Predicting Phase Equilibria Using COSMO-Based Thermodynamic Models and the VT-2004 Sigma-Profile Database

Richard Justin Oldland

Thesis submitted to the faculty of
the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE
In
Chemical Engineering

APPROVED:

Dr. Y.A. Liu, chairman
Dr. Rick M. Davis
Dr. Eva Marand

November 2004

Blacksburg, VA

Keywords: Sigma Profile, COSMO-SAC, COSMO, COSMO-RS, VT-2004, Solvation Thermodynamics

Predicting Phase Equilibria Using COSMO-Based Thermodynamic Models and the VT-2004 Sigma-Profile Database

Richard Justin Oldland

Dr. Y.A. Liu, chairman

Abstract

Solvation-thermodynamics models based on computational quantum mechanics, such as the conductor-like screening model (COSMO), provide a good alternative to traditional group-contribution methods for predicting thermodynamic phase behavior. Two COSMO-based thermodynamic models are COSMO-RS (real solvents) and COSMO-SAC (segment activity coefficient). The main molecule-specific input for these models is the sigma profile, or the probability distribution of a molecular surface segment having a specific charge density. Generating the sigma profiles represents 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 generating the sigma profiles. This thesis presents the first free, open-literature database of 1,513 self-consistent sigma profiles, together with two validation examples. The offer of these profiles will enable interested scientists and engineers to use the quantum-mechanics-based, COSMO methods without having to do quantum mechanics. This thesis summarizes the application experiences reported up to October 2004 to guide the use of the COSMO-based methods. Finally, this thesis also provides a FORTRAN program and a procedure to generate additional sigma profiles consistent with those presented here, as well as a FORTRAN program to generate binary phase-equilibrium predictions using the COSMO-SAC model.

Acknowledgement

First, I would like to thank my wonderful wife, Heather, for standing by me when we decided to return to Blacksburg for graduate school. She has been the one true constant in my life for a very long time and I could not have done this without her and her unwavering support.

Dr. Y.A. Liu deserves my undivided gratitude for providing the opportunity to continue my education at Virginia Tech. He was a great source of advice and reason as my undergraduate advisor and over the past two years as my graduate advisor. Dr. Liu is a very kind and understanding man who has put in many hours studying my work and teaching me many things that I will cherish throughout my career. Thank you very much Dr. Liu.

Next, I want to thank Dr. Eva Marand and Dr. Rick Davis for serving on my graduate committee. Both were very helpful and informative as instructors and very understanding as advisors toward completing my degree.

Finally, I want to thank some of my fellow graduate students. The former and current members of the design group, Kevin Seavey, Neeraj Khare, Bruce Lucas, Anthony Gaglione, and Eric Mullins. Each was a great sounding board and an overall wealth of advice concerning my graduate studies and the virtues of *Halo*. Also, Travis Gott and Will James were always willing to listen and put my golf game to the test.

Thank you everyone.

Richard Oldland

November 2004
Blacksburg, VA

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

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, based on computational quantum mechanics, enable us to predict thermophysical properties without any experimental data.

We begin with the *conductor-like screening model* or COSMO, which characterizes a molecule's surface-charge density. We then extend this model to include interactions between molecules in a condensed phase. Two such extensions are COSMO-RS[1,2,3,4] (*real solvents*) and COSMO-SAC[5,6] (*segment activity coefficient*), 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 can generate sigma profiles with only the molecular structure and quantum-mechanical calculations. Generating sigma profiles 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…").[7]

This work represents the first open-literature database containing sigma profiles for 1,513 chemicals (a paper based on this work is in press in *Industrial and Engineering Chemistry Research*). These chemicals have only ten elements H, C, N, O, F, P, S, Cl, Br, and I. Our database is available free of charge from our website (https://www.design.che.vt.edu/VT-2004.htm). We 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 two examples 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 results.

1.2 Thesis Overview

This work begins by reviewing the literature work on dielectric continuum-based solvation thermodynamic models and then continues with extending these models to the liquid-phase where we predict binary mixture behavior. We continue by outlining the VT-2004 sigma-profile database and our contribution to the field. We include several validation examples that show the usefulness of this approach and the VT-2004 database. Finally, we provide several guidelines from literature observations regarding the limitations and strengths of these models.

2 Background and Theory

2.1 Introduction

We give a brief overview of the COSMO model, sigma profiles, and the COSMO-SAC model. The COSMO model generates a surface-charge distribution for a single molecule. We plot the probability distribution of a molecular surface segment having a specific charge distribution as the sigma profile, which is used by COSMO-RS/SAC to compute activity coefficients.

2.2 The COSMO Model

The basis of the COSMO model is the "solvent accessible surface" of a solute molecule.[8,9] Conceptually, COSMO places 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 model.

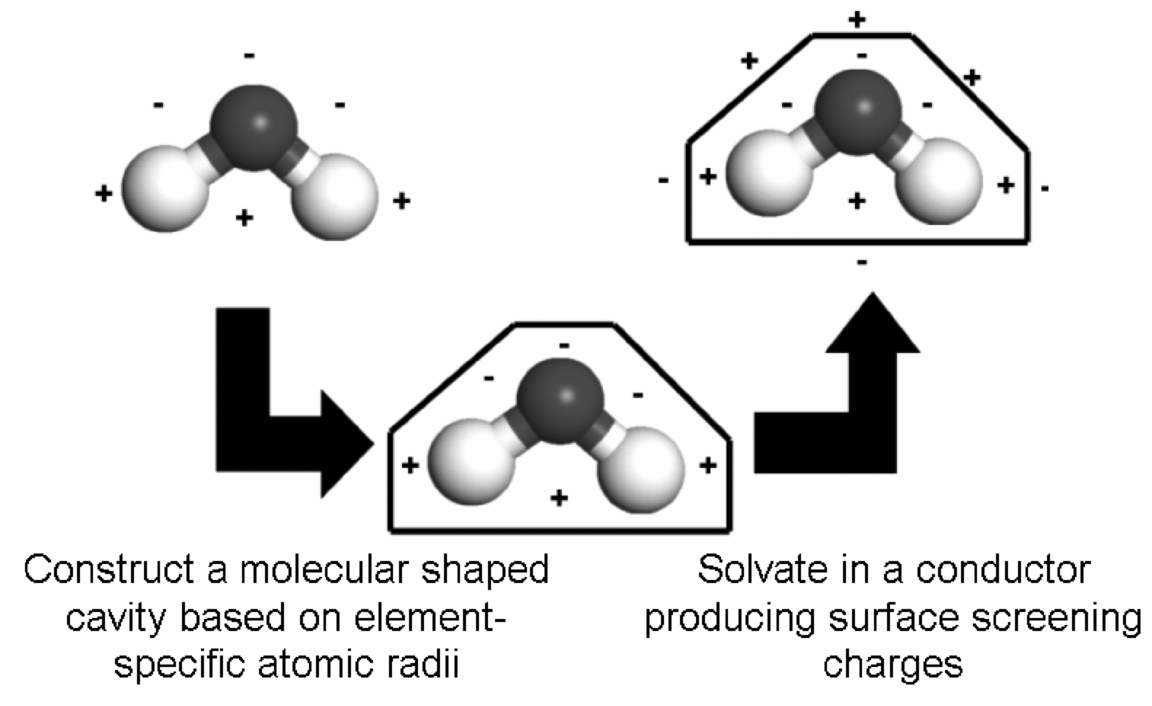


Figure 1. The ideal solvation process in the COSMO model places the molecule in a cavity and into a conducting medium. The molecule pulls charges from the conductor to the cavity surface. Then we represent this surface charge distribution as a sigma profile.

The model constructs the cavity within a perfect conductor according to a specific set of rules and atom-specific dimensions. Then, the molecule's dipole and higher moments 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

$$\mathbf{\Phi}_{\mathbf{tot}} = 0 = \mathbf{\Phi}_{\mathbf{sol}} + \mathbf{A}\mathbf{q}^{*} \tag{1}$$

where $\mathbf{\Phi}_{\mathbf{tot}}$ is the total potential on the cavity surface, $\mathbf{\Phi}_{\mathbf{sol}}$ is the potential due to the charge distribution of the solute molecule, and $\mathbf{q}^{*}$ is 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.[9] Eq. (1) is easier to solve than the corresponding equation in a medium of *finite* dielectric constant. The surface-charge distribution in a finite dielectric solvent is well approximated by a simple scaling of the surface-charge distribution σ^* in a conductor. In this way, COSMO greatly reduces the computational cost with a minimal loss of accuracy.[9]

2.3 Sigma Profiles

We average the COSMO output, σ^*, over a circular surface segment to obtain a new surface-charge distribution σ. Then we represent this charge distribution as the probability distribution of a molecular surface segment having a specific charge density. This probability distribution is called the sigma profile, $p(\sigma)$. Klamt[1] defines the sigma profile for a molecule *i* as

$$p_i(\sigma) = \frac{n_i(\sigma)}{n_i} = \frac{A_i(\sigma)}{A_i} \qquad (2)$$

$$n_i = \sum_{\sigma} n_i(\sigma) = \frac{A_i}{a_{eff}} \qquad (3)$$

$$A_i = \sum_{\sigma} A_i(\sigma) \qquad (4)$$

where $n_i(\sigma)$ is the number of segments with a discretized surface-charge density σ, A_i is the total cavity surface area, and $A_i(\sigma)$ is the total surface area of all of the segments with a particular charge density σ. Lin and Sandler[5] define $A_i(\sigma) = a_{eff}*n_i(\sigma)$, where a_{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[1] sets this adjustable parameter to 7.5Å^2.

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

$$p_s(\sigma) = \frac{\sum_i x_i n_i p_i(\sigma)}{\sum_i x_i n_i} = \frac{\sum_i x_i A_i p_i(\sigma)}{\sum_i x_i A_i} \tag{5}$$

We average the surface-charge densities from the COSMO output to find an effective surface-charge density using the following equation:[2]

$$\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{av}^2}{r_n^2 + r_{av}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{av}^2}\right)}{\sum_n \frac{r_n^2 r_{av}^2}{r_n^2 + r_{av}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{av}^2}\right)} \tag{6}$$

where σ_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. [2,5]

We use an averaging radius, r_{av}= 0.81764 Å, for the sigma-averaging algorithm that is different from the effective segment radius.[2] This corresponds to a_{av} = 7.5 au^2 =

2.10025 Å^2. The averaging radius directly affects the sigma profiles. Our averaging algorithm is identical to Lin and Sandler[10], Klamt[1], and Klamt et al.[2], except that we use a different value for r_{av}. Klamt and his coworkers report using averaging radii ranging between 0.5 and 1.0Å, stating that "[t]he best value for the averaging radius r_{av} turns out to be 0.5 Å. This is considerably less than the initially assumed value of about 1 Å."[2] Lin and Sandler use a similar algorithm involving the effective segment radius, but introduce another adjustable parameter, *c,* which they place in the exponential term in Eq. (6).[10] The VT-2004 sigma-profile database includes calculation results from the density-functional theory, thus enabling future work in optimization of the r_{av} value.

The sigma profile contains 50 segments, 0.001 e/Å^2 wide, in the range -0.025 to 0.025 e/Å^2. Figure 2 illustrates sigma profiles for water, acetone, n-hexane, and 1-octanol.

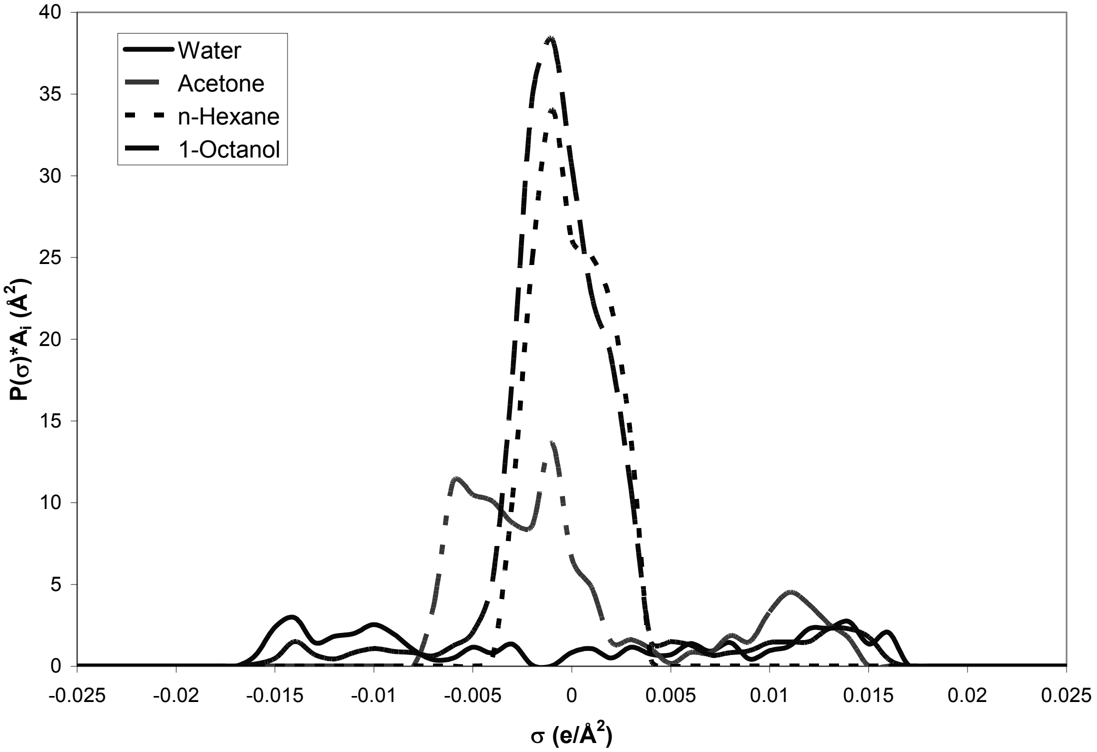


Figure 2. Sigma profiles for water, acetone, n-hexane, and 1-octanol

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.[1,11,12]

2.4 The COSMO-SAC Model

We outline the derivation and explanation of the COSMO-SAC model as published in Lin and Sandler[5] and the small variations we use in the VT-2004 programs. We first define the free energy of solvation, ΔG^{*sol}, as the energy of cavity formation, ΔG^{*cav} and the difference between the free energy of ideal solvation, ΔG^{*is}, and the restoring free energy, ΔG^{*res} (which removes the screening charges). The ideal solvation energy is identical for dissolving a solute in a solvent, *S*, or in pure solute, *i*.[13] We use the following definition of the activity coefficient

$$\ln \gamma_{i/s} = \frac{\Delta G^{*res}_{i/s} - \Delta G^{*res}_{i/i}}{RT} + \ln \gamma^{SG}_{i/s} \tag{7}$$

where $\gamma^{SG}{}_{i/S}$ is the Staverman-Guggenheim combinatorial term, which improves the calculations for the cavity-formation free energy according to Lin and Sandler.[14] They define this term as

$$\ln \gamma^{SG}_{i/s} = \ln\left(\frac{\phi_i}{x_i}\right) + \frac{z}{2} q_i \ln\left(\frac{\theta_i}{\phi_i}\right) + l_i - \frac{\phi_i}{x_i} \sum_j x_j l_j \tag{8}$$

where ϕ_i is the normalized volume fraction, θ_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[5] define the restoring free energy as the sum of the sigma profile times the natural log of the segment activity coefficients over all surface charges

$$\frac{\Delta G^{*res}_{i/s}}{RT} = \sum_{\sigma_m} \left[n_i\left(\sigma_m\right) \frac{\Delta G^{*res}_{\sigma_m/s}}{RT} \right] = n_i \sum_{\sigma_m} p_i\left(\sigma_m\right) \ln \Gamma_s\left(\sigma_m\right) \tag{9}$$

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

We calculate the segment activity coefficient using

$$\begin{aligned}\ln\Gamma_s\left(\sigma_m\right) &= -\ln\left\{\sum_{\sigma_n} p_s\left(\sigma_n\right)\Gamma_s\left(\sigma_n\right)\exp\left[\frac{-\Delta W\left(\sigma_m,\sigma_n\right)}{RT}\right]\right\}\\ \ln\Gamma_i\left(\sigma_m\right) &= -\ln\left\{\sum_{\sigma_n} p_s\left(\sigma_n\right)\Gamma_i\left(\sigma_n\right)\exp\left[\frac{-\Delta W\left(\sigma_m,\sigma_n\right)}{RT}\right]\right\}\end{aligned} \tag{10}$$

as derived rigorously using statistical mechanics.[5]

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

$$\Delta W\left(\sigma_m,\sigma_n\right) = \left(\frac{\alpha'}{2}\right)\left(\sigma_m+\sigma_n\right)^2 + c_{hb}\max\left[0,\sigma_{acc}-\sigma_{hb}\right]\min\left[0,\sigma_{don}+\sigma_{hb}\right] \tag{11}$$

where α' is the constant for the misfit energy which Klamt et al[2] and Klamt and Eckert[3] fit to experimental data, c_{hb} is a constant for hydrogen bonding, σ_{hb} is the sigma-value cutoff for hydrogen bonding.

Table 1 shows the values that we use for each of the adjustable parameters in the COSMO-SAC model.[2,5] These values are universal and independent of the atoms/molecules modeled.

Table 1. Parameter values used in the COSMO-SAC model[2,5]

Symbol	Units	Value	Meaning
r_{av}	Å	0.81764	sigma averaging radius
a_{eff}	$Å^2$	7.5	effective surface segment surface area
c_{hb}	kcal/mole*$Å^4$/e^2	85580.0	hydrogen-bonding constant
σ_{hb}	e/$Å^2$	0.0084	sigma cutoff for hydrogen-bonding
α'	$Å^4$*kcal/e^2*mol	9034.97	misfit energy constant
z	no units	10	coordination number
q	$Å^2$	79.53	standard area parameter
r	$Å^3$	66.69	standard volume parameter

We calculate the activity coefficient using the following expression

$$\ln \gamma_{i/s} = n_i \sum_{\sigma_m} p_i\left(\sigma_m\right)\left[\ln \Gamma_s\left(\sigma_m\right) - \ln \Gamma_i\left(\sigma_m\right)\right] + \ln \gamma_{i/s}^{SG} \tag{12}$$

2.5 VT-2004 Variation of the COSMO-SAC Model

We use the COSMO-SAC model proposed by Lin and Sandler[5] and 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.[3] 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-2004.exe (FORTRAN) program, is in our initial sigma profile calculation.

We use a slightly different definition of the sigma profile

$$p'\left(\sigma\right) = A_i\left(\sigma\right) = p\left(\sigma\right) * A_i \tag{13}$$

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 which incorporate the sigma profile.

The first place this presents an issue is the definition of the segment activity coefficient. Where Lin and Sandler use Eq. (10), we use[3]

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

which uses the same definition for each term as Eq. (10).

We also observe a change in the final form of the activity-coefficient equation. We use the following equation which is comparable to Eq. (12) as used by Lin and Sandler.

$$\ln\gamma_{i/s} = \frac{1}{a_{eff}}\sum_{\sigma_m} p'_i(\sigma_m)\left[\ln\left(\frac{\Gamma_s(\sigma_m)}{\Gamma_i(\sigma_m)}\right)\right] + \ln\gamma_{i/s}^{SG} \tag{15}$$

3 Method and Results

3.1 *Introduction*

The VT-2004 sigma-profile database contains 1513 compounds from a wide range of chemical structures. We outline how we generate the VT-2004 sigma-profile database and compare our sigma profiles to published profiles.

3.2 *Procedure to Compute Sigma Profiles*

First, we calculate the surface charges using the COSMO model and density-functional theory.[15] We use the DMol package in Accelrys' Material Studios v2.2 software suite.[16-19] Then, we average these charges to produce the sigma profile. We optimize the molecular geometry for each molecule with the DNP v4.0.0 basis set. We draw each 3-dimensional structure in Materials Studio. We use the "clean" tool to automatically adjust bond lengths and bond angles. To correctly place the hydroxyl hydrogen atoms and obtain the conformations of ethane, we manually place the hydroxyl hydrogen atom in a planar fashion (with respect to the C-C backbone) and rotate the ethane molecule to a staggered conformation.

Next, we obtain the ideal-gas phase equilibrium geometry with a density-functional theory (GGA/VWN-BP and DNP v.4.0.0) energy-minimization calculation that relaxes the geometry and obtains a low-energy state. Here, GGA stands for the generalized gradient approximation, and VWN-BP represents the Becke-Perdew version of the Volsko-Wilk-Nusair functional [15,20,21,22]. 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[15], DNP basis set is generally very reliable. We use the GGA/VWN-BP functional, with a real space cutoff of 5.5 Å.

Then we proceed with the solvation or COSMO calculation to generate the sigma profile at the same VWN-BP/DNP level.

We optimize the geometry under "fine" tolerances (1.0E-6 Hartree energy unit or *Ha*, for convergence of the self-consistent field equations, and 0.002 *Ha/Å* for the convergence of the geometry-optimization calculations). We include the atomic radii for the following ten elements in the COSMO calculations on the optimized geometries: H, C, N, O, F, P, S, Cl, Br, I (see Table 2). Klamt et al[2] and Klamt and Eckert[3] optimize these radii and suggest using 117% of the van der Waals bonding radius to approximate other elements. Sigma profiles with other atoms could contain some errors, since the COSMO-RS and COSMO-SAC approaches, as reported in the literature[1-6], are only optimized for these ten elements.[3,5] The VT-2004 database only includes compounds containing these ten elements.

Table 2. Table of elemental atomic radii for creating the COSMO molecular cavity[3]

Element	Cavity radius [*Å*]
H	1.30
C	2.00
N	1.83
O	1.72
F	1.72
S	2.16
P	2.12
Cl	2.05
Br	2.16
I	2.32

We calculate the surface charges on the optimized molecular geometry using the same tolerances and settings. Then, we average the COSMO output to yield the sigma profile.

The total calculation time for a single sigma profile depends on the molecule complexity (number of atoms and bonds) and the quality of the initial geometry. Small molecules like methane and ethane require twenty minutes on a 2.5-3.0 GHz Pentium IV-equipped PC. Other larger molecules require calculation times of five to seventy two hours on the same machine.

The VT-2004 sigma profiles represent a single conformation, but we acknowledge that large molecules may also exist in additional low-energy conformations. These additional conformations create slightly different sigma profiles from the conformations shown here.

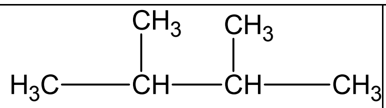
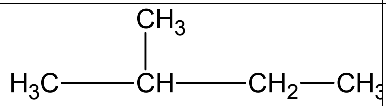
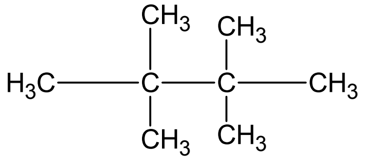
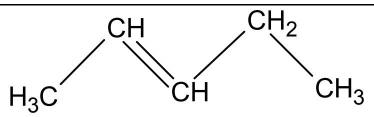
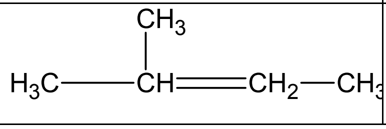
Section 3.5 illustrates the effect of conformational variations on the sigma profiles and energy calculations for small, medium-sized, and large molecules. We observe increased sigma-profile conformational variations with increasing molecule size. Also, solubility predictions appear to be more sensitive to these conformational variations than vapor-liquid equilibrium predictions.

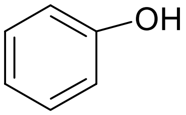
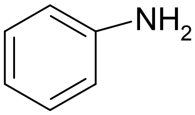
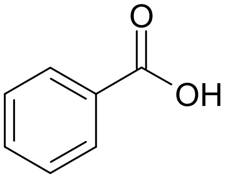
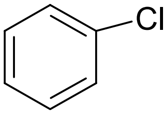
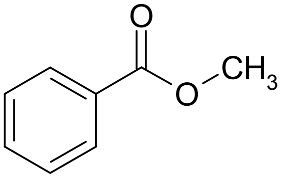
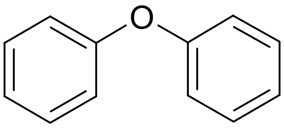
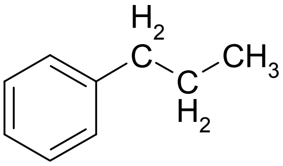
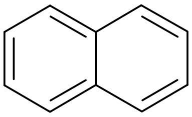
The molecular conformation used in the DFT calculations has a large effect on the sigma profile and as such, great care should be used in obtaining a low-energy geometry. We have verified over 400 molecules and an independent team at the University of Delaware has verified over 1200 molecules. In both cases, we observe less than 5% of our 1,513 compounds in the database initially had incorrect structures or non-low energy conformations. The current database includes the corrected structures and the low-energy conformations.

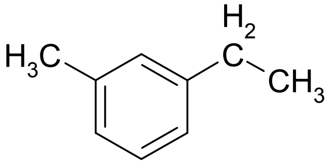
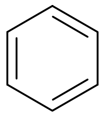
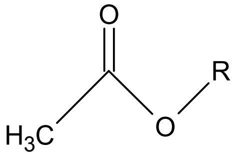
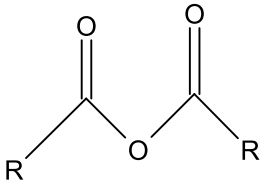
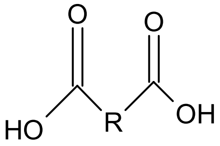
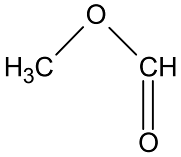
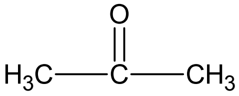
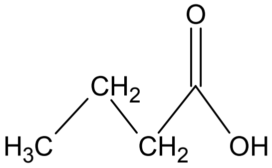
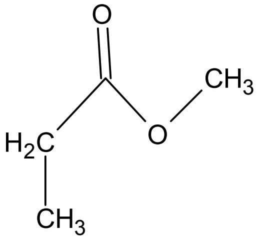
3.3 VT-2004 Database Summary

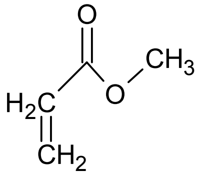
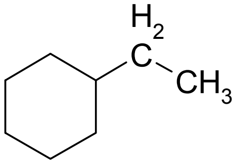
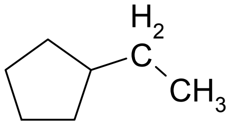
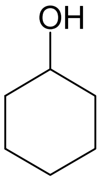
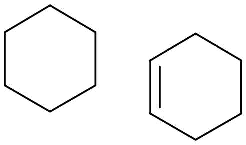
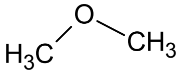
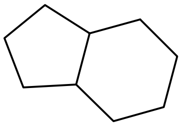
Table 3 summarizes the chemicals contained in the VT-2004 sigma-profile database. We identify each compound by a unique VT-2004 index number, its CAS registry number, chemical formula, and name in the VT-2004 index.

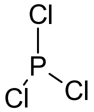
Table 3. Summary of the VT-2004 sigma-profile database

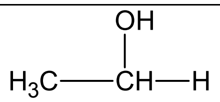
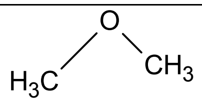
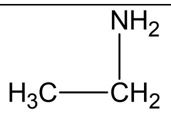
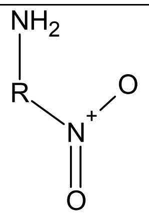
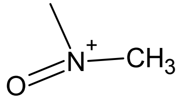
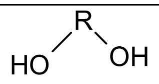
Chemical Family	**Characteristic Structure**	**Number of Compounds in VT-2004**	**Example Compounds**
	Alkanes		
Dimethyl-Alkanes	 $H_3C-\overset{CH_3}{CH}-\overset{CH_3}{CH}-CH_3$	21	2,2-dimethyl-propane
Methyl-Alkanes	 $H_3C-\overset{CH_3}{CH}-CH_2-CH_3$	17	Isobutane
n-Alkanes	H_3C-CH_3	27	Methane, Ethane, Propane
Other Alkanes	 $H_3C-\underset{CH_3}{\overset{CH_3}{C}}-\underset{CH_3}{\overset{CH_3}{C}}-CH_3$	25	3-Ethylpentane
	Alkenes		
Dialkenes	$H_2C=C=CH_2$	25	Propadiene (C_3H_4)
Ethyl/higher-Alkenes		11	2-ethyl-1-butene (C_6H_{12})
Methyl-Alkenes	 $H_3C-\overset{CH_3}{CH}=CH_2-CH_3$	22	Isobutylene

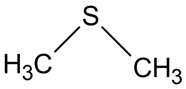
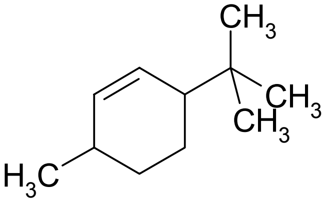
n-Alkenes	$H_2C{=}CH_2$	41	Ethylene, 1-Butene, cis-2-Butene
Alkynes			
n-Alkynes	$HC{\equiv}CH$	18	Acetylene (C_2H_2)
Aromatic structures			
Aromatic Alcohols		32	Phenol
Aromatic Amines		37	Pyridine and Aniline
Aromatic Carboxylic Acids		11	Benzoic Acid ($C_7H_6O_2$)
Aromatic Chlorides		15	Benzyl Chloride (C_7H_7Cl)
Aromatic Esters		23	Benzyl Acetate ($C_9H_{10}O_2$)
Diphenol/Polyaromtics		18	Diphenyl ($C_{12}H_{10}$)
n-Alkyl Benzenes		19	Toluene
Naphthalenes		15	1-Methylnaphthalene ($C_{11}H_{10}$)

Other Alkyl Benzenes		44	m-Xylene (C_8H_{10})
Other Monoaromatics		15	Benzene or Styrene
Carboxylic compounds			
Acetates		22	Methyl Acetate, Vinyl Acetate, Allyl Acetate
Anhydrides		9	Acetic Anhydride
Dicarboxylic Acids		14	Adipic Acid ($C_6H_{10}O_4$)
Formates		15	Ethyl Formate ($C_3H_6O_2$)
Ketones		33	Acetone
n-Aliphatic Acids		20	Formic Acid, Acetic Acid, Propionic Acid
Propionates and Butyrates		13	Methyl Propionate ($C_4H_8O_2$)

Unsaturated Aliphatic Esters		23	Methyl Acrylate ($C_4H_6O_2$)
Cyclic compounds			
Alkylcyclohexanes		16	Methylcyclohexane
Alkylcyclopentanes		11	Ethylcyclopentane
Cycloaliphatic Alcohols		10	Cyclohexanol
Cycloalkanes/alkenes		15	Cyclooctene (C_8H_{14}) or Cyclobutane (C_4H_8)
Epoxides	 (cyclic)	14	1,4-Dioxane ($C_4H_8O_2$)
Multiring-cycloalkanes		3	Cis-Decalin ($C_{10}H_{18}$)
Other Hydrocarbon Rings	No Common Structure	16	Adamantane ($C_{10}H_{16}$)
Other Condensed Rings	No Common Structure	10	Fluorene ($C_{13}H_{10}$)
Halogenated compounds			
Halogenated Alkanes	H_3C——X	98	Methyl Bromide, Ethyl Chloride, Difluoromethane

Multihalogenated Alkanes	$X—CH_2—X$	37	Chlorofluormethane
Inorganic compounds			
Inorganic Acids/Bases	No Common Structure	17	Sulfuric Acid, Nitric Acid, Ammonia
Inorganic Gases	No Common Structure	19	Ozone (O_3)
Inorganic Halides		7	Thionyl Chloride ($SOCl_2$)
Other Inorganics	No Common Structure	3	Water
Polyfunctional compounds			
Polyfunctional Acids	No Common Structure	17	Salicylic Acid ($C_7H_6O_3$)
Polyfunctional Amides/Amines	No Common Structure	28	Urea
Polyfunctional C,H,O,halides	No Common Structure	37	Chloroaniline (C_6H_6ClN)
Polyfunctional-C,H,N,halide,(O)	No Common Structure	12	Phosgene
Polyfunctional C,H,O,N	No Common Structure	30	Caffeine
Polyfunctional C,H,O,S	No Common Structure	13	Sulfolane ($C_4H_8O_2S$)
Polyfunctional Esters	No Common Structure	20	Ethyl Lactate ($C_5H_{10}O_3$)
Polyfunctional Nitriles	No Common Structure	7	Aminocapronitrile ($C_6H_{12}N_2$)
Other Polyfunctional C,H,O	No Common Structure	38	Dextrose
Other	No Common Structure	8	Malathion ($C_{10}H_{19}O_6PS_2$)

Polyfunctional Organics			
Other structures			
Aldehydes	 $H_3C-CH(OH)-H$	31	Formaldehyde
Aliphatic Ethers		33	Diethyl Ether
Elements	No Common Structure	7	Oxygen, Hydrogen, Nitrogen
Isocyanates/Diisocyanates	$R-N=C=O$	8	Tolene-diisocyanate ($C_9H_6N_2O_2$)
Mercaptans	$HS-R$	23	Methyl Mercaptan (CH_4S)
n-Alcohols	H_3C-OH	20	Methanol, Ethanol, 1-Octanol
n-Aliphatic Primary Amines	 $H_3C-CH_2-NH_2$	13	Ethyl Amine
Nitriles	$HC\equiv N$	29	Acetonitrile (C_2H_3N)
Nitroamines		5	o-Nitroaniline ($C_6H_6N_2O_2$)
Nitro-Compounds		21	Nitromethane (CH_3NO_2)
Organic Salts	No Common Structure	14	Ethylene Carbonate ($C_3H_4O_3$)
Peroxides	$HO-OH$	10	Benzoyl Peroxide ($C_{14}H_{10}O_4$)
Polyols		33	Glycerol

Sulfides/Thiophenes		21	Diethyl Sulfide ($C_4H_{10}S$)
Terpenes		6	Terpinolene ($C_{10}H_{16}$)
Other Aliphatic Acids	No Common Structure	19	Isobutyric Acid ($C_4H_8O_2$)
Other Aliphatic Alcohols	No Common Structure	26	Isobutanol
Other Aliphatic Amines	No Common Structure	19	Dimethylamine
Other Amines/Imines	No Common Structure	36	Pyrazine ($C_4H_4N_2$)
Other Ethers/Diethers	No Common Structure	17	Anethole ($C_{10}H_{12}O$)
Other Saturated Aliphatic Esters	No Common Structure	19	Capralactone ($C_6H_{10}O_2$)
Miscellaneous Other	No Common Structure	34	Methyl Hydrazine (CH_6N_2)

3.4 Sigma-Profile Comparison

We compare VT-2004 sigma profiles with published profiles for four compounds.[5] Figure 3 and Figure 4 compare the sigma profiles for water, hexane, acetone, and 1-octanol, scaled relative to their maximum values. We find an average difference of less than 2% between the VT-2004 and Lin and Sandler's sigma profiles.[5] We use the same calculation settings based on the density-functional theory. Therefore, we attribute the difference in sigma profiles either to variations in the DMol versions, in

Cerius and in Materials Studio (used here), or the difference in the sigma-averaging algorithm, as discussed in Section 2.3.

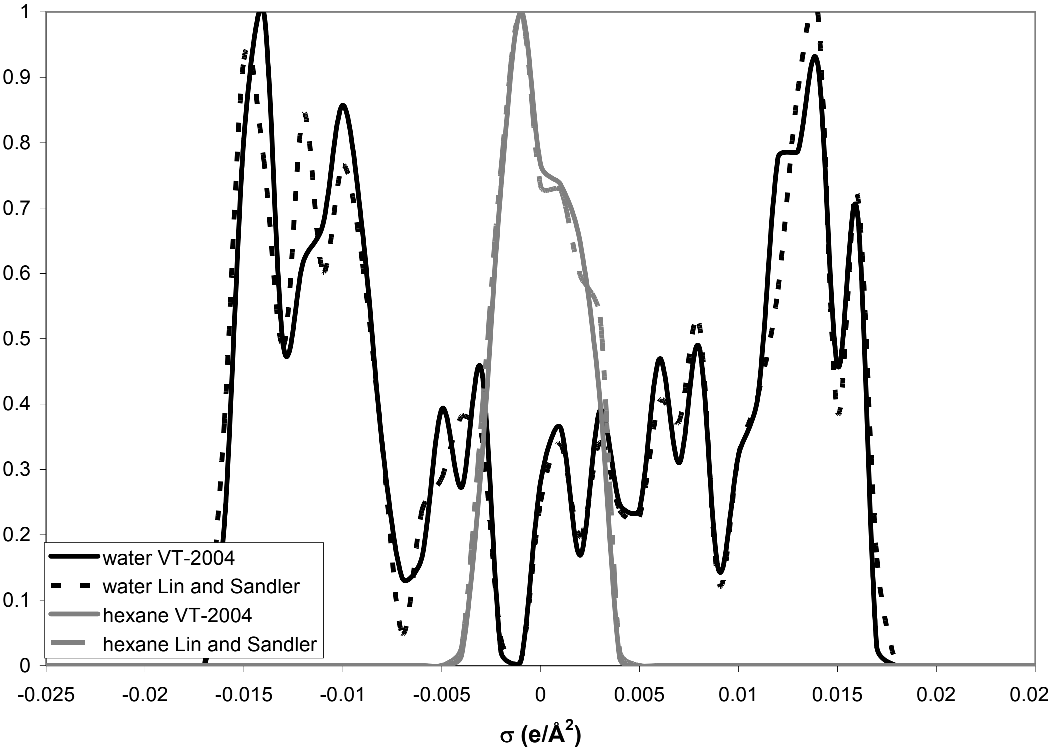


Figure 3. Scaled sigma-profile comparisons between VT-2004 and Lin and Sandler[5] for water and n-hexane. The solid curves represent VT-2004 profiles and dashed curves show the Lin and Sandler sigma profiles.

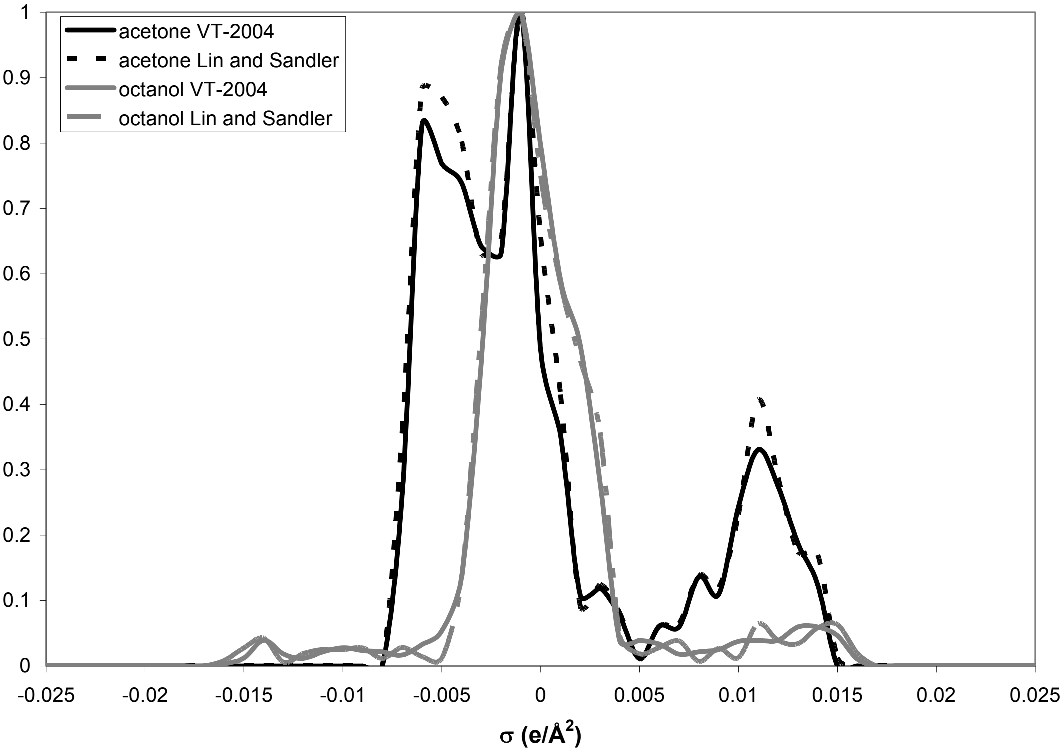


Figure 4. Scaled sigma profile comparisons between VT-2004 and Lin and Sandler[5] for acetone and 1-octanol. The solid curves represent VT-2004 profiles and dashed curves show the Lin and Sandler sigma profiles.

3.5 Conformational Variations

Molecules may exist in several low-energy structural conformations due to freedom in choosing dihedral angles. Each conformation results in a slightly different sigma profile and may affect property predictions. The VT-2004 sigma-profile database contains sigma profiles for one low-energy conformation for each molecule.

We examine the database in order to identify problematic structures. For some larger molecules, we investigate several conformations and utilize only the lowest-energy conformation in the database.

We provide four examples to quantify the effect of neglecting additional conformations on property predictions. These examples include comparisons of sigma profiles, and VLE and SLE predictions for four different sized molecules.

3.5.1 Small Molecules

First we examine the system nitroethane(1)/1-methoxy-2-propanol(2) reported in Klamt et al[28]. These authors run rigorous molecular dynamic calculations and identify two stable conformations for 1-methoxy-2-propanol; we compare three conformations depicted in Figure 5. Conformation A includes an internal hydrogen bond between the hydroxyl-hydrogen and the methoxy-oxygen atoms.

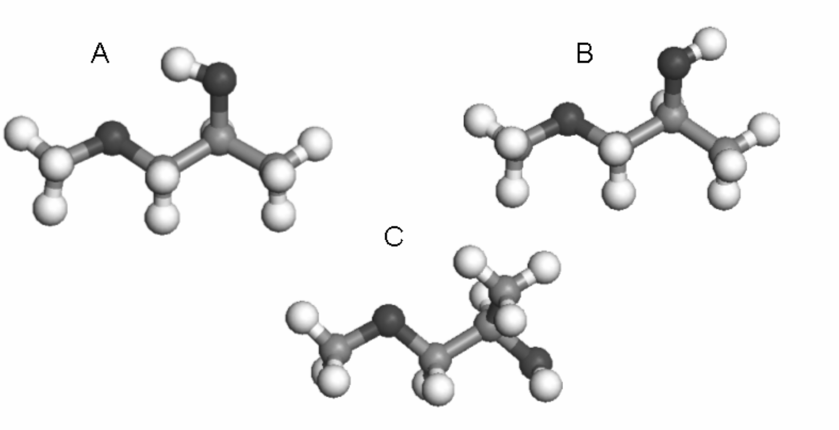


Figure 5. Three low-energy conformations for 1-methoxy-2-propanol.

Figure 6 shows the sigma profiles for all three conformations (the VT-2004 database contains Conformation A). We observe small differences between the sigma profiles, most notably in the external hydrogen-bonding region -0.015 ≤ sigma ≤ -0.010.

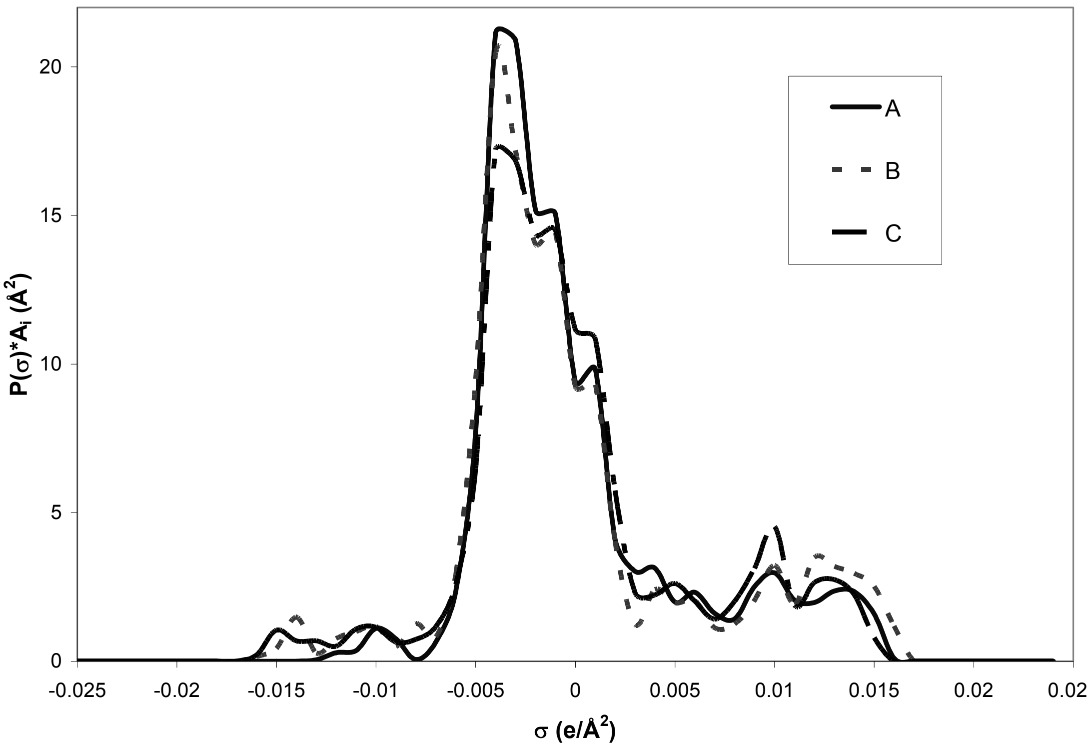


Figure 6. Sigma profile comparison between two conformations of 1-methoxy-2-propanol.

Figure 7 shows pressure-composition predictions for nitroethane(1)/1-methoxy-2-propanol(2), similar to Klamt et al, for 40^o and 80^o C.[28] We find minimal conformational effect on the VLE predictions for 40^{o}C, but an increased difference at 80^o C. The VT-2004 predictions show an average system-pressure error of 3.99% and the reported COSMO-RS predictions, which account for two combined conformations, show an error of 5%. We use pure-component vapor pressures from Aspen Plus database, version 12 (Aspen Technology, Inc.), while Klamt et al use COSMO-RS *a priori* vapor pressures.[28] The two predict the system behavior to the same level of accuracy, but the COSMO-SAC-VT-2004 predictions consider only conformation A, while Klamt et al account for two conformations.

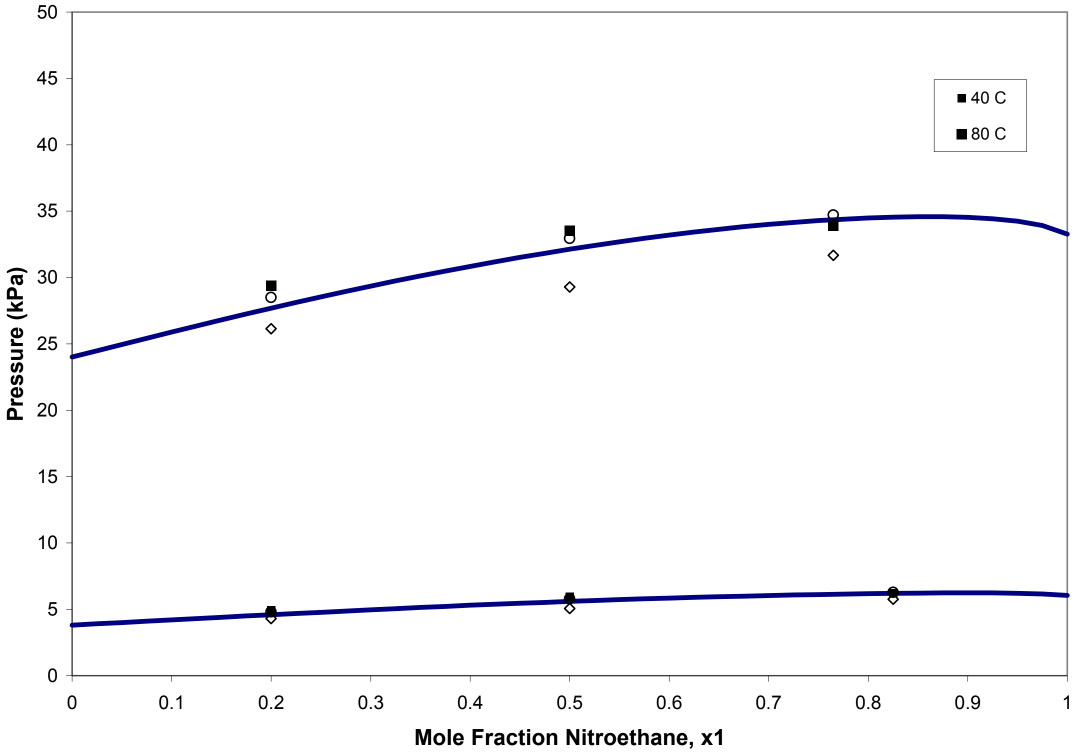


Figure 7. VLE predictions for nitroethane(1)/1-methoxy-2-propanol(2) at 313 and 353 K. The solid line shows the COSMO-SAC-VT-2004 prediction (conformation B), diamonds are conformation A, circles show conformation C, and the filled blocks represent experimental data.[28]

For small molecules, conformational differences appear to have only a negligible effect on sigma profiles and property predictions.

3.5.2 Medium-Sized Molecules

Now, we examine conformational effects on solvation calculations with medium-sized molecules in the VT-2004 database. These compounds contain more single bonds and allow more rotations, thus more conformations.

We identify three low energy states for benzyl benzoate (CAS-RN 120-51-4) and predict pressure-composition behavior using these conformations. Figure 8 shows the three conformations and Figure 9 depicts the resulting sigma profiles. The sigma profiles exhibit larger differences here, when compared to the small molecule, specifically in the region $0.000 \leq \text{sigma} \leq 0.010$.

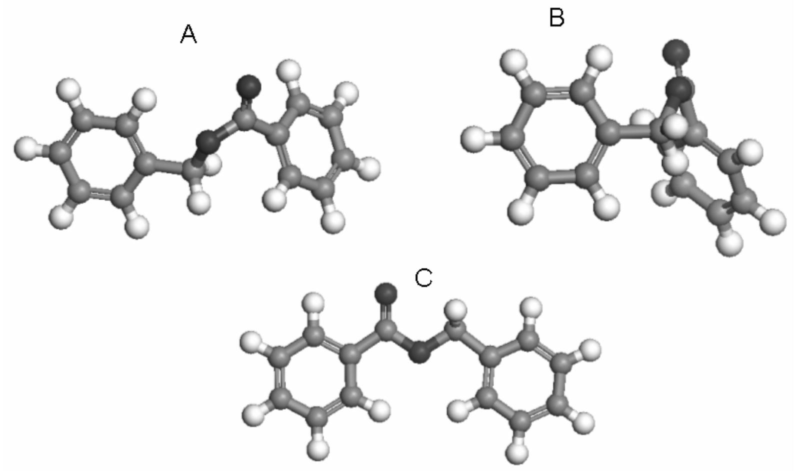


Figure 8. Three low-energy conformations for benzyl benzoate. Conformation C is the VT-2004 conformation.

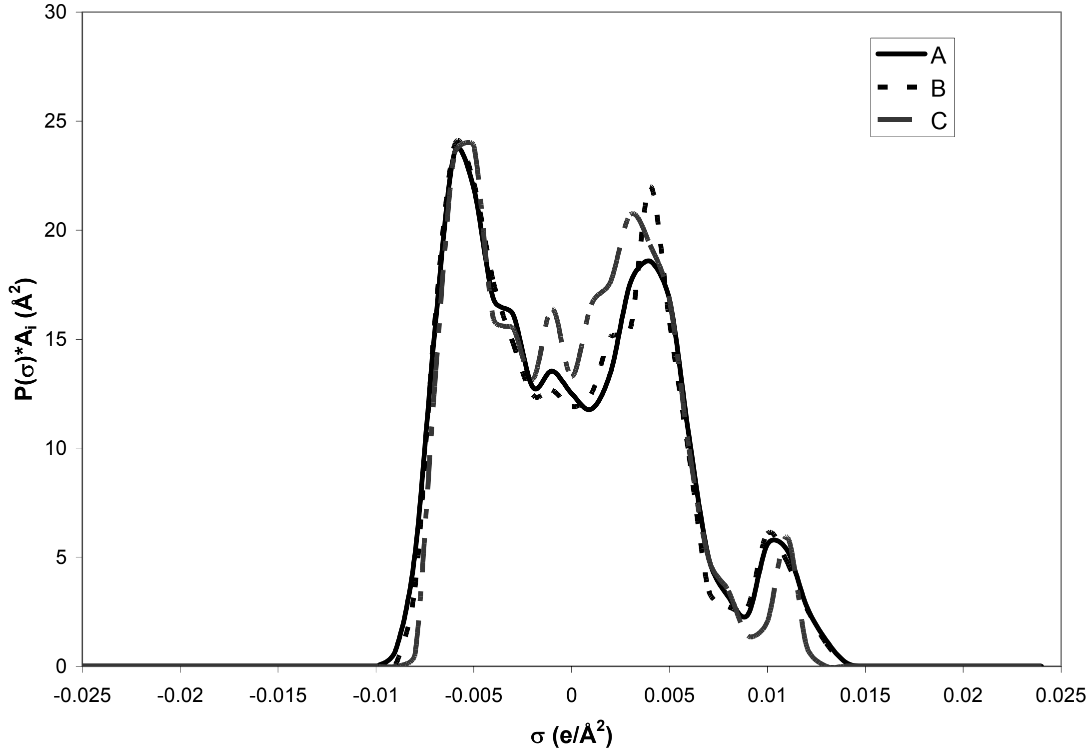


Figure 9. Sigma profile comparison for three conformations of benzy benzoate.

Figure 10 shows COSMO-SAC-VT-2004 predictions for the benzene(1)/benzyl benzoate(2) system at 453.25 K.[31] We observe more conformational sigma-profile variation in medium molecules than we observed in the small molecules. The VT-2004 predictions agree with the experimental data with an average RMS error (lnP) of 0.029 (0.034, 0.042, and 0.011 respectively) and with some variation between the three conformations. This system exhibits almost ideal liquid-phase behavior, but non-ideal systems may show more conformational variations. We suspect these variations would be more pronounced in a system that exhibits more non-idealities than the system we show here.

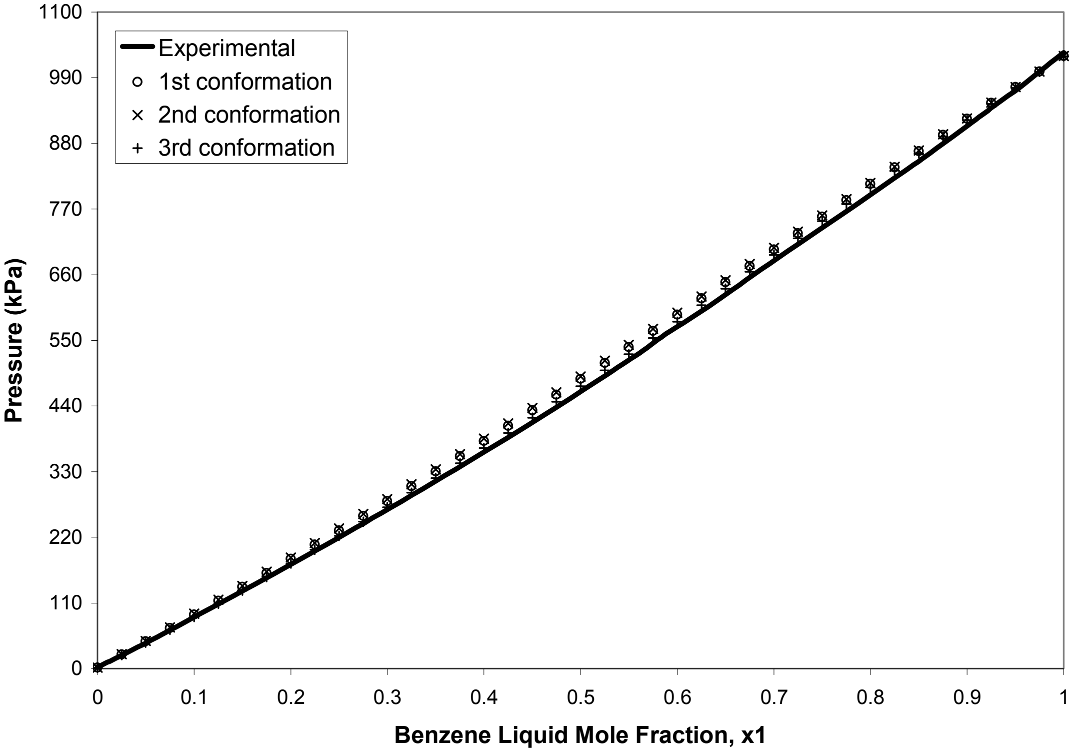


Figure 10. Pressure-composition COSMO-SAC predictions (symbols) for three benzyl benzoate conformations in the benzene(1)/benzyl benzoate(2) system at T = 453.25 K. The solid line represents experimental data.[31]

3.5.3 Large Molecules

The VT-2004 sigma-profile database includes many large molecules, which could occur in multiple conformations. We compare sigma profiles and solubility predictions for three conformations of ibuprofen and lidocaine. Figure 11 and Figure 12 illustrate the low-energy conformations we use for ibuprofen and lidocaine, respectively.

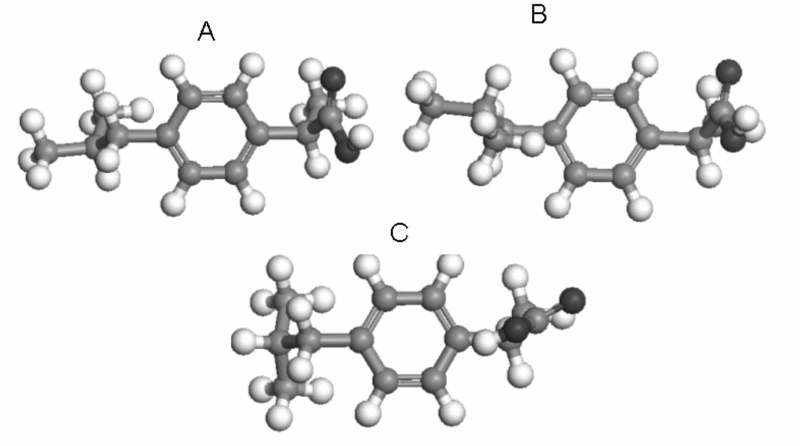


Figure 11. Three low-energy conformations for ibuprofen.

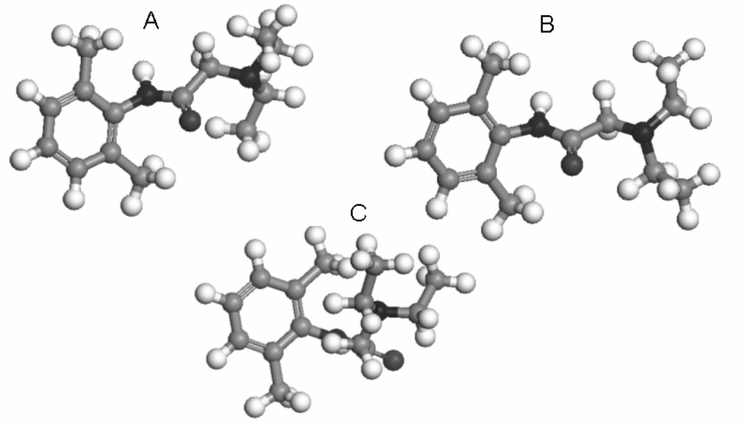


Figure 12. Three low-energy conformations for lidocaine.

Figure 13 and Figure 14 show the sigma profiles for ibuprofen and lidocaine, respectively. We observe small conformational effects on the sigma profiles, similar to those for smaller molecules.

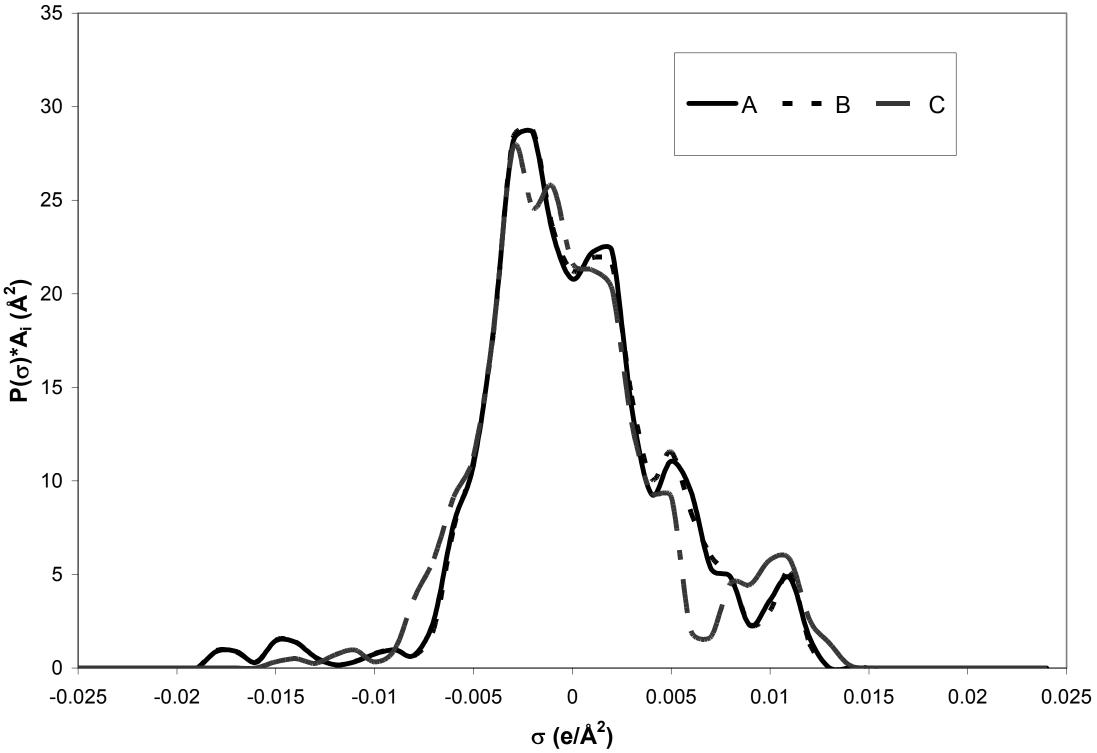


Figure 13. Sigma profile comparison for three ibuprofen conformations.

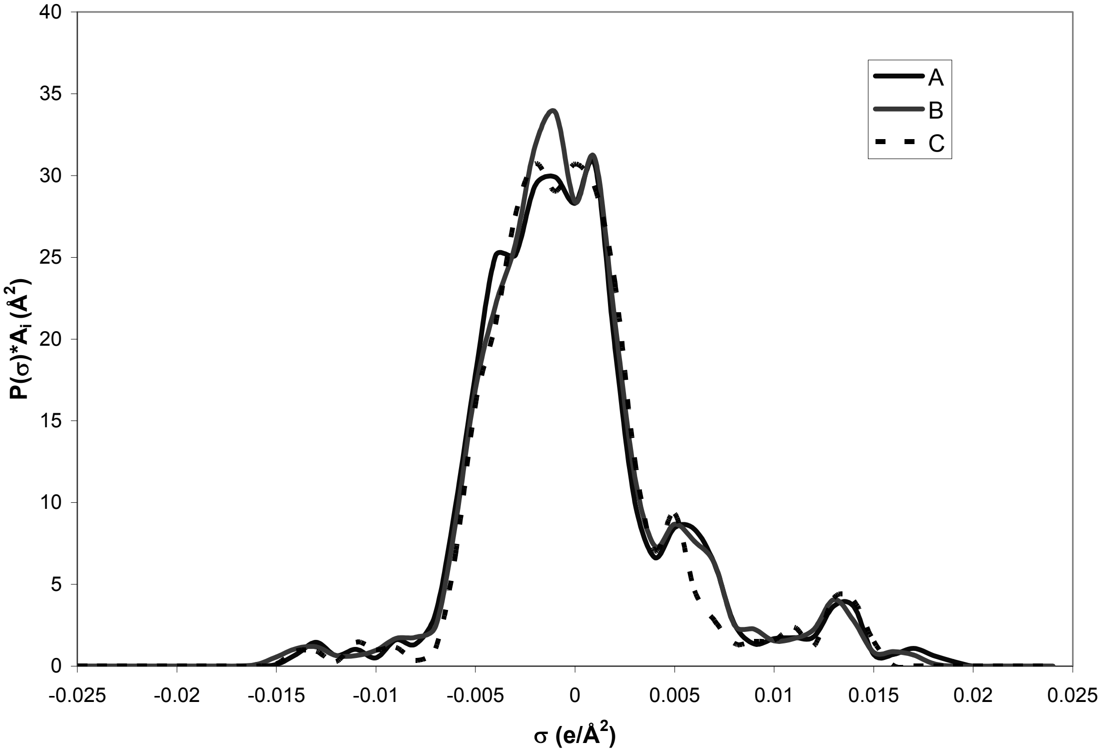


Figure 14. Sigma profile comparison for three lidocaine conformations.

Figure 15 and Figure 16 show solubility predictions for ibuprofen and lidocaine in various solvents. For clarity of both figures, we do not include the solvent names, since our focus here is on conformational variations and the actual solvent names are less important. The predictions for the ibuprofen system are very poor (RMS error of 2.69, 2.70, and 2.18 with respect to $\ln x_{sol}$, respectively, and the lidocaine predictions are much better (RMS 0.42, 0.37, 0.36), but both figures show some differences in solubility predictions among the three conformations. The large-molecule conformational effect appears more significant than that shown for smaller molecules. Klamt et al[29] note the effect of conformational variations on aqueous solubilities.

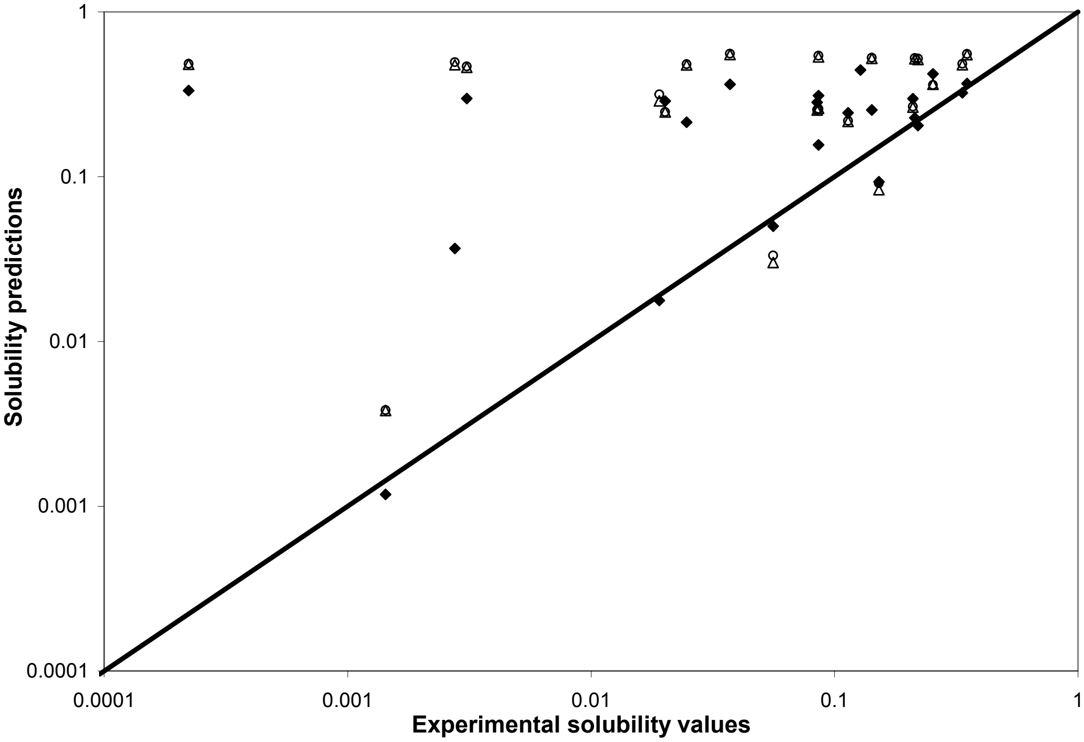


Figure 15. COSMO-SAC solubility predictions for three ibuprofen conformations in several solvents (A = triangles, B = circles, C = diamonds).

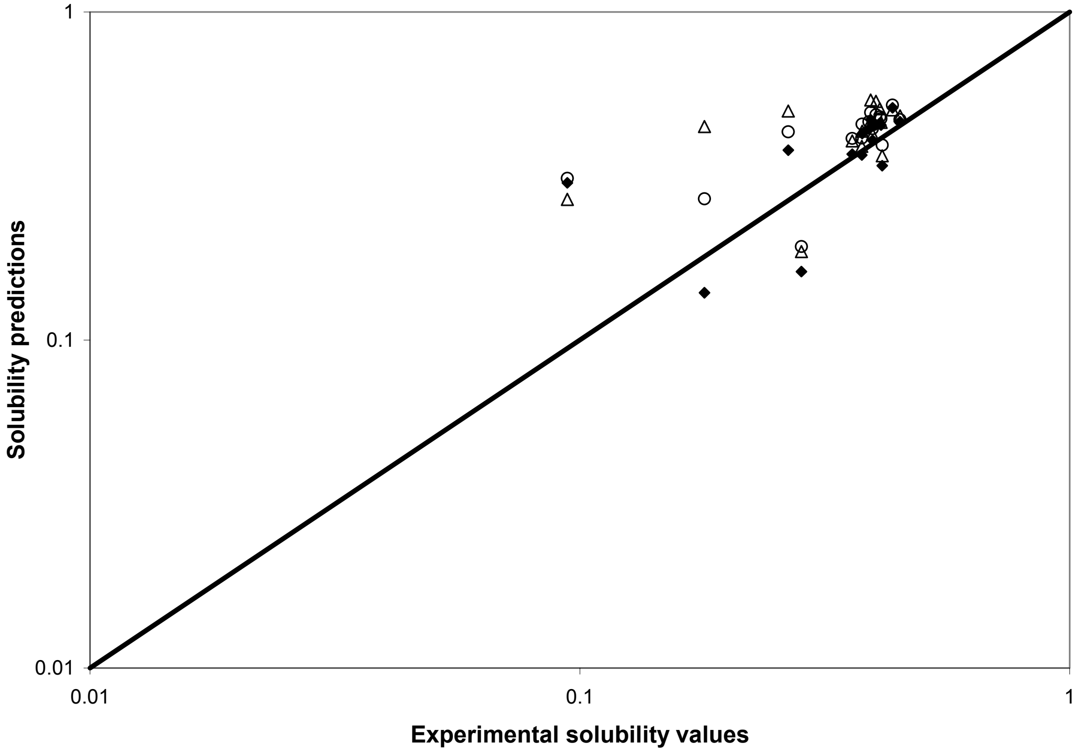


Figure 16. COSMO-SAC solubility predictions for three lidocaine conformations in several solvents (A = triangles, B = circles, C = diamonds).

Conformations appear to influence sigma profiles more so than VLE predictions for small and medium-sized molecules. This effect seems more significant in solubility predictions for large molecules, but the reliability of these predictions is not well documented and requires more work before they are of much use.

The effect of including multiple conformations does not appear to follow any predictable trends. While we see some variations in the property predictions and sigma profiles more work is required to quantify these effects and this is outside the scope of this thesis. Here we recognize some variations and alert our users that the VT-2004 sigma-profile database contains a single sigma profile, representing one conformation, for each molecule.

The COSMO-based thermodynamic methods are very useful in preliminary design and evaluation. The conformational variance is small and does not show an appreciable effect on preliminary design calculations. We should consider multiple conformations if we require a much higher accuracy level.

3.6 Resources

We provide, via http://www.design.che.vt.edu/VT-2004.htm, additional open-literature information as outlined below:

- The complete VT-2004 sigma-profile database of 1,513 compounds.
- Index of the VT-2004 database including CAS-RN, chemical formula, and compound name.
- Procedure for generating sigma profiles using DMol software by Accelrys Material Studio (including screen captures).
- VT-2004 sigma-profile averaging FORTRAN program, which takes the COSMO output from DMol and calculates the sigma profile.
- COSMO-SAC-VT-2004 FORTRAN program to do COSMO-SAC calculations for any binary mixture using either VT-2004 sigma profiles or new sigma profiles.
- Procedure for using the COSMO-SAC-VT-2004 program for predicting activity coefficients.
- COSMO output for each molecule for use in the averaging algorithm.

Appendix C illustrates the information, included in our database, for 1,4-dioxane that we use in our example in the following section.

4 Validation Examples

4.1 Introduction

We validate the VT-2004 sigma profiles and our implementation of the COSMO-SAC model. We compare predictions of activity coefficients, vapor-liquid equilibria and solubilities with published data.

4.2 Activity-Coefficient Predictions

Lin and Sandler[5] present activity-coefficient predictions for two binary mixtures, water/1,4-dioxane and methyl-acetate/water, using their COSMO-SAC model. Figure 17 shows the predicted activity coefficients for the water(1)/1,4-dioxane(2) system. We find very good agreement between the original COSMO-SAC and the VT-2004 predictions for this system. The VT-2004 predictions deviate from the Lin and Sandler data by less than 3% on average. The root-mean-square (RMS) errors are 0.0694 and 0.1374 for $\ln\gamma_1$ and $\ln\gamma_2$, respectively, for this system. We calculate the RMS error with

$$RMS = \sqrt{\frac{1}{n}\sum_{n}\left(\ln\gamma_{VT-2004} - \ln\gamma_{COSMO-SAC}\right)^2} \tag{16}$$

where n is the number of data points, $\ln\gamma_{\text{VT-2004}}$ are our current predictions, and $\ln\gamma_{\text{COSMO-SAC}}$ are the published predictions.[5]

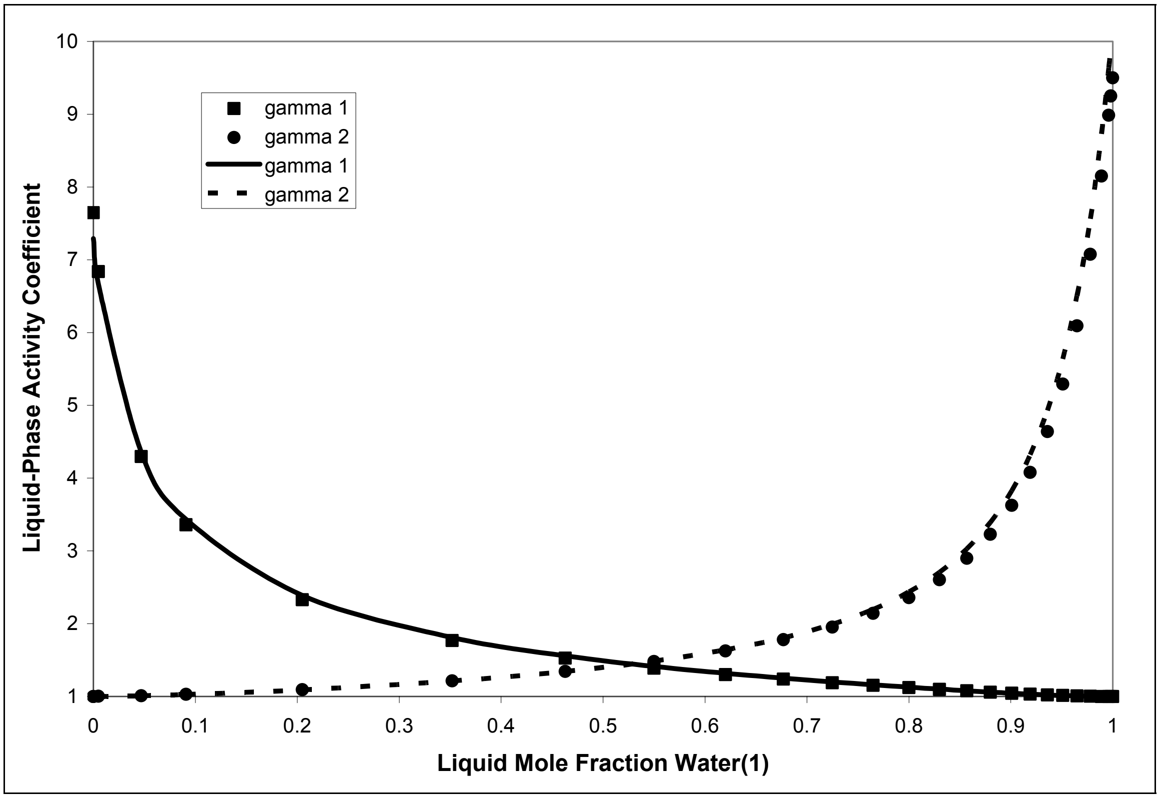


Figure 17. COSMO-SAC activity-coefficient predictions from published results (Lin and Sandler[5]) and VT-2004 for the water(1)/1,4-dioxane(2) system at T=308.15 K. The curves represent the VT-2004 predictions, and the squares and triangles show the Lin and Sandler data.

Figure 18 compares the VT-2004 predictions and the published Lin and Sandler data[5] for the methyl-acetate(1)/water(2) system at 330.15 K. The VT-2004 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 is as we approach infinite dilution with respect to methyl-acetate in water.

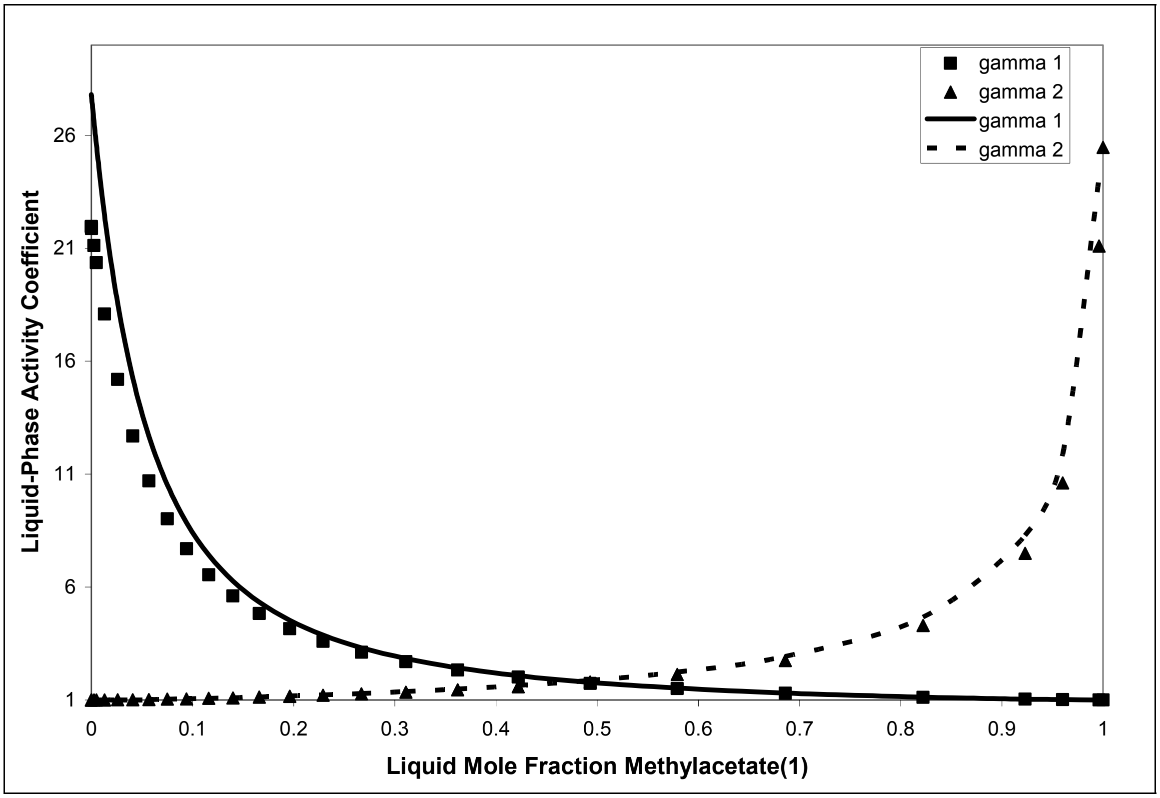
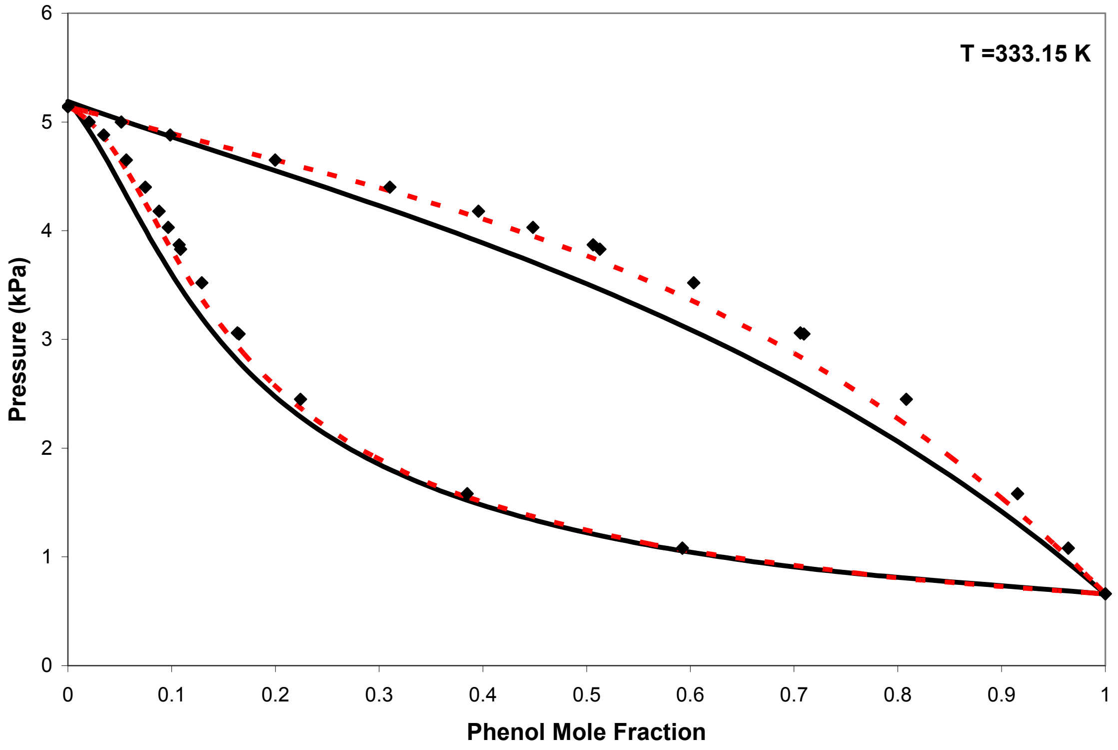


Figure 18. COSMO-SAC activity-coefficient predictions from published results (Lin and Sandler[5] shown as triangles and squares) and VT-2004 (curves) for the methyl-acetate(1)/water(2) system at T=330.15 K.

We now predict vapor-liquid equilibria for four binary mixtures. Figures 7 through 10 compare our predictions with experimental data and published COSMO-RS predictions.[11,23]



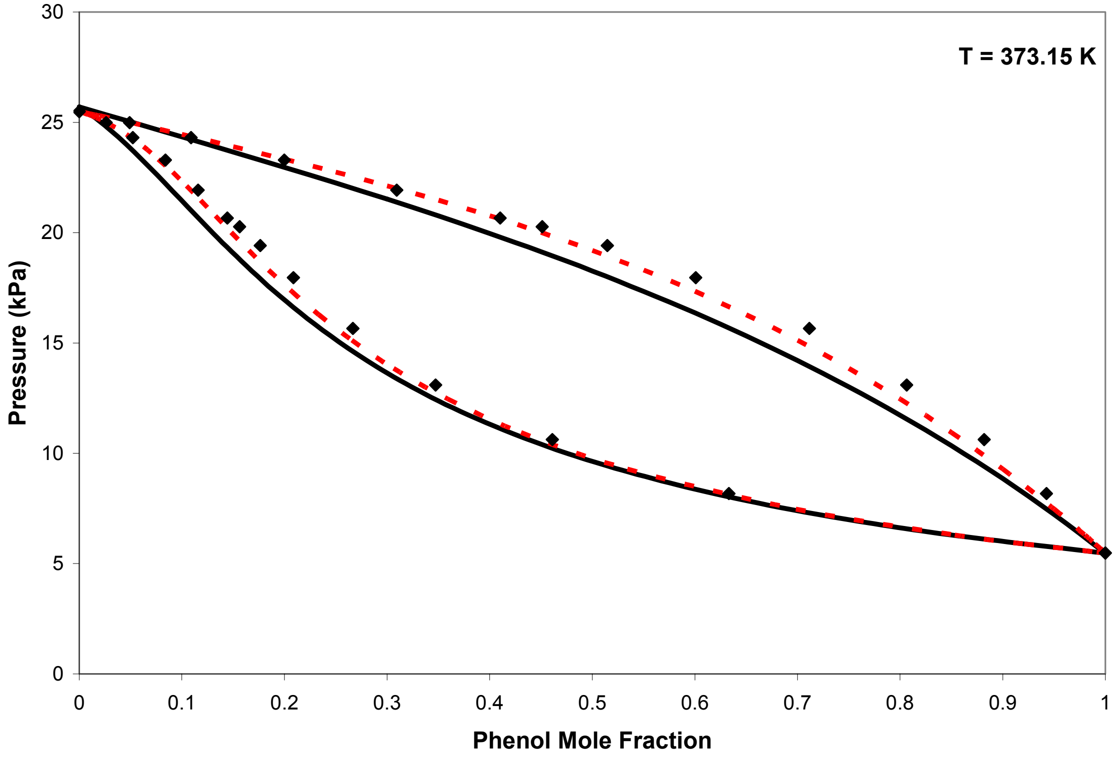


Figure 19. Vapor-Liquid Equilibrium for the phenol(1)/styrene(2) system for 333.15 K and 373.15 K. The solid curves show COSMO-SAC-VT-2004 predictions, dashed curves display COSMO-RS predictions, and symbols represent experimental data.[11,23]

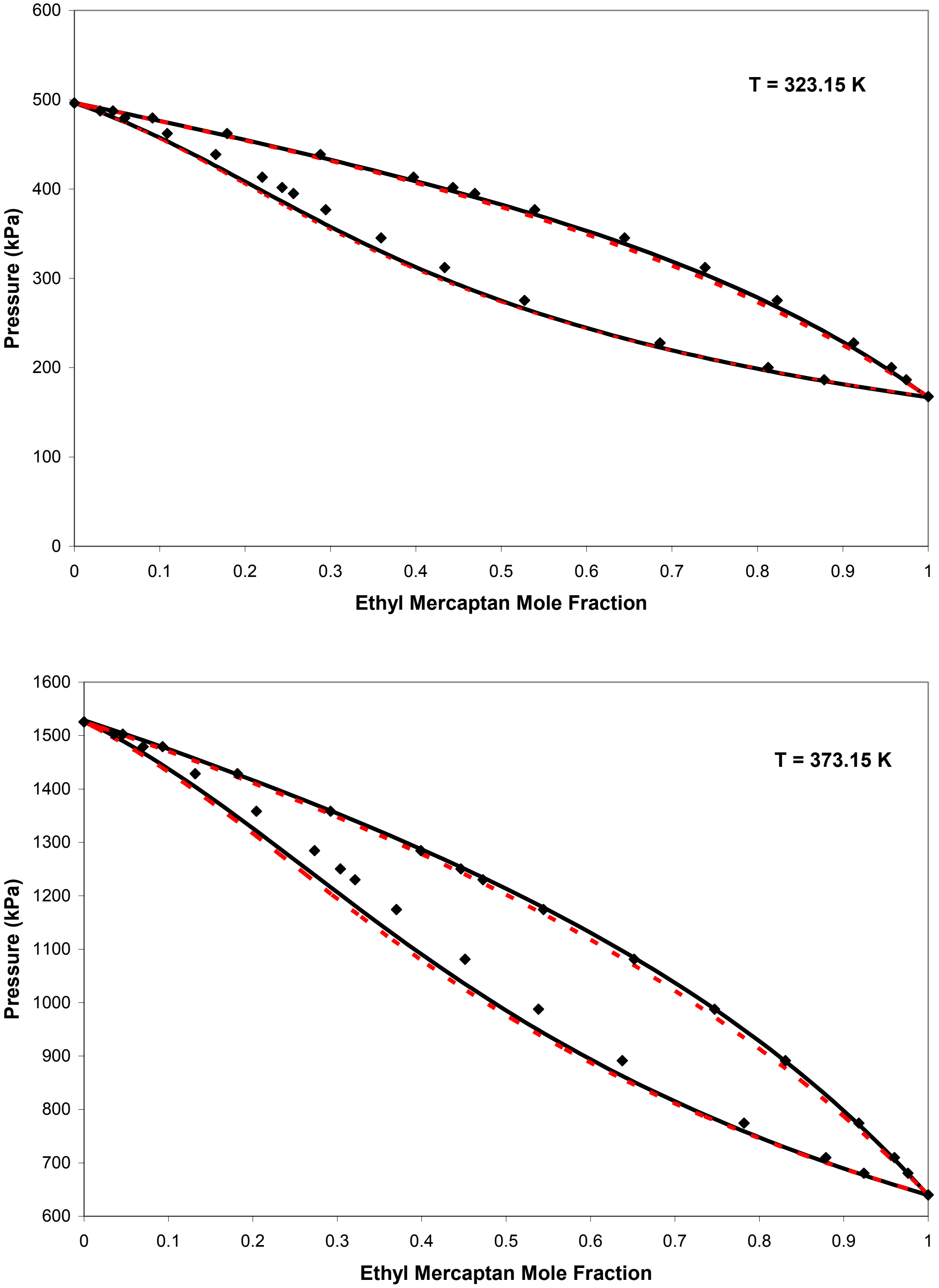


Figure 20. Vapor-Liquid Equilibrium for ethyl mercaptan(1)/n-butane(2) at 323.15 K and 373.15 K. The solid curves show COSMO-SAC-VT-2004 predictions, dashed curves display COSMO-RS predictions, and symbols represent experimental data.[11,23]

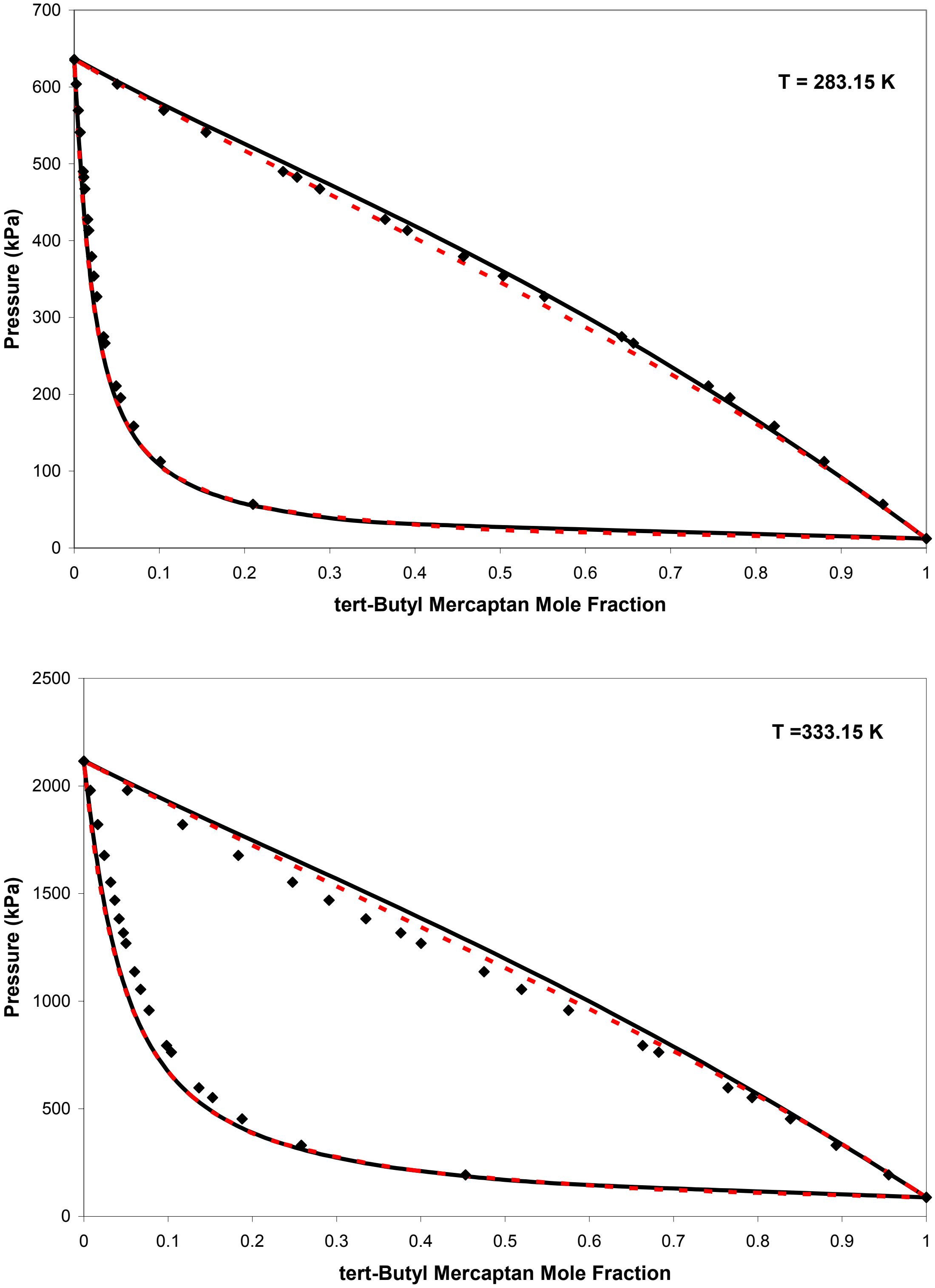
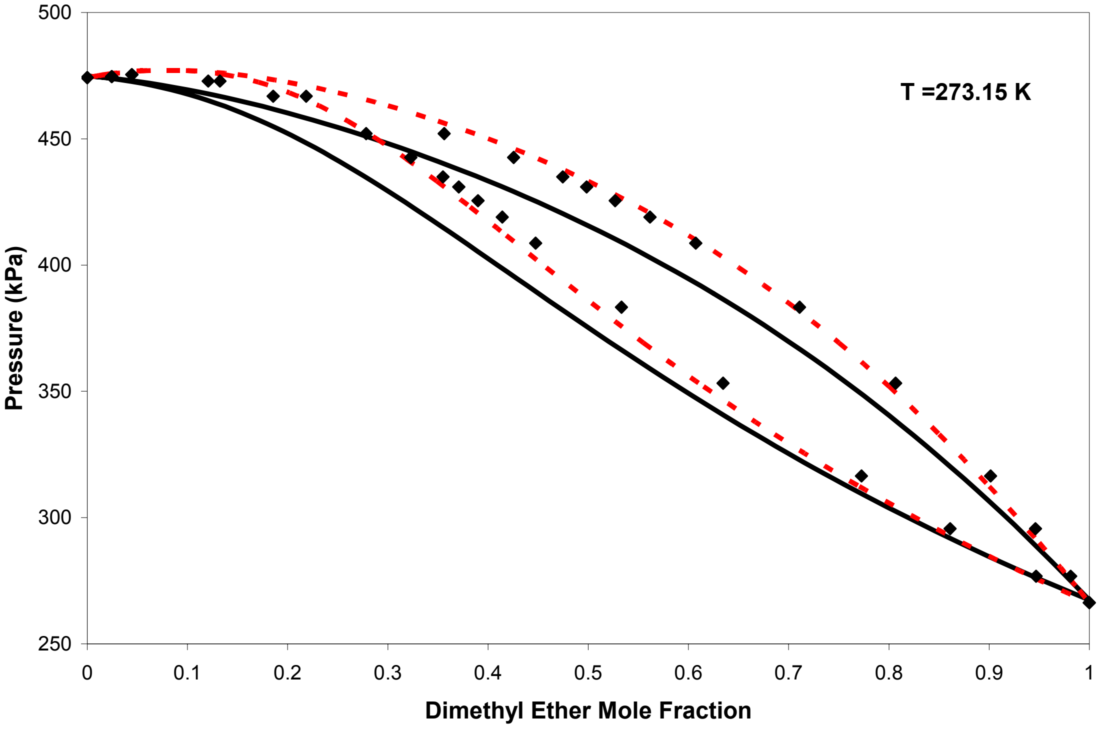


Figure 21. Vapor-Liquid Equilibrium for tert-butyl mercaptan(1)/propane(2) at 283.15 K and 333.15 K. The solid curves show COSMO-SAC-VT-2004 predictions, dashed curves display COSMO-RS predictions, and symbols represent experimental data.[11,23]



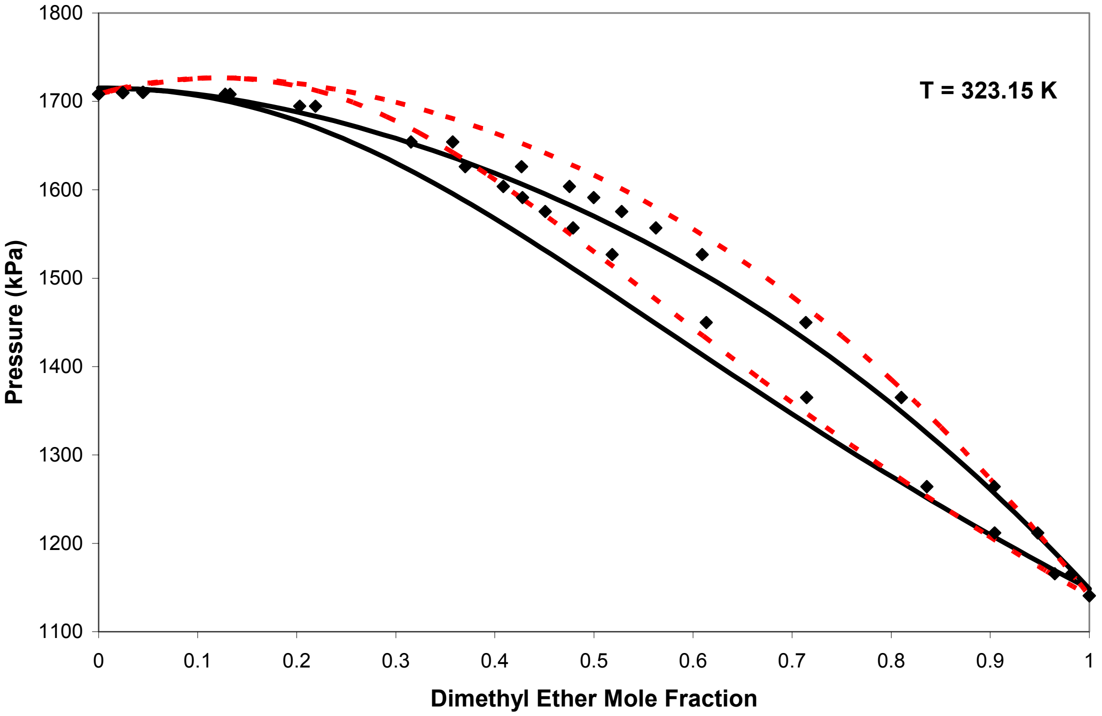


Figure 22. Vapor-Liquid Equilbrium for dimethyl ether(1)/propane(2) at 273.15 K and 323.15 K. The solid curves show COSMO-SAC-VT-2004 predictions, dashed curves show COSMO-RS predictions, and symbols represent experimental data.[11,23]

Figure 19 shows the VLE data for phenol(1)/styrene(2) at T= 333.15 K 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, but this discrepancy is less apparent at higher temperatures.

Figure 20 illustrates the VLE behavior for the ethyl mercaptan(1)/n-butane(2) system at T = 323.15 K and 373.15 K. Again, both models accurately predict the deviation from Raoult's law.

Figure 21 shows the VLE predictions for the tert-butyl mercaptan(1)/n-propane(2) system at T = 283.15 K 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[11] report that the UNIFAC model fails for this system because it lacks appropriate functional group parameters.

Figure 22 shows the dimethyl-ether(1)/n-propane(2) system at T = 273.15 K and 323.15 K. Both COSMO-SAC and COSMO-RS predict the minimum-boiling azeotropes in the dimethyl-ether-rich region for both temperatures. 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 7-10, the COSMO-SAC-VT-2004 predictions match reasonably well with experimental pressure and composition data. The COSMO-SAC-VT-2004 predictions for $\ln\gamma_1$ and $\ln\gamma_2$ agree with the reported calculated γ 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(1)/styrene(2) system.

4.3 Solubility Predictions

Equation (17) 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:[24]

$$\ln x_i = \frac{\Delta_{fus} S}{R}\left(1-\frac{T_m}{T}\right) - \ln \gamma_i^{sat} = \frac{\Delta_{fus} H}{RT_m}\left(1-\frac{T_m}{T}\right) - \ln \gamma_i^{sat} \tag{17}$$

for $T \leq T_m$. We predict the activity coefficient of the solute using the COSMO-SAC model and published pure solute properties (heat of fusion and melting point).[25]

Figure 23 compares COSMO-SAC-VT-2004 solubility predictions with published experimental values (solute mole fractions 0.0059 – 0.3348) for benzoic acid in nine solvents at 25 °C.[25] The VT-2004 predictions qualitatively agree with experimental data with a RMS error of 0.7092 for $\ln x_i$. Kolar et al.[26] report interesting results in using COSMO-RS for solubility calculations. Their COSMO-RS predictions for benzoic acid are consistent with the results shown here.

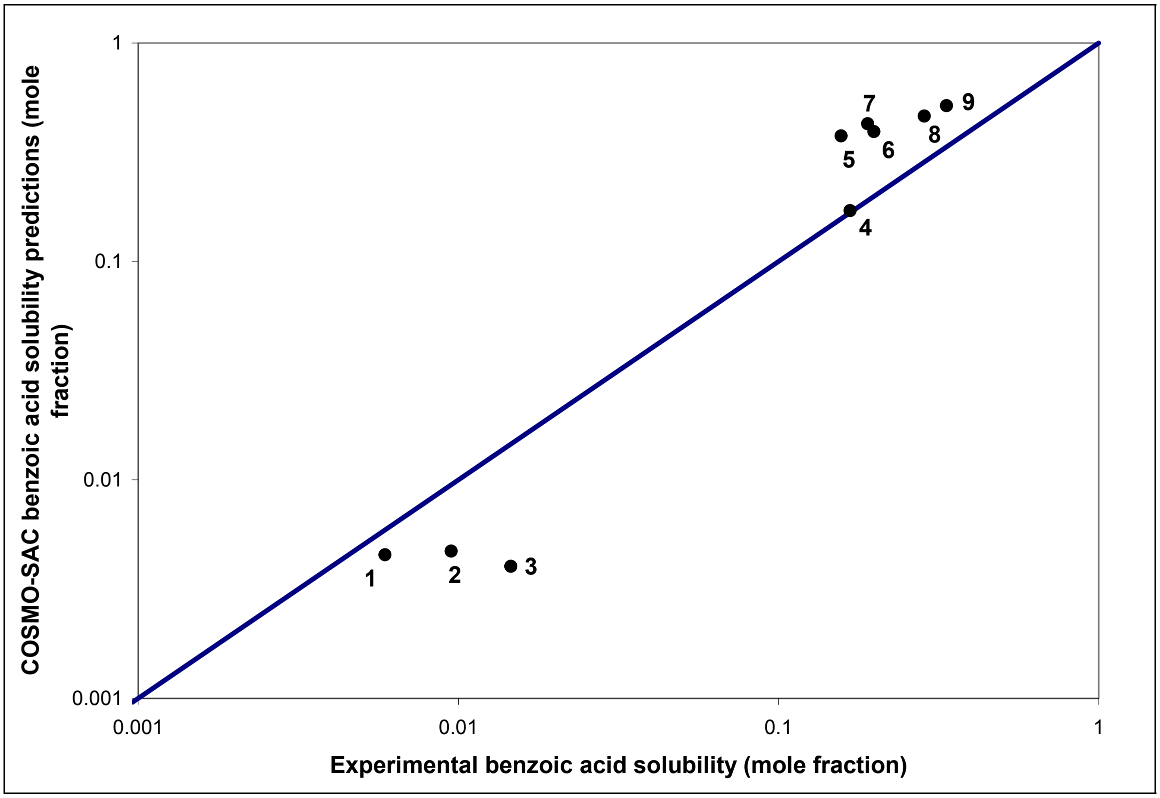


Figure 23. Benzoic acid solubility for various solvents at 298.15 K. Solvents 1-9 are pentane, n-hexane, cyclohexane, acetic acid, methanol, 1-octanol, 1-hexanol, 1,4-dioxane, and THF, respectively.

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. We review a number of reported experiences from the literature up to October 2004 to guide the reader with COSMO model applications. Our reference to COSMO-RS refers only to cited literature observations. Many groups are doing research on the methods and should broaden its applicability and improve its accuracy in the future.

1. COSMO-SAC and COSMO-RS predictions are sensitive to the sigma profiles. Sigma profiles used in the calculation should be consistent, i.e, generated with the same procedures and algorithms. We find some systematic differences, in the energy calculations, between density-functional theory calculations in DMol and Jaguar (recent work with Dr. Sandler)[30] that could lead to different sigma profiles. Each computational program uses a slightly different "zero energy" value and it is not accurate to compare finite energy calculations, but more appropriate to compare energy differences. These variations are small between the different software programs and should lead to only small variations in property predictions.
2. Avoid COSMO with amines. COSMO calculations, in general, show poor results for amines especially trialkylamines, because of the unpaired electron in the cavity.[5] Klamt[1], Klamt et al.[2], and Putnam et al.[12] report problems in applying COSMO calculations to highly polar groups with small surface area like amines.
3. Avoid COSMO with polymers. COSMO gives poor results for polymer-system thermodynamics ("…its inability to properly account for the thermodynamics of

polymer systems except for some rather limited cases" page 1495).[7] The literature as of October 2004 shows very little work with applying COSMO to polymer solutions. We believe this may result from the fact that COSMO-RS and COSMO–SAC neglect void space and intramolecular interactions within the system.

4. Avoid COSMO-SAC with mixtures of chloroform and ketones or alcohols. Lin and Sandler[5] report difficulties in accurately predicting interactions between chloroform and either ketones or alcohols. They suggest that "it is necessary to refine the COSMO calculation to provide a better σ-profile for chloroform." This is identical to the behavior that Klamt[1] and Klamt et al.[2] report regarding largely polar groups similar to nitriles and carbonyl groups.
5. 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.[5] Klamt and others optimize the COSMO-RS parameters with water/alkane solutions, while Lin and Sandler simply use the values from COSMO-RS without an independent optimization for use with COSMO-SAC.[1,5,6]
6. Avoid COSMO-SAC for highly hydrophobic solutes. Accuracy for COSMO-SAC predictions decreases with increasing hydrophobicity.[5]
7. Use COSMO predictions for molecules containing only H, C, N, O, F, P, S, Cl, Br, and I. Lin and Sandler report "It is worth mentioning that because other atomic radii have not been determined, the COSMO-SAC model is currently limited to compounds containing atoms of H, C, O, N, or Cl.".[5] Klamt et al.[2] optimize these parameters and suggest 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).[3] **Note:** We currently input atomic radii for H, C, N, O, F, P, S, Cl, Br, and I. Atomic radii for P via Klamt and Eckert, private communication. See Table 2.

8. Use COSMO-SAC or COSMO-RS for property predictions of systems with chemicals for which UNIFAC cannot accurately handle, for example, benzene/n-methylformamide.[5] Kolar et al.[26] show that UNIFAC only works for 82% of the 221 systems they studied.
9. Use COSMO-SAC and COSMO-RS for predicting activity coefficients, but avoid using these methods for predicting vapor pressures of pure components. Eckert and Klamt[11] find significant errors in the COSMO-RS vapor-pressure predictions. The vapor-pressure predictions are sufficient for initial guesses, but show significant error and should be avoided in detailed design work if experimental values are present. Several groups are working to improve vapor-pressure predictions.
10. 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.[5,6,3,11,26,27]

6 Conclusions

The dielectric-continuum COSMO-based thermodynamic models represent a new approach to a priori property predictions. These models represent a departure from traditional group contribution methods (UNIFAC) and empirical data-regressed approaches (NRTL) and allow engineers and scientist to accurately predict thermophysical behavior based on molecular structure and interactions. The COSMO-based models use a surface-charge density distribution, σ-profile, to characterize liquid-phase nonidealities.

This work presents the first open-literature sigma-profile database consisting of 1513 compounds. Each of these sigma profiles are calculated with the DNP v4.0.0 basis set and GGA/VWN-BP functional. This sigma-profile database reduces the time required to begin working with either of the COSMO-based thermodynamic models by almost 90% and allows groups to improve the existing models.

We validate our sigma profiles against four published sigma profiles for water, acetone, hexane, and 1-octanol and observe a difference less than 2%. We use these sigma profiles and the COSMO-SAC model to predict liquid-phase activity coefficients within 3% and 8% difference (RMS errors of 0.0478 to 0.1374, for $\ln\gamma$) of the published predictions for the water(1)/1,4-dioxane(2) and methyl-acetate(1)/water(2) systems, respectively.[5] This database, along with the COSMO-SAC-VT-2004 FORTRAN code, enables the user to further apply and advance this novel approach to predict phase-equilibrium behavior for process and product development.

7 Future Work

The COSMO-based thermodynamic models are relatively novel and still poses much room for improvement. The models need a standard averaging radius in the sigma-averaging algorithm. This value directly affects the sigma profile and thus the model accuracy. Future work could include using a large test case of chemicals, with varying chemical properties and molecular sizes, and optimizing the averaging radius based on property predictions for this test set. The VT-2004 database provides the starting point for this analysis in the form of calculation results from the density-functional theory.

The literature contains several examples of predicting VLE with both COSMO-RS and COSMO-SAC and comparing these predictions with UNIFAC. It would be very helpful to begin to explore other property predictions (i.e., heat of mixing, SLE, pure-component vapor pressure, etc.). A group from the University of Delaware seems to be making progress towards improving vapor-pressure predictions using VT-2004 sigma profiles.

Finally, the density-functional theory calculations for each sigma profile are dependent on the molecular conformation and show some variations according to various conformations. The VT-2004 sigma-profile database contains only a single conformation for each compound, but other low-energy conformations may exist and would have some effect on property predictions. This effect needs to be understood more fully to extend these models to predicting pharmaceutical properties and aiding in solvent selection. We include a small subsection to show the existence of these variations, but it would be interesting to see if any trends exist in regards to functional groups, molecular sizes, and conformational variations in sigma profiles.

8 Nomenclature

English Symbols

a_{eff} = effective surface area of a standard surface segment, 7.5 Å^2
A = Coulomb interaction matrix
A_i = total cavity surface area, Å^2
$A_i(\sigma_m)$ = surface area of all segments with surface charge density σ_m, Å^2
au = atomic unit; Bohr radius, 5.2918E-11 m
c_{hb} = a constant for hydrogen-bonding, Kcal/mole*Å^4/e^2
COSMO = conductor-like screening model
COSMO-RS = conductor-like screening model for real solvents
COSMO-SAC = conductor-like screening model with segment activity coefficients
d_{mn} = distance between surface segment m and n, Å
e = elementary charge, 1.6022E-19 Coulomb
GCM = group-contribution method
g^E = excess Gibbs free energy, kcal/mol
ΔG^{*is} = free energy of ideal solvation, kcal/mol
ΔG^{*res} = free energy of surface charge restoration, kcal/mol
ΔG^{*sol} = solvation free energy kcal/mol
Ha = Hartree energy unit, 4,3597E-18 J
h_{mix}^E = excess enthalpy of mixing, J/mol
$\Delta_{fus}H$ = enthalpy of fusion, kJ/kmol
$n_i(\sigma_m)$ = the number of segments with a surface charge density of σ_m
P = pressure, kPa
$p'(\sigma)$ = modified sigma profile, Å^2
$p(\sigma)$ = sigma profile
q = surface area constant, Å^2
q_i = normalized surface-area parameter
$\mathbf{q}^*$ = surface screening vector in the conductor
(r) = a specific position
r = volume constant, Å^3
R = ideal gas constant, 0.001987 kcal/mol*K or 8.314kJ/kmol*K
r_{av} = surface-segment averaging radius (adjustable parameter), Å
r_i = normalized volume parameter
r_n = effective radius of surface segment n, assuming circular surface segments, Å
$\Delta_{fus}S$ = entropy of fusion, kJ/kmol*K
SLE = solid-liquid equilibrium
T = temperature, K
T_m = melting point temperature, K
V_i = total cavity volume, Å^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^{sol} = solute solubility, mole fraction
z = coordination number

Greek symbols

α' = the constant for the misfit energy, Å^4*kcal/e^2*mol
γ_i = activity coefficient of component *i*
γ_i^{sat} = solute activity coefficient at saturation
$\gamma^{SG}{}_{i/S}$ = Staverman-Guggenheim combinatorial contribution to the activity coefficient
$\Gamma_s(\sigma_m)$ = segment activity coefficient of the solvent
$\Gamma_i(\sigma_m)$ = segment activity coefficient of the solute
$\mathbf{\Phi_{sol}}$ = potential due to the charge distribution of the solute molecule
$\mathbf{\Phi_{tot}}$ = total potential on the cavity surface
ϕ_i = normalized volume fraction
σ = surface-segment charge-density distribution, e/Å^2
$\boldsymbol{\sigma}$ = averaged charge on the conductor surface in vector form, e/Å^2
$\boldsymbol{\sigma}^*$ = induced-charge on the conductor surface in vector form, e/Å^2
σ_{hb} = the sigma-value cutoff for hydrogen-bonding, e/Å^2
$\sigma_n{}^*$ = surface-charge density for segment n from the COSMO output, e/Å^2
θ_I = normalized surface-area fraction

9 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 Parameterization 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 Equilibria* **2000***, 172*, 43.

(4) Eckert, F.; Klamt, A. Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE Journal* **2002**, *48***,** 369.

(5) Lin, S.T.; Sandler, S. A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res,* **2002***, 41*, 899
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. Eq.-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 Tans 2.* **1993**, 799.

(9) Klamt, A; COSMO and COSMO-RS, In *Encyclopedia of Computational Chemistry*: Schleyer, Paul von Rague Eds., Chichester, 1998

(10) Lin, S.T.; Sandler, S. A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. [Erratum for Volume 41, Number 5] *Ind. Eng. Chem. Research* **2004,** *43*, 1322.

(11) Eckert, F.; Klamt, A. Validation of the COSMO-RS Method: Six Binary Systems. *Ind. Eng. Chem. Res.* **2001**, *40*, 2371.

(12) 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.

(13) Ben-Naim, A.;*Solvation Thermodynamics*; Plenum Press: New York and London, 1987

(14) 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.

(15) Koch, W.; Holthausen, M. C. *Chemist's Guide to Density Functional Theory*, 2nd edition, Wiley0-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 Publ.: Amsterdam, 1995.

(20) Becke, A.D. Density Functional Calculations of Molecular Bond Energies. *J. Chem. Phys*., **1986**, *84*, 4524.

(21) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev*., **1986**, *33*, 8822.

(22) 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.

(23) Giles, N.; Wilson, G. Phase Equilibria on Seven Binary Mixtures. *J. Chem. Eng. Data*. **2000**, *45*, 146.

(24) Frank, T.C.; Downey, J.R.; Gupta, S.K. Quickly Screen Solvents for Organic Solids. *Chemical Engineering Progress*. **1999**, *December*, 41

(25) Marrero, J.; Abildskov, J. Solubility and Related Properties of Large Complex Chemicals, Part 1: Organic Solutes Ranging from C_4 to C_{40}. *Chemistry Data Series, XV*, DECHEMA, 2003.

(26) Kolar, P.; Shen, J.W.; Tsuboi, A.; Ishikawa, T. Solvent Selection of Pharmaceuticals. *Fluid Phase Equilibria* **2002**, *194-197*, 771.

(27) Eckert, F.; Klamt, A. Prediction of Halocarbon Thermodynamics with COSMO-RS. *Fluid Phase Equilibria.* **2003**, *210*, 117.

(28) Klamt, A.; Eckert, F. Prediction of Vapor Liquid Equilibria using COSMOtherm., *Fluid Phase Equilibria.,* **2004**, *217*, 53

(29) Klamt, A.; Eckert, F.; Hornig, M.; Beck, M.; Bürger, T. Prediction of Aqueous Solubility of Drugs and Pesticides with COSMO-RS. *Journal of Computational Chemistry*, **2002**, *23*, 275

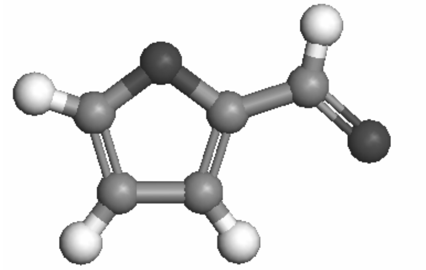
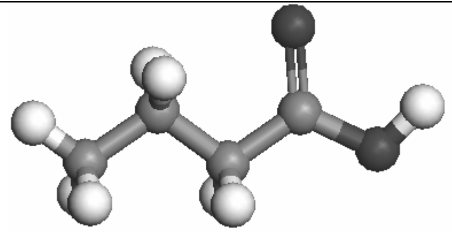
(30) Sandler, S. University of Delaware, Newark, Delaware. Private communications, August to October, **2004**.

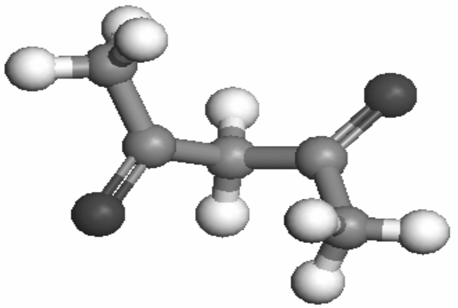
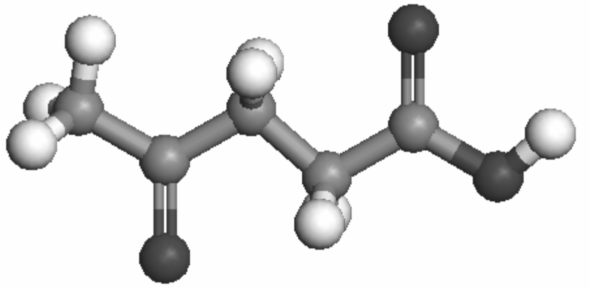
(31) Nienhaus,B.;Wittig,R.;Boelts,R.;De Haan,A.B.;Niemann,S.H.; Gmehling, J. Vapor-Liquid Equilibria at 453.25 K and Excess Enthalpies at 363.15 K and 413.15 K for Mixtures of Benzene, Toluene, Phenol, Benzaldehyde, and Benzyl Alcohol with Benzyl Benzoate, *J. Chem. Eng. Data.*, **1999**, *2*, 303

Appendix A: Energy Differences Between DMol Versions

We use DMol v2.2 for all of the VT-2004 calculations and find small discrepancies with recent work done by Professor Sandler's group at the University of Delaware using DMol v3.9.[30] Below, we present four examples of molecules where both groups work with the same structure, but the energies differ by small quantities.

Table 4. Comparison of computed energies for four molecules using different versions of DMol3.[30]

	VT-2004	**UD**
Furfural		
Gas Phase energy (kcal/mol)	-215528.7	-215524.2874
Conductor energy (kcal/mol)	-215535.6	-215531.2219
Energy Difference (kcal/mol)	-6.9027	-6.9345
% discrepancy	0.46%	
Butryic Acid		
Gas Phase energy	-193154.4	-193150.5166
Conductor energy	-193161.0	-193157.1743
Energy Difference	-6.6483	-6.6576
% discrepancy	0.139%	

2,4-Pentadedione		
Gas Phase energy	-217056.9920	-217006.948068422
Conductor energy	-217065.1665	-217015.0692362
Energy Difference	-8.1745	-8.12116777728079
% discrepancy	0.66%	
Levulinic Acid		
Gas Phase energy	-264295.9210	-264227.748809607
Conductor energy	-264306.6045	-264237.752167804
Energy Difference	-10.6836	-10.0033581965254
% discrepancy	6.8%	

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 6%. We recommend compiling sigma profiles with a consistent calculations algorithm and consistent software.

Appendix B: VT-2004 Example for 1,4-dioxane

Component Number		0728
Component Name		1,4-dioxane
Chemical Formula		C4H8O2
CAS #		123-91-1
Molecular Volume,	V (Å^3)	110.07874
Screening Charge Density,	σ_m (e/Å^2)	$A^2 p(\sigma_m)$ (Å^2)
	-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
	-0.007	0.728735793
	-0.006	9.406022106
	-0.005	11.74222472
	-0.004	18.93177401
	-0.003	17.99513336
	-0.002	10.37941775
	-0.001	9.212131231
	0.000	5.337467653
	0.001	3.566409281
	0.002	2.303757805
	0.003	2.241879818
	0.004	2.508549666
	0.005	2.357182743
	0.006	2.47906325
	0.007	1.137488854
	0.008	2.760834103
	0.009	3.540587672
	0.010	1.58058161
	0.011	3.344619557
	0.012	4.079974669
	0.013	3.797367601
	0.014	0.950486737
	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^2)	120.38169

Figure 24. VT-2004 entry for the 1,4-dioxane molecule as it appears in the searchable spreadsheet index.

VT-2004 COSMO output for 1,4-dioxane:

```
 text

  DMol3/COSMO Results


                COSMO input

 Dielectric Constant                 = infinity
 Basic Grid Size                   =  1082
 Number of Segments                   =    92
 Solvent Radius                    =  1.30
 A - Matrix Cutoff                  =  7.00
 Radius Increment                   =  0.00
 Non-Electrostatic Energy = A+B*area
 A   =  1.88219    B  =  0.01014



==================================================================

 Total energy (a.u.)              =      -307.756012

 Dielectric energy (a.u.)          =        -0.010036



 Total energy corrected (a.u. )    =      -307.755989

 Dielectric energy corr. (a.u.)    =        -0.010013



 Sum of polarization charges             =    -0.03004

 Sum of polarization charges(corr.)      =    -0.00108

 Total surface area of cavity (A**2)     =   120.26988

 Total volume of cavity (A**3)           =   110.07874



==================================================================
```

Molecular car file :
728.car

```
!BIOSYM archive 3
PBC=OFF

!DATE     Apr 07 10:51:31 2004
C1     -0.731714110   -0.207841653   -1.178769182 XXXX 1      xx      C   0.000
C2      0.731714110    0.207841653   -1.178769182 XXXX 1      xx      C   0.000
O1      1.406696102   -0.273884978    0.000000000 XXXX 1      xx      O   0.000
C3      0.731714110    0.207841653    1.178769182 XXXX 1      xx      C   0.000
C4     -0.731714110   -0.207841653    1.178769182 XXXX 1      xx      C   0.000
O2     -1.406696102    0.273884978    0.000000000 XXXX 1      xx      O   0.000
H1     -0.810840554   -1.311522300   -1.228770387 XXXX 1      xx      H   0.000
H2     -1.264385267    0.224701347   -2.038244288 XXXX 1      xx      H   0.000
H3      1.264385267   -0.224701347   -2.038244288 XXXX 1      xx      H   0.000
H4      0.810840554    1.311522300   -1.228770387 XXXX 1      xx      H   0.000
H5      0.810840554    1.311522300    1.228770387 XXXX 1      xx      H   0.000
H6      1.264385267   -0.224701347    2.038244288 XXXX 1      xx      H   0.000
H7     -1.264385267    0.224701347    2.038244288 XXXX 1      xx      H   0.000
H8     -0.810840554   -1.311522300    1.228770387 XXXX 1      xx      H   0.000
end
end
```

```
=====================================================
```

```
                COSMO-RS Atomic data

 Atom number and name    Radius      Charge        Surface     Charge Density


     1       C1      2.00      -0.00259      9.93942      -0.00026
     2       C2      2.00      -0.00237     10.15915      -0.00023
     3       O1      1.72       0.15018     11.81958       0.01271
     4       C3      2.00      -0.00283      9.90559      -0.00029
     5       C4      2.00      -0.00285     10.20569      -0.00028
     6       O2      1.72       0.15008     11.79430       0.01272
     7       H1      1.30      -0.02850      7.19526      -0.00396
     8       H2      1.30      -0.04435      6.94493      -0.00639
     9       H3      1.30      -0.04431      6.94501      -0.00638
    10       H4      1.30      -0.02845      7.13784      -0.00399
    11       H5      1.30      -0.02875      7.15605      -0.00402
    12       H6      1.30      -0.04388      6.98418      -0.00628
    13       H7      1.30      -0.04374      6.98427      -0.00626
```

```
     14       H8      1.30       -0.02872        7.09861       -0.00405

=====================================================================



                         Segment information:

          total number of segments:    551
          n           - segment number
          atom          - atom associated with segment n
          position      - segment coordinates
          charge        - segment charge
          area          - segment area
          potential     - solute potential on a segment



  n   atom       position (X, Y, Z) [au]        charge      area      charge/area  potential


   1    1     -4.84057  -1.81446  -2.78097     0.00037     0.27874     0.00134    -0.02004
   2    1     -2.35409  -2.02783  -5.49363     0.00030     0.37165     0.00080     0.02212
   3    1     -2.08155   3.23104  -3.04242     0.00028     0.23228     0.00118    -0.00717
   4    1     -2.72344   3.06693  -2.94683     0.00014     0.13937     0.00101    -0.01473
   5    1     -2.29477  -2.81901  -4.97814     0.00029     0.27874     0.00104     0.01903
   6    1     -4.42965  -2.54950  -2.81821     0.00023     0.23228     0.00098    -0.00890
   7    1     -4.54862  -2.43382  -1.91877     0.00034     0.27874     0.00121    -0.02453
   8    1     -4.51165  -1.81594  -3.79873     0.00048     0.32519     0.00148    -0.00266
   9    1     -3.37360  -2.26376  -4.83906     0.00078     0.32519     0.00238     0.01184
  10    1     -1.17019  -2.49826  -5.35899    -0.00001     0.32519    -0.00004     0.02518
```

See our website for remaining 541 data points (n=11 to 551) for the 1,4-dioxane molecule.

Appendix C: Sigma-Averaging Algorithm FORTRAN Code

```
PROGRAM SIGMAPROFILE
!****************************************************************************************
!CREATED USING DIGITAL VISUAL FORTRAN 6.0 (2003)
!
!        This program reads the modified COSMO output file (in text format) from dmol3
! and averages the surface segment charge densities per Klamt (1995), Klamt et al (1998)
! Lin and Sandler (2002) to establish the segment charges for the "sigma-profile".  This
! program creates a text file that MS Excel can read and plot.
!
!        THIS PROGRAM WRITTEN BY:
!                RICHARD OLDLAND (roldland@vt.edu)              MIKE ZWOLAK (zwolak@caltech.edu)
!                DEPARTMENT OF CHEMICAL ENGINEERING   PHYSICS DEPARTMENT
!                VIRGINIA TECH                                  CALIFORNIA INSTITUTE OF TECHNOLOGY
!                BLACKSBURG, VA 24060                           PASADENA, CA 91125
!
!        VALUES READ FROM THE DATA FILE:
!        ATOM = ATOM NUMBER IN MOLECULE
!        POSXAU = X-CORDINATE OF THE SEGMENT POSITION IN ATOMIC UNITS
!        POSYAU = Y-CORDINATE OF THE SEGMENT POSITION IN ATOMIC UNITS
!        POSZAU = Z-CORDINATE OF THE SEGMENT POSITION IN ATOMIC UNITS
!        POSXA = X-CORDINATE OF THE SEGMENT POSITION IN ANGSTROMS
!        POSYA = Y-CORDINATE OF THE SEGMENT POSITION IN ANGSTROMS
!        POSZA = Z-CORDINATE OF THE SEGMENT POSITION IN ANGSTROMS
!        A = SURFACE SEGMENT AREA; APPROXIMATED AS CIRCULAR (ANGSTROMS SQUARED)
!        CHG = CHARGE OF THE SURFACE SEGMENT (SIGMA, e)
!        SIGMA = RATIO OF SURFACE CHARGE TO AREA (SIGMA/A, e/A**2)
!        POTENT = SURFACE POTENTIAL
!
!        FROM SEGMENT CHARGE AVERAGING:
!        REFF = RADIUS OF THE AREA THAT AFFECTS THE SURFACE SEGMENT CHARGE
(ANGSTROMS)
!    DMN = DISTANCE BETWEEN CALCULATED SEGEMENT AND SEGEMENTS AFFECTING IT
(ANGSTROMS)
!        RAD = RADIUS OF THE PARTICULAR SURFACE SEGMENT (ASSUMED CIRCULAR,
ANGSTROMS SQUARED)
!        SIGMANEW = NEW AVERAGED SIGMA VALUE
!        SIGMASUM = SUMMATION OF EFFECTS FROM OTHER SURFACE SEGMENTS
!        NORMDIST = NORMALIZATION FACTOR
!        NORMSUM = SUMMATION OF ALL NORMALIZATION FACTORS
!
!        REQUIRED INPUT (ON PROMPT):
!                NAME OF FILE (INCLUDING LOCATION AND EXTENSION)
!                NAME OF CHEMICAL (THIS WILL APPEAR IN THE OUTPUT FILE)
!                NUMSEGMENT = THE NUMBER OF SURFACE SEGMENTS
!
! THE OUTPUT FILE "'CHEMICAL'SIGMA-PROFILE.TXT" IS THE SORTED SIGMA PROFILE
!
! LITERATURE CITED:
! 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.
! Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. Refinement and Parameterization of COSMO-RS.
!    J. Phys. Chem A 1998, 102, 5074.
! Lin, S.T.; Sandler, S. A Priori Phase Equilibrium Prediction from a Segment
!    Contribution Solvation Model. Ind. Eng. Chem. Res, 2002, 41, 899
! Lin, S.T.; PhD. Dissertation, University of Delaware, Newark, DE, 2000
!****************************************************************************************
IMPLICIT NONE
```

```
CHARACTER(35):: FILENAME
CHARACTER(10):: CHEMICAL
INTEGER :: I, J, K, DUMBI, TMP, NUMSEGMENT
INTEGER, DIMENSION (:), ALLOCATABLE :: ATOM
REAL*8 :: REFF, PI
REAL*8, DIMENSION (:), ALLOCATABLE :: POSXAU, POSYAU, POSZAU, POSXA, POSYA, POSZA, A
REAL*8, DIMENSION (:), ALLOCATABLE :: CHG, SIGMA, POTENT, SIGMANEW, SIGMASUM, RAD,
NORMDIST
REAL*8, DIMENSION (:), ALLOCATABLE :: NORMSUM, DMN
REAL*8, DIMENSION(0:49) :: CHGDEN,SP

!ESTABLISH CONSTANTS
PI = 3.141592653D0
REFF = 0.817642D0

!ESTABLISHING THE COSMO FILE TO READ
WRITE(*,*) "TYPE THE NAME OF THE FILE YOU WISH TO READ IN,"
WRITE(*,*) "INCLUDING LOCATION (MAX 35 CHARACTERS), AND HIT ENTER"
READ (*,*) FILENAME

!ESTABLISH THE CHEMICAL NAME
WRITE(*,*) "TYPE IN THE NAME OF THE CHEMICAL (MAX 10 CHARACTERS), AND HIT ENTER"
READ (*,*) CHEMICAL

!OPEN THE COSMO OUTPUT FILE
OPEN(UNIT=10, FILE = FILENAME, STATUS = "OLD", ACTION = "READ", POSITION = "REWIND")

!ESTABLISH THE NUMBER OF SURFACE SEGMENTS AND ALLOCATE THE ARRAYS
WRITE(*,*) "TYPE THE NUMBER OF SURFACE SEGMENTS, FROM THE COSMO OUTPUT, AND HIT
ENTER"
READ (*,*) NUMSEGMENT
ALLOCATE(ATOM(NUMSEGMENT), POSXAU(NUMSEGMENT), POSYAU(NUMSEGMENT),
POSZAU(NUMSEGMENT), &
        POSXA(NUMSEGMENT), POSYA(NUMSEGMENT), POSZA(NUMSEGMENT), A(NUMSEGMENT),
&
        CHG(NUMSEGMENT), SIGMA(NUMSEGMENT), POTENT(NUMSEGMENT),
SIGMANEW(NUMSEGMENT), &
        SIGMASUM(NUMSEGMENT), RAD(NUMSEGMENT), NORMDIST(NUMSEGMENT),
NORMSUM(NUMSEGMENT), &
        DMN(NUMSEGMENT))

!READ THE COSMO FILE AND ESTABLISH THE DATA ARRAYS
DO I = 1, NUMSEGMENT
        READ(10,*) DUMBI,ATOM(I),POSXAU(I),POSYAU(I),POSZAU(I),CHG(I),A(I),SIGMA(I),POTENT(I)
        !CONVERT THE POSITIONS FROM ATOMIC UNITS TO ANGSTROMS AND ASSIGN NEW
ARRAYS
        POSXA(I) = POSXAU(I) * 0.529177
        POSYA(I) = POSYAU(I) * 0.529177
        POSZA(I) = POSZAU(I) * 0.529177
        RAD(I) = SQRT(A(I)/PI)
END DO
CLOSE(10)

!BEGIN AVERAGING SURFACE CHARGES
DO J=1, NUMSEGMENT
        SIGMANEW(J) = 0.D0
        NORMSUM(J)=0.D0
        DO K=1, NUMSEGMENT
                DMN(K) = SQRT((POSXA(K)-POSXA(J))**2+(POSYA(K)-POSYA(J))**2+ &
                        (POSZA(K)-POSZA(J))**2)
                SIGMASUM(K)=SIGMA(K)*(RAD(K)**2*REFF**2)/(RAD(K)**2+REFF**2)* &
```

```
                    DEXP(-(DMN(K)**2)/(RAD(K)**2+REFF**2))
            NORMDIST(K) =(RAD(K)**2*REFF**2)/(RAD(K)**2+REFF**2)* &
                    DEXP(-(DMN(K)**2)/(RAD(K)**2+REFF**2))
            NORMSUM(J) = NORMSUM(J) + NORMDIST(K)
            SIGMANEW(J) = SIGMANEW(J) + SIGMASUM(K)
      END DO
      SIGMANEW(J) = SIGMANEW(J)/NORMSUM(J)
END DO

!CONTAINS AVERAGED SIGMA-PROFILE
OPEN (12,FILE='C:\PROFILES\'//CHEMICAL//'SIGMA-PROFILE.TXT')

!SETTING CHGDEN MATRIX
DO J=0,49
      SP(J)=0.D0
      CHGDEN(J) = -0.025D0+0.001D0*DBLE(J)
END DO

!SIGMA PROFILE SORTING TAKEN FROM LIN DISSERTATION**
DO J=1,NUMSEGMENT
      TMP=INT((SIGMANEW(J)-CHGDEN(0))/0.001D0)
      SP(TMP)=SP(TMP)+A(J)*(CHGDEN(TMP+1)-SIGMANEW(J))/0.001D0
      SP(TMP+1)=SP(TMP+1)+A(J)*(SIGMANEW(J)-CHGDEN(TMP))/0.001D0
END DO

DO J=0,49
      WRITE(12,*) CHGDEN(J),SP(J)
END DO
CLOSE(12)

END PROGRAM SIGMAPROFILE
```

Appendix D: COSMO-SAC-VT-2004 FORTRAN Code

```
PROGRAM GAMMA1
!********************************************************************************
!        This program uses the sigma profiles of two pure components to calculate
!  the liquid-phase activity coefficients in a solution.  This is the first
!  step in predicting VLE for mixtures.
!
!        This program uses the COSMO-SAC model as published (Lin, S.T.,
!        S.I. Sandler, Ind. Eng. Chem. Res. 41, (2002), 899-913).
!
!        THIS PROGRAM WRITTEN BY:
!          RICHARD OLDLAND (roldland@vt.edu)    MIKE ZWOLAK (mzwolak@caltech.edu)
!          DEPARTMENT OF CHEMICAL ENGINEERING   PHYSICS DEPARTMENT
!          VIRGINIA TECH                        CALIFORNIA INSTITUTE OF TECHNOLOGY
!          BLACKSBURG, VA 24060                 PASADENA, CA 91125
!
!        PHYSICAL CONSTANTS AND PARAMETERS:
!          EO = PERMITTIVITY IN A VACUUM (e**2*mol/Kcal*Angstrom)
!          AEFFPRIME = EFFECTIVE SURFACE AREA (ANGSTROMS**2) --FROM LIN
!          RGAS = IDEAL GAS CONSTANT (Kcal/mol*K)
!          VNORM = VOLUME NORMALIZATION CONSTANT (A**3) --FROM LIN
!          ANORM = AREA NORMALIZATION CONSTANT (A**2) --FROM LIN
!          COORD = THE COORIDINATION NUMBER --FROM LIN
!          CHB = HYDROGEN BONDING COEFFICIENT (Kcal/mole*Angstroms**4/e**2)
!          SIGMAHB = CUTOFF VALUE FOR HYDROGEN BONDING (e/Angstrom**2)
!          EPS = RELATIVE PERMITTIVITY --FROM LIN
!          ALPHAPRIME = A CONSTANT USED IN THE MISFIT ENERGY CALCULATION
!
!        INPUT PARAMETERS:
!          SYSTEMP = THE SYSTEM TEMPERATURE (K)
!          COMP = NUMBER OF COMPONENTS IN THE SYSTEM --SET TO 2 FOR BINARY
!          SYSCOMP = NAMES OF COMPONENTS IN THE SYSTEM
!          VCOSMO = CAVITY VOLUME FROM COSMO OUTPUT (A**3)
!          ACOSMO = MOLECULAR SURFACE AREA FROM COSMO OUTPUT (A**2) --THE SUM
!                OF THE INDIVIDUAL PROFILE.
!
!
!        LITERATURE CITED:
! 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.
! Lin, S.T.; Sandler, S. A Priori Phase Equilibrium Prediction from a Segment
!     Contribution Solvation Model. Ind. Eng. Chem. Res, 2002, 41, 899
! Lin, S.T.; PhD. Dissertation, University of Delaware, Newark, DE, 2000
!
!
!        PROGRAM CURRENTLY SETUP FOR BINARY MIXTURES ONLY
!*******************************************************************************
IMPLICIT NONE
REAL,PARAMETER:: EO = 2.395*10.0**-4, AEFFPRIME = 7.5, RGAS = 0.001987
REAL,PARAMETER:: VNORM = 66.69, ANORM = 79.53
REAL :: FPOL, ALPHA, ALPHAPRIME, COORD, EPS, SYSTEMP, SIGMAHB, CHB, FRAC1, FRAC2
REAL :: SYSPRES, SIGMAACC, SIGMADON, SUMMATION, BOTTHETA, BOTPHI, PHI1, PHI2
REAL :: THETA1, THETA2, L1, L2, GAMMASG1, GAMMASG2, GAMMA, GAMMA2, SUMGAMMA1
REAL :: SUMGAMMA2, N1, N2, LNGAMMA, LNGAMMA2
INTEGER :: I, J, K, L, M, COMPSEG, COUNT, COMP
REAL, DIMENSION(2):: VCOSMO, ACOSMO, RNORM, QNORM
REAL, DIMENSION(:), ALLOCATABLE:: COUNTER, DENOM, PROFILE, NUMER, SEGGAMMA
```

```
REAL, DIMENSION(:), ALLOCATABLE:: SEGGAMMAOLD, CONVERG
REAL, DIMENSION(:,:), ALLOCATABLE:: SIGMA, DELTAW, SEGGAMMAPR, SEGGAMMAOLDPR, CONPR
CHARACTER (13), DIMENSION(2):: SYSCOMP
CHARACTER (35), DIMENSION(2):: FILENAME
CHARACTER :: ANSWER
CHARACTER (35):: OUTPUT

COMPSEG = 50 !NUMBER OF INTERVALS FOR THE SIGMA PROFILE
EPS = 3.667 !(LIN AND SANDLER USE A CONSTANT FPOL WHICH YEILDS EPS=3.68)
COORD = 10.0 !(KLAMT USED 7.2)
SIGMAHB = 0.0084
CHB = 85580.0
COMP =2

FPOL = (EPS-1.0)/(EPS+0.5)
ALPHA = (0.3*AEFFPRIME**(1.5))/(EO)
ALPHAPRIME = FPOL*ALPHA


ALLOCATE(SIGMA(COMPSEG,COMP), COUNTER(COMPSEG), PROFILE(COMPSEG),
NUMER(COMPSEG),&
        DENOM(COMPSEG), DELTAW(COMPSEG,COMPSEG), SEGGAMMA(COMPSEG),
SEGGAMMAOLD(COMPSEG), &
        CONVERG(COMPSEG), SEGGAMMAPR(COMPSEG,COMP),
SEGGAMMAOLDPR(COMPSEG,COMP), &
        CONPR(COMPSEG,COMP))


!DEFINE SYSTEM TEMPERATURE (K)
WRITE(*,*) "ENTER IN THE SYSTEM TEMPERATURE (K)"
READ(*,*) SYSTEMP

!DEFINE THE SYSTEM AS WELL AS THE AREA AND VOL FROM THE COSMO CALCULATION
DO I=1, COMP
        WRITE(*,*) "ENTER THE NAME OF COMPONENT", I
        READ(*,*) SYSCOMP(I)
        WRITE(*,*) "ENTER THE NAME AND LOCATION OF THE  ", SYSCOMP(I), "SIGMA PROFILE (35
character max)"
        READ(*,*) FILENAME(I)
        WRITE(*,*) "ENTER THE CAVITY VOLUME FOR COMPONENT",I
        READ(*,*) VCOSMO(I)
END DO

!OPEN THE SIGMA PROFILES FOR THE PURE COMPONENTS
OPEN(UNIT=1, FILE=FILENAME(1), STATUS="OLD", ACTION="READ", POSITION="REWIND")
OPEN(UNIT=2, FILE=FILENAME(2), STATUS="OLD", ACTION="READ", POSITION="REWIND")

!READ INDIVIDUAL SIGMA PROFILES; COUNTER IS THE SIGMA VALUE, SIGMA IS P(SIGMA)
DO K = 1, COMP
        ACOSMO(K) =0.0
        DO J=1, COMPSEG
                READ(K,*)COUNTER(J), SIGMA(J,K)
                ACOSMO(K)=ACOSMO(K)+SIGMA(J,K)
        END DO
END DO
CLOSE(1)
CLOSE(2)

!ESTABLISH OUTPUT FILES.  THE PROGRAM CREATES A FILE FOR THE GAMMA-X DATA
WRITE(*,*) "ENTER THE NAME FOR THE OUTPUT FILE, INCLUDE EXTENSION. (25 CHARACTER MAX)"
READ (*,*) OUTPUT
OPEN(UNIT=11, FILE='C:\PROFILES\GAMMAS\'//OUTPUT, STATUS="NEW")
```

```
WRITE(11,*) "SYSTEMP", SYSTEMP, "KELVIN"
5  FORMAT (1X,A10,5X,A12,5X,A12,5X,A12,5X,A12)
6  FORMAT (1X,A10,5X,A9,5X,A10,7X,A10,5X,A10)
WRITE(11,6) "MOLE FRAC", "GAMMA1", "GAMMA2", "LNGAMMA1", "LNGAMMA2"
WRITE(11,5) "X1", SYSCOMP(1), SYSCOMP(2), SYSCOMP(1), SYSCOMP(2)



!VARYING MOLE FRACTIONS // ONLY WORKS FOR BINARY MIXTURE
DO FRAC1 = 0.005, 0.995, 0.01
FRAC2 = 1.0 - FRAC1
!CALCULATE THE MIXTURE SIGMA PROFILE
        DO J =1,COMPSEG
                NUMER(J) = FRAC1*SIGMA(J,1) + FRAC2*SIGMA(J,2)
                DENOM(J) = FRAC1*ACOSMO(1) + FRAC2*ACOSMO(2)
                PROFILE(J) = NUMER(J)/DENOM(J)

        END DO
        DO I = 1, COMPSEG
                DO K = 1, COMPSEG
                        IF (COUNTER(I)>=COUNTER(K)) THEN
                                SIGMAACC = COUNTER(I)
                                SIGMADON = COUNTER(K)
                        END IF
                        IF (COUNTER(I)<COUNTER(K)) THEN
                                SIGMADON = COUNTER(I)
                                SIGMAACC = COUNTER(K)
                        END IF

                DELTAW(I,K) = (ALPHAPRIME/2.0)*(COUNTER(I)+COUNTER(K))**2.0 + CHB *   &
                        MAX(0.0,(SIGMAACC - SIGMAHB))*MIN(0.0,(SIGMADON + SIGMAHB))
                END DO
        END DO


!ITERATION FOR SEGMENT ACTIVITY COEF. (MIXTURE)
SEGGAMMA = 1.0
DO
        SEGGAMMAOLD = SEGGAMMA
        DO I = 1, COMPSEG
                SUMMATION = 0.0
                DO K = 1, COMPSEG
                        SUMMATION = SUMMATION + PROFILE(K)* SEGGAMMAOLD(K) * &
                                EXP(-DELTAW(I,K)/(RGAS*SYSTEMP))
                END DO
                SEGGAMMA(I)=EXP(-LOG(SUMMATION))
                SEGGAMMA(I)=(SEGGAMMA(I)+SEGGAMMAOLD(I))/2.0
        END DO
        DO I=1, COMPSEG
                CONVERG(I)=ABS((SEGGAMMA(I)-SEGGAMMAOLD(I))/SEGGAMMAOLD(I))
        END DO
IF (MAXVAL(CONVERG) <=0.000001) EXIT
END DO

!ITERATION FOR SEGMENT ACITIVITY COEF (PURE SPECIES)
DO L = 1, COMP
        SEGGAMMAPR (:,L) = 1.0
        DO
                SEGGAMMAOLDPR (:,L) = SEGGAMMAPR (:,L)
                DO I = 1, COMPSEG
                        SUMMATION = 0.0
                        DO K = 1, COMPSEG
```

```
                                        SUMMATION = SUMMATION +
(SIGMA(K,L)/ACOSMO(L))*SEGGAMMAOLDPR(K,L) * &
                                                  EXP(-DELTAW(I,K)/(RGAS*SYSTEMP))
                              END DO
                              SEGGAMMAPR(I,L)=EXP(-LOG(SUMMATION))
                              SEGGAMMAPR(I,L)=(SEGGAMMAPR(I,L)+SEGGAMMAOLDPR(I,L))/2.0
                    END DO
                    DO I=1, COMPSEG
                              CONPR(I,L)=ABS((SEGGAMMAPR(I,L)-
SEGGAMMAOLDPR(I,L))/SEGGAMMAOLDPR(I,L))
                    END DO

          IF (MAXVAL(CONPR) <=0.000001) EXIT
          END DO
END DO

!THE STAVERMAN-GUGGENHEIM EQUATION
DO I = 1,COMP
          RNORM(I) = VCOSMO(I)/VNORM
          QNORM(I) = ACOSMO(I)/ANORM
END DO

BOTTHETA = FRAC1*QNORM(1) + FRAC2*QNORM(2)
BOTPHI = FRAC1*RNORM(1) + FRAC2*RNORM(2)

THETA1 = (FRAC1*QNORM(1))/BOTTHETA
THETA2 = (FRAC2*QNORM(2))/BOTTHETA

PHI1 = (FRAC1*RNORM(1))/BOTPHI
PHI2 = (FRAC2*RNORM(2))/BOTPHI

L1 = (COORD/2.0)*(RNORM(1)-QNORM(1))-(RNORM(1)-1.0)
L2 = (COORD/2.0)*(RNORM(2)-QNORM(2))-(RNORM(2)-1.0)

!GAMMASG1 AND GAMMASG2 ARE ACTUALLY LNGAMMASG
GAMMASG1 = LOG(PHI1/FRAC1)+(COORD/2)*QNORM(1)*LOG(THETA1/PHI1)+L1-(PHI1/FRAC1)* &
          (FRAC1*L1 + FRAC2*L2)
GAMMASG2 = LOG(PHI2/FRAC2)+(COORD/2)*QNORM(2)*LOG(THETA2/PHI2)+L2-(PHI2/FRAC2)* &
          (FRAC1*L1 + FRAC2*L2)

!CALCULATION OF GAMMAS
SUMGAMMA1 = 0.0
SUMGAMMA2 = 0.0
DO I = 1, COMPSEG
          SUMGAMMA1 = SUMGAMMA1
+((SIGMA(I,1)/AEFFPRIME)*(LOG(SEGGAMMA(I)/(SEGGAMMAPR(I,1)))))
          SUMGAMMA2 = SUMGAMMA2
+((SIGMA(I,2)/AEFFPRIME)*(LOG(SEGGAMMA(I)/(SEGGAMMAPR(I,2)))))
END DO

GAMMA =EXP(SUMGAMMA1 + (GAMMASG1))
GAMMA2=EXP(SUMGAMMA2 + (GAMMASG2))
LNGAMMA = LOG(GAMMA)
LNGAMMA2 = LOG(GAMMA2)

WRITE(11,*) FRAC1,GAMMA, GAMMA2, LNGAMMA, LNGAMMA2


END DO


END PROGRAM GAMMA1
```

10 Vita

Richard Oldland was born on July 22, 1979 to Brent and Linda Oldland. He attended Virginia Polytechnic Institute and State University in the fall of 1997 and earned a Bachelors of Science in Chemical Engineering in 2001. Between June 2001 and December 2002 he worked as a Process Engineer for Honeywell International, Inc., in Richmond, VA. He returned to Virginia Tech in the spring of 2003 and enrolled in the Chemical Engineering Masters program. He completed his Masters of Science in December, 2004 and currently works as a Process Engineer with The Dow Chemical Company near Houston, TX.