

# Sigma-Profile Database for Using COSMO-Based Thermodynamic Methods

Eric Mullins,<sup>†</sup> Richard Oldland,<sup>†</sup> Y. A. Liu,<sup>\*,†</sup> Shu Wang,<sup>‡</sup> Stanley I. Sandler,<sup>‡</sup> Chau-Chyun Chen,<sup>§</sup> Michael Zwolak,<sup>†</sup> and Kevin C. Seavey<sup>†</sup>

*Honeywell Center of Excellence in Computer-Aided Design and SINOPEC/FPCC/AspenTech Center of Excellence in Process Systems Engineering, Department of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, Center for Molecular and Engineering Thermodynamics, Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716, and Aspen Technology, Inc., Ten Canal Park, Cambridge, Massachusetts 02141*

Thermodynamic methods based on conductor-like screening models (COSMO) originated from the use of solvation thermodynamics and computational quantum mechanics. These methods rely on sigma profiles specific to each molecule. A sigma profile is the probability distribution of a molecular surface segment having a specific charge density. Two COSMO-based thermodynamic models are COSMO-RS (realistic solvation) developed by Klamt and his colleagues, and COSMO-SAC (segment activity coefficient) published by Lin and Sandler. Quantum mechanical calculations for generating the sigma profiles represent the most time-consuming and computationally expensive aspect of using COSMO-based methods. A growing number of scientists and engineers are interested in the COSMO-based thermodynamic models but are intimidated by the complexity of performing quantum mechanical calculations. This paper presents the first free, web-based sigma profile database of 1432 compounds. We describe the procedure for sigma profile generation, and we have validated our database by comparing COSMO-based predictions of activity coefficients, normal boiling point and solubility with experimental data and thermodynamic property database. We discuss improvements which include using supplemental geometry optimization software packages to provide good initial guesses for molecular conformations as a precursor to the COSMO calculation. Finally, this paper provides a FORTRAN program and a procedure to generate additional sigma profiles, as well as a FORTRAN program to generate binary phase-equilibrium predictions using the COSMO-SAC model. Our sigma profile database will facilitate predictions of thermodynamic properties and phase behaviors from COSMO-based thermodynamic models.

## 1. Introduction

In process and product development, chemists and engineers frequently need to perform phase-equilibrium calculations and account for liquid-phase nonidealities resulting from molecular interactions. They often use group-contribution methods such as UNIFAC, or activity-coefficient models such as NRTL. These methods require binary interaction parameters regressed from experimental data, and thus have little or no applicability to compounds with new functional groups (in the case of UNIFAC) or new compounds (in the case of NRTL) without a substantial experimental database and data analysis.

An alternative approach is to use solvation-thermodynamics methods to characterize molecular interactions and account for liquid-phase nonideality. These models, which are based on computational quantum mechanics, allow us to predict thermophysical properties without any experimental data. Two such models are conductor-like screening models—realistic solvation (COSMO-RS)<sup>1–4</sup> and conductor-like screening models—segment activity coefficient (COSMO-SAC),<sup>5,6</sup> which predict intermolecular interactions based on only molecular structure and a few adjustable parameters. COSMO-RS is the first extension of a dielectric continuum-solvation model to liquid-phase thermodynamics, and COSMO-SAC is a variation of COSMO-RS.

These solvation-thermodynamics methods require sigma profiles in a manner similar to the way UNIFAC requires

parameter databases, with one exception: sigma profiles are molecule-specific, whereas UNIFAC binary interaction parameters are specific to functional groups. We generate sigma profiles from the molecular structure only with quantum-mechanical calculations. Doing the quantum mechanical calculations is the most time-consuming task of applying COSMO-based methods, representing over 90% of the computational effort involved in property predictions. The lack of a comprehensive, open-literature sigma-profile database hinders the ability to apply or improve the COSMO approach (“...available database is, relatively nonextensive...”).<sup>7</sup>

This paper presents the first open-literature database that contains sigma profiles for 1432 chemicals. These chemicals have only 10 elements: hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine. Our database is available free of charge from our website ([www.design.che.vt.edu](http://www.design.che.vt.edu)). We will continue to update the sigma profiles as new results become available. Our website also includes a detailed procedure for generating additional sigma profiles for any compound, along with FORTRAN programs for the sigma-averaging algorithm and the COSMO-SAC model. We present several examples, including a study of conformational effects, to validate the accuracy of our sigma profile database and our implementation of the COSMO-SAC model by comparing our predictions of activity coefficients, vapor-liquid equilibria, and solubilities with literature values.

## 2. Background and Theory

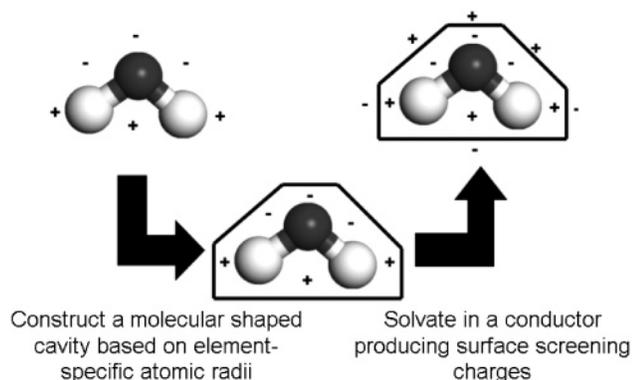
We give a brief overview of COSMO-based thermodynamic models, sigma profiles, and the COSMO-SAC model. COSMO-based models generate a surface-charge distribution. We plot

\* To whom correspondence should be addressed. Tel.: (540) 231-7800. Fax: (540) 231-5022. E-mail: [design@vt.edu](mailto:design@vt.edu).

<sup>†</sup> Virginia Polytechnic Institute and State University.

<sup>‡</sup> University of Delaware.

<sup>§</sup> Aspen Technology, Inc.



**Figure 1.** Schematics showing that the ideal solvation process in the COSMO-based model places the molecule in a cavity and into a conducting medium. The molecule pulls charges from the conductor to the cavity surface. We then represent this surface-charge distribution as a sigma profile.

**Table 1. Parameter Values Used in the COSMO–SAC Model<sup>a</sup>**

symbol	value	description
$r_{av}$	0.81764 Å	sigma averaging radius
$a_{eff}$	7.5 Å <sup>2</sup>	effective surface segment surface area
$c_{hb}$	85580.0 (kcal Å <sup>4</sup> )/(mol e <sup>2</sup> )	hydrogen-bonding constant
$\sigma_{hb}$	0.0084 e/Å <sup>2</sup>	sigma cutoff for hydrogen bonding
$\alpha'$	9034.97 (kcal Å <sup>4</sup> )/(mol e <sup>2</sup> )	misfit energy constant
$z$	10	coordination number
$q$	79.53 Å <sup>2</sup>	standard area parameter
$r$	66.69 Å <sup>3</sup>	standard volume parameter

<sup>a</sup> Data taken from refs 2 and 5.

the probability distribution of a molecular surface segment that has a specific charge density as the sigma profile, which is used by COSMO–RS/SAC to compute activity coefficients.

**2.1. The COSMO-Based Thermodynamic Model.** The basis of the COSMO-based model is the “solvent-accessible surface”

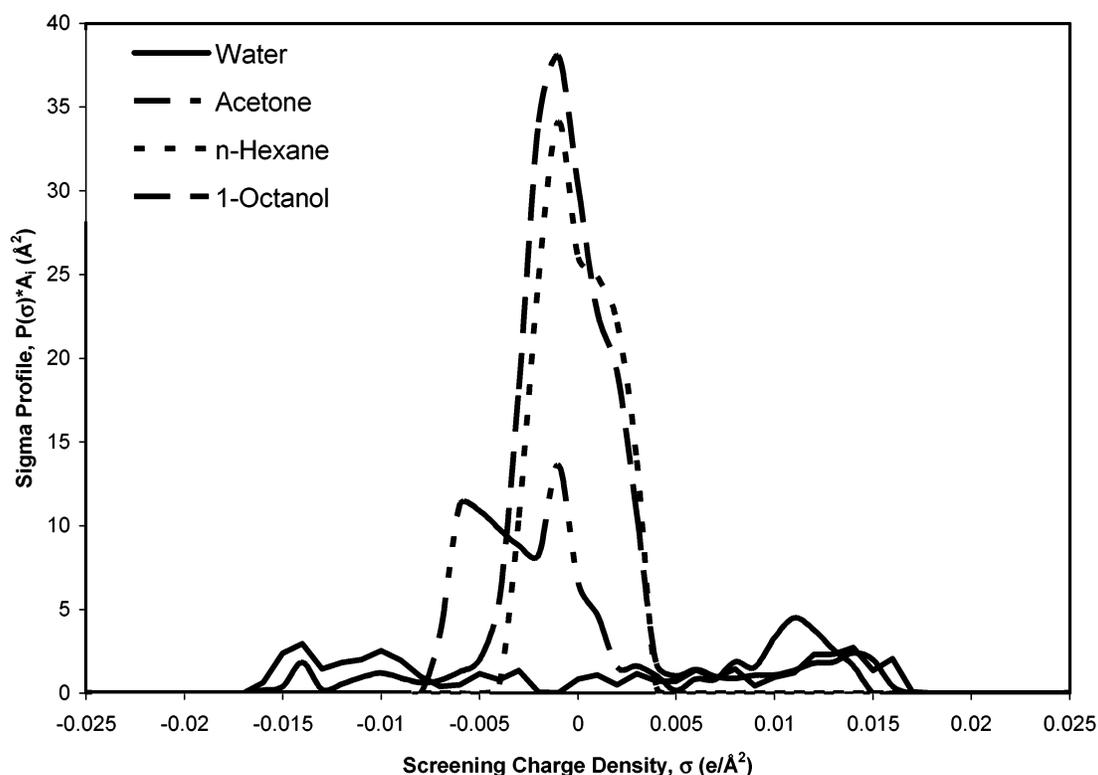
of a solute molecule.<sup>8,9</sup> Conceptually, COSMO-based models place the molecule inside a cavity formed within a homogeneous medium, taken to be the solvent. Figure 1 illustrates the ideal solvation process in the COSMO-based model.

The model constructs the cavity within a perfect conductor, according to a specific set of rules and atom-specific dimensions. The molecule’s dipole and higher moments then draw charge from the surrounding medium to the surface of the cavity to cancel the electric field both inside the conductor and tangential to the surface. We find the induced surface charges in a discretized space with eq 1:

$$\Phi_{tot} = 0 = \Phi_{sol} + Aq^* \quad (1)$$

where  $\Phi_{tot}$  is the total potential on the cavity surface,  $\Phi_{sol}$  the potential due to the charge distribution of the solute molecule, and  $q^*$  the surface screening charge in the conductor.  $A$  is the “coulomb interaction matrix”, which describes potential interactions between surface charges and is a function of the cavity geometry.<sup>9</sup> The surface-charge distribution in a finite dielectric solvent is well-approximated by a simple scaling of the surface-charge distribution in a conductor ( $\sigma^*$ ). In this way, COSMO greatly reduces the computational cost with a minimal loss of accuracy.<sup>9</sup>

**2.2. The COSMO–SAC Model.** We outline the derivation and explanation of the COSMO–SAC model as published in Lin and Sandler,<sup>5</sup> and for the small variations, we use in the VT-2005 programs. The activity coefficient defined in eq 2 is the result of two contributions. The first of which is defined as the difference in the free energies of restoring the charges around the solute molecule in a solution  $S$ ,  $\Delta G_{i/S}^{*res}$ , and restoring the charges in a pure liquid  $i$ ,  $\Delta G_{i/i}^{*res}$ . The restoring free energy is part of the charging free energy,  $\Delta G^{*chg}$ , which is defined as the sum of the ideal solvation free energy ( $\Delta G^{*is}$ ) and the free energy required to remove the screening charges on the solute. The ideal solvation energy is identical for dissolving a solute

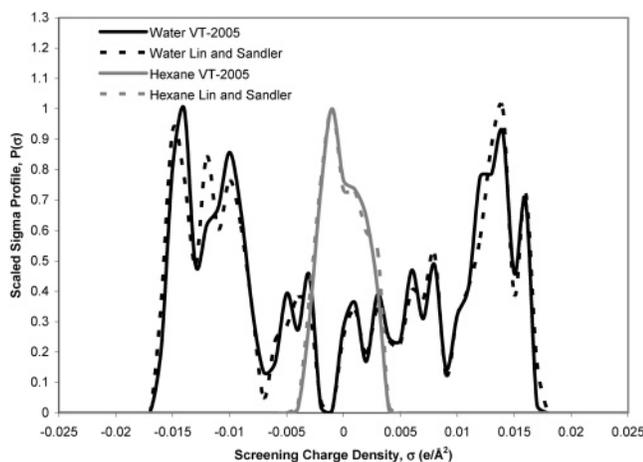


**Figure 2.** Sigma profiles for water, acetone, *n*-hexane, and 1-octanol.

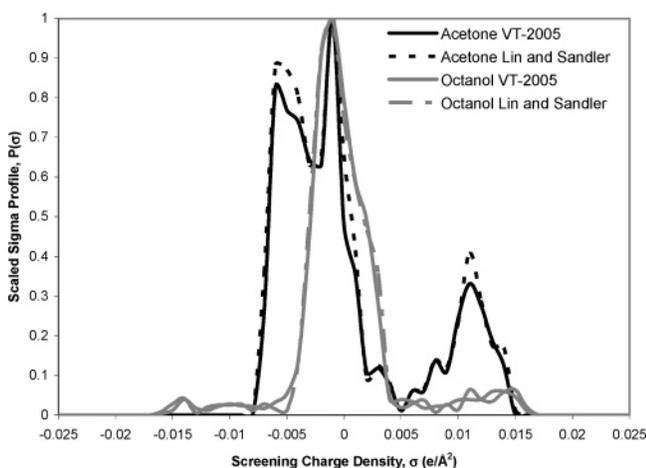
**Table 2.** COSMO Keywords Used in Calculating Surface Segment Charges in DMol<sup>a</sup>

keyword name	default value	description
Cosmo	on	turns on COSMO solvation procedure
Cosmo_Grid_Size	1082	tells DMol <sup>3</sup> how many basic grid points per atom to consider
Cosmo_Segments	92	specifies the maximum number of segments on each atomic surface
Cosmo_Solvent_Radius	1.300000	solvent probe radius
Cosmo_A-Matrix_Cutoff	7.000000	determines the accuracy of the electrostatic interactions on the COSMO surface
Cosmo_Radus_Incr	0.000000	specifies the increment to the atomic radii used in the construction of the COSMO cavity
Cosmo_RadCorr_Incr	0.150000	used to construct the outer cavity for the outlying charge correction
Cosmo_A-Constraint	1.882190	used to approximate the non-electrostatic contribution to the solvation energy within the COSMO model
Cosmo_B-Constraint	0.010140	used to approximate the non-electrostatic contribution to the solvation energy within the COSMO model

<sup>a</sup> Parameter descriptions are available in the Accelrys Materials Studio Software documentation.<sup>20</sup>



**Figure 3.** Comparison of scaled sigma profiles between VT-2005 and the Lin and Sandler<sup>5</sup> data for water and *n*-hexane, scaled relative to their maximum values. The solid curves represent VT-2005 profiles, and the dashed curves show the Lin and Sandler<sup>5</sup> sigma profiles.



**Figure 4.** Comparison of scaled sigma profiles between VT-2005 and Lin and Sandler<sup>5</sup> data for acetone and 1-octanol, scaled relative to their maximum values. The solid curves represent VT-2005 profiles, and the dashed curves show the Lin and Sandler sigma profiles.

in a solvent *S* or in pure solute *i* and, therefore, the charging free energy is reduced to the restoring free energy.<sup>10</sup> The second contribution to the activity coefficient is the Staverman–Guggenheim combinatorial term,  $\gamma_{i/s}^{SG}$ , which improves the calculations for the cavity-formation free energy, according to Lin and Sandler.<sup>11</sup>

$$\ln \gamma_{i/s} = \frac{\Delta G_{i/s}^{*res} - \Delta G_{i/i}^{*res}}{RT} + \ln \gamma_{i/s}^{SG} \quad (2)$$

**Table 3.** Table of Elemental Atomic Radii for Creating the COSMO Molecular Cavity<sup>2,3,24</sup>

element	cavity radius(Å)	element	cavity radius(Å)
H	1.30	S	2.16
C	2.00	P	2.12
N	1.83	Cl	2.05
O	1.72	Br	2.16
F	1.72	I	2.32

They define the Staverman–Guggenheim combinatorial term as

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

where  $\phi_i$  is the normalized volume fraction,  $\theta_i$  is the normalized surface-area fraction,  $l_i = (z/2)(r_i - q_i) - (r_i - 1)$ ,  $z$  is the coordination number (value taken as 10),  $x_i$  is the mole fraction,  $r_i$  and  $q_i$  are the normalized volume and surface-area parameters, i.e.,  $q_i = A_i/q$  and  $r_i = V_i/r$ , where  $A_i$  is the cavity surface area and  $V_i$  is the cavity volume, both from the COSMO calculation.

Lin and Sandler<sup>5</sup> have defined the restoring free energy as the sum of the sigma profile times the natural logarithm of the segment activity coefficients over all surface charges:

$$\frac{\Delta G_{i/s}^{*res}}{RT} = \sum_{\sigma_m} \left[ n_i(\sigma_m) \frac{\Delta G_{\sigma_m/s}^{*res}}{RT} \right] = n_i \sum_{\sigma_m} p_i(\sigma_m) \ln \Gamma_s(\sigma_m) \quad (4)$$

where  $\Gamma_s(\sigma_m)$  is the activity coefficient for a segment of charge density  $\sigma$ .

We calculate the segment activity coefficient using

$$\ln \Gamma_s(\sigma_m) = -\ln \left\{ \sum_{\sigma_n} p_s(\sigma_n) \Gamma_s(\sigma_n) \exp \left[ \frac{-\Delta W(\sigma_m, \sigma_n)}{RT} \right] \right\}$$

$$\ln \Gamma_i(\sigma_m) = -\ln \left\{ \sum_{\sigma_n} p_i(\sigma_n) \Gamma_i(\sigma_n) \exp \left[ \frac{-\Delta W(\sigma_m, \sigma_n)}{RT} \right] \right\} \quad (5)$$

as derived rigorously, using statistical mechanics.<sup>5</sup>

The exchange energy,  $\Delta W(\sigma_m, \sigma_n)$ , is

$$\Delta W(\sigma_m, \sigma_n) = \left( \frac{\alpha'}{2} \right) (\sigma_m + \sigma_n)^2 + c_{hb} \max[0, \sigma_{acc} - \sigma_{hb}] \min[0, \sigma_{don} + \sigma_{hb}] \quad (6)$$

where  $\alpha'$  is the constant for the misfit energy, which Klamt et al.<sup>2</sup> and Klamt and Eckert<sup>3</sup> fit to experimental data,  $c_{hb}$  is a constant for hydrogen bonding, and  $\sigma_{hb}$  is the sigma-value cutoff for hydrogen bonding.

Table 4. Summary of the VT-2005 Sigma Profile Database

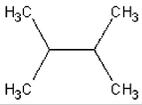
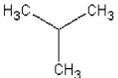
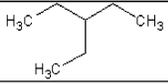
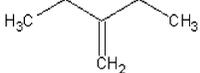
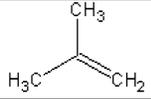
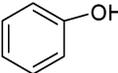
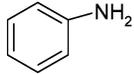
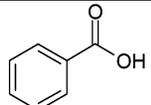
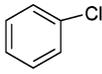
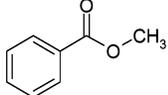
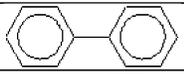
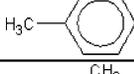
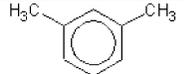
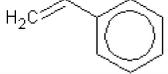
Chemical Family	Characteristic Structure	Number of Compounds in VT-2005	Example Compounds
Alkanes			
Dimethyl-Alkanes		21	2,3-dimethyl-butane
Methyl-Alkanes		17	Isobutane
n-Alkanes		27	Propane
Other Alkanes		27	3-Ethylpentane
Alkenes			
Dialkenes		25	1,4-pentadiene
Ethyl/higher-Alkenes		11	2-ethyl-1-butene (C <sub>6</sub> H <sub>12</sub> )
Methyl-Alkenes		22	Isobutylene
n-Alkenes	$H_2C=CH_2$	41	Ethylene, 1-Butene, cis-2-Butene
Alkynes			
n-Alkynes	$R-C\equiv C-R$	18	1-Butyne (C <sub>4</sub> H <sub>6</sub> )
Aromatic structures			
Aromatic Alcohols		31	Phenol
Aromatic Amines		37	Pyridine and Aniline
Aromatic Carboxylic Acids		9	Benzoic Acid (C <sub>7</sub> H <sub>6</sub> O <sub>2</sub> )
Aromatic Chlorides		15	Benzyl Chloride (C <sub>7</sub> H <sub>7</sub> Cl)
Aromatic Esters		19	Benzyl Acetate (C <sub>9</sub> H <sub>10</sub> O <sub>2</sub> )
Diphenol/Polyaromatics		17	Biphenyl (C <sub>12</sub> H <sub>10</sub> )
n-Alkyl Benzenes		19	Toluene
Naphthalenes		15	1-Methylnaphthalene (C <sub>11</sub> H <sub>10</sub> )
Other Alkyl Benzenes		44	m-Xylene (C <sub>8</sub> H <sub>10</sub> )
Other Monoaromatics		15	Styrene

Table 4 (Continued)

Chemical Family	Characteristic Structure	Number of Compounds in VT-2005	Example Compounds
Carboxylic compounds			
Acetates		22	Methyl Acetate, Vinyl Acetate, Allyl Acetate
Anhydrides		8	Acetic Anhydride
Dicarboxylic Acids		11	Maleic Acid (C <sub>4</sub> H <sub>4</sub> O <sub>4</sub> )
Formates		13	Ethyl Formate (C <sub>3</sub> H <sub>6</sub> O <sub>2</sub> )
Ketones		33	Acetone
n-Aliphatic Acids		20	Formic Acid, Acetic Acid, Propionic Acid
Propionates and Butyrates		13	Methyl Propionate (C <sub>4</sub> H <sub>8</sub> O <sub>2</sub> )
Unsaturated Aliphatic Esters		23	Methyl Acrylate (C <sub>4</sub> H <sub>6</sub> O <sub>2</sub> )
Cyclic compounds			
Alkylcyclohexanes		19	Methylcyclohexane
Alkylcyclopentanes		22	Ethylcyclopentane
Cycloaliphatic Alcohols		10	Cyclohexanol
Cycloalkanes/alkenes		15	Cyclooctene (C <sub>8</sub> H <sub>14</sub> ) or Cyclobutane (C <sub>4</sub> H <sub>8</sub> )
Epoxides		14	1,3-Dioxane (C <sub>4</sub> H <sub>8</sub> O <sub>2</sub> ), Furan, Ethylene Oxide
Multiring-cycloalkanes		3	Cis-Decalin (C <sub>10</sub> H <sub>18</sub> )
Other Hydrocarbon Rings	No Common Structure	16	Indene (C <sub>10</sub> H <sub>16</sub> )
Halogenated compounds			
F, Cl, Br, I Compounds	H <sub>3</sub> C—X	94	Methyl Bromide, Ethyl Chloride, Difluoromethane
Multihalogenated Alkanes	X—CH <sub>2</sub> —X	37	Chlorofluoromethane
Inorganic compounds			
Inorganic Acids/Bases	No Common Structure	9	Sulfuric Acid, Nitric Acid, Ammonia
Inorganic Gases	No Common Structure	8	Ozone (O <sub>3</sub> )
Inorganic Halides		5	Thionyl Chloride (SOCl <sub>2</sub> )
Other Inorganics	No Common Structure	3	Water

Table 4 (Continued)

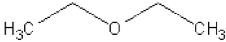
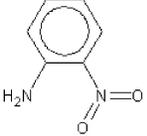
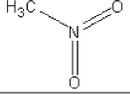
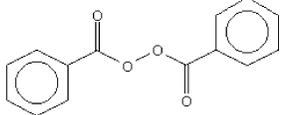
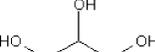
Chemical Family	Characteristic Structure	Number of Compounds in VT-2005	Example Compounds
Polyfunctional compounds			
Polyfunctional Acids	No Common Structure	16	Glycolic Acid (C <sub>2</sub> H <sub>4</sub> O <sub>3</sub> )
Polyfunctional Amides/Amines	No Common Structure	26	Formamide
Polyfunctional C, H, O, halides	No Common Structure	36	Chloroacetic Acid
Polyfunctional- C, H, N, halide, (O)	No Common Structure	12	O-Chloroaniline
Polyfunctional C, H, O, N	No Common Structure	27	Niacin
Polyfunctional C, H, O, S	No Common Structure	13	Sulfolane (C <sub>4</sub> H <sub>8</sub> O <sub>2</sub> S)
Polyfunctional Esters	No Common Structure	21	Ethyl Lactate (C <sub>5</sub> H <sub>10</sub> O <sub>3</sub> )
Polyfunctional Nitriles	No Common Structure	7	Aminocapronitrile (C <sub>6</sub> H <sub>12</sub> N <sub>2</sub> )
Other Polyfunctional C, H, O	No Common Structure	35	Furfural
Other Polyfunctional Organics	No Common Structure	4	Malathion (C <sub>10</sub> H <sub>19</sub> O <sub>6</sub> PS <sub>2</sub> )
Other structures			
Aldehydes		31	Formaldehyde
Aliphatic Ethers		33	Diethyl Ether
Elements	No Common Structure	5	Hydrogen, Bromine, Iodine
Isocyanates/ Diisocyanates	$R-N=C=O$	6	n-butyl-diisocyanate (C <sub>5</sub> H <sub>9</sub> NO)
Mercaptans	$HS-CH_3$	22	Methyl Mercaptan (CH <sub>4</sub> S)
n-Alcohols	$R-OH$	20	Methanol, Ethanol, 1-Octanol
n-Aliphatic Primary Amines		13	Ethyl Amine
Nitriles	$H_3C-C\equiv N$	28	Acetonitrile (C <sub>2</sub> H <sub>3</sub> N)
Nitroamines		4	o-Nitroaniline (C <sub>6</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub> )
C, H, NO <sub>2</sub> Compounds		20	Nitromethane (CH <sub>3</sub> NO <sub>2</sub> )
Sulfates, Nitrates, Phosphates, Carbonates	No Common Structure	11	Ethylene Carbonate (C <sub>3</sub> H <sub>4</sub> O <sub>3</sub> )
Peroxides		10	Benzoyl Peroxide (C <sub>14</sub> H <sub>10</sub> O <sub>4</sub> )
Polyols		33	Glycerol

Table 4 (Continued)

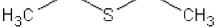
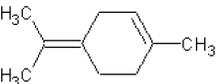
Chemical Family	Characteristic Structure	Number of Compounds in VT-2005	Example Compounds
Other structures			
Sulfides/ Thiophenes		21	Diethyl Sulfide (C <sub>4</sub> H <sub>10</sub> S)
Terpenes		6	Terpinolene (C <sub>10</sub> H <sub>16</sub> )
Other Aliphatic Acids	No Common Structure	19	Isobutyric Acid (C <sub>4</sub> H <sub>8</sub> O <sub>2</sub> )
Other Aliphatic Alcohols	No Common Structure	26	Isobutanol
Other Aliphatic Amines	No Common Structure	16	Dimethylamine
Other Amines/Imines	No Common Structure	36	Pyrazine (C <sub>4</sub> H <sub>4</sub> N <sub>2</sub> )
Other Ethers/Diethers	No Common Structure	17	Anethole (C <sub>10</sub> H <sub>12</sub> O)
Other Saturated Aliphatic Esters	No Common Structure	19	Caprolactone (C <sub>6</sub> H <sub>10</sub> O <sub>2</sub> )
Miscellaneous Other	No Common Structure	3	Ethanesulfonyl-Chloride (C <sub>2</sub> H <sub>5</sub> ClO <sub>2</sub> S)
Condensed Rings	No Common Structure	8	Anthracene (C <sub>14</sub> H <sub>10</sub> ), Pyrene (C <sub>16</sub> H <sub>10</sub> )

Table 1 shows the values that we use for each of the adjustable parameters in the COSMO-SAC model.<sup>2,5</sup>

We calculate the activity coefficient using the following expression:

$$\ln \gamma_{i/s} = n_i \sum_{\sigma_m} p_i(\sigma_m) [\ln \Gamma_s(\sigma_m) - \ln \Gamma_i(\sigma_m)] + \ln \gamma_{i/s}^{SG} \quad (7)$$

**2.3. Sigma Profiles.** We average the screening charge densities from the COSMO calculation output,  $\sigma^*$ , over a circular surface segment to obtain a new surface-charge density,  $\sigma$ . We then represent this charge distribution as the probability distribution of a molecular surface segment that has a specific charge density. We call this probability distribution the sigma profile,  $p(\sigma)$ . Klamt<sup>1,4</sup> has defined the sigma profile for a molecule  $i$ ,  $p_i(\sigma)$ , as

$$p_i(\sigma) = \frac{n_i(\sigma)}{n_i} = \frac{A_i(\sigma)}{A_i} \quad (8)$$

$$n_i = \sum_{\sigma} n_i(\sigma) = \frac{A_i}{a_{\text{eff}}} \quad (9)$$

$$A_i = \sum_{\sigma} A_i(\sigma) \quad (10)$$

where  $n_i(\sigma)$  is the number of segments with a discretized surface-charge density  $\sigma$ ,  $A_i$  the total cavity surface area, and  $A_i(\sigma)$  the total surface area of all of the segments with a particular charge density  $\sigma$ . Lin and Sandler<sup>5</sup> have defined  $A_i(\sigma) = a_{\text{eff}} n_i(\sigma)$ , where  $a_{\text{eff}}$  is the effective surface area of a standard surface segment that represents the contact area between different molecules, e.g., a theoretical bonding site. Klamt and co-workers<sup>2</sup> set this adjustable parameter,  $a_{\text{eff}}$ , to 7.1 Å<sup>2</sup>.

We calculate the sigma profile for a mixture as the weighted average of sigma profiles of pure components:

$$p_s(\sigma) = \frac{\sum_i x_i n_i p_i(\sigma)}{\sum_i x_i n_i} = \frac{\sum_i x_i A_i p_i(\sigma)}{\sum_i x_i A_i} \quad (11)$$

We average the surface-charge densities from the COSMO output to find an effective surface-charge density using the following equation:<sup>2</sup>

$$\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{\text{av}}^2}{r_n^2 + r_{\text{av}}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{\text{av}}^2}\right)}{\sum_n \frac{r_n^2 r_{\text{av}}^2}{r_n^2 + r_{\text{av}}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{\text{av}}^2}\right)} \quad (12)$$

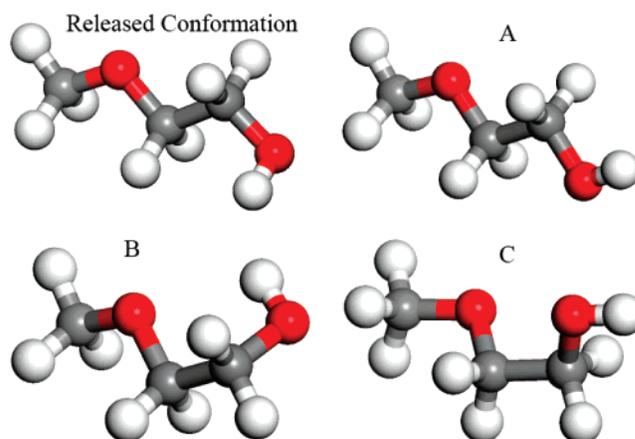
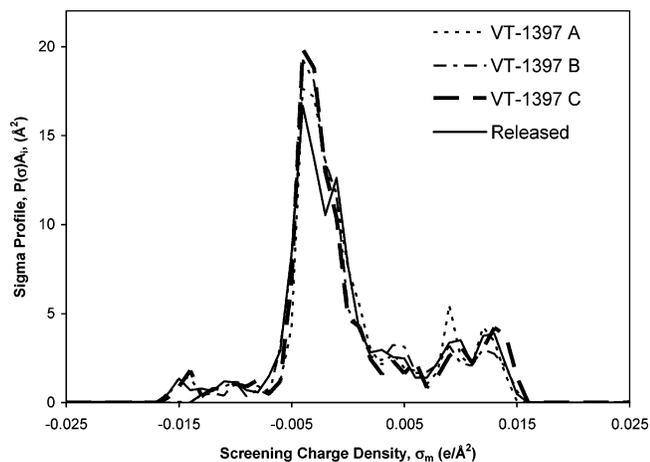


Figure 5. Low-energy conformations for 2-methoxy-ethanol (VT-1397).

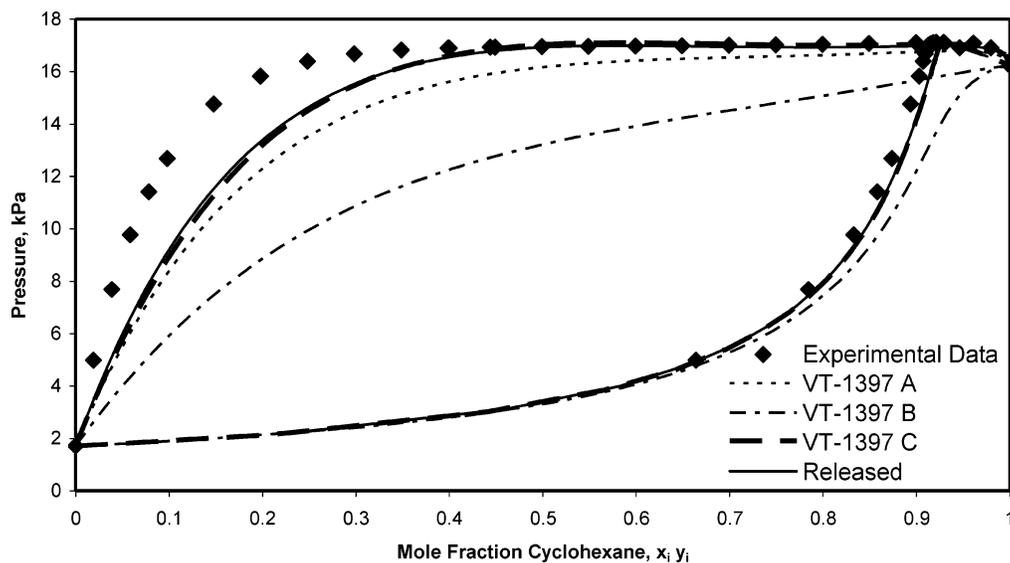


**Figure 6.** Sigma profiles for various conformations of 2-methoxy-ethanol (VT-1397). The released conformation is included in the VT-2005 Sigma Profile Database.

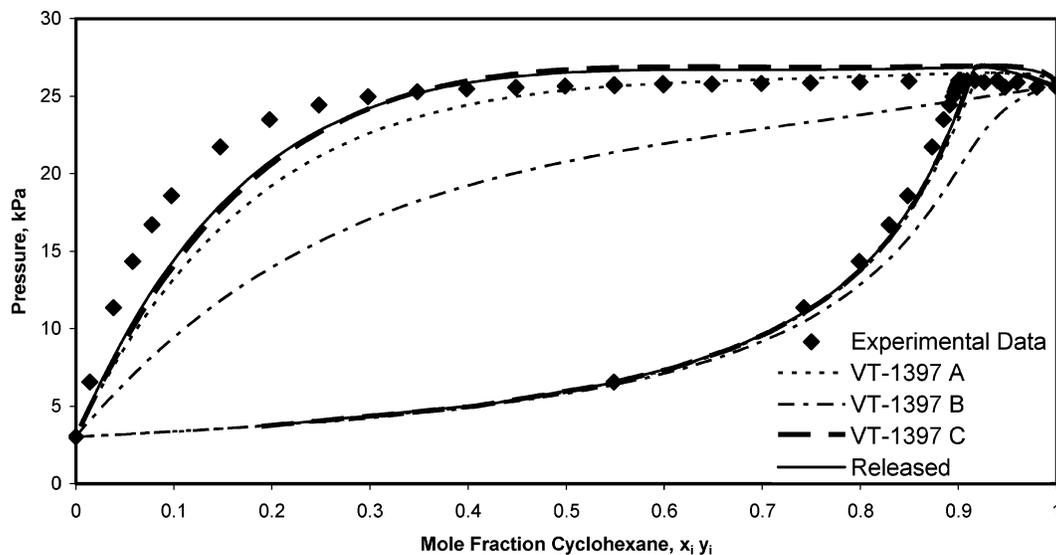
where  $\sigma_m$  is the average surface-charge density on segment  $m$ , the summation is over  $n$  segments from the COSMO output,  $r_n$  is the radius of the actual surface segment (assuming circular

segments),  $r_{av}$  is the averaging radius (an adjustable parameter), and  $d_{mn}$  is the distance between the two segments.<sup>2,5</sup> The paired segments  $m$  and  $n$  have segment charge densities  $\sigma_m$  and  $\sigma_n$ , respectively.

We use an averaging radius,  $r_{av} = 0.81764 \text{ \AA}$ , for the sigma-averaging algorithm that is different from the effective segment radius  $r_{eff}$  defined by Klamt.<sup>1</sup> This corresponds to the average segment surface area of  $a_{av} = 7.5 \text{ a.u.}^2 = 2.10025 \text{ \AA}^2$ . The averaging radius directly affects the sigma profiles. Our averaging algorithm is identical to those determined by Lin and Sandler,<sup>12</sup> Klamt,<sup>1</sup> and Klamt et al.,<sup>2</sup> except that we use a different value for  $r_{av}$ . Klamt and co-workers reported using averaging radii ranging between 0.5 and 1.0  $\text{\AA}$ , stating that “the best value for the averaging radius  $r_{av}$  turns out to be 0.5  $\text{\AA}$ . This is less than the initially assumed value of about 1  $\text{\AA}$ .”<sup>2</sup> Lin and Sandler used a similar algorithm that involved the effective segment radius but introduced another adjustable parameter,  $c$ , which they place in the exponential term in eq 12.<sup>12</sup> The VT-2005 sigma-profile database includes calculation results from the density-functional theory (DFT), thus enabling future work in optimization of the  $r_{av}$  value.



**Figure 7.** Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/cyclohexane system at 303.15 K.



**Figure 8.** Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/cyclohexane system at 313.15 K.

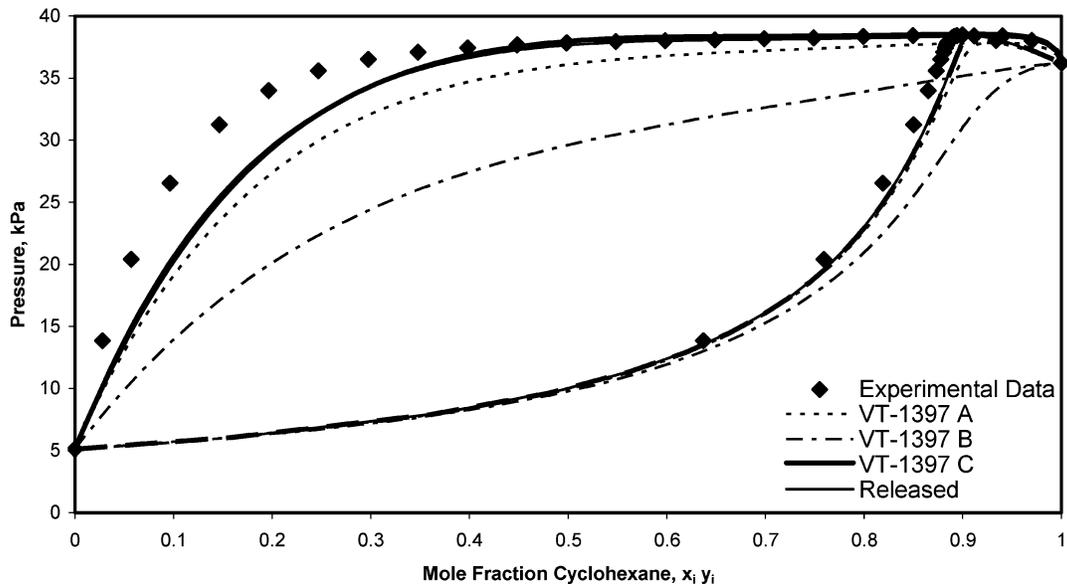


Figure 9. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/cyclohexane system at 323.15 K.

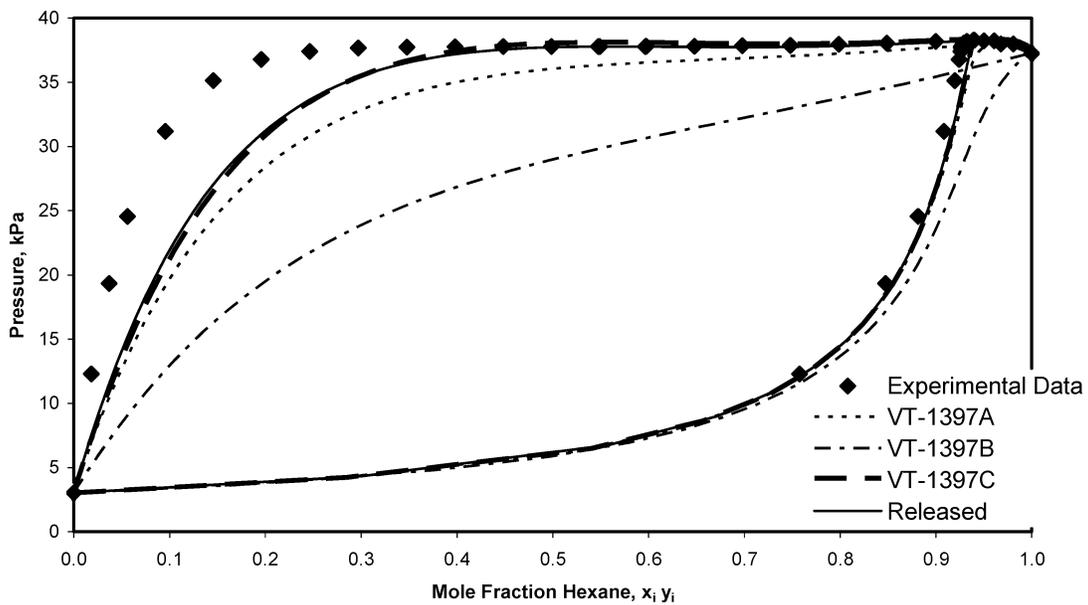


Figure 10. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/*n*-hexane system at 313.15 K.

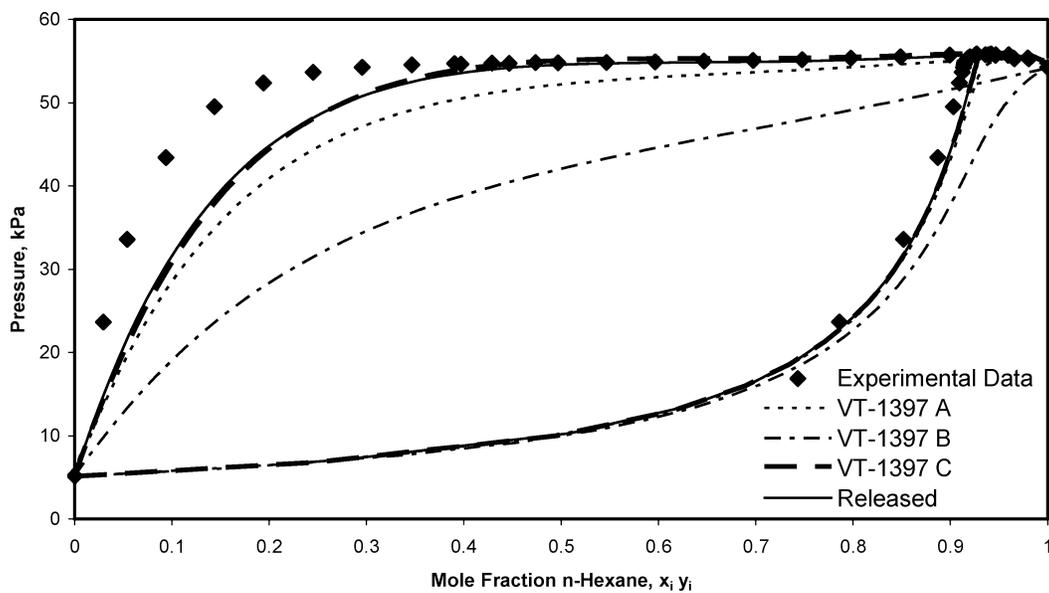


Figure 11. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/*n*-hexane system at 323.15 K.

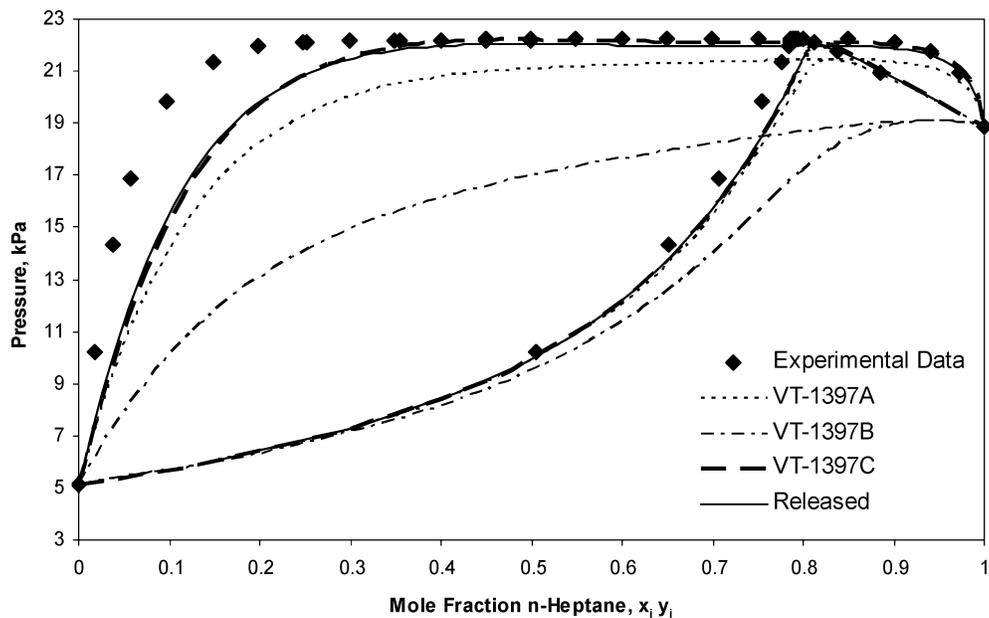


Figure 12. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/*n*-heptane system at 323.15 K.

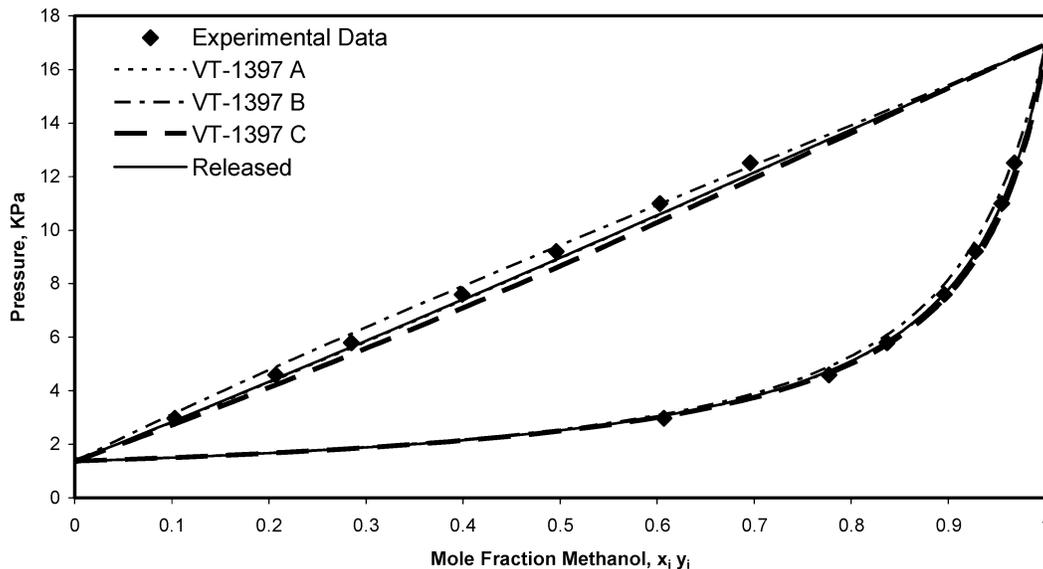


Figure 13. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/methanol system at 298.15 K.<sup>26</sup>

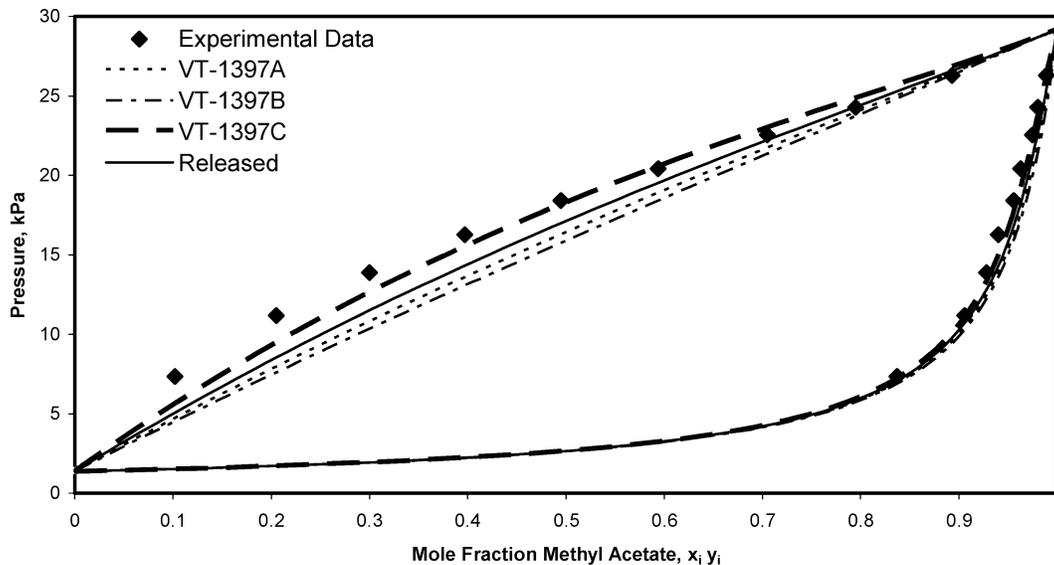


Figure 14. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/methyl acetate system at 298.15 K.<sup>26</sup>

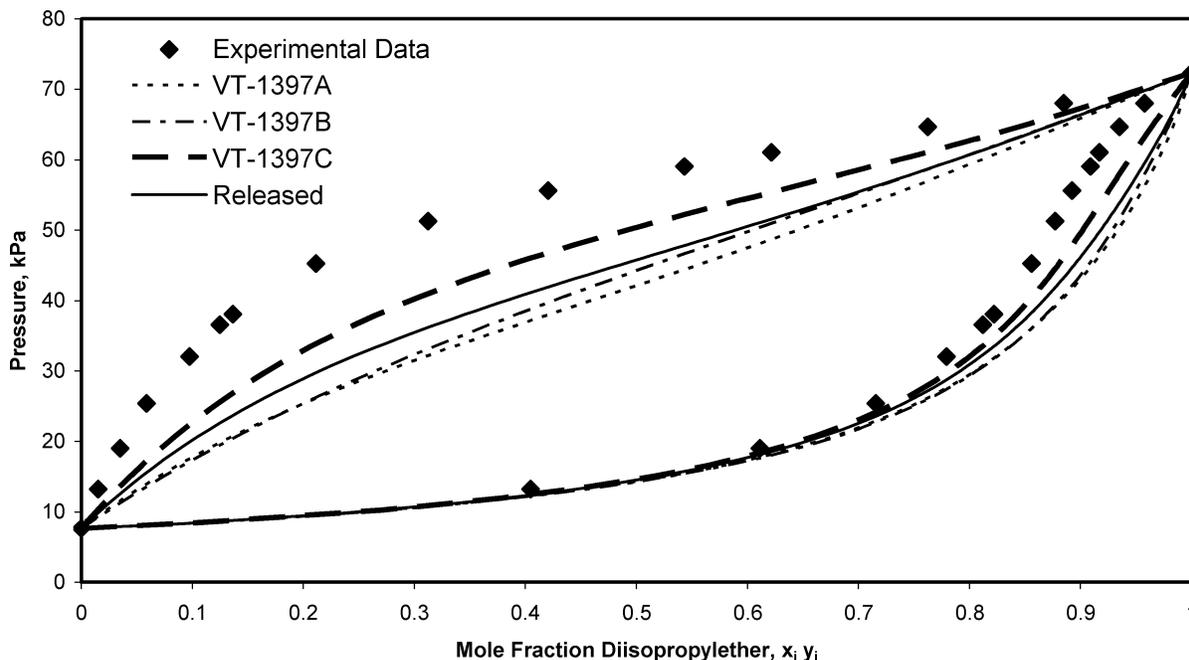


Figure 15. Pressure–composition data and COSMO–SAC–VT-2005 predictions for the 2-methoxy-ethanol/diisopropyl ether system at 331.02 K.<sup>27</sup>

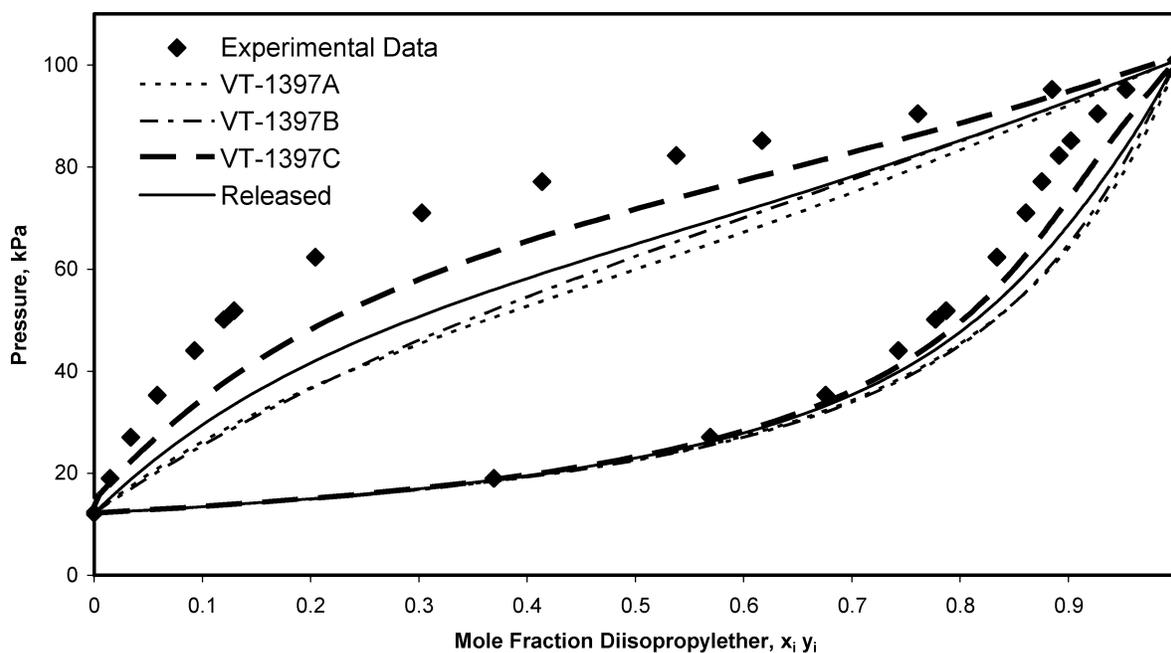


Figure 16. Pressure–composition data and COSMO–SAC–VT-2005 predictions for the 2-methoxy-ethanol/diisopropyl ether system at 341.01 K.<sup>27</sup>

The sigma profile contains 50 segments,  $0.001 \text{ e}/\text{\AA}^2$  wide, in the range of  $-0.025 \text{ e}/\text{\AA}^2$  to  $0.025 \text{ e}/\text{\AA}^2$ . Figure 2 illustrates sigma profiles for water, acetone, *n*-hexane, and 1-octanol.

Given the sigma profiles and the COSMO–RS/SAC models (including the fitted parameters), we can compute various physical properties, including partition coefficients, infinite-dilution activity coefficients, and phase equilibrium, etc.<sup>1,13,14</sup>

**2.4. VT-2005 Variation of the COSMO–SAC Model.** We use the COSMO–SAC model proposed by Lin and Sandler.<sup>5</sup> We modify some of the equation structures to better suit our calculation scheme, and these changes are very similar to the work of Klamt and Eckert.<sup>3</sup> These differences do not alter the mathematics of the model but reflect minor calculation changes in our FORTRAN program. The most obvious difference, shown in the COSMO–SAC–VT-2005.exe (FORTRAN) program, is in our initial sigma-profile calculation.

We use a slightly different definition of the sigma profile:

$$p'(\sigma) = A_i(\sigma) = p(\sigma)A_i \quad (13)$$

where we do not calculate the modified profile as a probability, but actually as a finite value for the surface area with a specific charge. Mathematically, this only presents a few minor differences in the equations that incorporate the sigma profile.

The first place this presents an issue is the definition of the segment activity coefficient. Although Lin and Sandler used eq 5, we use<sup>3</sup>

$$\ln \Gamma_s(\sigma_m) = -\ln \left\{ \sum_i \frac{p'_s(\sigma_n)}{A_i} \Gamma_s(\sigma_n) \exp \left[ \frac{-\Delta W(\sigma_m, \sigma_n)}{RT} \right] \right\} \quad (14)$$

which uses the same definition for each term as eq 5.

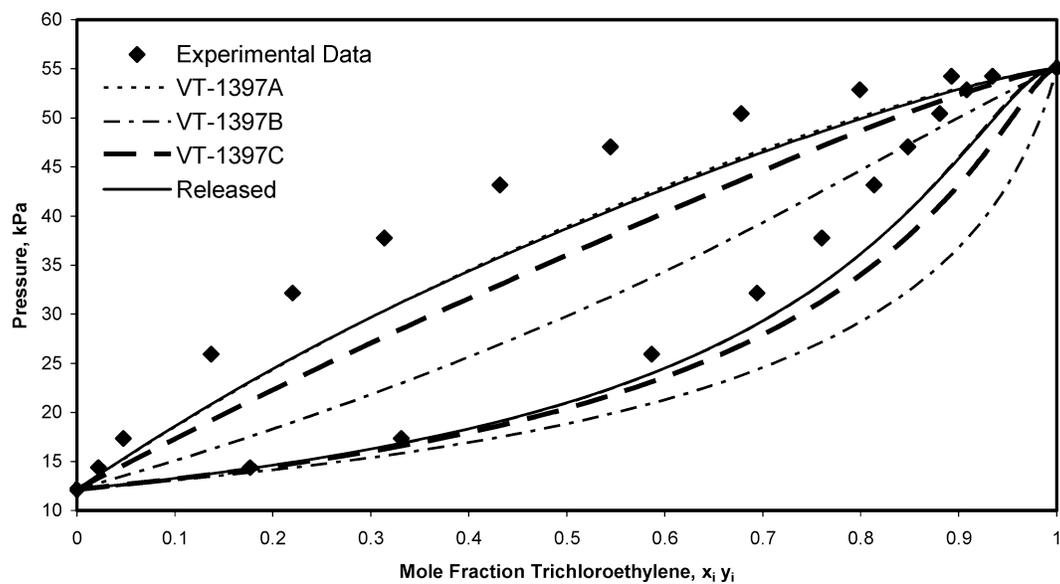


Figure 17. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/trichloroethylene system at 341.01 K.<sup>27</sup>

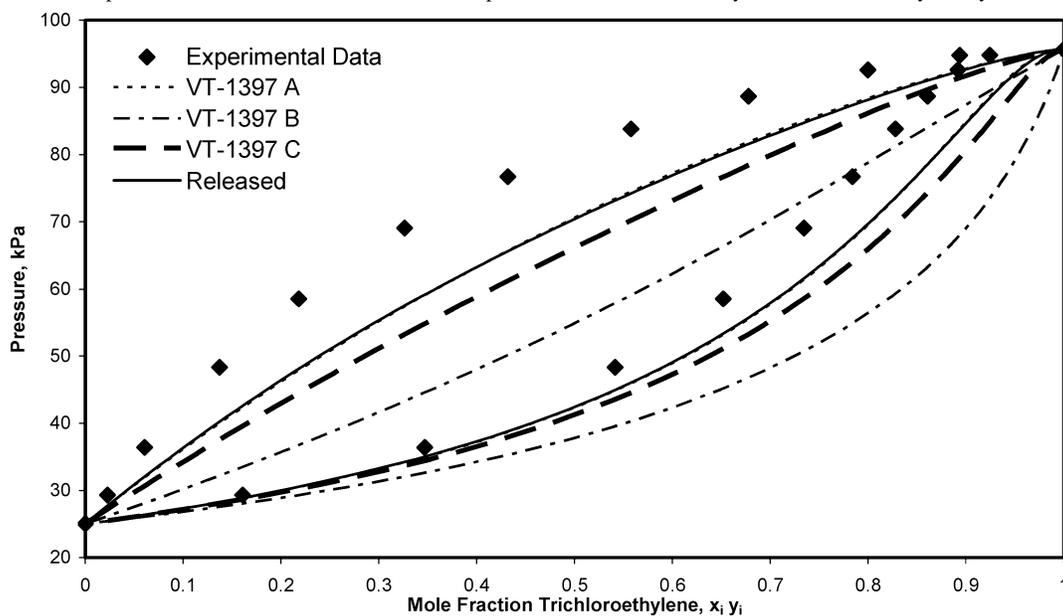


Figure 18. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the 2-methoxy-ethanol/trichloroethylene system at 357.98 K.<sup>27</sup>

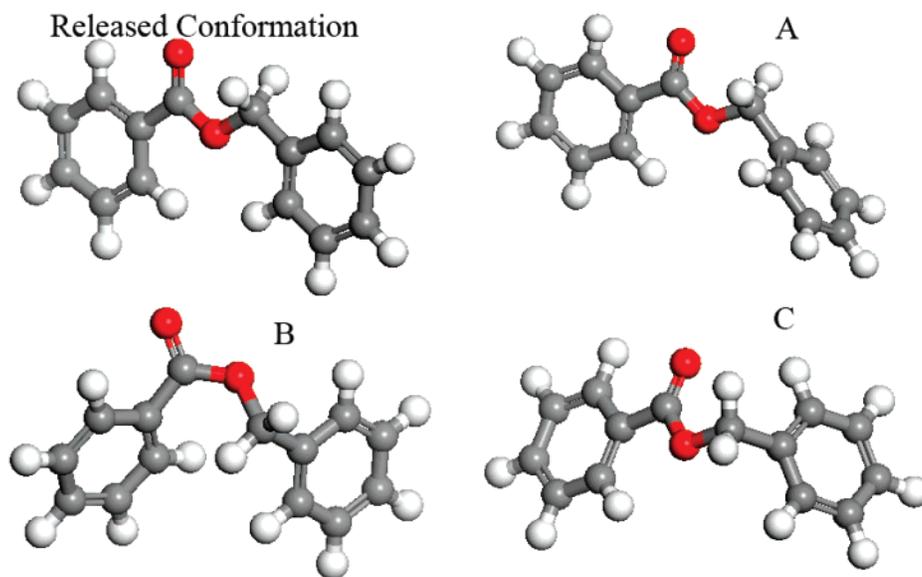
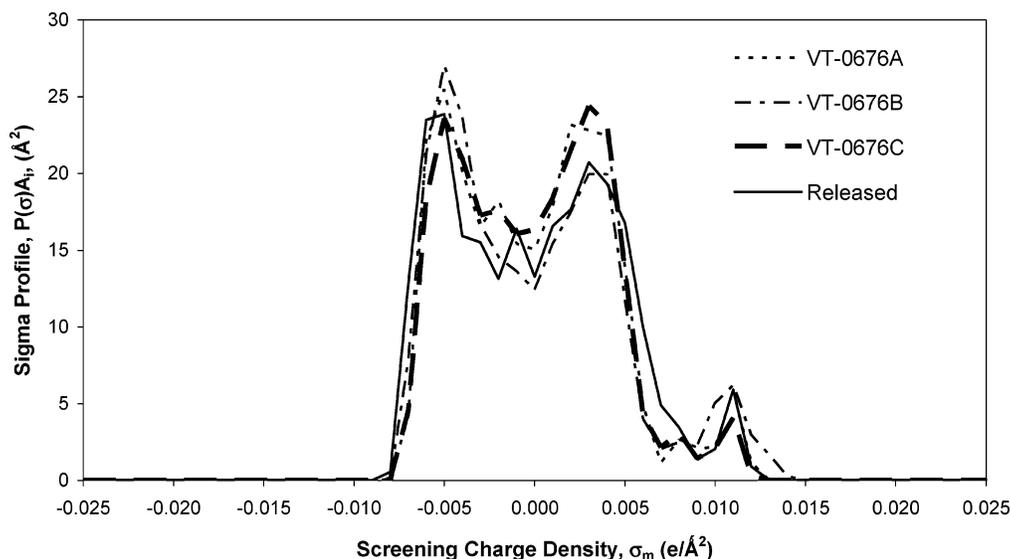
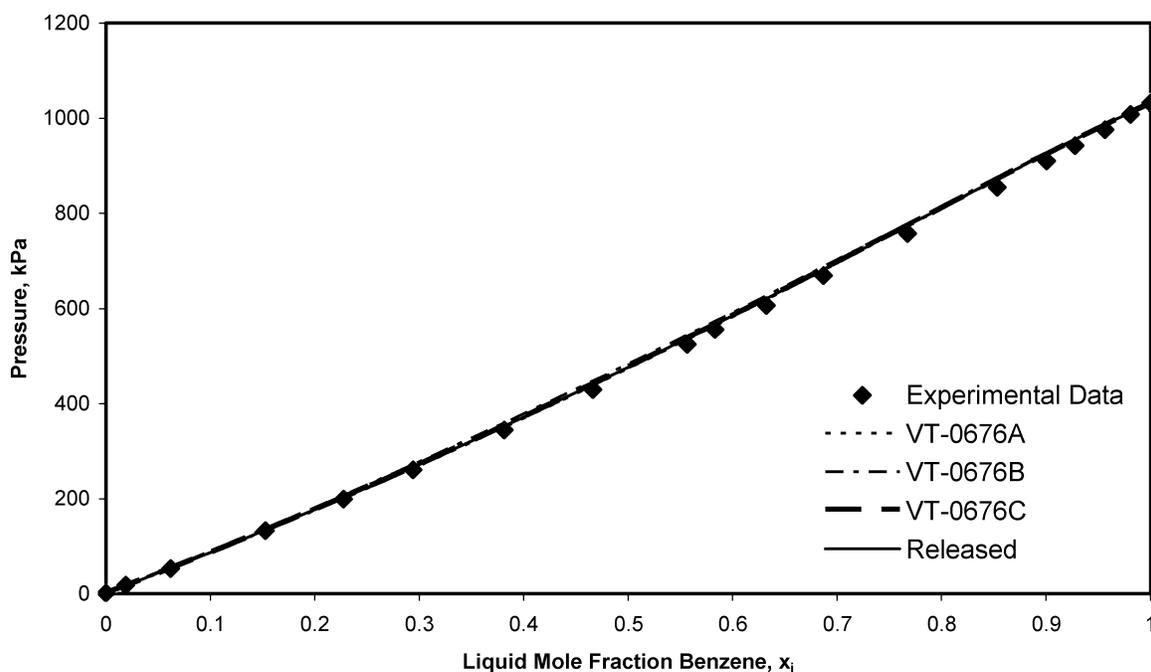


Figure 19. Four low-energy conformations for benzyl benzoate (VT-0676).



**Figure 20.** Sigma-profile comparison for three conformations of benzyl benzoate and the released conformation included in the VT-2005 Sigma Profile Database.



**Figure 21.** Pressure-composition data and COSMO-SAC-VT-2005 predictions for the benzene/benzyl benzoate system at 453.23 K.<sup>28</sup>

We also observe a change in the final form of the activity-coefficient equation. We use the following equation, which is comparable to eq 7, as used by Lin and Sandler.

$$\ln \gamma_{i/s} = \frac{1}{a_{\text{eff}} \sigma_m} \sum p'_i(\sigma_m) [\ln(\Gamma_s(\sigma_m)) - \ln(\Gamma_i(\sigma_m))] + \ln \gamma_{i/s}^{\text{SG}} \quad (15)$$

### 3. Method and Results

In this section, we outline how we generate the VT-2005 sigma-profile database, and we present our validation results.

**3.1. Procedure To Compute Sigma Profiles.** The sigma profile is the molecule-specific output from a series of calculations using the DMol module, which incorporates both the density-functional theory (DFT),<sup>15</sup> available in Accelrys' Materials Studio,<sup>16-19</sup> and the COSMO-based models.<sup>1-6</sup> The first step is to correctly draw each molecule in three dimensions,

checking for proper structure and connectivity. An alternative to manually drawing each molecule is downloading pre-drawn structures from one of several online databases, e.g., the NIST database (<http://webbook.nist.gov/chemistry/>), Scifinder (<http://www.cas.org/SCIFINDER/SCHOLAR/index.html>), or the National Library of Medicine (<http://chem.sis.nlm.nih.gov/chemidplus>). These databases are also useful to check the correctness of the manually drawn molecule.

After the molecule is drawn in Materials Studio or imported from an online source, we use Materials Studio's "clean" tool to automatically adjust bond lengths and bond angles. The "clean" tool arranges each atom's position with fairly good accuracy, but it is important to note that it only provides a rough geometry optimization. The initial geometry does affect the computational time required for the DFT calculations. A key assumption in this procedure is that each molecule is in a global energy minimum after the geometry optimization and not stuck

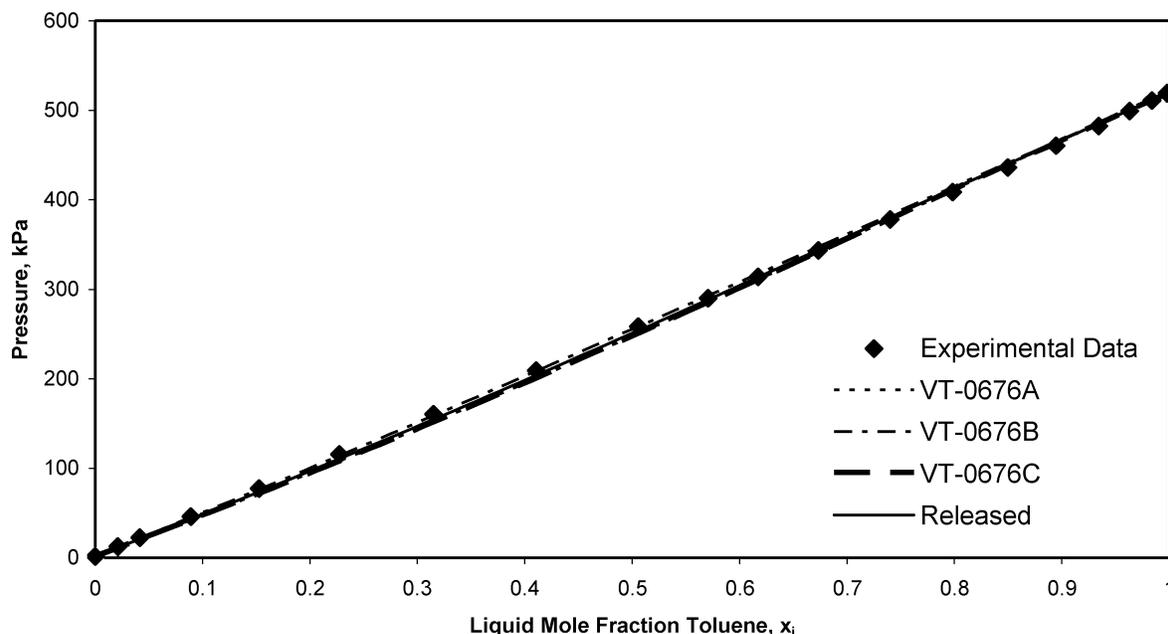


Figure 22. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the toluene/benzyl benzoate system at 453.25 K.<sup>28</sup>

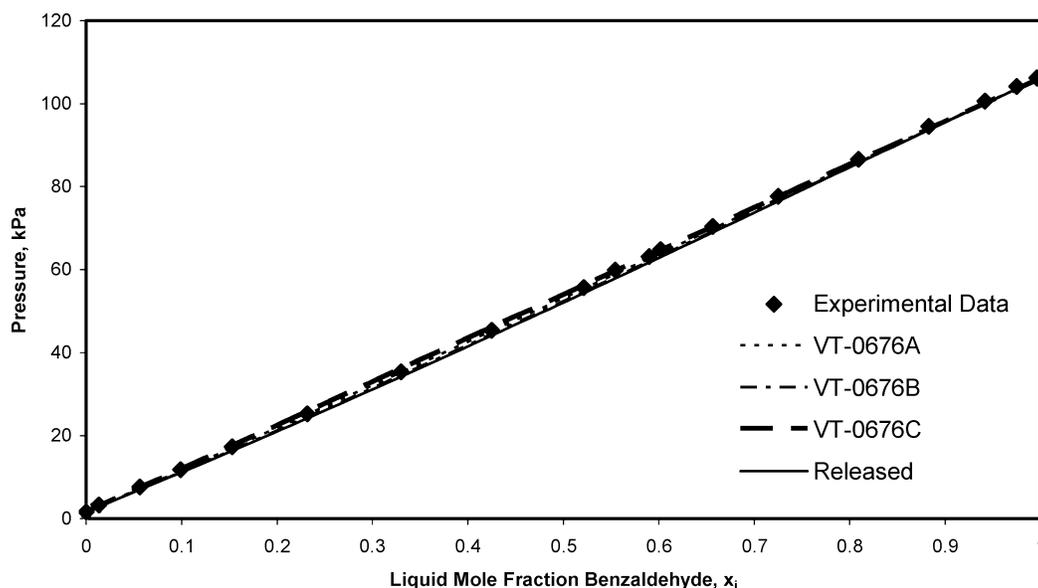


Figure 23. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the benzaldehyde/benzyl benzoate system at 453.25 K.<sup>28</sup>

in a local minimum. To ensure this for larger molecules, we draw several different randomized conformations and relax the geometry to determine if the conformations return to the low-energy form.

Because the “clean” tool only provides a rough estimate of optimum geometry, we must evaluate each structure’s conformation. The “clean” tool will “optimize” atomic positions based on how they are initially drawn in Materials Studio, therefore it is imperative to start with a good initial guess for the conformation. An issue arises when a molecule’s conformation is flexible and could exist in multiple stable conformations. For example, diols experience both intramolecular and intermolecular hydrogen bonding. (Section 4.1 gives more details concerning conformational variations and their effects on geometry optimization.)

Next, we optimize the molecular geometry for each molecule with the DNP v4.0.0 basis set, which is recommended by Accelrys’ Materials Studio for COSMO applications.<sup>20</sup> DNP refers to Double Numerical basis with Polarization functions,

i.e., functions with angular momentum one higher than that of the highest occupied orbital in free atom. According to Koch and Holthausen,<sup>15</sup> the DNP basis set is generally very reliable.

We obtain the ideal-gas phase-equilibrium geometry by running a “Geometry Optimization” task in a DMOL module. We use the GGA/VWN-BP functional setting with a real space cutoff of 5.5 Å for all DMOL calculations. Here, GGA represents the generalized gradient approximation, and VWN-BP represents the Becke-Perdew version of the Volsko-Wilk-Nusair functional.<sup>15,21-23</sup> We optimize the geometry under “fine” tolerances, with  $1.0 \times 10^{-6}$  hartree energy units (or Ha), for convergence of the self-consistent field (SCF) equations, and 0.002 Ha/Å for the convergence of the geometry-optimization calculations.

When the geometry optimization is complete, we calculate the surface screening charges surrounding the molecule in the condensed phase by performing an “Energy Calculation” in the DMol module with the addition of several keywords to turn on the built-in COSMO program in Materials Studio. This step

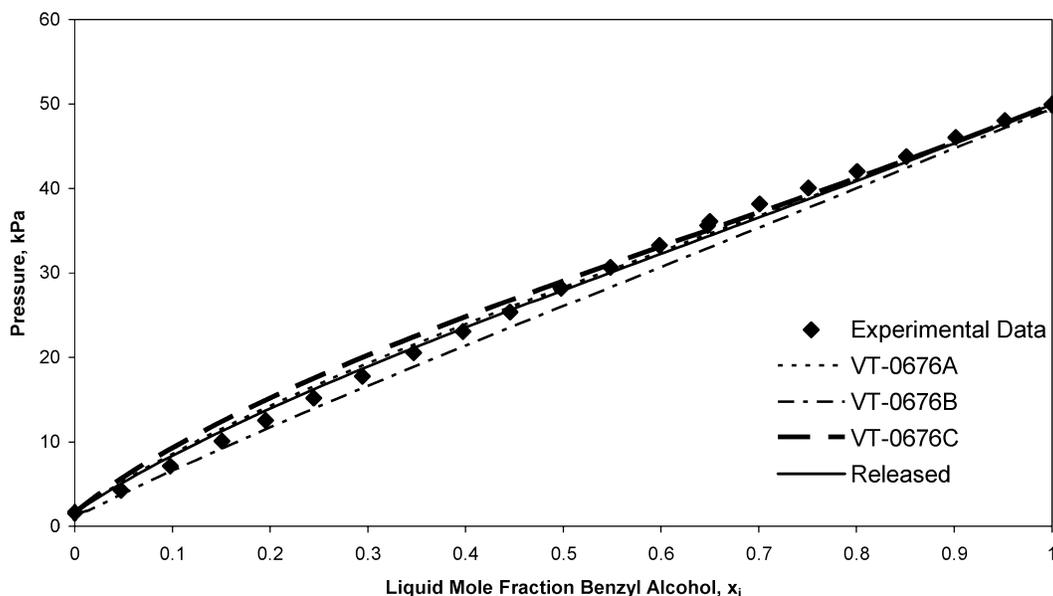


Figure 24. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the benzyl alcohol/benzyl benzoate system at 453.23 K.<sup>28</sup>

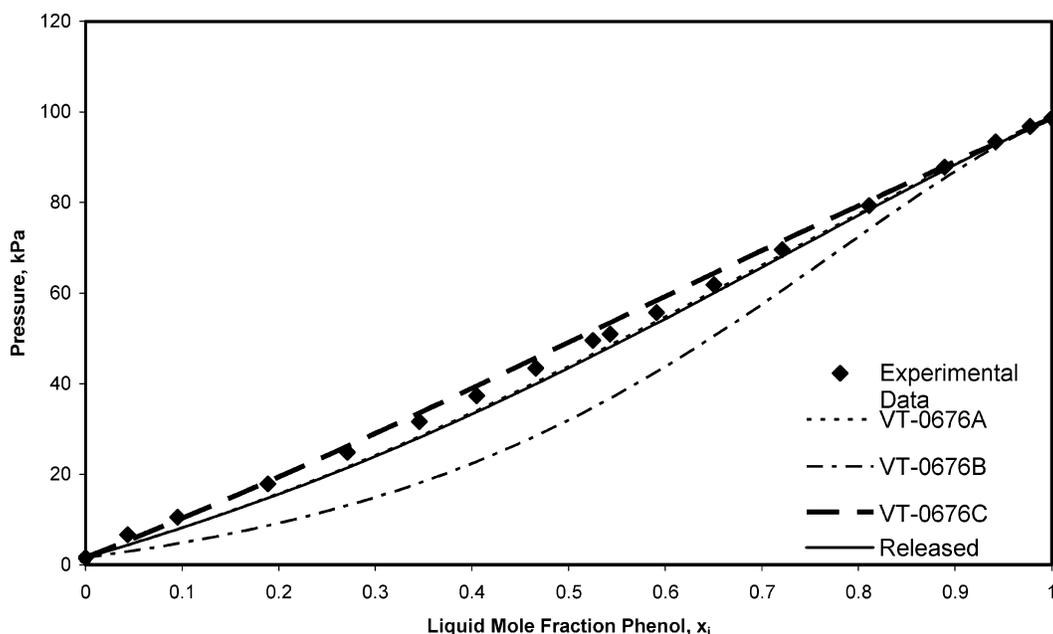


Figure 25. Pressure-composition data and COSMO-SAC-VT-2005 predictions for the phenol/benzyl benzoate system at 453.26 K.<sup>28</sup>

assumes that the optimized gas-phase conformation is identical to the condensed-phase conformation. These keywords are added to the \*.input file prior to running the Energy Calculation. We use the default values for the built-in COSMO program in Materials Studio (See Table 2).

The COSMO keywords also include the atomic radii for the following 10 elements in the COSMO calculations on the optimized geometries: hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine. Klamt and co-workers<sup>2,3</sup> have optimized these radii and suggested using 117% of the van der Waals bonding radius to approximate other elements. (See Table 3.) Sigma profiles with other atoms could contain some errors, because the COSMO-RS and COSMO-SAC approaches, as reported in the literature,<sup>1-6</sup> are only optimized for these 10 elements.<sup>3,5</sup> The VT-2005 database only includes compounds that contain these 10 elements. Recently, Klamt<sup>24</sup> published a table of 11 optimized atomic radii, the difference being silicon (Si).

Using the calculated surface screening charges generated from the COSMO calculation, we average the segments using eq 12 to yield the sigma profile using a FORTRAN program. The total calculation time for a single sigma profile is dependent on the molecule complexity (the number of atoms and bonds) and the quality of the initial geometry. Small molecules such as methane and ethane require <20 min on a 3.6 GHz Pentium IV-equipped personal computer (PC). Other larger molecules can require calculation times from 1 h to 48 h on the same machine. The VT-2005 sigma profiles represent a single conformation for each compound, but we acknowledge that large molecules may also exist in additional low-energy conformations. These additional conformations create slightly different sigma profiles. Section 4.1 illustrates the effect of conformational variations on the sigma profiles and energy calculations for small and medium-sized molecules. We observe increased sigma-profile conformational variations with increasing molecule size.

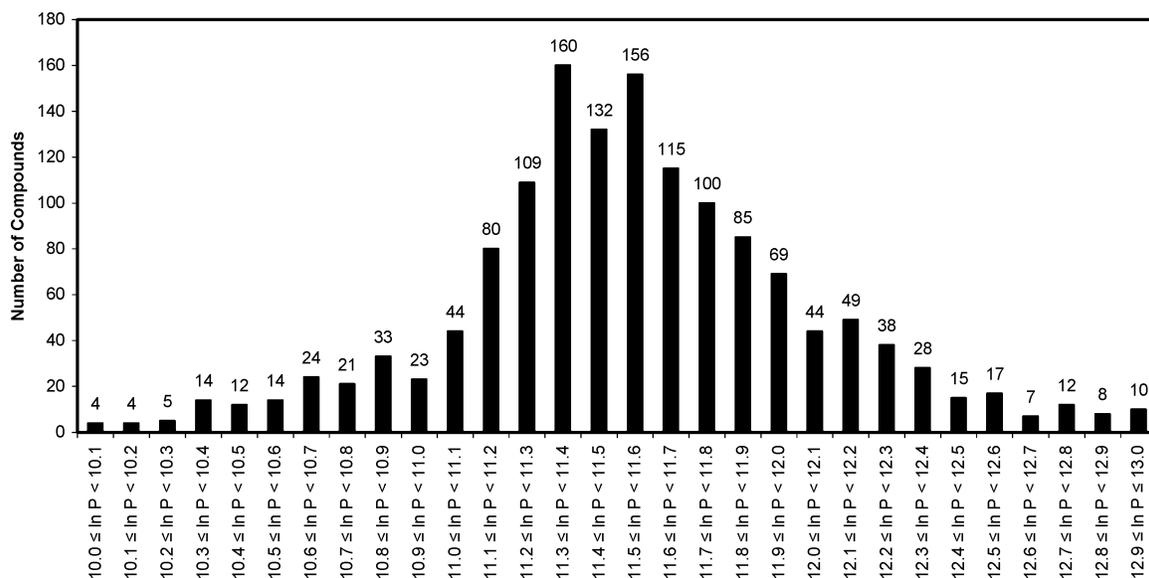


Figure 26. Distribution of vapor pressure predictions in the  $\ln P_{\text{vap}}$  value range of 10.0–13.0. The theoretical value is 11.5261.

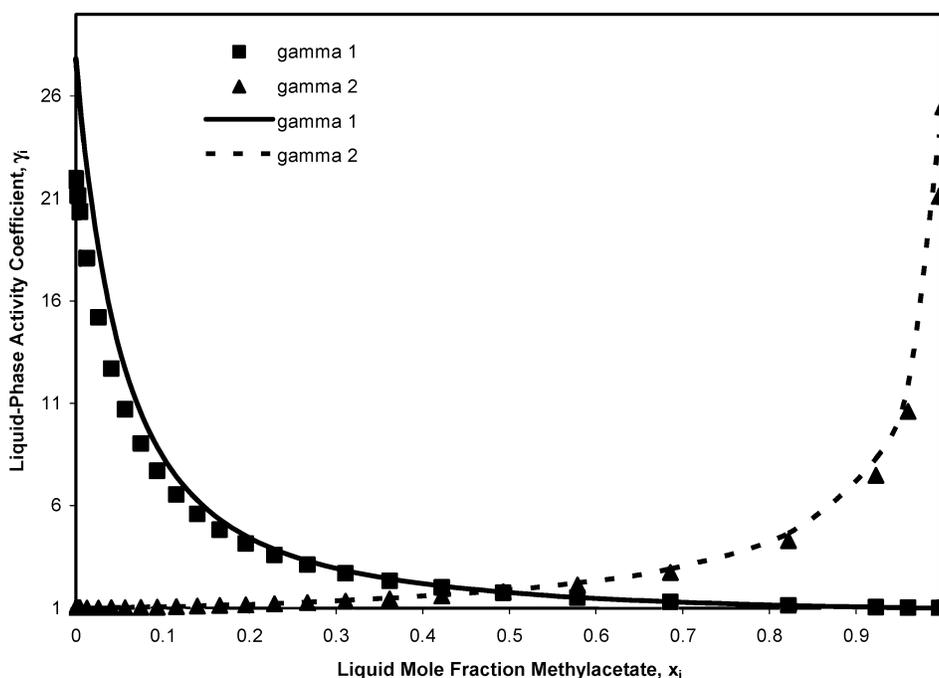


Figure 27. COSMO-SAC activity-coefficient predictions from published results (Lin and Sandler<sup>5</sup> shown as triangles and squares) and VT-2005 (curves) for the methyl-acetate/water system at 330.15 K.

The molecular conformation used in the DFT calculation has a large effect on the sigma profile, and, therefore, great care should be taken in obtaining a low-energy geometry. We have verified all database entries by comparing COSMO-based model predictions of activity coefficients and normal boiling points with experimental data and a thermodynamic property database.

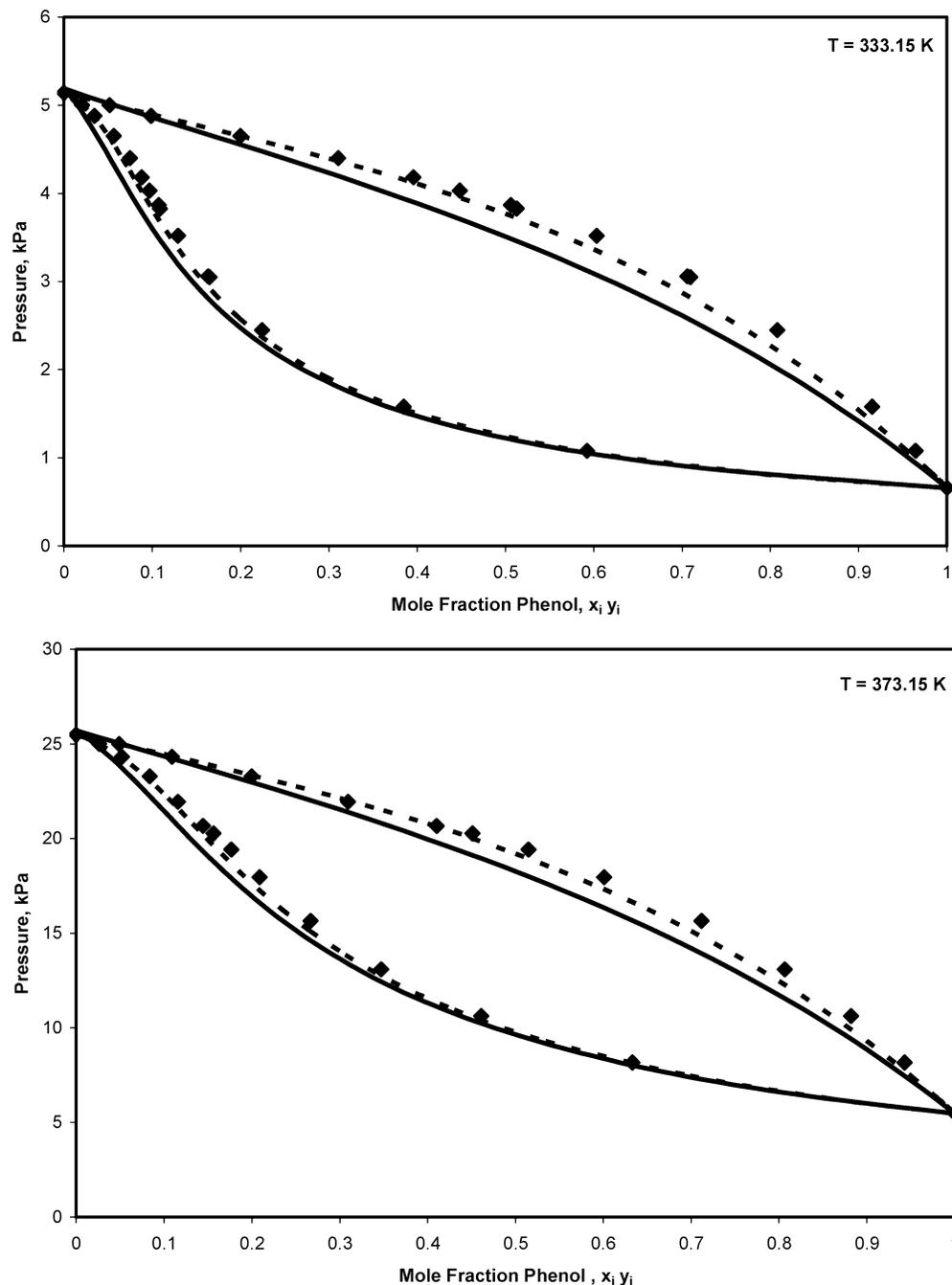
**3.2. VT-2005 Database Summary.** We identify each compound by a unique VT-2005 index number, its CAS registry number, chemical formula, and name in the VT-2005 index. We recommend searching the database by CAS-RN. The chemical family classification in the spreadsheet is a reference, but many molecules fit into more than one grouping. The functional groups represented are acetates, alcohols, aldehydes, alkanes, alkenes, alkynes, anhydrides, aromatics, carboxylic acids, cyclic compounds, cyanates, elements, epoxides, esters, ethers, formates, halogenated compounds, inorganic acids and bases, ketones, mercaptans, nitriles, nitro compounds, peroxides,

sulfides, thiophenes, etc. (See Table 4.) The large majority of molecules are small (containing 10 carbons or less).

**3.3. Sigma-Profile Comparison.** We compare VT-2005 sigma profiles with published profiles for four compounds.<sup>5</sup> Figures 3 and 4 compare the sigma profiles for water, hexane, acetone, and 1-octanol, scaled relative to their maximum values. We find an average difference of <2% between the VT-2005 and Lin and Sandler's sigma profiles.<sup>5</sup> We use the same calculation settings based on the DFT. Therefore, we believe that the difference in sigma profiles is likely due to variations in the DMol versions (Cerius versus Materials Studio). (See Appendix A.)

**3.4. Resources.** We provide, via [www.design.che.vt.edu](http://www.design.che.vt.edu), additional open-literature information as outlined below:

(1) The complete VT-2005 sigma-profile database of 1423 compounds, and an index of the VT-2005 database including CAS-RN, chemical formula, compound name, normal boiling point, and predicted pure component vapor pressures using a



**Figure 28.** Pressure–composition data for the phenol(1)/styrene(2) system for 333.15 and 373.15 K. The solid curves show COSMO–SAC–VT-2005 predictions, dashed curves display COSMO–RS predictions, and symbols represent experimental data.<sup>13,29</sup>

revised COSMO–SAC–BP model<sup>37</sup> (see VT-2005\_Sigma\_Profile\_Database\_Index\_v1.xls).

(2) Procedure for generating sigma profiles using DMol software by Accelrys' Material Studio (including screen captures).

(3) VT-2005 sigma-profile averaging FORTRAN program, which takes the COSMO output from DMol and calculates the sigma profile (see Sigma-average.exe).

(4) COSMO–SAC–VT-2005 FORTRAN program to perform COSMO–SAC calculations for any binary mixture using either VT-2005 sigma profiles or new sigma profiles.

(5) Procedure for using the COSMO–SAC–VT-2005 program for predicting activity coefficients.

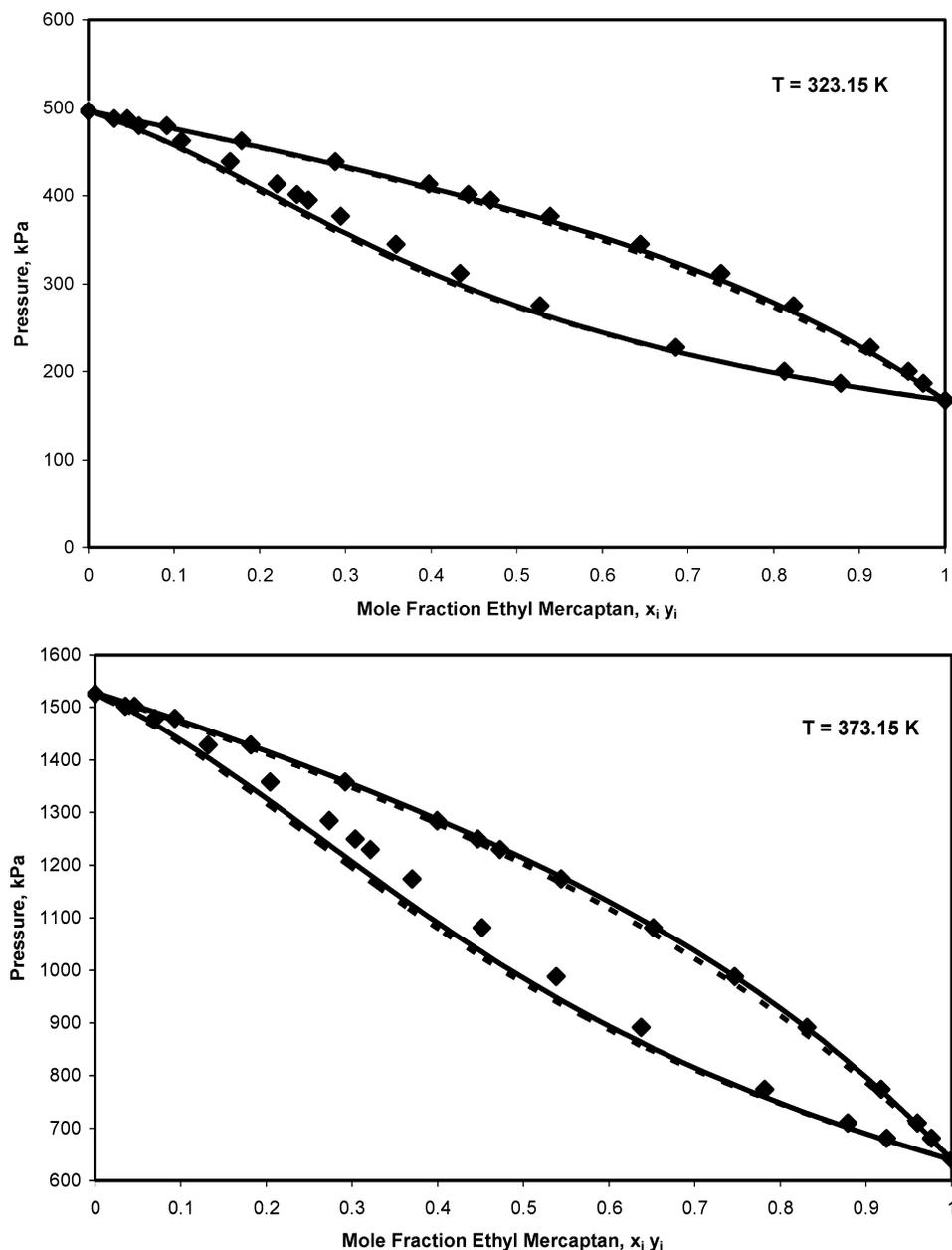
(6) COSMO calculation output for each molecule for use in the averaging algorithm.

Appendix B illustrates the information for methyl acetate (VT-0638), which we use for a validation example in the following section.

#### 4. Conformational Analysis and Validation Examples

In this section, we first describe the effects of structural conformations of molecules on the geometry optimization, and then we validate the VT-2005 Sigma profiles and our implementation of the COSMO–SAC model. We compare predictions of activity coefficients, vapor–liquid equilibria (VLE), and solubilities with published data.

**4.1. Conformational Analysis.** The COSMO–SAC model assumes that the optimized geometry from the DFT calculation



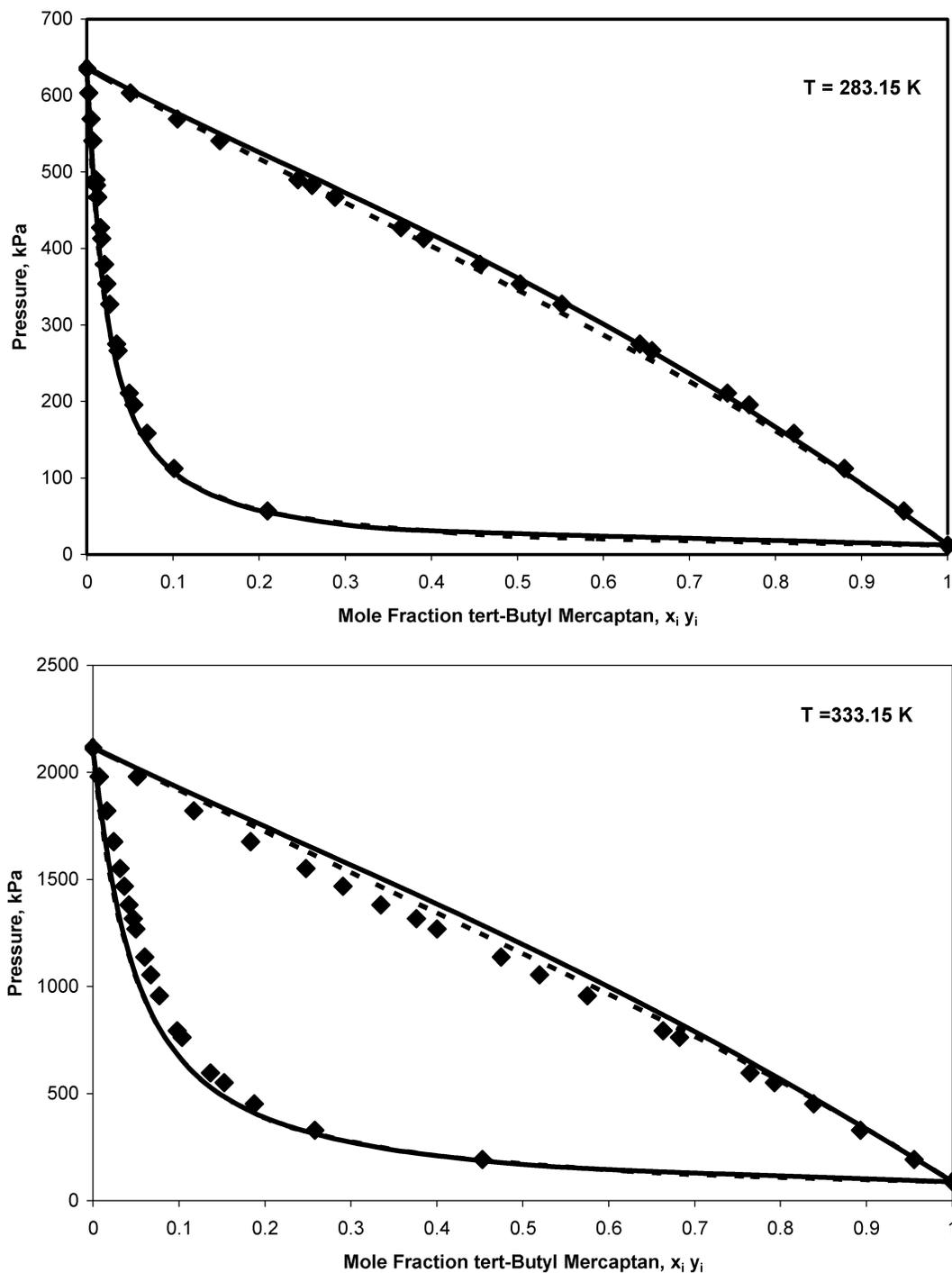
**Figure 29.** Pressure–composition data for ethyl mercaptan(1)/*n*-butane(2) at 323.15 and 373.15 K. The solid curves show COSMO–SAC–VT-2005 predictions, the dashed curves display COSMO–RS predictions, and the symbols represent experimental data.<sup>13,29</sup>

in the vapor phase is identical to that in the condensed phase, which is not necessarily true in all instances. Factors such as solvent polarity, size, and solvent–solute interactions could affect a solute molecule’s structural conformation. The COSMO–SAC model also assumes that the molecule is in a low-energy conformation, but several low-energy structural conformations may exist, because of the freedom in choosing dihedral angles. Each conformation results in a slightly different sigma profile and may affect property predictions. There is no guarantee that our geometry-optimization DFT calculations will yield the lowest energy state possible, simply because energy optimizations can result in local minima instead of global minima. The VT-2005 Sigma Profile database contains sigma profiles for only one low-energy conformation per molecule. However, we explore the effects of multiple conformations in flexible molecules, such as molecules with a high probability to form strong interactions in solution.

We present two examples, with comparison and analysis relative to published data. We use binary systems that contain

2-methoxy-ethanol (VT-1397) with several polar and nonpolar solvents over a range of temperatures as an example of a small molecule capable of having multiple structural conformations. We also use binary systems that contain benzyl benzoate (VT-0676) and other aromatic solvents (benzene, toluene, etc.) as an example of a medium-sized molecule with the same capabilities. We quantify the effect multiple conformations have on property predictions. These examples include comparisons of sigma profiles and VLE predictions.

**4.1.1. Small Molecule Example: 2-Methoxy-ethanol.** We use the COSMO–SAC–VT-2005 model to predict activity coefficients for binary systems of *n*-hexane (VT-0009), *n*-heptane (VT-0014), cyclohexane (VT-0099), methanol (VT-0477), methyl acetate (VT-0638), diisopropyl ether (VT-0713), and trichloroethylene (VT-0802) with 2-methoxy-ethanol (VT-1397) at various temperatures. We generate  $P$ – $x$ – $y$  data using the modified Raoult’s Law and compare these predictions to literature data. We use pure-component vapor pressure data or predictions from the Antoine equation if that data are unavail-



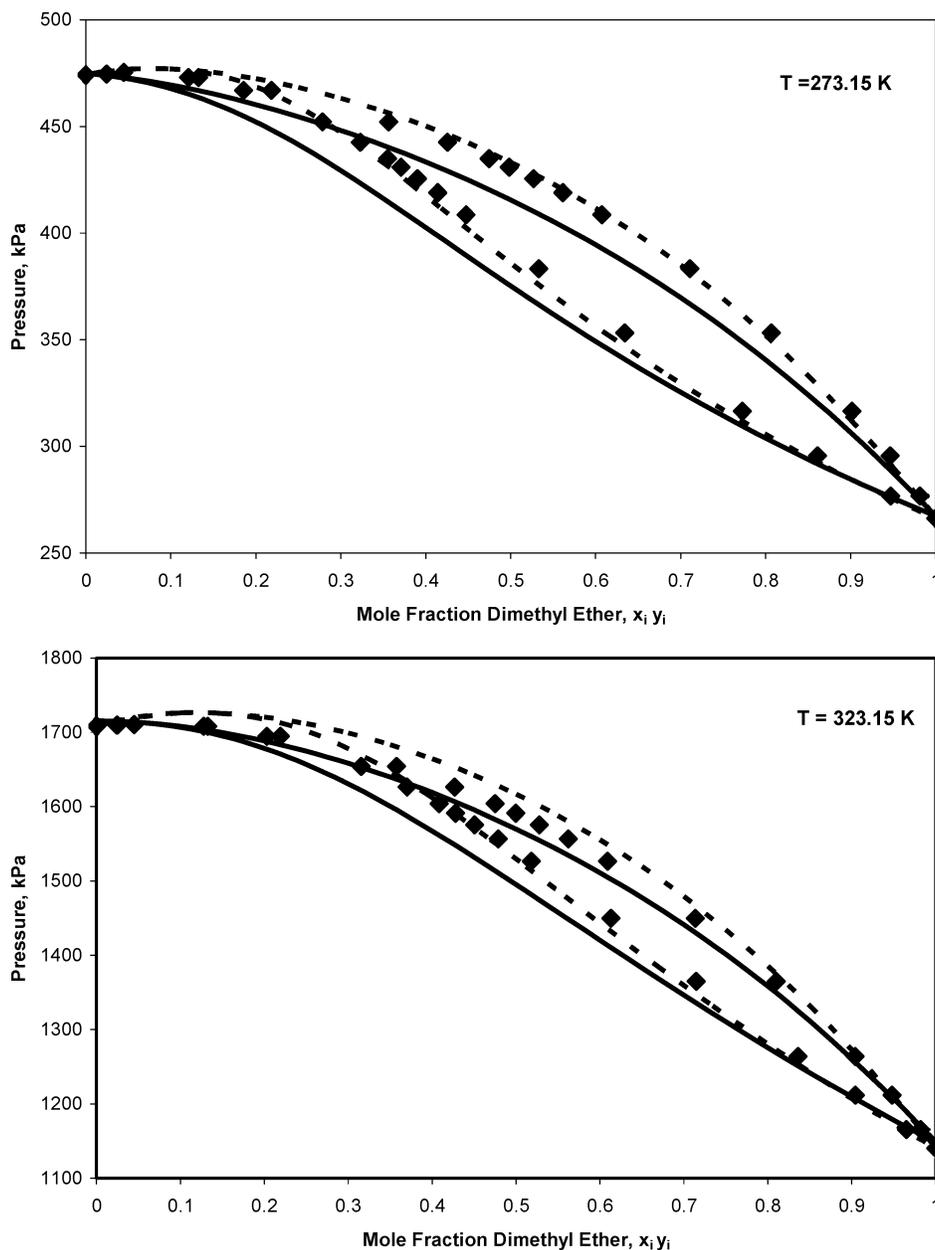
**Figure 30.** Pressure–composition data for *tert*-butyl mercaptan(1)/propane(2) at 283.15 and 333.15 K. The solid curves show COSMO–SAC–VT-2005 predictions, the dashed curves display COSMO–RS predictions, and the symbols represent experimental data.<sup>13,29</sup>

able. Figure 5 shows the released conformation from the VT-2005 Sigma Profile Database and three alternative conformations (labeled A–C in the figure) for comparison. Conformations A, B, and C were varied manually from the released conformation; however, all underwent geometry optimization following the procedure outlined in Section 3.1.

Figure 6 shows that each of these structures produces a slightly different sigma profile. Conformation B contains an intramolecular hydrogen bond, whereas the other structures would be more likely to form intermolecular hydrogen bonds in solution. There is no open-literature data available concerning what percentage of each structure actually exists in solution.

We examine pressure compositions data for systems with cyclohexane. Figures 7–9 compare the COSMO–SAC–VT-2005 predictions with experimental data<sup>25</sup> at temperatures of 303.15, 313.15, and 323.15 K, respectively. Figures 10–12 compare our predictions for binary systems with other nonpolar solvents, *n*-hexane and *n*-heptane with experimental data.<sup>25</sup> Figures 13–18 compare COSMO–SAC–VT-2005 predictions with polar solvents (i.e., methanol, methyl acetate, diisopropyl ether, and trichloroethylene) to published data.

The COSMO–SAC–VT-2005 predictions improve as the temperature increases for each conformation, based on RMS error calculations, with the exception of diisopropyl ether. Table 5 sum-



**Figure 31.** Pressure–composition data for dimethyl ether(1)/propane(2) at 273.15 and 323.15 K. The solid curves show COSMO–SAC–VT-2005 predictions, the dashed curves show COSMO–RS predictions, and the symbols represent experimental data.<sup>13,29</sup>

**Table 5. Summary of the Root Mean Square (RMS) Error for Each Conformation and System**

solvent	temperature (K)	RMS Error for VT-1397			
		A	B	C	released
cyclohexane	303.15	0.0532	0.1003	0.0473	0.0427
cyclohexane	313.15	0.0488	0.0943	0.0429	0.0388
cyclohexane	323.15	0.0429	0.0871	0.0363	0.0332
<i>n</i> -hexane	313.15	0.0531	0.1011	0.0469	0.0426
<i>n</i> -hexane	323.15	0.0367	0.0771	0.0310	0.0283
<i>n</i> -heptane	323.15	0.0645	0.1281	0.0551	0.0504
methanol	298.15	0.0166	0.0038	0.0246	0.0152
methyl acetate	298.15	0.0361	0.0408	0.0200	0.0294
diisopropyl ether	331.02	0.1282	0.1345	0.0737	0.0983
diisopropyl ether	341.01	0.1279	0.1357	0.0740	0.0986
trichloroethylene	341.01	0.0489	0.1153	0.0690	0.0481
trichloroethylene	357.98	0.0451	0.1096	0.0627	0.0440

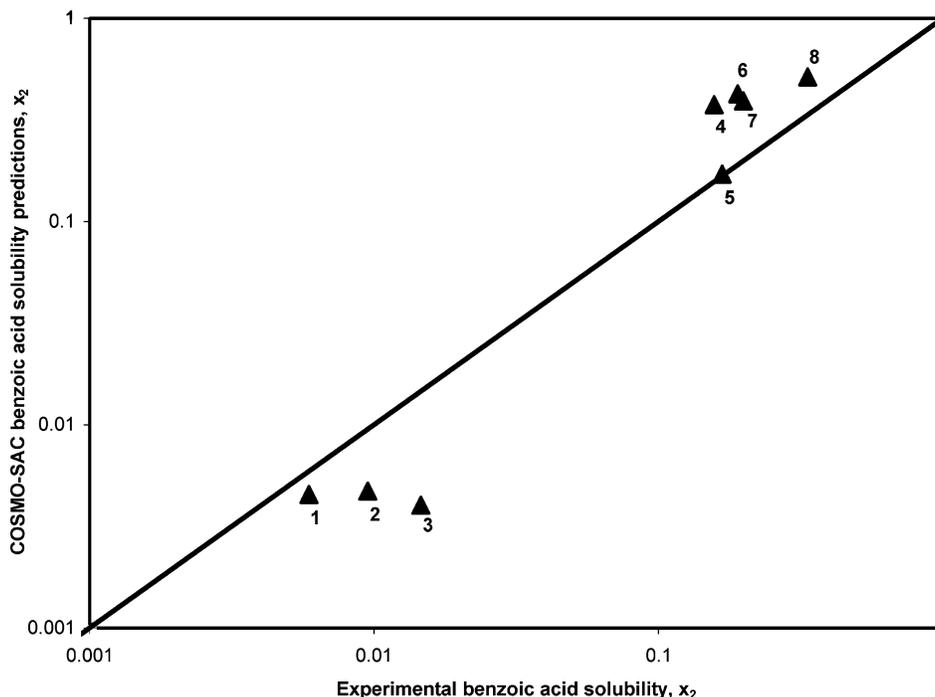
marizes the root mean square (RMS) error for each system and conformation. We calculated the RMS error using eq 16, where  $n$

represents the number of data points and  $y_{VT-2005}$  is the predicted vapor fraction from the COSMO–SAC–VT-2005 model.

$$RMS = \sqrt{\frac{1}{n} \sum_i^N (y_{\text{exp}_i} - y_{VT-2005_i})^2} \quad (16)$$

The released conformation included with the database produces the best overall predictions for all of the solvents, with an average RMS error of 0.04746. Conformation C also adequately predicts pressure–composition data, with a similar average RMS error value of 0.04863. The difference in RMS error between the released conformation and conformation C is  $\sim 1\%$ , which is insignificant. The COSMO–SAC–VT-2005 predictions are most accurate for the methanol/2-methoxy-ethanol system, which seems consistent with Lin and Sandler’s observation that the COSMO–SAC model accuracy decreases for highly hydrophobic systems.<sup>5</sup>

Each of the six nonpolar solvents forms an azeotrope with 2-methoxy-ethanol. Table 6 compares the predicted azeotropic



**Figure 32.** Benzoic acid (VT-0610) solubility for various solvents at 298.15 K. Solvents 1–8 are pentane, *n*-hexane, cyclohexane, methanol, acetic acid, 1-hexanol, 1-octanol, and tetrahydrofuran (THF), respectively.

**Table 6. Predicted Azeotropic Compositions for Nonpolar Solvents with 2-Methoxy-ethanol**

solvent	temperature (K)	A	B	C	released	experimental result
cyclohexane	303.15	0.9440	N/A	0.9307	0.9316	0.9204
cyclohexane	313.15	0.9393	N/A	0.9251	0.9264	0.9112
cyclohexane	323.15	0.9273	N/A	0.9123	0.9136	0.9001
<i>n</i> -hexane	313.15	0.9645	N/A	0.9519	0.9538	0.9408
<i>n</i> -hexane	323.15	0.9587	N/A	0.9447	0.9470	0.9349
<i>n</i> -heptane	323.15	0.8298	0.9443	0.8124	0.8116	0.7939

composition with the experimental value.<sup>25</sup> We note that conformation B does not predict an azeotrope in five of the six systems and gives a poor prediction in the case of *n*-heptane. We interpolated the azeotropic composition from experimental data. In most cases, conformation C generates the most-accurate azeotropic composition.

#### 4.1.2. Medium-Sized Molecule Example: Benzyl Benzoate.

Next, we examine conformational effects due to varying structures with a medium-sized molecule from the VT-2005 database. Medium-sized molecules generally contain more single bonds and allow more rotations and, thus, more conformations.

We identify four low-energy states for benzyl benzoate (VT-0676) and predict pressure–composition behavior using these conformations. Figure 19 shows the released conformation from the VT-2005 Sigma Profile Database and three other low-energy conformations, and Figure 20 depicts the resulting sigma profiles. As with the previous example, conformations A, B, and C were varied from the released conformation and underwent a geometry optimization, following the procedure in Section 3.1. The benzyl benzoate sigma profiles exhibit larger variations, based on standard deviation and variance calculations, when compared to 2-methoxy-ethanol sigma profiles. This suggests that conformational differences in larger molecules created larger variations in sigma profiles than smaller molecules.

We compare predictions for binary systems of benzyl benzoate and five solvents: benzene (VT-0242), toluene (VT-

0243), benzaldehyde (VT-0432), benzyl alcohol (VT-0541), and phenol (VT-0542). Figures 21–25 show COSMO–SAC–VT-2005 predictions for those systems in comparison with experimental data.

All of these binary systems exhibit ideal or almost-ideal behavior. With a few exceptions, the COSMO–SAC model predicts ideal or almost-ideal behavior, to varying degrees of accuracy. We quantify the error in our predictions using a RMS calculation, according to eq 17:

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_i^N (P_{\text{exp}_i} - P_{\text{VT-2005}_i})^2} \quad (17)$$

Table 7 summarizes the error calculations for each conformation of benzyl benzoate with the five solvents studied. There is an insignificant difference in the RMS error values concerning each prediction for conformations A and C, and the released conformation. Conformation B is only marginally worse, with the exception of the prediction for the phenol system, which was significantly different from the published values and the other predictions. We observe the largest deviation from experimental data for the phenol/benzyl benzoate system predictions for all conformations. Overall, conformation A predicts the pressure–composition data with the most accuracy, based on a simple average of the RMS error values.

**4.2. Validation Examples.** We present several examples of validated predictions of activity coefficients, VLE, and solubilities using the COSMO–SAC–VT-2005 model.

**4.2.1. Pure-Component Vapor-Pressure Predictions.** Each compound in the VT-2005 Sigma Profile Database has a predicted pure component vapor pressure calculated at its respective normal boiling temperature with the VT-2005 Sigma Profile, using a revised COSMO–SAC–BP model developed by Wang et al.<sup>37</sup> Here, we summarize the predicted values for each compound using the natural logarithm of the vapor pressure. Theoretically, each compound's vapor pressure should

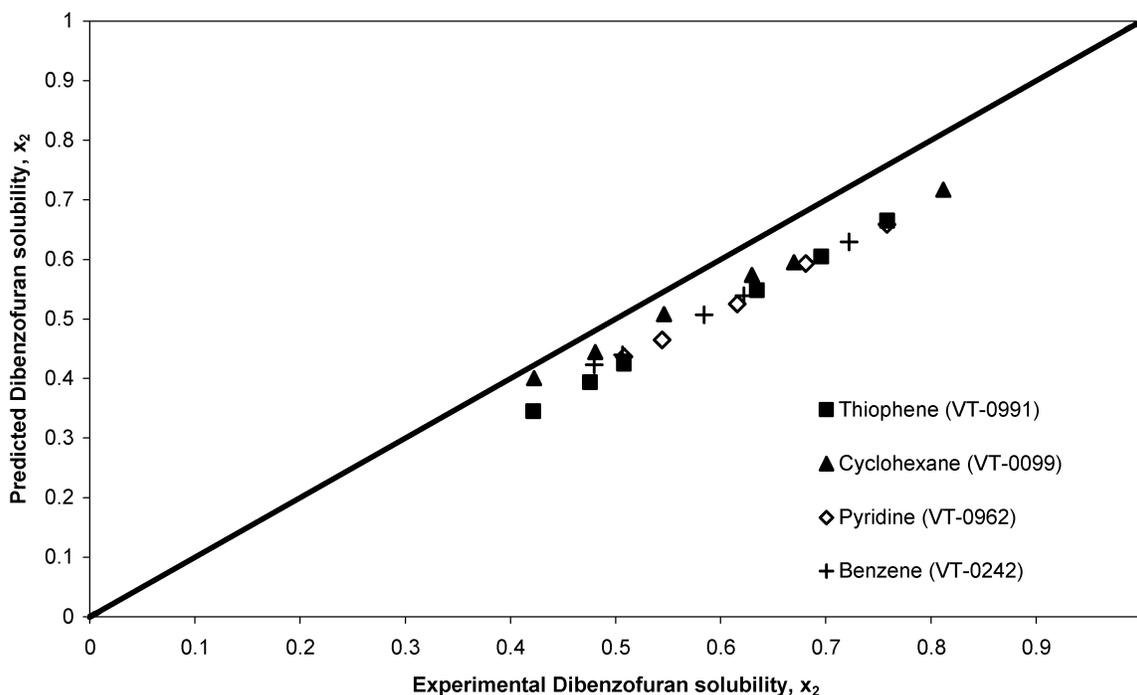


Figure 33. Dibenzofuran (VT-0770) solubility for various solvents over a range of temperatures (314–345 K).<sup>31</sup>

Table 7. RMS Error Calculated for Each System and Conformation

solvent	temperature (K)	RMS Error for VT-0676 (kPa)			
		A	B	C	released
benzene	453.23	9.054	13.646	10.877	10.258
toluene	453.25	4.678	2.544	4.455	3.164
phenol	453.26	2.203	9.807	1.612	2.413
benzaldehyde	453.25	0.522	0.895	0.415	1.091
benzyl alcohol	453.23	0.992	1.416	1.375	0.955
average error		3.490	5.662	3.747	3.576

be 101 325 Pa (or  $\ln(101\,325) = 11.5261$ ). Figure 26 summarizes the distribution of vapor-pressure predictions for every compound in the database. The average predicted  $\ln(P_{\text{vap}})$  value is 11.5405, with a standard deviation of 0.4435.

**4.2.2. Activity-Coefficient Predictions.** Lin and Sandler<sup>5</sup> have presented activity-coefficient predictions for the methyl-acetate/water system, using their COSMO–SAC model. Figure 27 compares the VT-2005 predictions and the published Lin and Sandler data<sup>5</sup> for the methyl-acetate/water system at 330.15 K. The VT-2005 predictions deviate slightly from the Lin and Sandler data, with an average error of 8% (RMS errors are 0.1349 and 0.0478 for  $\ln \gamma_1$  and  $\ln \gamma_2$ , respectively). The most notable deviation seems to appear as we approach infinite dilution, with respect to methyl-acetate in water. We calculate the RMS error with eq 18, where the  $\ln(\gamma_{\text{COSMO-SAC}})$  values are the published predictions.<sup>5</sup>

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_n (\ln \gamma_{\text{VT-2005}} - \ln \gamma_{\text{COSMO-SAC}})^2} \quad (18)$$

We now predict the VLE for four binary mixtures. Figures 28–31 compare our predictions with experimental data and published COSMO–RS predictions.<sup>13,29</sup>

Figure 28 shows the VLE data for the phenol/styrene system at 333.15 and 373.15 K. Both COSMO–RS and COSMO–SAC predict a positive deviation from Raoult's law. The reported COSMO–RS predictions better agree with the experimental data at lower temperatures than the COSMO–SAC

predictions; however, this discrepancy is less apparent at higher temperatures.

Figure 29 illustrates the VLE behavior for the ethyl mercaptan/*n*-butane system at 323.15 and 373.15 K. Again, both models accurately predict the deviation from Raoult's law.

Figure 30 shows the VLE predictions for the *tert*-butyl mercaptan/*n*-propane system at 283.15 and 333.15 K. The COSMO–SAC model accurately predicts both the positive and negative deviations from Raoult's law for this system and agrees very well with the COSMO–RS predictions. Eckert and Klamt<sup>13</sup> have reported that the UNIFAC model fails for this system because it lacks appropriate functional group parameters.

Figure 31 shows the dimethyl-ether/*n*-propane system at 273.15 and 323.15 K. Both COSMO–SAC and COSMO–RS predict the minimum-boiling azeotropes in the dimethyl-ether-rich region for both temperatures. In this case, the COSMO–RS model fits the lower-temperature data better and both methods compare well with the higher-temperature data.

For the systems reported in Figures 27–30, the COSMO–SAC–VT-2005 predictions match reasonably well with experimental pressure-composition data. The COSMO–SAC–VT-2005 predictions for  $\ln(\gamma_1)$  and  $\ln(\gamma_2)$  agree with the reported calculated  $\gamma$  values, with an average RMS error of 0.0867. The eight data sets vary between RMS errors of 0.0164 and 0.1973. The highest deviations correspond to the phenol/styrene system.

**4.2.3. Solubility Predictions.** Equation 19 defines the solubility (in terms of solute mole fraction dissolved in the solvent) as a function of pure solute properties (heat of fusion and melting temperature) and the activity coefficient of the solute in the solvent:<sup>30</sup>

$$\begin{aligned} \ln x_i &= \frac{\Delta_{\text{fus}} S}{R} \left(1 - \frac{T_m}{T}\right) - \ln \gamma_i^{\text{sat}} \\ &= \frac{\Delta_{\text{fus}} H}{RT_m} \left(1 - \frac{T_m}{T}\right) - \ln \gamma_i^{\text{sat}} \end{aligned} \quad (19)$$

Equation 19 is valid for  $T \leq T_m$ . We predict the activity coefficient of the solute using the COSMO–SAC model and

published pure solute properties (the heat of fusion  $\Delta_{\text{fus}}H$  and the melting point  $T_m$ ).<sup>31</sup>

Figure 32 compares COSMO-SAC-VT-2005 solubility predictions with published experimental values with solute mole fractions in the range of 0.0059–0.3348 for benzoic acid in nine solvents at 25 °C. The VT-2005 predictions qualitatively agree with the experimental data, with a RMS error of 0.7092 for  $\ln x_2$ . Kolar et al.<sup>32</sup> have reported interesting results in using COSMO-RS for solubility calculations. Their COSMO-RS predictions for benzoic acid are consistent with the results shown here.

In Figure 33, we use the COSMO-SAC model to predict solubility values for dibenzofuran (VT-0770) in thiophene (VT-0991), cyclohexane (VT-0099), pyridine (VT-0962), and benzene (VT-0242) over a range of temperatures (314–345 K). The solute experimental mole fractions are in the range of 0.4220–0.8120. The predicted solubility is consistently low, by an average error of 11.7% for all solvents and temperatures.

## 5. Application Guidelines for Using COSMO-RS and COSMO-SAC

The literature has identified molecules and systems for which the COSMO approach is not suited or suggested without proper research. We have reviewed many reported experiences from the literature up to December 2005 to provide the reader with a guide for COSMO model applications. Our experience is mainly with COSMO-SAC and references to COSMO-RS refer only to cited literature observations. Many groups are researching these methods, and it should broaden its applicability and improve the model's accuracy in the future.

(1) One should use COSMO-SAC and COSMO-RS to predict liquid–liquid equilibrium (LLE), vapor–liquid equilibrium (VLE), and solid–liquid equilibrium (SLE) properties. The COSMO-RS/SAC methods are generally applicable and show promise in predicting equilibrium behavior.<sup>3,5,6,13,32,33</sup>

(2) COSMO-SAC and COSMO-RS predictions are sensitive to variations in sigma profiles; therefore, one should use a consistent procedure or algorithm when generating sigma profiles. We find some systematic differences, in the energy calculations, between DFT calculations in DMol and Jaguar<sup>34</sup> that could lead to different sigma profiles. However, the extent of error introduced by these variations is not quantified as yet. (See Appendix A.)

(3) One should exercise caution when using COSMO-based models with amines. COSMO calculations generally show poor results for amines, especially trialkylamines, because of the unpaired electron in the cavity.<sup>5</sup> Klamt and co-workers<sup>1,2,14</sup> had reported problems in applying COSMO calculations to highly polar groups with a small surface area, such as amines.

(4) One should exercise caution when using COSMO-based models with polymers systems. COSMO-based models are not well-proven for polymer-system thermodynamics (Panayiotou<sup>7</sup> has commented on "...its inability to properly account for the thermodynamics of polymer systems except for some rather limited cases"). Klamt<sup>24</sup> reported to have successfully modeled gas solubility in polymers by creating partial sigma profiles based on the repeating unit. We note that COSMO-RS models predict solubilities well but with the use of a fitted polymer-specific correction constant.<sup>24</sup>

(5) One should avoid COSMO-SAC with mixtures of chloroform and ketones or alcohols. Lin and Sandler<sup>5</sup> have reported difficulties in accurately predicting interactions between chloroform and either ketones or alcohols. They suggested that "it is necessary to refine the COSMO calculation to provide a

better  $\sigma$ -profile for chloroform." This is identical to the behavior that Klamt and co-workers<sup>1,2</sup> have reported, regarding largely polar groups that are similar to nitriles and carbonyl groups.

(6) One should review which COSMO approach is best for the chemical system. For example, COSMO-RS provides better predictions for alkanes in water, when compared to COSMO-SAC.<sup>5</sup> Klamt and others have optimized the COSMO-RS parameters with water/alkane solutions, whereas Lin and Sandler simply use the values from COSMO-RS without an independent optimization for use with COSMO-SAC.<sup>1,5,6</sup>

(7) One should avoid COSMO-SAC for highly hydrophobic solutes. The accuracy for COSMO-SAC predictions decreases as the hydrophobicity increases.<sup>5</sup> Klamt<sup>35</sup> has reported satisfactory solubility predictions of hydrocarbons in aqueous binary systems using the COSMOtherm program developed by COSMOlogic and released in 2002. The test set included aqueous systems of alkanes, alkylbenzenes, alkylcyclohexanes, and alkenes.

(8) One should use COSMO predictions for molecules that contain only hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine. Klamt et al.<sup>2</sup> have optimized these parameters and suggested using 117% of the van der Waals radii for other atoms, but they do not validate this assumption past the stated five atoms (others later extended it to nine atoms).<sup>3</sup> Note: We currently input atomic radii for hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine.<sup>24</sup>

(9) One should use COSMO-SAC or COSMO-RS for property predictions of systems with chemicals for which UNIFAC cannot properly handle, for example, benzene/*n*-methylformamide.<sup>5</sup> Kolar et al.<sup>32</sup> have shown that UNIFAC only worked for 82% of the 221 systems in their study.

(10) COSMO-RS<sup>1,4,13</sup> and COSMO-SAC-BP,<sup>36,37</sup> a variation of the COSMO-SAC model developed by Lin and Sandler, have predicted pure-component vapor pressures and enthalpies of vaporization.

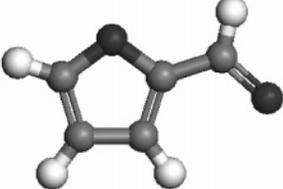
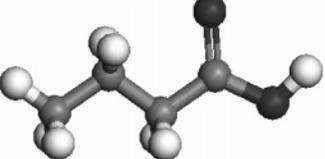
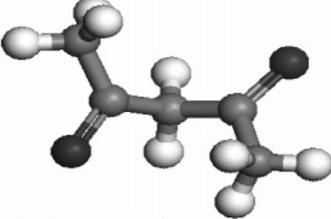
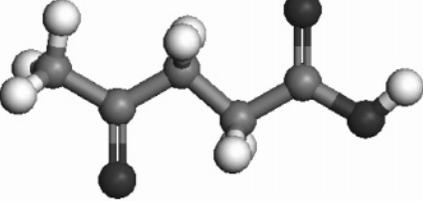
(11) Because molecules can exist in multiple stable conformations, carefully analyze the system to determine effects of conformational changes on thermodynamic property predictions, as suggested by Klamt<sup>1</sup> and in section 4.1.

(12) Constantinescu et al.<sup>38</sup> have expanded COSMO-based models to predict VLE thermodynamic data at high temperatures and pressure from infinite-dilution activity coefficients, using the Huron–Vidal mixing rule.

## 6. Conclusions

We have presented a consistent sigma-profile database, VT-2005, that contains sigma profiles for 1432 compounds. The sigma profiles for water, acetone, hexane, and 1-octanol agree within 2% with the published results.<sup>5</sup> We have used these sigma profiles and the COSMO-SAC model to predict liquid-phase activity coefficients within an 8% difference (root mean square (RMS) error of 0.1374, for  $\ln \gamma$ ) of the published predictions for the methyl-acetate/water systems.<sup>5</sup> We have concluded from our study of conformational effects that changes in the sigma profile from small to large structural conformational variations can have a significant effect on property predictions. Changes in the sigma profile resulted in the prediction of an azeotrope in some conformations but not in others. Consideration should be given to possible conformational variations in all species, although, in our study, we only present data for variations in one species. COSMO-based models can predict solubility for both smaller solutes (such as benzoic acid) and larger solutes (such as dibenzofuran). Our database, along with the COSMO-SAC-VT-2005 FORTRAN code, enables the user to further

Table A1. Comparison of Computed Energies for Four Molecules Using Different Versions of DMol<sup>a</sup>

	VT-2005	UD
<b>Furfural</b>		
<b>Gas Phase energy (kcal/mol)</b>	-215528.7	-215524.2874
<b>Conductor energy (kcal/mol)</b>	-215535.6	-215531.2219
<b>Energy Difference (kcal/mol)</b>	-6.9027	-6.9345
<b>% discrepancy</b>	0.46%	
<b>Butyric Acid</b>		
<b>Gas Phase energy</b>	-193154.4	-193150.5166
<b>Conductor energy</b>	-193161.0	-193157.1743
<b>Energy Difference</b>	-6.6483	-6.6576
<b>% discrepancy</b>	0.139%	
<b>2,4-Pentanedione</b>		
<b>Gas Phase energy</b>	-217056.9920	-217006.948068422
<b>Conductor energy</b>	-217065.1665	-217015.0692362
<b>Energy Difference</b>	-8.1745	-8.1211677728079
<b>% discrepancy</b>	0.66%	
<b>Levulinic Acid</b>		
<b>Gas Phase energy</b>	-264295.9210	-264227.748809607
<b>Conductor energy</b>	-264306.6045	-264237.752167804
<b>Energy Difference</b>	-10.6836	-10.0033581965254
<b>% discrepancy</b>	6.8%	

<sup>a</sup> From ref 34.

apply and advance this novel approach to predict phase-equilibrium behavior for process and product development.

#### Acknowledgment

We thank Alliant Techsystems (particularly Ken Dolph, Vice President), Aspen Technology (particularly Ron Langill, Director of Worldwide University Programs, Larry Evans, Founder, and Mark Fusco, President), China Petroleum and Chemical Cor-

poration (particularly Tianpu Wang, President, and Xianghong Cao, Senior Vice President), Formosa Petrochemical Corporation (particularly Wilfred Wang, President), Milliken Chemical (particularly John Rekers, President, and Ed Zhao, Director of R&D), and Honeywell Specialty Materials (particularly Tom Williams, Senior Program Manager) and Honeywell International Foundation for supporting our educational programs in computer-aided design and process systems engineering at

Component Number	0638
Component Name	METHYL-ACETATE
Chemical Formula	C3H6O2
CAS #	79-20-9
Molecular Volume, V (Å <sup>3</sup> )	97.00036
Screening Charge Density, σ <sub>m</sub> (e/Å <sup>2</sup> )	Sigma Profile - p(σ)*A(σ),(Å <sup>2</sup> )
-0.025	0
-0.024	0
-0.023	0
-0.022	0
-0.021	0
-0.020	0
-0.019	0
-0.018	0
-0.017	0
-0.016	0
-0.015	0
-0.014	0
-0.013	0
-0.012	0
-0.011	0
-0.010	0
-0.009	0
-0.008	0.439810777
-0.007	4.364406899
-0.006	9.853749198
-0.005	11.73903343
-0.004	8.996995407
-0.003	12.20087449
-0.002	15.15980019
-0.001	9.703940341
0.000	5.318819766
0.001	2.703970156
0.002	2.08499745
0.003	1.875127284
0.004	3.332828662
0.005	1.783528908
0.006	2.722914978
0.007	2.871962706
0.008	2.91787053
0.009	3.352985278
0.010	1.58240072
0.011	4.741989323
0.012	4.302562662
0.013	1.410363745
0.014	0.093727103
0.015	0
0.016	0
0.017	0
0.018	0
0.019	0
0.020	0
0.021	0
0.022	0
0.023	0
0.024	0
0.025	0
Molecular Surface Area, A (Å <sup>2</sup> )	113.5546

Figure A1. VT-0638 methyl acetate sigma profile.

Virginia Tech. We thank Dr. Andreas Klamt (COSMOlogic) and Dr. Yuhua Song (Aspen Technology), who provided sound critiques of our work. We also thank Dr. Frank Eckert for providing the original COSMO-RS data for Figures 28–31. M.Z. gratefully appreciates support from a National Science Foundation Graduate Fellowship.

## Nomenclature

### English Symbols

$a_{\text{eff}}$  = effective surface area of a standard surface segment;  $a_{\text{eff}} = 7.5 \text{ \AA}^2$

Table B1. DMol3/COSMO Results: COSMO Input<sup>a</sup>

parameter	value
dielectric constant	∞ (infinity)
basic grid size	1082
number of segments	92
solvent radius	1.30
A – matrix cutoff	7.00
radius increment	0.00
nonelectrostatic energy expression	$A + B \times (\text{area})$
A	1.88219
B	0.01014
total energy	-268.480147 au
dielectric energy	-0.010172 au
total energy, corrected	-268.480173 au
dielectric energy, corrected	-0.010197 au
sum of polarization charges	-0.02581
sum of polarization charges, corrected	-0.00080
total surface area of cavity	113.55453 Å <sup>2</sup>
total volume of cavity	97.00036 Å <sup>3</sup>

<sup>a</sup> VT-2005 COSMO output for methyl acetate is text.

Table B2

Molecular car file 638.car					
!BIOSYM archive 3					
PBC = OFF					
!DATE: Dec 11 08:16:57 2003					
C1	-1.725099231	-0.656894195	-0.033562033	XXXX 1	xx C 0.000
C2	-0.526609885	0.253782039	0.054419124	XXXX 1	xx C 0.000
O1	0.638324874	-0.436992955	-0.156446303	XXXX 1	xx O 0.000
C3	1.844409725	0.367173782	-0.037237431	XXXX 1	xx C 0.000
O2	-0.567826024	1.452751547	0.303857916	XXXX 1	xx O 0.000
H1	-1.521962849	-1.548493114	-0.641407542	XXXX 1	xx H 0.000
H2	-2.576197435	-0.095565319	-0.440486457	XXXX 1	xx H 0.000
H3	-1.994344730	-0.985571828	0.983450259	XXXX 1	xx H 0.000
H4	1.831494550	1.186334067	-0.768875124	XXXX 1	xx H 0.000
H5	2.672270229	-0.321779671	-0.239260568	XXXX 1	xx H 0.000
H6	1.925540776	0.785255644	0.975548160	XXXX 1	xx H 0.000
end					
end					

Table B3. COSMO-RS Atomic Data

atom number	atom name	radius (Å)	charge	surface	charge density
1	C1	2.00	-0.00967	17.10870	-0.00056
2	C2	2.00	-0.02069	9.25218	-0.00224
3	O1	1.72	0.08278	10.04567	0.00824
4	C3	2.00	-0.02151	17.94297	-0.00120
5	O2	1.72	0.22447	18.95295	0.01184
6	H1	1.30	-0.04129	6.61363	-0.00624
7	H2	1.30	-0.04243	6.89064	-0.00616
8	H3	1.30	-0.04938	6.81745	-0.00724
9	H4	1.30	-0.03756	6.57735	-0.00571
10	H5	1.30	-0.04762	6.73608	-0.00707
11	H6	1.30	-0.03792	6.61690	-0.00573

A = coulomb interaction matrix

$A_i$  = total cavity surface area (Å<sup>2</sup>)

$A_i(\sigma_m)$  = surface area of all segments with a surface charge density  $\sigma_m$  (Å<sup>2</sup>)

au = atomic unit; Bohr radius,  $5.2918 \times 10^{-11} \text{ m}$

$c_{\text{hb}}$  = a constant for hydrogen-bonding ((kcal/mol)Å<sup>4</sup>/e<sup>2</sup>)

COSMO = conductor-like screening model

COSMO-RS = conductor-like screening model for realistic solvation

COSMO-SAC = conductor-like screening model with segment activity coefficients

$d_{mn}$  = distance between surface segment  $m$  and  $n$  (Å)

$e$  = elementary charge;  $e = 1.6022 \times 10^{-19}$  coulomb

$g^E$  = excess Gibbs free energy (kcal/mol)

$\Delta G^{*\text{is}}$  = free energy of ideal solvation (kcal/mol)

$\Delta G^{*\text{res}}$  = free energy of surface charge restoration (kcal/mol)

$\Delta G^{*\text{sol}}$  = solvation free energy (kcal/mol)

Table B4. Segment Information<sup>a</sup>

segment number, <i>n</i>	atom <sup>b</sup>	Position <sup>c</sup> (au)			segment charge	segment area	charge/area	solute potential on a segment
		X	Y	Z				
1	1	-3.22451	-4.65548	1.55726	0.00049	0.23228	0.00210	0.02479
2	1	-6.30262	-3.36512	-0.78171	0.00032	0.37165	0.00087	0.02205
3	1	-4.59184	-1.26848	-3.60031	0.00080	0.41810	0.00190	0.01713
4	1	-4.92524	0.76440	2.67300	0.00034	0.41810	0.00081	-0.00927
5	1	-5.74393	0.08261	2.45873	0.00031	0.18582	0.00169	0.00688
6	1	-5.49464	1.30009	1.61930	0.00052	0.27874	0.00186	-0.01598
7	1	-4.26113	-2.30312	-3.54976	0.00026	0.27874	0.00093	0.02269
8	1	-5.53288	-3.97042	-1.35575	-0.00003	0.41810	-0.00008	0.02801
9	1	-2.83750	-4.88573	0.84441	0.00035	0.37165	0.00094	0.02285
10	1	-6.57686	-0.92323	1.72015	0.00047	0.37165	0.00125	0.01777
11	1	-6.72000	-0.42134	1.21717	0.00034	0.23228	0.00148	0.01664
12	1	-1.27643	-4.10431	1.40399	0.00067	0.18582	0.00363	0.00458
13	1	-2.09742	-4.35021	1.74425	0.00052	0.32519	0.00159	0.01649
14	1	-4.35167	-4.71366	0.95410	0.00073	0.51102	0.00142	0.02653
15	1	-5.61945	-4.19210	0.03669	0.00047	0.41810	0.00113	0.02403
16	1	-5.99019	-3.05950	-1.94074	0.00051	0.32519	0.00156	0.02275
17	1	-5.42540	-2.03148	-3.05854	0.00066	0.37165	0.00177	0.02167
18	1	-3.42445	-1.44757	-3.83366	0.00055	0.32519	0.00169	0.01506
19	1	-2.30139	-1.23360	-3.71929	0.00022	0.23228	0.00093	0.01437
20	1	-6.58403	-2.42239	1.29299	0.00015	0.37165	0.00041	0.02530

<sup>a</sup> Total number of segments = 459. See our website (<http://www.design.che.vt.edu>) for all 459 data points for methyl acetate. <sup>b</sup> Atom associated with segment *n*. <sup>c</sup> Segment coordinates.

Ha = Hartree energy unit; Ha =  $4.3597 \times 10^{-18}$  J

$h_{\text{mix}}^E$  = excess enthalpy of mixing (J/mol)

$\Delta_{\text{fus}}H$  = enthalpy of fusion (kJ/kmol)

$n_i(\sigma_m)$  = the number of segments with a surface charge density of  $\sigma_m$

*P* = pressure (kPa)

$p'(\sigma)$  = modified sigma profile ( $\text{\AA}^2$ )

$p(\sigma)$  = sigma profile

*q* = surface area constant ( $\text{\AA}^2$ )

$q_i$  = normalized surface-area parameter

$\mathbf{q}^*$  = surface screening vector in the conductor

(*r*) = a specific position

*r* = volume constant ( $\text{\AA}^3$ )

*R* = ideal gas constant; *R* = 0.001987 kcal mol<sup>-1</sup> K<sup>-1</sup> or *R* = 8.314 kJ kmol<sup>-1</sup> K<sup>-1</sup>

$r_{\text{av}}$  = surface-segment averaging radius (adjustable parameter) ( $\text{\AA}$ )

$r_i$  = normalized volume parameter

$r_n$  = effective radius of surface segment *n*, assuming circular surface segments ( $\text{\AA}$ )

SLE = solid-liquid equilibrium

$\Delta_{\text{fus}}S$  = entropy of fusion (kJ kmol<sup>-1</sup> K<sup>-1</sup>)

*T* = temperature (K)

$T_m$  = melting point temperature (K)

$V_i$  = total cavity volume ( $\text{\AA}^3$ )

VLE = vapor-liquid equilibrium

$\Delta W(\sigma_m, \sigma_n)$  = exchange energy (kcal/mol)

$x_i$  = mole fraction of component *i* in the liquid phase

$x_i^{\text{sol}}$  = solute solubility, mole fraction

*z* = coordination number

#### Greek Symbols

$\alpha'$  = the constant for the misfit energy ( $\text{\AA}^4 \cdot \text{kcal}/\text{e}^2 \cdot \text{mol}$ )

$\gamma_i$  = activity coefficient of component *i*

$\gamma_i^{\text{sat}}$  = solute activity coefficient at saturation

$\gamma_{iS}^{\text{SG}}$  = Staverman-Guggenheim combinatorial contribution to the activity coefficient

$\Gamma_s(\sigma_m)$  = segment activity coefficient of the solvent

$\Gamma_s(\sigma_m)$  = segment activity coefficient of the solute

$\Phi_{\text{sol}}$  = potential due to the charge distribution of the solute molecule

$\Phi_{\text{tot}}$  = total potential on the cavity surface

$\phi_i$  = normalized volume fraction

$\sigma$  = surface-segment charge-density distribution ( $\text{e}/\text{\AA}^2$ )

$\sigma$  = averaged charge on the conductor surface in vector form ( $\text{e}/\text{\AA}^2$ )

$\sigma^*$  = induced-charge on the conductor surface in vector form ( $\text{e}/\text{\AA}^2$ )

$\sigma_{\text{hb}}$  = the sigma-value cutoff for hydrogen bonding ( $\text{e}/\text{\AA}^2$ )

$\sigma_n^*$  = surface-charge density for segment *n* from the COSMO output ( $\text{e}/\text{\AA}^2$ )

$\theta_l$  = normalized surface-area fraction

#### Appendix A: Energy Differences between DMol Versions

We used DMol v2.2 for most (1266) of the VT-2005 calculations, and we found small discrepancies with recent work done by Professor Sandler's group at the University of Delaware using DMol v3.9.<sup>34</sup> We used DMol v3.2 for 166 compounds, for which we also performed calculations with Amber8<sup>39</sup> and Forcite Plus (a Materials Studio annealing module) to generate good initial guesses for optimal molecular geometry prior to the DMol calculations. Initial vapor pressure predictions for these 166 compounds were not within the designated limits ( $10.0 \leq \ln P_{\text{vap}} \leq 13.0$ ), but after using a pre-optimization tool, the property predictions satisfy these limits. In Table A1, we present four examples of molecules where both groups work with the same structure, but the energies differ by small quantities. All VT-2005 calculations in Table A1 used DMol v2.2.1. These examples vary by an average of 2%, with respect to the energy difference, with Professor Sandler's recent work using DMol v3.9. Over the full range of compounds, we find an average error of 9%. We recommend compiling sigma profiles with the same calculation algorithm and software.

#### Appendix B: VT-2005 Example for Methyl Acetate (VT-0638)

Data specific to the VT-2005 example for methyl acetate (VT-0638) are given in Figure B1 and in Tables B1–B4.

**Note Added after ASAP Publication.** Because of a production error, the version of this paper that was published on the

Web 5/3/2006 was published without final review from the authors. The correct version of this paper was published 5/16/2006.

### Literature Cited

- (1) 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.
- (2) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074.
- (3) 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.
- (4) Eckert, F.; Klamt, A. Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE J.* **2002**, *48*, 369.
- (5) (a) Lin, S. T.; Sandler, S. A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res.* **2002**, *41*, 899. (b) Lin, S. T.; Sandler, S. Reply to Comments on A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res.* **2002**, *41*, 2332.
- (6) Lin, S. T. Quantum Mechanical Approaches to the Prediction of Phase Equilibria: Solvation Thermodynamics and Group Contribution Methods, PhD. Dissertation, University of Delaware, Newark, DE, 2000.
- (7) Panayiotou, C. Equation-of-State Models and Quantum Mechanics Calculations. *Ind. Eng. Chem. Res.* **2003**, *42*, 1495.
- (8) Klamt, A.; Schüürmann, 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.
- (9) Klamt, A. COSMO and COSMO-RS. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: Chichester, U.K., 1998.
- (10) Ben-Naim, A. *Solvation Thermodynamics*, Plenum Press: New York and London, 1987.
- (11) Lin, S. T.; Sandler, S. Prediction of Octanol-Water Partition Coefficients Using a Group Contribution Solvation Model. *Ind. Eng. Chem. Res.* **1999**, *38* (10), 4081.
- (12) Lin, S. T.; Sandler, S. A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. [Erratum] *Ind. Eng. Chem. Res.* **2004**, *43*, 1322.
- (13) Eckert, F.; Klamt, A. Validation of the COSMO-RS Method: Six Binary Systems. *Ind. Eng. Chem. Res.* **2001**, *40*, 2371.
- (14) Putnam, R.; Taylor, R.; Klamt, A.; Eckert, F.; Schiller, M. Prediction of Infinite Dilution Activity Coefficients Using COSMO-RS. *Ind. Eng. Chem. Res.* **2003**, *42*, 3635.
- (15) Koch, W.; Holthausen, M. C. *Chemist's Guide to Density Functional Theory*, 2nd Edition; Wiley-VCH: Weinheim, Germany, 2001.
- (16) Andzelm, J.; Kolmel, C.; Klamt, A. Incorporation of Solvent Effects into Density Functional Calculations of Molecular Energies and Geometries. *J. Chem. Phys.* **1995**, *103*, 9312.
- (17) Delley, B. An All-Electron Numerical Method For Solving the Local Density Functional for Polyatomic Molecules. *J. Chem. Phys.* **1990**, *92*, 508.
- (18) Delley, B. Analytic energy derivatives in the numerical local-density-functional approach. *J. Chem. Phys.* **1991**, *94*, 7245.
- (19) Delley, B. In *Modern Density Functional Theory: A Tool for Chemistry*; Seminario, J. M., Politzer, P., Eds.; Theoretical and Computational Chemistry, Vol. 2; Elsevier Science Publishers: Amsterdam, 1995.
- (20) Accelrys, *Materials Studio Online Help*, Release 3.2, Accelrys Software, Inc.: San Diego, CA, 2005.
- (21) Becke, A. D. Density Functional Calculations of Molecular Bond Energies. *J. Chem. Phys.* **1986**, *84*, 4524.
- (22) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev.* **1986**, *33*, 8822.
- (23) Vosko, S. J.; Wilk, L.; Nusiar, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200.
- (24) Klamt, A. *COSMO-RS, From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier BV: Amsterdam, The Netherlands, 2005.
- (25) Carmona, F. J.; et al. Thermodynamic properties of (*n*-alkoxyethanols + organic solvents). XII. Total vapour pressure measurements for (*n*-hexane, *n*-heptane, or cyclohexane + 2-methoxyethanol) at different temperatures. *J. Chem. Thermodyn.* **2001**, *33*, 47-59.
- (26) Martin, M. C.; Cocero, M. J.; Mato, R. B. Vapor-Liquid Equilibrium Data at 298.15 K for Binary Systems Containing Methyl Acetate or Methanol with 2-Methoxyethanol or 2-Ethoxyethanol. *J. Chem. Eng. Data* **1994**, *39*, 535-537.
- (27) Jonasson, A.; Savoia, M.; Persson, O.; Fredenslund, A. Isothermal Vapor-Liquid Equilibrium Data for Ether + Glycol, Chloroalkene + Glycol, Epoxy Ether + Alkane, Epoxy Ether + Alkene, and Epoxy Ether + Chloroalkane Systems. *J. Chem. Eng. Data* **1994**, *39*, 134-139.
- (28) Nienhaus, B. et al. Vapor-Liquid Equilibria at 453.25 K and Excess Enthalpies at 363.15 and 413.15 K for Mixtures of Benzene, Toluene, Phenol, Benzaldehyde, and Benzyl Alcohol with Benzyl Benzoate. *J. Chem. Eng. Data* **1999**, *44*, 303-308.
- (29) Giles, N.; Wilson, G. Phase Equilibria on Seven Binary Mixtures. *J. Chem. Eng. Data* **2000**, *45*, 146.
- (30) Frank, T. C.; Downey, J. R.; Gupta, S. K. Quickly Screen Solvents for Organic Solids. *Chem. Eng. Prog.* **1999**, (December), 41.
- (31) Marrero, J.; Abildskov, J. Solubility and Related Properties of Large Complex Chemicals, Part 1: Organic Solutes Ranging from C4 to C40. *Chemistry Data Series, XV*; DECHEMA: Frankfurt, Germany, 2003.
- (32) Kolar, P.; Shen, J. W.; Tsuboi, A.; Ishikawa, T. Solvent Selection of Pharmaceuticals. *Fluid Phase Equilib.* **2002**, *194-197*, 771.
- (33) Eckert, F.; Klamt, A. Prediction of Halocarbon Thermodynamics with COSMO-RS. *Fluid Phase Equilib.* **2003**, *210*, 117.
- (34) Sandler, S. University of Delaware, Newark, DE. Private communications, August to October, 2004.
- (35) Klamt, A. Prediction of the mutual solubilities of hydrocarbons and water with COSMO-RS. *Fluid Phase Equilib.* **2003**, *206*, 223-235.
- (36) Lin, S. T.; Chang, J.; Wang, S.; Goddard, W. A., III; Sandler, S. I. Prediction of Vapor Pressures and Enthalpies of Vaporization Using COSMO Solvation Model. *J. Phys. Chem. A* **2004**, *108*, 7429-7439.
- (37) Wang, S.; Lin, S.-T.; Chang, J.; Goddard, W. A., III; Sandler, S. I. Application of COSMO-SAC-BP Solvation Model to Predictions of Normal Boiling Temperatures for Environmentally Significant Substances. *Ind. Eng. Chem. Res.* **2005**, in press [DOI: 10.1021/ie050352k].
- (38) Constantinescu, D.; Klamt, A.; Geana, D. Vapor-liquid equilibrium prediction at high pressures using activity coefficients at infinite dilution from COSMO-type methods. *Fluid Phase Equilib.* **2005**, *231*, 231-238.
- (39) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. Amber8, University of California, San Francisco, CA, 2004.

Received for review March 25, 2006

Accepted March 27, 2006

IE060370H