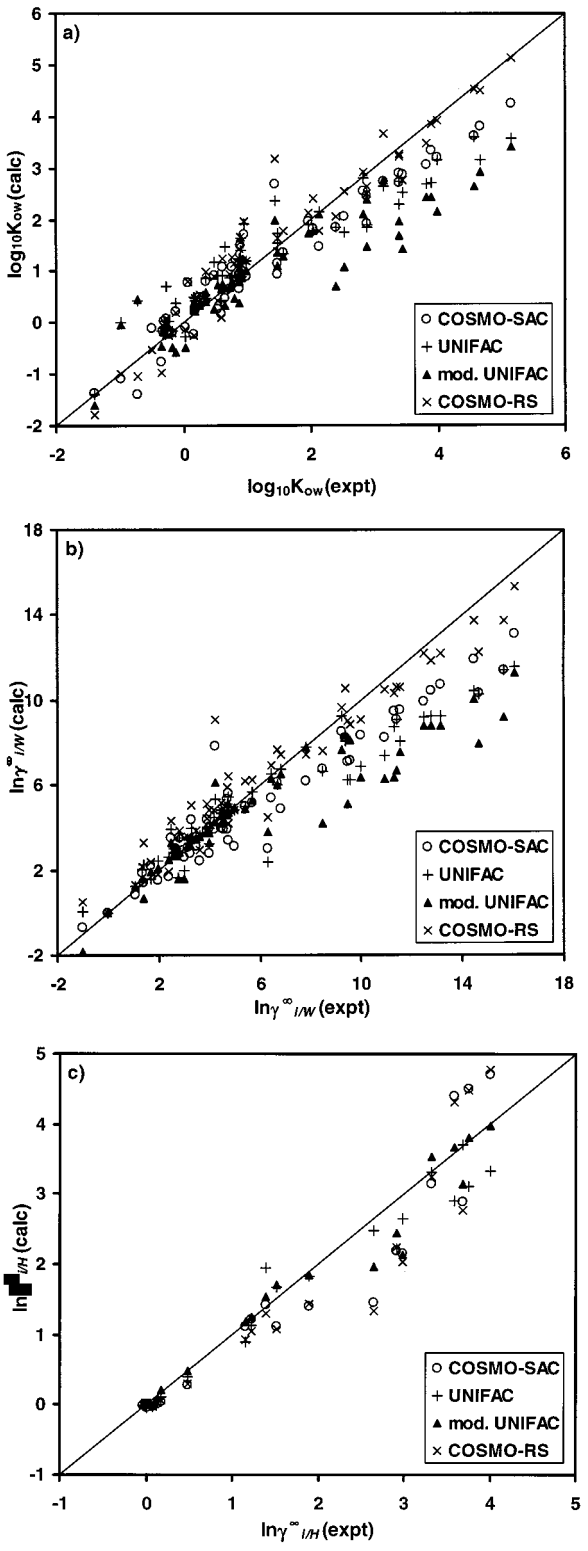

Figure 6. Comparison of calculated and experimental values for (a) octanol/water partition coefficient and infinite dilution activity coefficients in (b) water and (c) hexane from COSMO-SAC, UNIFAC, modified UNIFAC, and COSMO-RS models at 298.15 K.

necessary. This may require further refinement of model parameters (the atomic radii, a_{eff}, and c_{hb}) by including phase equilibrium data of water/alkane mixtures in the regression of these parameters or by a better description for the segment dispersion interactions. The predictions from COSMO-SAC for γ^{∞} in hexane compares reasonably well with the other two models; rms deviations of ln γ^{∞} are 0.50, 0.33, 0.29, corresponding to 65%, 40%, and 34% errors in γ^{∞} for COSMO-SAC, UNIFAC, and modified UNIFAC, respectively. The consistent results from COSMO-SAC for γ^{∞} in water and in hexane suggest that this model may also be applicable for predictions of mutual solubility.

A comparison of the COSMO-SAC and COSMO-RS models for the octanol/water partition coefficients and the infinite dilution activity coefficient in water and hexane is also shown in Figure 6 and summarized in Table 2. The COSMO-RS results are obtained with the DMol3 program in Cerius2.[17] (Interested readers are referred to ref 16 and the users' manual of DMol3 for the exact procedure and parameters used in the COSMO-RS calculations.) In general, the predictions from the COSMO-SAC and COSMO-RS models are comparable, with COSMO-RS being exceptionally good for alkanes in water. The superiority of COSMO-RS over the COSMO-SAC model in these cases may be because (1) the COSMO-RS model parameters were optimized using data for partition coefficients, whereas VLE data were used for the COSMO-SAC model, (2) there are additional parameters (such as an averaging radius r_{av}, a correlation factor f_{corr}, and an empirical λ parameter) in the COSMO-RS model but not in COSMO-SAC, and (3) we used several COSMO-RS default parameters that were not optimized for COSMO-SAC calculations. However, it is also interesting to point out that the maximum absolute deviations from COSMO-SAC are smaller than those from COSMO-RS. This may be a weak indication that COSMO-SAC is a more reliable model for new compounds as compared to the COSMO-RS model and that with all parameters optimized (including atomic radii), the accuracy of the COSMO-SAC model would be improved in a consistent manner.

With the utility of the COSMO-SAC model established, in the remainder of this paper, we compare the theoretical bases of the COSMO-SAC, COSMO-RS, and other models.

Comparison of the Segment Chemical Potential in the COSMO-SAC and COSMO-RS Models. The central equation in COSMO-RS is, for the chemical potential of charged surface segments,[4]

$$\mu^{\mathrm{COSMO-RS}}(\sigma_m) = -kT \ln\left\{ N^2 \sum_{\sigma_n} p(\sigma_n) \exp\left[\frac{-E_{\mathrm{pair}}(\sigma_m, \sigma_n) + \mu^{\mathrm{COSMO-RS}}(\sigma_n)}{kT}\right]\right\} \quad (17)$$

where N is the total number of segments in the mixture. This equation provides a completely new prospective by treating molecules in condensed phases as paired surface segments and allows determination of the segment interactions from fast quantum mechanical COSMO calculations. However, in the derivation of eq 17, all segments are labeled and, consequently, taken to be distinguishable in solution. Because segments with the same charge density are indistinguishable, eq 17 is exact only in the limit when each segment in the mixture has a different charge density. Therefore, in the general case, eq 17 overcounts the degeneracy of the ensemble of segments and leads to the following problems. First, $\mu^{\mathrm{COSMO-RS}}(\sigma)$ diverges to negative infinity as the number of molecules increases so that the total number of segments in the system goes to infinity. However, the chemical potential is an intensive property

and should not depend on the size of the system. In addition, $\mu^{\mathrm{COSMO-RS}}(\sigma)$ for a segment in the infinite dilution limit (i.e., as "$p(\sigma) \rightarrow 0$") is finite while this chemical potential should diverge to negative infinity in this limit. Moreover, $\mu^{\mathrm{COSMO-RS}}(\sigma)$ does not satisfy the proper boundary condition in the "pure" segment limit (i.e., all identical segments with charge density σ). The partition function of a system composed of N identical segments with charge density σ in this case is $Q(N, V, T) = e^{-\beta(N/2)E_{\mathrm{pair}}(\sigma, \sigma)}$, with $\beta = 1/kT$. Therefore, the chemical potential $\mu^0(\sigma)$ of a segment in such pure fluid is

$$\mu^0(\sigma) = -kT \left.\frac{\partial \ln Q(N, V, T)}{\partial N}\right|_{T,V} = {}^1/_2 E_{\mathrm{pair}}(\sigma, \sigma) \quad (18)$$

whereas it is $(N/2)E_{\mathrm{pair}}(\sigma, \sigma)$ from eq 17. In contrast, the chemical potential for segments derived here (eq 9) satisfies all these necessary boundary conditions. Equation 9 converges to eq 18 in the limit when all segments are identical (i.e., $p(\sigma_m) = 1$) and becomes the same as COSMO-RS (eq 17) in the limit when all segments are distinct (i.e., $p(\sigma_m) = 1/N$). Furthermore, eq 9 does not depend on the total number of segments in the system and diverges to negative infinity at the infinite dilution limit $p(\sigma_m) \rightarrow 0$.

These problems in the COSMO-RS model has been masked by the ad hoc introduction of a normalization $\mu'^{\mathrm{COSMO-RS}}(\sigma) = \mu^{\mathrm{COSMO-RS}}(\sigma) + kT \ln N$. However, this normalization leads to an unrealistically large $n_i kT \ln N$ dependence of the chemical potential of molecules. This term has been misinterpreted as the combinatorial contribution, and the incorrect system number dependence has been attributed to the decoupling of the segments. In recent modifications of COSMO-RS,[8,16] this term has been replaced with $\lambda kT \ln A_S$, where A_S is the mole fraction weighted surface area of all of the species in the mixture and λ is an adjustable parameter. However, the activity coefficient derived from this term, $\lambda kT \ln A_S/A_i$, does not satisfy the Gibbs−Duhem relation, as demonstrated with examples in Appendix 2. (For binary mixtures, this term gives $\lambda kT \ln A_2/A_1$ for one infinite dilution limit and $-\lambda kT \ln A_2/A_1$ for the other.)

Therefore, when COSMO-RS has been used for phase equilibrium predictions, either a small value of λ (0.14) has been used,[16] this term has been neglected completely,[23] or it has been diluted with the arbitrary addition of an engineering combinatorial term.[8] As has been explained previously, there is no net combinatorial contribution for the equally sized surface segments. The contribution from molecular size and shape effects should appear from the insertion of molecule into the solution in the ideal solvation process not from the restoring process.

In the latest version of COSMO-RS,[8] the activity coefficient is calculated as follows:

$$\ln \gamma_{i/S}{}^{\mathrm{COSMO-RS}} = n_i \sum_{\sigma_m} p_i(\sigma_m) \exp\left[\frac{\mu_S'^{\mathrm{COSMO-RS}}(\sigma_m) - \mu_i'^{\mathrm{COSMO-RS}}(\sigma_m)}{kT}\right] - \lambda \ln(A_S/A_i) + \ln \gamma_{i/S}{}^{\mathrm{SG}} \quad (19)$$

Interestingly, with the definition for segment activity coefficient $\ln \Gamma^{\mathrm{COSMO-RS}}(\sigma) = \mu'^{\mathrm{COSMO-RS}}(\sigma)/RT$ and setting λ to zero, the molecular activity coefficient from COSMO-RS becomes identical to that from COSMO-SAC. This coincidence results from the expedient and ad hoc introduction of a Staverman−Guggenheim term to COSMO-RS, whereas the derivation we present here is rigorous and coherent.

Comparison of Segment Activity Coefficient from COSMO-SAC and Lattice Theory-Based Models. It is interesting to note that the activity coefficient of a surface segment can also be determined from lattice theory-based models. Each surface segment with its charge density can be treated as a functional group. The formation of paired segments is equivalent to setting the coordination number z to 1. Because all surface segments are of the same size, the surface area parameter q also equals 1 for all segments. The interaction energy $U(\sigma_m, \sigma_n)$ between two segments can then be calculated as follows:

$$U(\sigma_m, \sigma_n) = E_{\mathrm{pair}}(\sigma_m, \sigma_n) - E_{\mathrm{pair}}(\sigma_m, 0) - E_{\mathrm{pair}}(\sigma_n, 0) \quad (20)$$

Substituting these relations into the formulas of the UNIQUAC or Wilson models,[24] we obtain the following expression for the activity coefficient of surface segments in solution:

$$\ln \Gamma^{\mathrm{UNIQUAC}}(\sigma_m) = \left[1 - \ln \sum_{\sigma_n} p(\sigma_n)\, \tau_{nm} - \sum_{\sigma_n} \frac{p(\sigma_n)\, \tau_{mn}}{\sum_{\sigma_k} p(\sigma_k)\, \tau_{kn}}\right] \quad (21)$$

with

$$\tau_{mn} = \exp\left[-\frac{E_{\mathrm{pair}}(\sigma_m, \sigma_n) - E_{\mathrm{pair}}(\sigma_n, \sigma_n) + E_{\mathrm{pair}}(\sigma_n, 0) - E_{\mathrm{pair}}(\sigma_m, 0)}{2kT}\right] \quad (22)$$

It should be noted that the reference state used in eq 21 is a pure segment fluid consisting of identical charge densities. To compare eq 21 to the model we have developed here, it is necessary to change the reference state used in eq 10 from a fluid of neutral segments to that in eq 21, that is, $\ln p(\sigma_m)\, \Gamma'(\sigma_m) = (\mu(\sigma_m) - \mu^0(\sigma))/kT$, with $\mu^0(\sigma) = {}^1/_2 E_{\mathrm{pair}}(\sigma, \sigma)$.

Therefore, eq 10 becomes

$$\ln \Gamma'(\sigma_m) = -\ln \sum_{\sigma_n} p(\sigma_n)\, \Gamma'(\sigma_n) \tau'_{mn} \quad (23)$$

with

$$\tau'_{mn} = \exp\left[-\frac{E_{\mathrm{pair}}(\sigma_m, \sigma_n) - E_{\mathrm{pair}}(\sigma_n, \sigma_n) + E_{\mathrm{pair}}(\sigma_m, \sigma_n) - E_{\mathrm{pair}}(\sigma_m, \sigma_m)}{2kT}\right] \quad (24)$$

Although the UNIQUAC model was derived from a very different basis (local composition assumption), the activity profiles from the two models have a striking similarity. Figure 7 shows the activity coefficient profiles for water and hexane calculated from eqs 21 and 23, respectively. The two profiles are almost identical for hexane, while for water the UNIQUAC model deviates from the exact solution at large charge densities. We have examined such profiles for other solvents, and the results are similar: the UNIQUAC expression tends to

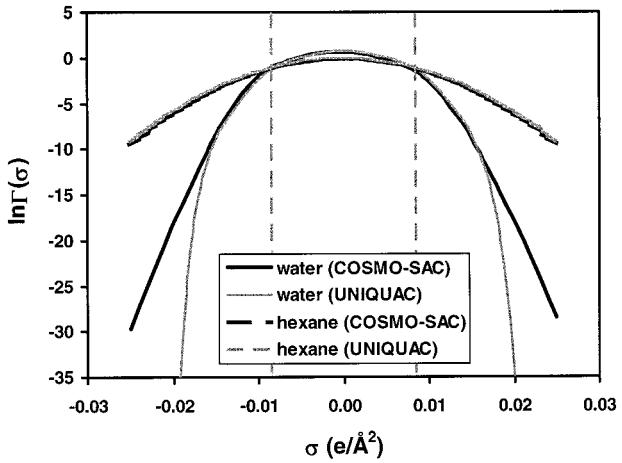

Figure 7. Comparison of segment chemical potential from eq 21 and the exact model from eq 23 at 298.15 K.

exaggerate hydrogen-bonding effects. Because such exaggeration is consistent, errors may cancel when the compounds in a mixture are similar. However, for mixtures containing dissimilar species, such differences accumulate when summing over contributions from the segments. This may explain why the UNIQUAC model, while very successful in correlating VLE data, is less successful for modeling mutual solubility and simultaneous VLE−LLE phase equilibrium predictions.

Comparison of Methodology in UNIFAC, COSMO-RS, and COSMO-SAC. It is useful to compare and summarize the calculation methods used in UNIFAC, COSMO-RS, and COSMO-SAC. Interestingly, there is a similarity between the three approaches as shown in Figure 8. The UNIFAC model breaks an isolated molecule into functional groups and uses group frequencies (i.e., the number of each type of functional group contained in a molecule) for each pure compound. This group frequency distribution is in some ways analogous to the σ profiles in COSMO-RS and COSMO-SAC, which are computed from partitioning an ideally screened molecule into surface segments. The group frequency distribution and the σ profile are the unique characteristics of each molecule. The sum of these profiles, each weighted by the mole fraction for each component, produces the profile of a mixture. In the UNIFAC model, based on the UNIQUAC equation, the

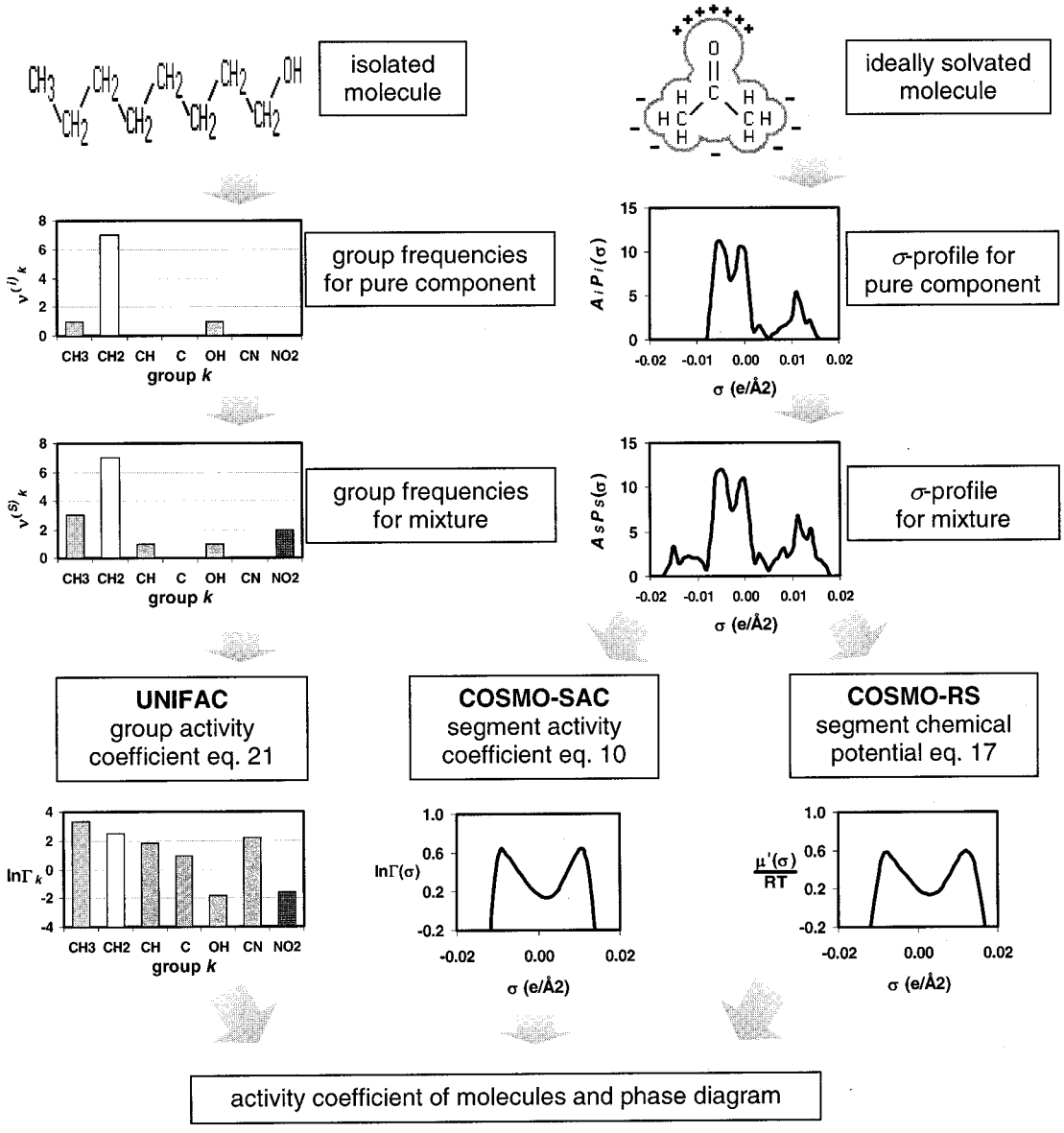

Figure 8. Methodology comparison between UNIFAC, COSMO-RS, and COSMO-SAC.

group activity coefficient is computed using the mixture group frequencies and a predetermined set of energy parameters. In contrast, the COSMO-SAC (eq 10) and COSMO-RS (eq 17) models determine the segment activity coefficient or chemical potential from the mixture σ profile with segment interactions calculated from the charge density on each segment. Finally, all three models sum the contributions from groups or segments to give molecular activity coefficients and generate the phase diagram. One advantage of COSMO-SAC and COSMO-RS is the use of a more detailed group partition method (i.e., the σ profile). Consequently, these models are able to distinguish between isomers and resolve problems of intramolecular proximity effects.[3] More importantly, the segment interactions can be determined from the charge densities of the segments; therefore, considerably fewer parameters are necessary. Although similar to COSMO-RS, the COSMO-SAC model uses an exact statistical mechanical model for the segment activity coefficient (eq 10), while the segment chemical potential (eq 17) used in COSMO-RS is exact only in the limit when all segments are distinguishable.

There is one important feature of both COSMO-RS and COSMO-SAC that should be emphasized. That is, unlike the UNIQUAC, UNIFAC, Wilson, NRTL, and other activity coefficient models,[24] no composition dependence of G^{ex} is presupposed. The composition dependence arises naturally here and depends on the species involved and, therefore, changes from mixture to mixture. This suggests that the COSMO-RS and COSMO-SAC are of a new class of models that is universally applicable, and of similar accuracy, for all mixtures, in which case it may not be necessary for a scientist or engineer to use different models for different types of mixtures.

In this work, we have used values for most parameters in the COSMO calculation from the previous work for the COSMO-RS model (i.e., the level of theory and the atomic radii in the quantum mechanical calculation). Although the results are satisfactory, it is likely that the accuracy of COSMO-SAC would be improved by optimizing these parameters using a very large data bank, as well as considering other methods for computing the σ profile. Also, the VLE or partition coefficient data may not be the best to use for such parameter estimations. This is because the ideal solvation free energy ΔG^{*is}, which is most sensitive to the atomic radii and the level of quantum mechanical calculations, cancels out of such phase equilibrium calculations. One possible way to directly examine ΔG^{*is} is to consider the vapor pressure determined from the free energy

$$\ln P_i^{vap} = \frac{\Delta G^{*sol}_{i/i}}{RT} + \ln c_i RT = \frac{\Delta G^{*is}_{i/i} + \Delta G^{*res}_{i/i} + \Delta G^{*cav}_{i/i}}{RT} + \ln c_i RT \quad (25)$$

When combined with a cavity formation model for pure fluid, such as the scaled particle theory (SPT),[25,26] the Boublík−Mansoori−Carnahan−Starling−Leland model (BMCSL),[27,28] or information theory (IT),[29,30] experimental vapor pressures can be used to directly check the calculated ideal solvation free energy. Once the parameters have been optimized in this manner, it may be possible to predict accurate vapor pressures from this method.

Conclusion

A modification to the COSMO-based activity coefficient model has been developed here that has a well-defined theoretical basis. In this model, molecules are considered to consist of a collection of equally sized surface segments, with their interactions determined from the screening charges they acquire on ideal solvation. The molecular activity coefficient is then obtained by summing the contributions from the segment activity coefficients. A statistical mechanical model has been derived for the segment activity coefficients based on the assumptions that the segments are paired in the solution, and segment interactions are confined within each pair (i.e., there is no interaction among segments of different pairs). To compute segment interactions, this model requires only atomic radii for calculating the screening charges that we have taken from previous publications and two universal parameters for computing the segment interactions that we have fit to VLE data. The consistent accuracy of the predictions for new compounds from this model has been demonstrated by calculations for VLE, octanol/water partition coefficients, and infinite dilution activity coefficients in water and hexane. This model provides a valuable a priori estimate of the thermophysical properties of new or previously unmeasured species and of the phase equilibria for new mixtures. In addition, the exact statistical mechanical expression for the activity coefficient of segments in a mixture of segment pairs developed here provides a necessary boundary condition for other lattice theory-based models and can be valuable for the future development of more complicated models that take into account different segment/group sizes and varying coordination numbers.

Acknowledgment

The authors would like to thank the National Science Foundation (Grants CTS-9521406 and CTS-0083709) and the Department of Energy (Grant DE-FG02-85ER13436) for financial support of this research. S.-T.L. is grateful to Professor J. Gmehling for his generous offer and kind help on the usage of the Dortmund Data Bank during his stay at the Universität Oldenburg.

Appendix I. Derivation of the Chemical Potential of Charged Surface Segments

[Note: To assist in understanding the following derivation, one can imagine that segments with the same charge density σ are one type of molecule and that the ensemble is a mixture that contains many different types of molecules. In this picture, the probability $p(\sigma)$ of finding a segment with charge density σ is equivalent to the mole fraction of that type of molecule in the solution.]

Consider a mixture consisting of n_1 segments with charge density σ_1, n_2 segments with charge density σ_2, and so on to n_f segments with charge density σ_f (or equivalently, n_1 molecules of species 1, n_2 molecules of species 2, ..., n_f molecules of species f). Because the segments are assumed to form pairs in the liquid phase, the mixture contains $(1 + f)(f/2)$ types of pairs (n_{11} pairs formed by segments with charge densities σ_1 and σ_1, n_{12} pairs formed by segments with charge densities σ_1 and σ_2, n_{mn} pairs formed by segments with charge densities

σ_m and σ_n, ..., and n_{ff} pairs formed by segments with charge densities σ_f and σ_f). Because the total number of each type of segment must be conserved, we have f independent conservation equations

$$\begin{aligned} 2n_{11} + n_{12} + ... + n_{1f} &= n_1 \\ n_{21} + 2n_{22} + ... + n_{2f} &= n_2 \\ &\vdots \\ n_{f1} + n_{f2} + ... + 2n_{ff} &= n_f \end{aligned} \qquad \text{(A1.1)}$$

There are a total of $N_p = (n_1 + n_2 + ... + n_f)/2$ pairs formed in the mixture and the probability $p(\sigma_m, \sigma_n)$ of finding a pair formed by segments with charge densities σ_m and σ_n is

$$p(\sigma_m, \sigma_n) = \frac{n_{mn}}{(n_1 + n_1 + ... n_f)/2} = \frac{n_{mn}}{N_p} \qquad \text{(A1.2)}$$

This probability can be calculated exactly based on statistical mechanics. Suppose that the partition function $Q(N)$ for the N segments is known. The Helmholtz free energy $A(N)$ of the system is then $A(N) = -kT \ln Q(N)$. Because the segment pairs are assumed to be independent from each other, the reduction of the Helmholtz free energy, in the limit of large N, of removing one (σ_m, σ_n) pair is equivalent to a decrease due to the chemical potentials of the two segments (i.e., $A(N-2) = -kT \ln Q(N-2) = A(N) - \mu(\sigma_m) - \mu(\sigma_n)$. Therefore

$$\frac{Q(N-2)}{Q(N)} = \exp\left[\frac{\mu(\sigma_m) + \mu(\sigma_n)}{kT}\right] \qquad \text{(A1.3)}$$

The partition function is the sum of the Boltzmann factors for all possible states of the system so that the probability $p(\sigma_m, \sigma_n)$ of finding a (σ_m, σ_n) pair in an N-segment system is calculated from the summation of the probabilities of finding a (σ_m, σ_n) pair in states that contain (σ_m, σ_n) pairs

$$p(\sigma_m, \sigma_n) = \frac{2^t \exp\left[-\frac{E_{\text{pair}}(\sigma_m, \sigma_n)}{kT}\right] Q(N-2)}{Q(N)} = \frac{2^t \exp\left[\frac{-E_{\text{pair}}(\sigma_m, \sigma_n) + \mu(\sigma_m) + \mu(\sigma_n)}{kT}\right]}{} \qquad \text{(A1.4)}$$

with $t = 0$ when $m = n$ and $t = 1$ when $m \neq n$. The product of the Boltzmann factor for the (σ_m, σ_n) pair and $Q(N-2)$ includes all the possible states that contain at least one (σ_m, σ_n) pair. When the two segments are distinct ($m \neq n$), there will be an increase of degeneracy by a factor of 2 because interchanging the positions of the segments results in different states. When the two segments are the same, they are indistinguishable and do not result in an increase of degeneracy when interchanging positions.

There are a total of $(3 + f)(f/2)$ unknowns $[(1 + f)(f/2)$ from the pairings n_{mn} and f from the chemical potentials $\mu(\sigma_m))$, and we have $(3 + f)(f/2)$ relations to solve for all of the unknowns (f conservation equations in eq A1.1 and $(1 + f)(f/2)$ relations by combining eqs A1.2 and A1.4). To obtain the chemical potential of segments, using the conservation equations, we consider the probability of finding a segment with charge density σ_m paired with any other segment. This is equivalent to finding a segment with charge density σ_m in the mixture (i.e., $p(\sigma_m)$)

$$\begin{aligned} &\sum_{n=1}^{f} \frac{1}{2^t} p(\sigma_m, \sigma_n) \\ &= \frac{n_{m1}/2 + n_{m2}/2 + ... + n_{mm} + ... + n_{mf}/2}{N_p} \\ &= \frac{n_{m1} + n_{m2} + ... + 2n_{mm} + ... + n_{mf}}{n_1 + n_2 + ... + n_f} = \frac{n_m}{n_1 + n_2 + ... + n_f} \\ &= p(\sigma_m) \end{aligned} \qquad \text{(A1.5)}$$

From eq A1.4, we have

$$\sum_{n=1}^{f} \frac{1}{2^t} p(\sigma_m, \sigma_n) = \sum_{n=1}^{f} \exp\left[\frac{-E_{\text{pair}}(\sigma_m, \sigma_n) + \mu(\sigma_m) + \mu(\sigma_n)}{kT}\right] \qquad \text{(A1.6)}$$

When eqs A1.5 and A1.6 are combined and the terms are rearranged, we obtain the following expression for chemical potential of the charged surface segments:

$$\mu(\sigma_m) = -kT \ln\left\{\sum_{n=1}^{f} \exp\left[\frac{-E_{\text{pair}}(\sigma_m, \sigma_n) + \mu(\sigma_n)}{kT}\right]\right\} + kT \ln p(\sigma_m) \qquad \text{(A1.7)}$$

which is eq 9 in the text.

Appendix II. Thermodynamic Consistency of the COSMO-SAC and COSMO-RS Models

In this appendix, we consider COSMO-SAC and COSMO-RS calculations for two arbitrarily chosen systems to compare the differences in predictions and consider the question of thermodynamic consistency. The COSMO-RS predictions were made using the program DMol3 in Cerius2, and the COSMO-SAC predictions were made as described here. Figures 9 and 10 show the predictions of both methods for these systems. The experimental data are taken from the Dortmund Data Bank.[21] The difference in the predictions emphasizes the fact that the COSMO-RS and COSMO-SAC, while based on the same physical picture, are different models. Tables 3 and 4 contain the activity coefficients predicted from the two models, together with the results of a Gibbs−Duhem point-to-point consistency test for the two models. There, we test consistency using

$$x_1 \frac{\partial \ln \gamma_1}{\partial x_1} + x_2 \frac{\partial \ln \gamma_2}{\partial x_1} = 0 \qquad \text{(A2.1)}$$

which is applicable because the points are calculated along an isotherm and the product of the pressure and the excess volume term is negligible. For thermodynamic consistency, this term in the tables should be zero at each point; however, because the value is calculated by numerical differentiation rather than analytically, small fluctuations about zero are acceptable. From these tables, we see that the COSMO-SAC model satisfies the thermodynamic consistency test, while COSMO-RS does

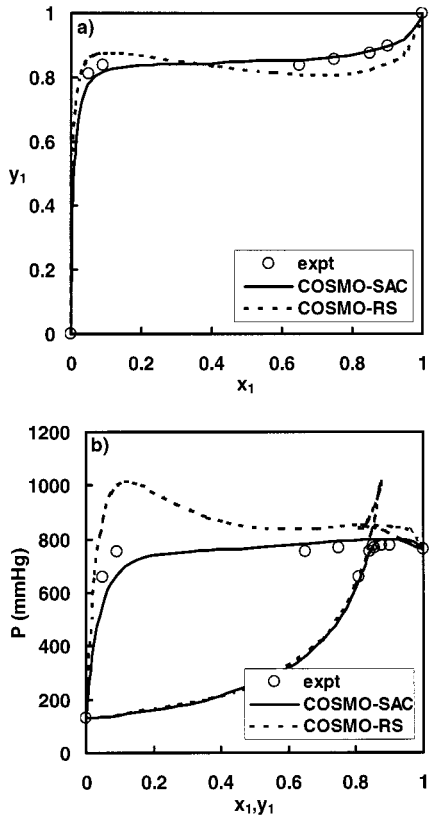

Figure 9. VLE prediction from COSMO-SAC and COSMO-RS for methyl acetate (1)/water (2) binary mixture at 330.05 K.

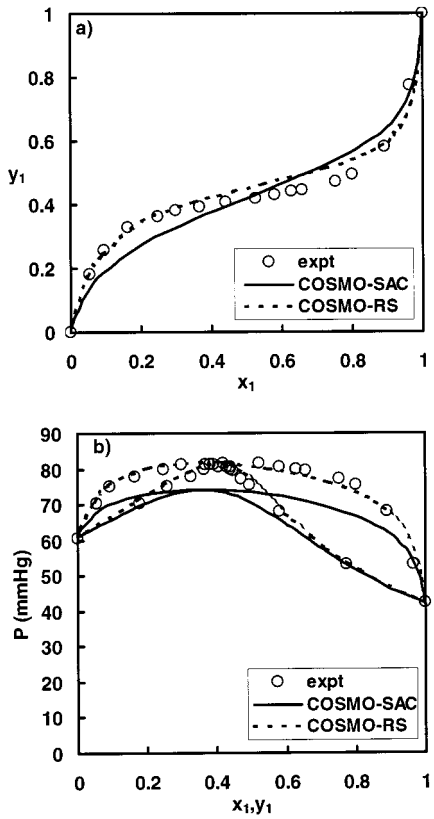

Figure 10. VLE prediction from COSMO-SAC and COSMO-RS for water (1)/1,4-dioxane (2) binary mixture at 308.15 K.

Table 3. Comparison of Activity Coefficient and Gibbs−Duhem Term from COSMO-RS and COSMO-SAC Models for Methyl Acetate (1)/Water (2) Mixture at 330.05 K

	COSMO-RS				COSMO-SAC		
x_1	ln γ_1	ln γ_2	GD[a]	GD[b] (eq A2.2)	ln γ_1	ln γ_2	GD[a]
2.40×10^{-9}	3.7380	0.0000	0.000	0.229	3.0894	0.0000	0.000
2.40×10^{-6}	3.7380	0.0000	−0.003	0.229	3.0893	0.0000	−0.002
2.40×10^{-4}	3.7330	0.0000	−0.024	0.228	3.0856	0.0000	0.001
2.50×10^{-3}	3.6940	0.0000	0.347	0.228	3.0504	0.0001	−0.003
4.90×10^{-3}	3.6500	0.0010	0.223	0.226	3.0137	0.0002	0.000
1.30×10^{-2}	3.5200	0.0040	0.135	0.222	2.8953	0.0013	−0.001
2.60×10^{-2}	3.3080	0.0100	0.182	0.217	2.7209	0.0047	−0.002
4.10×10^{-2}	3.1010	0.0200	0.154	0.212	2.5409	0.0109	−0.002
5.70×10^{-2}	2.8990	0.0330	0.152	0.207	2.3702	0.0197	−0.001
7.50×10^{-2}	2.7000	0.0500	0.145	0.201	2.2000	0.0317	−0.002
9.40×10^{-2}	2.5050	0.0710	0.141	0.195	2.0410	0.0463	−0.002
1.16×10^{-1}	2.3130	0.0970	0.149	0.189	1.8786	0.0653	−0.002
1.40×10^{-1}	2.1230	0.1290	0.158	0.183	1.7229	0.0882	−0.001
1.66×10^{-1}	1.9340	0.1680	0.109	0.177	1.5745	0.1149	−0.002
1.96×10^{-1}	1.7440	0.2140	0.095	0.170	1.4244	0.1480	−0.001
2.29×10^{-1}	1.5550	0.2690	0.139	0.163	1.2800	0.1869	−0.001
2.67×10^{-1}	1.3670	0.3380	0.113	0.155	1.1350	0.2347	−0.001
3.11×10^{-1}	1.1800	0.4210	0.113	0.148	0.9894	0.2938	−0.001
3.62×10^{-1}	0.9920	0.5250	0.112	0.140	0.8441	0.3674	−0.001
4.22×10^{-1}	0.8060	0.6560	0.085	0.131	0.6981	0.4614	−0.001
4.93×10^{-1}	0.6200	0.8240	0.090	0.122	0.5521	0.5844	0.000
5.79×10^{-1}	0.4380	1.0510	0.083	0.113	0.4051	0.7541	0.002
6.86×10^{-1}	0.2620	1.3780	0.110	0.103	0.2571	1.0094	0.012
8.22×10^{-1}	0.1060	1.9170	0.092	0.094	0.1129	1.4584	0.026
9.23×10^{-1}	0.0280	2.5240	0.076	0.090	0.0347	2.0135	0.013
9.60×10^{-1}	0.0090	2.8780	0.107	0.088	0.0136	2.3613	0.062
9.96×10^{-1}	0.0000	3.4510	0.056	0.087	0.0004	3.0492	0.006
$1.00 \times 10^{+0}$	0.0000	3.5620	0.001	0.087	0.0000	3.2376	0.001

[a] GD = $x_1(\partial \ln \gamma_1/\partial x_1) + x_2(\partial \ln \gamma_2/\partial x_1)$ is calculated from numerical differentiation. [b] GD = $(-\lambda(A_1 - A_2))/(x_1A_1 + x_2A_2)$, with $\lambda = -0.14$, $A_1 = 114.03$ Å, and $A_2 = 43.27$ Å.

Table 4. Comparison of Activity Coefficient and Gibbs−Duhem Term from COSMO-RS and COSMO-SAC Models for Water (1)/1,4-Dioxane (2) Mixture at 308.15 K

	COSMO-RS				COSMO-SAC		
x_1	ln γ_1	ln γ_2	GD[a]	GD[b] (eq A2.2)	ln γ_1	ln γ_2	GD[a]
4.90×10^{-8}	2.2580	0.0000	−0.001	−0.090	2.0341	0.0000	−0.001
4.90×10^{-5}	2.2570	0.0000	−0.032	−0.090	2.0328	0.0000	−0.003
4.90×10^{-3}	2.1940	0.0000	−0.019	−0.091	1.9223	0.0003	−0.037
4.70×10^{-2}	1.8260	0.0090	−0.082	−0.094	1.4575	0.0110	−0.009
9.10×10^{-2}	1.5850	0.0230	−0.114	−0.099	1.2115	0.0289	−0.018
2.05×10^{-1}	1.1690	0.0800	−0.072	−0.109	0.8454	0.0901	−0.008
3.52×10^{-1}	0.8150	0.2020	−0.096	−0.121	0.5701	0.1947	−0.001
4.63×10^{-1}	0.6100	0.3250	−0.111	−0.133	0.4235	0.2953	0.000
5.50×10^{-1}	0.4730	0.4460	−0.099	−0.144	0.3294	0.3919	0.001
6.20×10^{-1}	0.3760	0.5660	−0.115	−0.153	0.2631	0.4856	0.001
6.77×10^{-1}	0.3030	0.6820	−0.122	−0.163	0.2136	0.5771	0.001
7.25×10^{-1}	0.2460	0.7960	−0.136	−0.172	0.1742	0.6695	0.001
7.65×10^{-1}	0.2000	0.9090	−0.125	−0.180	0.1428	0.7616	0.001
8.00×10^{-1}	0.1630	1.0220	−0.166	−0.188	0.1160	0.8582	0.001
8.30×10^{-1}	0.1310	1.1360	−0.128	−0.195	0.0935	0.9573	0.002
8.57×10^{-1}	0.1050	1.2540	−0.128	−0.202	0.0738	1.0639	0.001
8.80×10^{-1}	0.0830	1.3770	−0.202	−0.209	0.0575	1.1718	0.002
9.01×10^{-1}	0.0630	1.5010	−0.164	−0.215	0.0432	1.2884	0.001
9.19×10^{-1}	0.0470	1.6300	−0.138	−0.221	0.0316	1.4055	0.002
9.36×10^{-1}	0.0340	1.7640	−0.168	−0.227	0.0216	1.5345	0.001
9.51×10^{-1}	0.0230	1.9030	−0.178	−0.232	0.0137	1.6662	0.002
9.65×10^{-1}	0.0140	2.0490	−0.117	−0.238	0.0076	1.8072	0.002
9.78×10^{-1}	0.0080	2.2000	−0.213	−0.243	0.0032	1.9563	0.001
9.89×10^{-1}	0.0030	2.3560	−0.180	−0.246	0.0009	2.0978	−0.002
9.96×10^{-1}	0.0010	2.4530	−0.451	−0.248	0.0001	2.1956	0.004
9.98×10^{-1}	0.0000	2.4850	0.018	−0.249	0.0000	2.2247	0.000
$1.00 \times 10^{+0}$	0.0000	2.5140	0.002	−0.250	0.0000	2.2513	0.002

[a] GD = $x_1(\partial \ln \gamma_1/\partial x_1) + x_2(\partial \ln \gamma_2/\partial x_1)$ is calculated from numerical differentiation. [b] GD = $(-\lambda(A_1 - A_2))/(x_1A_1 + x_2A_2)$, with $\lambda = -0.14$, $A_1 = 43.27$ Å, and $A_2 = 120.46$ Å.

not. It should be noted that the inconsistency of the COSMO-RS model is a result of

$$x_1\frac{\partial \ln \gamma_1^{\text{COSMO-RS}}}{\partial x_1} + x_2\frac{\partial \ln \gamma_2^{\text{COSMO-RS}}}{\partial x_1} = \frac{-\lambda(A_1 - A_2)}{x_1A_1 + x_2A_2} \quad \text{(A2.2)}$$

where A_i is the surface area of component i in the binary mixture. Equation A2.2 is not zero unless the surface areas (A_1 and A_2) are identical. Furthermore, the inconsistency problem becomes more severe for systems containing components that are very different in size.

Literature Cited

(1) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-Contribution Estimation of Activity-Coefficients in Nonideal Liquid-Mixtures. *AIChE J.* **1975**, *21*, 1086.

(2) Gmehling, J.; Li, J. D.; Schiller, M. A Modified Unifac Model. 2. Present Parameter Matrix and Results for Different Thermodynamic Properties. *Ind. Eng. Chem. Res.* **1993**, *32*, 178.

(3) Lin, S. T.; Sandler, S. I. Multipole Corrections to Account for Structure and Proximity Effects in Group Contribution Methods: Octanol−Water Partition Coefficients. *J. Phys. Chem.* **2000**, *104*, 7099.

(4) Klamt, A. Conductor-Like Screening Model for Real Solvents—a New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224.

(5) Klamt, A.; Schuurmann, G. Cosmo—a New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.

(6) Tanpipat, N.; Klamt, A.; Clausen, I.; Andzelm, J. Dmol/Cosmo-Rs: A Bridge between Chemistry and Chemical Engineering. *Abstr. Pap.—Am. Chem. Soc.* **1999**, *217*, U671.

(7) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. Cosmo Implementation in Turbomole: Extension of an Efficient Quantum Chemical Code Towards Liquid Systems. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187.

(8) Klamt, A.; Eckert, F. Cosmo-Rs: A Novel and Efficient Method for the a Priori Prediction of Thermophysical Data of Liquids. *Fluid Phase Equilib.* **2000**, *172*, 43.

(9) Lin, S. T.; Sandler, S. I. Infinite Dilution Activity Coefficients from Ab Initio Solvation Calculations. *AIChE J.* **1999**, *45*, 2606.

(10) Tomasi, J.; Persico, M. Molecular-Interactions in Solution—an Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027.

(11) Amovilli, C.; Mennucci, B. Self-Consistent-Field Calculation of Pauli Repulsion and Dispersion Contributions to the Solvation Free Energy in the Polarizable Continuum Model. *J. Phys. Chem. B* **1997**, *101*, 1051.

(12) Ben-Naim, A. *Solvation Thermodynamics*. Plenum Press: New York, 1987.

(13) Lin, S. T.; Sandler, S. I. Prediction of Octanol−Water Partition Coefficients Using a Group Contribution Solvation Model. *Ind. Eng. Chem. Res.* **1999**, *38*, 4081.

(14) Staverman, A. J. The Entropy of High Polymer Solutions. *Recl. Trav. Chim. Pays-Bas* **1950**, *69*, 163.

(15) Guggenheim, E. A. *Mixtures*. Clarendon Press: Oxford, U.K., 1952.

(16) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. Refinement and Parametrization of Cosmo-Rs. *J. Phys. Chem. A* **1998**, *102*, 5074.

(17) Cerius2, Dmol3, version 4.0; Molecular Simulations Inc.: San Diego, CA, 1999.

(18) Delley, B. An All-Electron Numerical-Method for Solving the Local Density Functional for Polyatomic-Molecules. *J. Chem. Phys.* **1990**, *92*, 508.

(19) Andzelm, J.; Sosa, C.; Eades, R. A. Theoretical-Study of Chemical-Reactions Using Density-Functional Methods with Non-local Corrections. *J. Phys. Chem.* **1993**, *97*, 4664.

(20) Abrams, D. S.; Ausnitz, J. M. Statistical Thermodynamics of Liquid-Mixtures—New Expression for Excess Gibbs Energy of Partly or Completely Miscible Systems. *AIChE J.* **1975**, *21*, 116.

(21) Gmehling, J.; Fischer, K.; Menke, J.; Rarey, J. The Dortmund Data Bank and the Integrated Software Package. *Fluid Phase Equilib.* In preparation, **2000**.

(22) Wu, H. S.; Sandler, S. I. Use of Ab Inito Quantum Mechanics Calculations in Group Contribution Methods. 2. Test of New Groups in Unifac. *Ind. Eng. Chem. Res.* **1991**, *30*, 889.

(23) Clausen, I.; Arlt, W. A Priori Calculation of Phase Equilibria for Thermal Process Engineering with the Aid of Cosmo-Rs. *Chem.-Ing.-Tech.* **2000**, *72*, 727.

(24) Prausnitz, J. M.; Lichtenthaler, R. N.; Azevedo, E. G. d. *Molecular Thermodynamics of Fluid-Phase Equilibria*, 2nd ed.; Prentice Hall: New York, 1986.

(25) Reiss, H. *Adv. Chem. Phys.* **1966**, *9*, 1.

(26) Reiss, H.; Fish, H. L.; Lebowitz, J. L. Statistical Mechanics of Rigid Spheres. *J. Chem. Phys.* **1959**, *31*, 369.

(27) Boublík, T. Hard-Sphere Equation of State. *J. Chem. Phys.* **1970**, *53*, 471.

(28) Mansoori, G. A.; Carnahan, N. F.; Starling, K. E.; Leland, J. T. W. Equilibrium Thermodynamic Properties of the Mixture of Hard Spheres. *J. Chem. Phys.* **1971**, *54*, 1523.

(29) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. Hydrophobic Effects on a Molecular Scale. *J. Phys. Chem. B* **1998**, *102*, 10469.

(30) Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, L. R. An Information Theory Model of Hydrophobic Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8951.

Received for review December 1, 2000
Revised manuscript received July 10, 2001
Accepted November 2, 2001

IE001047W