AIChE WILEY InterScience® DISCOVER SOMETHING GREAT

Group Contribution Prediction of Surface Charge Density Profiles for COSMO-RS(Ol)

Tiancheng Mu
Dept. of Industrial Chemistry, University of Oldenburg, D-26111 Oldenburg, Germany
Dept. of Chemistry, Renmin University of China, 100872 Beijing, China

Jürgen Rarey and Jürgen Gmehling
Dept. of Industrial Chemistry, University of Oldenburg, D-26111 Oldenburg, Germany

DOI 10.1002/aic.11338

A new method for predicting the surface charge density distribution (σ profile) and cavity volume of molecules based on group contributions was developed. The original σ profiles used for the regression were obtained using Gaussian 03 B3LYP/6-311G(d,p). In total 1363 σ profiles were used for the regression of group parameters. Group definitions are identical to those used previously for boiling point estimation. Original and estimated σ profiles were used to predict activity coefficients at infinite dilution and VLE data of binary systems using the COSMO-RS(Ol) model. The results were compared with the experimental data stored in the Dortmund Data Bank. In many cases the results were of comparable accuracy. However, for a few compounds, poor results were obtained, in particular for conjugated components like nitrobenzenes. The method offers a fast and reliable generation of σ profiles to be used with COSMO-RS(Ol) within its range of applicability. © 2007 American Institute of Chemical Engineers *AIChE J,* 53: 3231–3240, 2007

Keywords: COSMO-RS, GC-COSMO, σ profile, activity coefficient at infinite dilution, vapor-liquid equilibrium

Introduction

A reliable knowledge of thermodynamic properties of real mixtures is a prerequisite for the design and simulation of chemical, biochemical, and environmental systems. Since the literature data are limited and experimental determination is costly, time consuming or sometimes even impossible, various approaches for the estimation of the real behavior of fluid mixtures have been developed (group contribution equations of states or G^E models, quantitative structure property relationship (QSPR) methods[1] and so on). These methods usually contain parameters derived from available experimental data. Because of the rapid progress in quantum chemistry and computer technology, Monte Carlo simulation, molecular dynamics methods, and conductor-like screening model for real solvent (COSMO-RS)[2–4] methods based on force field, molecular mechanics, ab initio or density functional theory (DFT) calculations are considered as alternatives, especially if no data for chemically similar systems are available.

Until now, the most successful methods used for calculation of thermophysical properties of mixtures are group contribution methods for the estimation of the real mixture behavior. In group contribution methods, a molecule is considered as a collection of appropriately defined independent functional groups and a mixture is described as a mixture of these functional groups, thus the interaction energy of a system can be expressed by the sum of group interaction energies

---

This article contains Supplementary Material available via the Internet at http://www.interscience.wiley.com/jpages/0001-1541/suppmat.

Correspondence concerning this article should be addressed to J. Gmehling at gmehling@tech.chem.uni-oldenburg.de.

based on the previously defined functional groups. Once the group interaction parameters were determined by fitting the available experimental data of systems containing the same structural groups, the properties of any system of interest can be calculated.

The UNIFAC[5] and modified UNIFAC(Do)[6,7] methods are the most successful group contribution methods for the estimation of the real mixture behavior. They have been used for the calculation of vapor–liquid equilibria, liquid–liquid equilibria, solid–liquid equilibria, polymer solutions, and vapor pressures of pure components, excess enthalpies, activity coefficients at infinite dilution, azeotropic data, solvent effects on reaction rate, and so on.

However, some drawbacks and deficiencies of group contribution methods are well known:

- The effect of neighbor groups on the electronic structure of a group is not described (proximity effect).
- Only the type and frequency of groups is taken into account. Thus the relative position in the molecule is not considered (isomeric components cannot be distinguished).
- All groups and group interaction parameters need to be available.

The quality of group contribution predictions depends on the similarity of the new system to the database used in its parameterization. However, the prediction results can be improved by adding additional functional groups, and increasing the database used to fit group interaction parameters.

In the last few years, Klamt and Schüürmann extended the dielectric continuum solvation model[8,9] to COSMO-RS (COnductor like Screening MOdel for Real Solvents). Within the COSMO-RS framework, the COSMO method treats the solvent as an ideal conductor, so that quantum chemical calculations are not required for different solvent environments. In COSMO-RS molecules are treated as collections of interacting surface segments. The chemical potential of segments in a pure component or mixture can be calculated from the segment interaction energies. The chemical potential of each molecule is obtained by summation of the contributions of all segments.

COSMO-RS depends on a small number of adjustable parameters, some of which are physically predetermined, while others have to be derived from experiment data. Once the parameters were obtained, no experimental data are required to perform the COSMO-RS calculations. For this reason, COSMO-RS has attracted much attention from both chemists and chemical engineers and many research groups in the world work on COSMO-RS based thermodynamic models. The COSMO method has been incorporated into several quantum chemistry software packages such as Gaussian,[10,11] TURBOMOLE,[12] DMol,[13] GAMESS,[14] and so on. Lin and Sandler presented a variant version of COSMO-RS named COSMO-SAC (COSMO Segment Activity Coefficient).[15] Grensemann and Gmehling proposed another version of COSMO-RS called COSMO-RS(Ol),[16,17] which is available in the Dortmund Data Bank (DDB)[18] software package DDBSP.

COSMO-RS has been extended to solve a very wide variety of problems, for example to predict many thermodynamic properties such as vapor pressures, partition coefficients, activity coefficients, solubilities, excess Gibbs energies, and excess enthalpies, etc. It should in principle be able to describe isomeric, tautomeric, and conformational differences (This could not be verified in a previous work[16]). COSMO-RS has been applied to ionic liquids, electrolyte solutions, polymers, surfactant micelles, biomembranes, and so on.[19]

To perform a COSMO-RS calculation, a quantum chemical COSMO calculation for each component in the mixture is required to obtain the surface charge density distribution, the σ profile. Generating the σ profile is the most time consuming step of a COSMO-RS calculation. Different model chemistries can be used to generate the σ profile. Until now only one openly published and free database of σ profiles is available (total 1423 compounds).[20] Computations were performed using BP-DFT in the quantum chemical software DMol3. Only one conformer of each component was considered.

In this work, a group contribution method for the prediction of the σ profile and cavity volume was developed. The quality of the predicted profiles was examined in different ways and found to be satisfactory. The new method was named group contribution COSMO method (GC-COSMO method). Using GC-COSMO, the time-consuming quantum mechanic calculations are no further required and the σ profiles of molecules can be obtained by simple algebraic calculation.

Hornig and Klamt[21] have previously published a very different procedure (COSMO*frag*) to estimate missing σ profiles. This procedure is based on atomic similarity codes and requires a large profile database. As the commercial COSMO*frag* software was not available to the authors and COSMOfrag represents a very different methodological approach, it was not further considered in this work.

Theory and Model

Calculation of the σ profile

After the quantum mechanical calculation of the surface charge density distribution and cavity volume of the molecules, an ensemble averaging procedure is applied over a region of radius $r_{\alpha v}$ using the following averaging algorithm[3]:

$$\sigma_v = \frac{\sum_\mu \sigma_\mu^* \frac{r_\mu^2 r_{av}^2}{r_\mu^2 + r_{av}^2} \exp\left(-\frac{d_{\mu v}^2}{r_\mu^2 + r_{av}^2}\right)}{\sum_\mu \frac{r_\mu^2 r_{av}^2}{r_\mu^2 + r_{av}^2} \exp\left(-\frac{d_{\mu v}^2}{r_\mu^2 + r_{av}^2}\right)} \tag{1}$$

where $d_{\mu v}$ is the distance between the segments μ and v, r_μ is the mean radius of segment μ calculated from the area of segment μ, and $r_{\alpha v}$ is an adjustable parameter.

From the resulting charge densities of the molecular surface segments the σ profile ($p_i(\sigma)$) is calculated. For segment of equal size, Klamt has defined the σ profile for a molecule i as[2,20,22]

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

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

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

where $n_i(\sigma)$ is the number of segments with a discretized surface charge density σ, A_i is the total cavity surface area,

$A_i(\sigma)$ is the total surface area of all the segments with a particular charge density σ. σ profiles are molecule-specific. In this work a σ profile contains 80 segments, 0.001 e/Å^2 wide, in the range -0.040 to 0.040 e/Å^2.

In COSMO-RS(Ol),[16] a further averaging is applied to obtain the final σ profile:

$$p_i(\sigma_n)_{\text{COSMO-RS(Ol)}} = \frac{1}{3}\sum_{n-1}^{n+1} p_i(\sigma_n) \qquad (5)$$

Figure 1 shows the σ profiles of four representative compounds calculated by Gaussian B3LYP[23,24]/6-311G(d,p). For simplicity, Gaussian B3LYP/6-311G(d,p) is abbreviated as G03 B3LYP further on.

In COSMO-RS(Ol), the interaction energy between two segments is composed of two terms, a charge misfit and a hydrogen bonding interaction term. Hydrogen bonding is always assumed between segments of opposite charge density if the absolute value of both charge densities exceeds a predefined cutoff value (0.0082 e/A^2). Hydrogen bonds are strong polar interactions that require special steric configurations and the treatment of these bonds in COSMO-RS is in many cases over-simplified. To suppress the ability of the ether oxygen atom to act as a hydrogen bond acceptor, two separate profiles are used in COSMO-RS(Ol). The second profile contains the charge density distribution of the segments belonging to the ether-oxygen while the first profile describes the distribution for all other segments. During the calculation of the chemical potentials, no hydrogen bonding energy is used for interactions with and between segments of the second profile. This way, the ether-oxygen does not act as a hydrogen bond acceptor.

A detailed description on how to calculate the real mixture behavior (activity coefficients) from the σ profiles and cavity volume using COSMO-RS(Ol) was given by Grensemann and Gmehling.[16]

Theory of the GC-COSMO method

A group contribution method is generally based on the assumption that the effect of a certain group on a physical property is independent of the other groups in the molecule and that the contributions of the individual groups are additive.

In GC-COSMO, the total σ profile of a molecule is described as a vector $\vec{P}$ with 80 elements whereby the element j contains the total area of all segments with an averaged surface charge density between $-0.040 + (j-1)\cdot 0.001$ e/Å^2 and $-0.040 + j\cdot 0.001$ e/Å^2. On the basis of the assumption of group additivity the σ profile $\vec{P}$ of the molecule can be calculated from the contributions $\vec{\mathbf{P}}_i$ of the individual groups i in the molecule:

$$\vec{P} = \sum v_i \vec{P}_i \qquad (6)$$

where v_i is the frequency of the groups i. $\vec{P}_i$ describes the effect of group i on the σ profile $\vec{P}$ of the molecule and usually differs from the σ profile of the structural group. The reason is that $\vec{\mathbf{P}}_i$ also contains the effect of a group on the screening charge density distribution of the neighboring groups. For this reason, negative values in the vector $\vec{\mathbf{P}}_i$ are physically realistic.

Neutral molecules should have a total surface charge of zero. The quantumchemical results used in this work lead to surface charges that deviate from neutrality by typically less than 0.02% (max. 0.2%) of the sum of all absolute charges. The averaged profiles show a higher deviation of 0.17% (average over 1300 components). For the same group of components GC-COSMO leads to 1.09%. This higher deviation results from relatively few estimated profiles, which should not be used for COSMO-RS calculations. Significant deviations from electro neutrality were observed only for highly substituted fluoro-chloro-hydrocarbons.

In several cases adding the group profile contributions resulted in a charge density profile with small negative probability values at one ore more charge density values. These values were set to zero and the negative value was added to the neighboring lower charge interval.

To derive the group contribution vectors $\vec{\mathbf{P}}_i$, a multi-linear regression was performed for each of the 80 elements of the σ profile vectors of the 1363 components in the data bank. Separate regressions were performed for both the hydrogen bonding and nonhydrogen bonding parts of the profiles and a further single regression for the total cavity volume.

Structural groups and correction groups were taken from the method of Nannoolal et al.[25] This method also employs so-called group-interaction contributions to model nonadditive intermolecular interactions in the liquid phase. For the estimation of σ profiles these contributions were not taken into account.

The list of structural groups, group contribution vectors, and cavity volume contributions are available in the complementary material of this article.

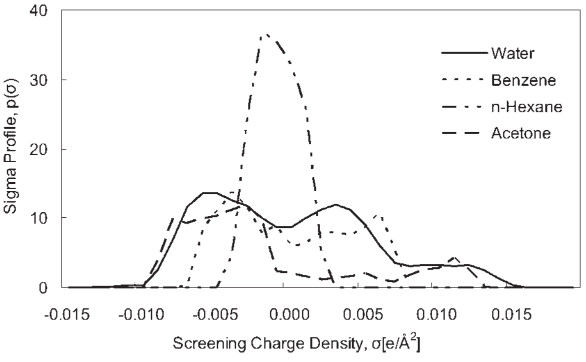

Figure 1. σ profiles of four representative compounds.

Computational details

The original molecular structures of the compounds used in COSMO calculations were taken from the chemical structure database (ChemDB) of the DDB. A MOPAC molecular mechanics optimization was performed for every compound to guarantee a reliable starting geometry. The COSMO method implemented in Gaussian 03[26] was used for the quantum chemical calculation. In our earlier research,[27] the performance of several model chemistries was compared and the results showed that generally the σ profiles based on G03 B3LYP provided the best results. So in this research, σ profiles from G03 B3LYP were chosen for the regression. The

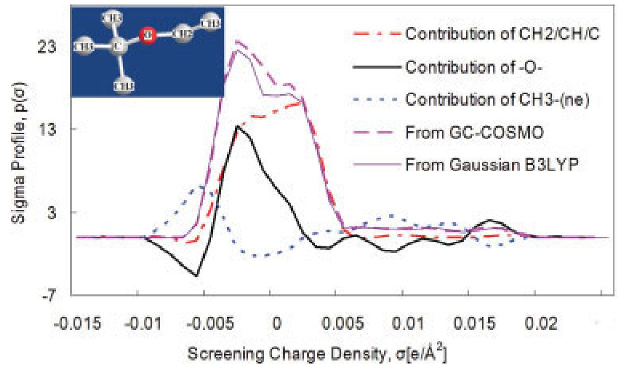

Figure 2. Example for the estimation of the nonhydrogen bonding part of the σ profile of ethyl *tert*-butyl ether by GC-COSMO.

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

COSMO calculations were carried out using the DFT level B3LYP/6-311G(d, p).

The atomic radii of Klamt as delivered with the Gaussian program package were used directly for the cavity construction, since these radii are not sensitive to the quantum mechanical method used.[4]

Results and Discussion

σ profile

Figure 2 shows the calculated and estimated hydrogen bonding σ profiles of ethyl *tert*-butyl ether together with the group contribution vectors of the individual groups.

The list and frequency of the group in the molecule is given in Table 1.

Figures 3–6 contain further typical examples for calculated and estimated averaged σ profiles.

Although the estimated σ profiles shown here are very similar to the calculated profiles, this does not mean that pure component and mixture behavior calculated from these profiles is also very close in every case. The results of COSMO-RS(Ol) calculations for γ^{∞} and VLE data will therefore be used as a criterion for the performance of GC-COSMO.

Table 1. The Definition and Description of Groups Used in Ethyl *tert*-Butyl Ether

Group	Description	Name	Frequency
—O—	O connected to two neighbors which are each either C or Si (ethers)	(C,Si)—O—(C,Si)(z)	1
$-CH_3$	CH_3 not connected to either N, O, F or Cl	CH_3-(ne)	4
>C<	CH2/CH/C in a chain attached to F,Cl,O,N	$>C_{(c)}<$(e)	2

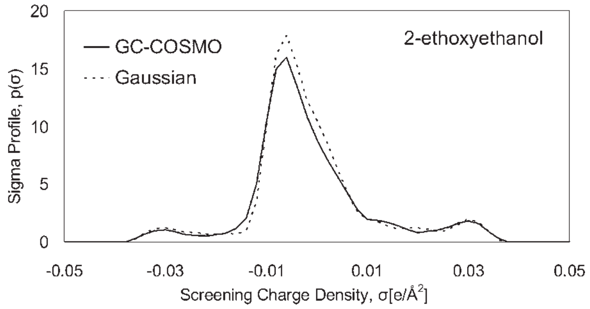

Figure 3. The nonhydrogen bonding σ profile of 2-ethoxyethanol calculated by GC-COSMO and G03 B3LYP.

Results for activity coefficients at infinite dilution

Activity coefficients at infinite dilution usually represent the largest deviation from ideal behavior and are of great importance for the design of distillation columns, the selection of entrainers and environmental protection. The DDB contains ~47,000 experimental values, which were used to test COSMO-RS(Ol) using σ profiles from quantum chemical calculations and GC-COSMO. γ^{∞}-values larger than 500 were excluded from the test set as these values are often of questionable accuracy. For a number of systems, σ profiles were not available so that a total number of 13,254 γ^{∞}-values were used for the test.

The relative average deviations (RAD) between estimated and experimental data were calculated by the following equation:

$$\Delta\gamma_{\mathrm{rel}}\% = \frac{1}{n}\left(\sum \frac{|\gamma^{\infty}_{\mathrm{calc}} - \gamma^{\infty}_{\mathrm{exp}}|}{\gamma^{\infty}_{\mathrm{exp}}}\right) \times 100 \qquad (7)$$

where n is the number of data points, $\gamma^{\infty}_{\mathrm{calc}}$ and $\gamma^{\infty}_{\mathrm{exp}}$ are the calculated and experimental activity coefficients at infinite dilution. The RAD for activity coefficients at infinite dilution predicted by COSMO-RS(Ol) using calculated and estimated σ profiles and also by UNIFAC and modified UNIFAC(Do) are given in Figure 7 and selected deviations for combinations of different compound classes are listed in Table 2.

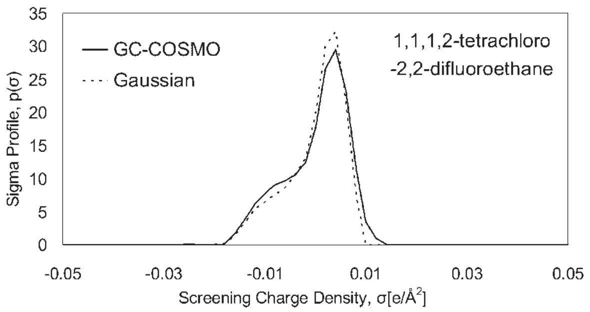

Figure 4. The σ profile of 1,1,1,2-tetrachloro-2,2-difluoroethane [R112a] calculated by GC-COSMO and G03 B3LYP.

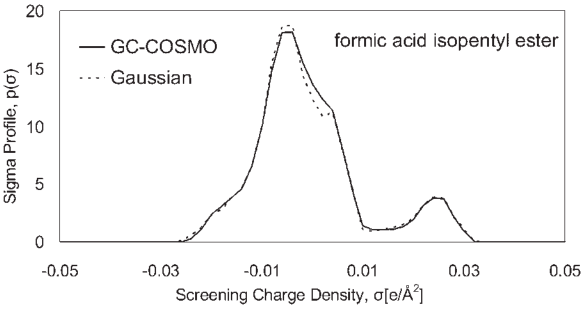

Figure 5. The nonhydrogen bonding σ profile part of formic acid isopentyl ester calculated by GC-COSMO and G03 B3LYP.

From Figure 7, it can be concluded that the mod. UNIFAC(Do) method provides the best results while the results from GC-COSMO are only slightly worse than the results of COSMO-RS(Ol) with calculated σ profiles.

Table 2 shows that in case of nonpolar compounds, σ profiles from GC-COSMO lead to similar results than original profiles. In case of mixtures of nonpolar and polar compounds (nonpolar solutes in polar solvents and vice versa), estimated σ profiles lead to larger deviations for example for hydrocarbons in alcohols or chlorinated hydrocarbons. In several cases though, the results using estimated σ profiles are better than those achieved with the original calculations. In these cases, group contribution regression might have smoothed out differences in individual profiles.

In mixtures of polar compounds, no significant differences were observed but the errors are mostly large for both types of profiles. Especially large deviations were found for amines in water.

Vapor–liquid equilibrium results

The reliability to predict the vapor–liquid equilibrium behavior of binary and multicomponent mixtures is of great importance for every new estimation method for the real liquid mixture behavior. Therefore the performance of COSMO-RS(Ol) with σ profiles obtained from G03 B3LYP in comparison to σ profiles from GC-COSMO methods was investigated.

For the comprehensive comparison of the two methods, experimental data stored in DDB were used. First the consistency of VLE data was checked by the Redlich-Kister area test and the point-to-point test proposed by van Ness et al.[28] and only complete consistent VLE data sets were selected. Furthermore VLE data sets for which the pure component vapor pressures were not given by the authors were excluded. At the end, 1623 binary VLE data were used for the model comparison.

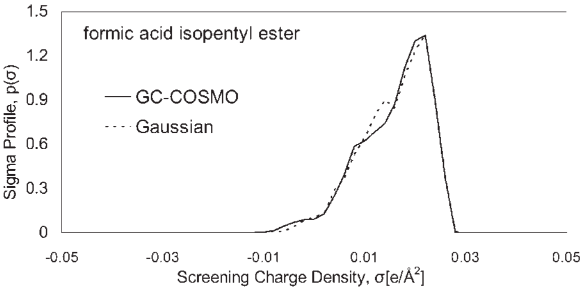

Figure 6. The hydrogen bonding σ profile part of formic acid isopentyl ester calculated by GC-COSMO and G03 B3LYP.

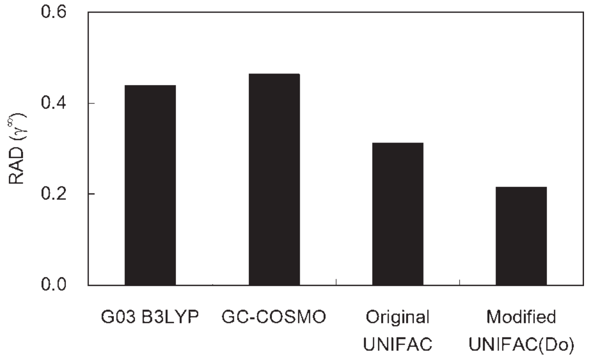

Figure 7. RAD between experimental and calculated γ^∞ based on different models (13,254 data points).

The relative deviation in activity coefficient, the absolute deviation in vapor phase composition and the relative deviation in system pressure or absolute deviation in temperature were calculated using Eqs. 7–10 for σ profiles from G03 B3LYP and GC-COSMO. Overall results for all types of binary mixtures are given in Table 3.

$$\Delta y_{\mathrm{abs}} = \frac{1}{2}\sum\frac{1}{n}\left(\sum\left|y_{k,i} - y_{i,i,\mathrm{calc}}\right|\right) \tag{8}$$

$$\Delta P_{\mathrm{rel}}\% = \frac{1}{n}\left(\sum\left|\frac{P_k - P_{k,\mathrm{calc}}}{P_k}\right|\right)\times 100 \tag{9}$$

$$\Delta T_{\mathrm{abs}} = \frac{1}{n}\left(\sum\left|T_k - T_{k,\mathrm{calc}}\right|\right) \tag{10}$$

Table 3 shows that GC-COSMO is a reliable method for the estimation of the required σ profiles.

In Table 4, the $\Delta\gamma_{\mathrm{rel}}\%$, Δy_{abs}, $\Delta P_{\mathrm{rel}}\%$ and ΔT_{abs} for several types of mixtures (combinations of component classes) based on G03 B3LYP and GC-COSMO are listed. In case both calculations lead to very similar deviations, only the results for GC-COSMO are presented in Table 5.

Tables 4 and 5 show results similar to those observed in case of activity coefficients at infinite dilution. GC-COSMO provides reliable results with a moderately increased deviation.

Results for the predicted cavity volumes

The cavity volume of the molecules is required to calculate the combinatorial part of the activity coefficients. The

Table 2. $\Delta\gamma_{rel}$% **for Mixtures Composed of Different Compound Classes (Based on 17,774 Experimental** γ^{∞}**-Values)**

Solute	Solvent	n_{data}*	G03 B3LYP	GC-COSMO
Nonpolar-nonpolar systems				
n-Alkanes	*n*-Alkanes	528	5.27	5.20
Saturated HC[†]	Saturated HC	1047	7.75	7.47
Alkanes (noncyclic)	Alkanes (noncyclic)	782	5.27	5.19
Alkanes (noncyclic)	Alkanes (cyclic)	36	17.20	16.13
Alkanes (cyclic)	Alkanes (noncyclic)	218	15.20	14.31
Alkanes (cyclic)	Alkanes (cyclic)	11	5.67	5.85
Aromatic HC	Aromatic HC	34	13.32	13.11
Aromatic HC	Saturated HC	347	33.24	24.50
Saturated HC	Aromatic HC	194	16.04	20.52
Alkenes (HC)	Alkenes (HC)	18	8.50	9.39
Alkenes (HC)	Alkenes (cyclic C=C)	9	10.48	11.75
Alkenes (cyclic C=C)	Alkenes (HC)	9	11.52	12.57
Alkenes (cyclic C=C)	Alkenes (cyclic C=C)	4	16.47	18.01
Hydrocarbons (HC)	Aromatic HC	281	16.63	19.91
Aromatic HC	Hydrocarbons (HC)	392	30.77	23.01
Hydrocarbons (HC)	Ethers (HC)	185	19.33	20.79
Ethers (HC)	Hydrocarbons (HC)	127	12.48	10.55
Aromatic HC	Ethers (HC)	22	10.90	9.13
Ethers (HC)	Aromatic HC	16	17.44	19.01
Nonpolar solutes in polar solvents				
Hydrocarbons (HC)	Alcohols (HC)	1306	37.89	49.37
Hydrocarbons (HC)	Ketones (HC, unconj.)	401	38.98	42.34
Hydrocarbons (HC)	Ketones (HC, conj.)	93	42.48	46.13
Hydrocarbons (HC)	Esters (HC, unconj.)	268	19.60	24.28
Hydrocarbons (HC)	Fluorinated (HC)	43	58.25	63.62
Hydrocarbons (HC)	Chlorinated (HC)	256	24.53	36.74
Hydrocarbons (HC)	Halogenated Compounds	601	41.82	51.31
Hydrocarbons (HC)	Amines	700	53.34	60.01
Hydrocarbons (HC)	Nitroalkanes	278	44.71	36.59
Aromatic HC	Alcohols (HC)	272	41.26	46.19
Aromatic HC	Ketones (HC, unconj.)	62	27.82	28.28
Aromatic HC	Ketones (HC, conj.)	16	23.34	22.58
Aromatic HC	Esters (HC, unconj.)	48	8.87	12.74
Aromatic HC	Fluorinated (HC)	3	48.05	50.43
Aromatic HC	Chlorinated (HC)	47	15.16	20.77
Aromatic HC	Halogenated Compounds	90	23.61	27.40
Aromatic HC	Amines	140	36.52	40.71
Aromatic HC	Nitroalkanes	53	20.59	17.02
Ethers (HC)	Alcohols (HC)	64	18.82	21.99
Ethers (HC)	Ketones (HC, unconj.)	3	32.87	33.69
Polar solutes in nonpolar solvents				
Alcohols (HC)	Hydrocarbons (HC)	575	58.35	72.55
Alcohols (HC)	Aromatic HC	102	61.17	69.86
Alcohols (HC)	Ethers (HC)	24	29.16	36.61
Ketones (HC, unconj.)	Ethers (HC)	5	26.58	25.94
Ketones (HC, unconj.)	Hydrocarbons (HC)	291	32.53	39.75
Ketones (HC, unconj.)	Aromatic HC	39	36.54	40.47
Amines	Hydrocarbons (HC)	55	46.60	36.23
Nitroalkanes	Hydrocarbons (HC)	77	47.46	50.48
Nitroalkanes	Aromatic HC	14	16.35	21.58
Esters (HC, unconj.)	Hydrocarbons (HC)	203	18.75	12.78
Esters (HC, unconj.)	Aromatic HC	15	15.37	19.83
Fluorinated (HC)	Hydrocarbons (HC)	31	33.98	39.49
Chlorinated (HC)	Hydrocarbons (HC)	338	22.67	20.09
Halogenated Compounds	Hydrocarbons (HC)	400	25.63	24.12
Fluorinated (HC)	Aromatic HC	5	35.49	38.45
Chlorinated (HC)	Aromatic HC	56	16.36	17.63
Halogenated Compounds	Aromatic HC	64	18.60	20.14
Polar-polar systems				
Chlorinated (HC)	1-Alcohols (HC)	168	39.46	36.81
Halogenated Compounds	1-Alcohols (HC)	221	44.32	44.01
1-Alcohols (HC)	1-Alcohols (HC)	34	7.23	8.23
Alcohols (HC)	Alcohols (HC)	64	12.60	12.72
Esters (HC, unconj.)	Alcohols (HC)	36	46.67	47.89
Alcohols (HC)	Esters (HC, unconj.)	27	32.49	33.11
Ketones (HC, unconj.)	Alcohols (HC)	110	45.47	48.42
Alcohols (HC)	Ketones (HC, unconj.)	44	39.69	40.36
Nitroalkanes	Alcohols (HC)	14	58.65	64.01

Table 2. Continued

Solute	Solvent	n_{data}*	G03 B3LYP	GC-COSMO
Alcohols (HC)	Nitroalkanes	6	69.56	75.59
Amines	Alcohols (HC)	15	102.20	87.33
Alcohols (HC)	Amines	20	50.15	46.57
Nitroalkanes	Ketones(HC, unconj.)	8	12.35	15.08
Ketones (HC, unconj.)	Nitroalkanes	4	34.50	24.67
Aldehydes (HC, unconj.)	Alcohols (HC)	13	142.08	144.44
Compounds in Water				
1-Alcohols (HC)	Water	196	50.79	46.35
Alcohols (HC)	Water	288	70.27	68.98
Secondary alcohols (HC)	Water	43	143.49	211.72
tertiary alcohols (HC)	Water	13	322.61	321.03
Diols, triols	Water	20	268.72	337.80
Ketones (HC, unconj.)	Water	94	33.20	38.34
Carboxylic Acids (HC, unconj.)	Water	11	50.58	54.97
Amines	Water	32	849.91	559.11
Epoxides (HC)	Water	11	70.87	32.81
Aldehydes (HC, unconj.)	Water	48	48.94	58.63
Aldehydes (HC, conj.)	Water	14	84.65	88.39
Esters (HC, unconj.)	Water	140	73.55	74.91
Esters (HC, conj.)	Water	19	89.41	90.17
Water in other compounds				
Water	1-Alcohols (HC)	29	19.20	18.12
Water	Alcohols (HC)	43	20.10	18.55
Water	Secondary Alcohols (HC)	7	23.86	19.12
Water	Ketones(HC, unconj.)	18	60.22	59.57
Water	Amines	5	50.13	47.95
Water	Diols, triols	35	29.24	28.90
Water	Esters (HC, unconj.)	11	70.82	70.41

*Number of data points.
†In this paper, HC denotes components that contain only the further elements H and C.

cavity volumes V_{COSMO} of all compounds in the training set were calculated via GC-COSMO and compared to the original values in Figure 8. The RMSD and the relative deviation of V_{COSMO} were calculated with the following equations.

$$\mathrm{RMSD} = \sqrt{\frac{1}{n}\sum_{n}\left(V_{\mathrm{COSMO}}^{\mathrm{GC}} - V_{\mathrm{COSMO}}^{\mathrm{Gaussian}}\right)^2} \tag{11}$$

$$\Delta V_{\mathrm{COSMO}}^{\mathrm{rel}}\% = \frac{1}{n}\left(\sum \frac{\left|V_{\mathrm{COSMO}}^{\mathrm{Gaussian}} - V_{\mathrm{COSMO}}^{\mathrm{GC}}\right|}{V_{\mathrm{COSMO}}^{\mathrm{Gaussian}}}\right)\times 100\% \tag{12}$$

where the $V_{\mathrm{COSMO}}^{\mathrm{Gaussian}}$ and $V_{\mathrm{COSMO}}^{\mathrm{GC}}$ denote the cavity volume of molecules calculated with the quantum mechanics method G03 B3LYP and GC-COSMO method respectively. The RMSD of the V_{COSMO} is 3.23 and the $\Delta V_{\mathrm{COSMO}}^{\mathrm{rel}}\%$ is 1.74% for all 1325 compounds taken into account. It can be concluded from Figure 8 and the deviation results calculated that the predicted V_{COSMO} results are in good agreement with the values calculated using quantum mechanics calculated results.

Limitations of GC-COSMO method

In a number of cases, poor results were obtained when applying GC-COSMO profiles. One typical example is nitrobenzene. Figure 9 shows both the calculated and estimated profiles. It can be seen that significant differences are observed.

In nitrobenzene, all atoms are connected by a mesomeric π-electron system and at the same time the hetero-atoms N and O have an inductive effect on the delocalized electrons. The database for the regression of GC-COSMO included 33 components in which a nitro-group is connected to an aromatic system. Their σ profiles did not follow a simple scheme of group additivity.

During the careful analysis of the σ profiles obtained with G03 B3LYP and GC-COSMO, we found that using GC-COSMO poor results for chlorosilanes, thiols, sulfides, and some highly conjugated compounds are observed, e.g. for 2-butyne-1,4-diol (Figure 10), etc. One should be careful when applying GC-COSMO to systems containing these components.

Final groups used and parameters

The final groups used and group parameters for the nonhydrogen bonding σ profile part and hydrogen bonging σ profile part are given as supporting information.

Conclusions

In this research, a group contribution method for the estimation of averaged σ profiles for COSMO-RS(Ol) (GC-

Table 3. Comparison of the Mean Deviation of the Activity Coefficients and Vapor–Liquid Equilibria

	G03 B3LYP	GC-COSMO
$\Delta\gamma_{\mathrm{rel}}\%$ (1623 data sets)	10.40	11.88
Δy_{abs} (1623 data sets)	0.0203	0.0243
$\Delta P_{\mathrm{rel}}\%$ (837 data sets)	6.27	7.25
ΔT_{abs} (786 data sets)	1.75	2.01

Table 4. $\Delta\gamma_{rel}\%$, $\Delta y_{abs}\%$, $\Delta P_{rel}\%$, ΔT_{abs} **for Mixtures Composed of Different Compound Classes,** n_{data} **is the Number of Data Sets**

	$\Delta\gamma_{rel}\%$		$\Delta y_{abs}\%$		$\Delta P_{rel}\%$		ΔT_{abs}	
	G03 B3LYP	GC-COSMO	G03 B3LYP	GC-COSMO	G03 B3LYP	GC-COSMO	G03 B3LYP	GC-COSMO
Nonpolar-Nonpolar Systems								
		Aromatic HC				Saturated HC		
n_{data}	98		44		98		54	
	1.98	2.87	1.48	2.21	0.45	0.60	0.28	0.44
		Hydrocarbons (HC)				Aromatic HC		
n_{data}	147		65		147		82	
	2.14	2.78	1.40	1.96	0.42	0.53	0.24	0.35
		Aromatic HC				Ethers (HC)		
n_{data}	5		4		5		1	
	4.88	4.56	3.16	2.92	0.89	0.83	1.02	0.98
Nonpolar-Polar Systems								
		Hydrocarbons (HC)				Alcohols (HC)		
n_{data}	216		126		216		90	
	11.18	15.90	7.93	12.57	2.82	4.27	2.58	3.39
		Aromatic HC				Alcohols (HC)		
n_{data}	72		23		72		49	
	13.97	16.10	10.69	12.29	3.73	4.38	2.67	3.26
		Ethers (HC)				Alcohols (HC)		
n_{data}	26		8		26		18	
	8.58	11.63	8.01	10.20	2.36	3.27	1.47	2.10
		Hydrocarbons (HC)				Esters (HC, unconj.)		
n_{data}	45		32		45		13	
	3.81	6.40	1.48	2.87	0.79	1.24	0.75	1.13
		Aromatic HC				Esters (HC, unconj.)		
n_{data}	3		1		3		2	
	4.16	5.13	2.16	2.37	0.63	0.87	0.89	1.15
		Hydrocarbons (HC)				Chlorinated (HC)		
n_{data}	113		87		113		26	
	3.68	5.65	2.44	3.80	0.68	1.08	0.62	1.07
		Aromatic HC				Chlorinated (HC)		
n_{data}	56		43		56		13	
	3.77	5.09	2.76	3.42	0.62	0.84	0.58	0.98
		Hydrocarbons (HC)				Nitroalkanes		
n_{data}	12		12		12		0	
	10.90	6.93	9.25	5.89	2.71	1.65		
		Aromatic HC				Nitroalkanes		
n_{data}	5		5		5		0	
	4.47	3.37	2.67	2.12	1.08	0.67		
Polar-Polar Systems								
		Alcohols (HC)				Amines		
n_{data}	14		11		14		3	
	20.74	18.36	7.80	6.99	2.75	2.42	2.43	1.85
		Alcohols (HC)				Esters (HC, unconj.)		
n_{data}	63		18		63		45	
	11.72	12.75	8.05	9.06	2.72	2.96	2.05	2.23

Table 5. $\Delta\gamma_{rel}\%$, $\Delta y_{abs}\%$,$\Delta P_{rel}\%$, ΔT_{abs} for Mixtures Composed of Different Compound Classes Based on GC-COSMO, n_{data} is the Number of Data Sets

		$\Delta\gamma_{rel}\%$	$\Delta y_{abs}\%$	$\Delta P_{rel}\%$	ΔT_{abs}
n-Alkanes	n_{data}	7	5	7	2
	n-Alkanes	3.53	2.46	0.77	0.62
Saturated HC	n_{data}	32	20	32	12
	Saturated HC	2.14	1.36	0.48	0.28
Alkanes (noncyclic)	n_{data}	16	14	16	2
	Alkanes (noncyclic)	2.40	1.31	0.53	0.62
Alkanes (noncyclic)	n_{data}	15	5	15	10
	Alkanes (cyclic)	1.97	1.75	0.44	0.21
Aromatic HC	n_{data}	38	12	38	26
	Aromatic HC	0.85	0.39	0.22	0.09
Hydrocarbons (HC)	n_{data}	31	16	31	15
	Ethers (HC)	2.87	2.39	0.53	0.28
Ethers (HC)	n_{data}	5	4	5	1
	Ketones (HC, unconj.)	11.35	7.05	3.07	7.17
Hydrocarbons (HC)	n_{data}	69	41	69	28
	Ketones (HC, unconj.)	11.43	9.25	2.54	2.14
Hydrocarbons (HC)	n_{data}	4	1	4	3
	Ketones (HC, conj.)	12.32	19.88	1.81	1.63
Aromatic HC	n_{data}	29	16	29	13
	Ketones (HC, unconj.)	9.49	7.07	1.82	1.72
Aromatic HC	n_{data}	3	0	3	3
	Ketones (HC, conj.)	9.83		2.12	1.63
Hydrocarbons (HC)	n_{data}	36	31	36	5
	Fluorinated (HC)	8.28	6.62	2.65	0.60
Aromatic HC	n_{data}	25	20	25	5
	Fluorinated (HC)	5.61	3.91	1.23	0.60
Hydrocarbons (HC)	n_{data}	22	18	22	4
	Amines	6.70	5.19	1.59	0.80
Aromatic HC	n_{data}	11	9	11	2
	Amines	7.09	6.95	1.72	0.92
Alcohols (HC)	n_{data}	36	14	36	22
	Alcohols (HC)	4.49	1.61	0.85	0.61
Ketones (HC, unconj.)	n_{data}	8	0	8	8
	Ketones (HC, unconj.)	2.66		0.44	0.29
Alcohols (HC)	n_{data}	14	1	14	13
	Ketones (HC, unconj.)	19.47	9.93	3.30	2.53
Alcohols (HC)	n_{data}	4	2	4	2
	Nitroalkanes	22.14	18.93	6.35	4.08

COSMO) was proposed. The theory and procedures of application were described and predicted thermophysical properties were compared to those calculated using profiles from quantum chemical calculations. In general, the GC-COSMO method provides reliable σ profiles and cavity volumes.

Good results were obtained for alkanes, aromatics, alkenes, esters, ethers, ketones, epoxides, anhydrides, alcohols, carboxylic acids, etc. However, poor results were found for the σ profiles of chlorosilanes, thiols, sulfides, and some highly conjugated compounds.

GC-COSMO presents a shortcut for the generation of σ profiles and is of great value especially in case of large molecules, for which the quantum chemical calculations are very time consuming. In case of small molecules that mostly do not follow the trend in their respective homologous series and where the rigorous calculation is relatively inexpensive, GC-COSMO should not be employed. Thus GC-COSMO can

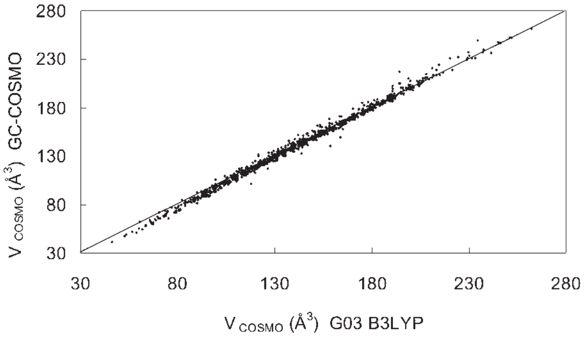

Figure 8. V_{COSMO} **based on GC-COSMO vs. G03 B3LYP (1325 data points).**

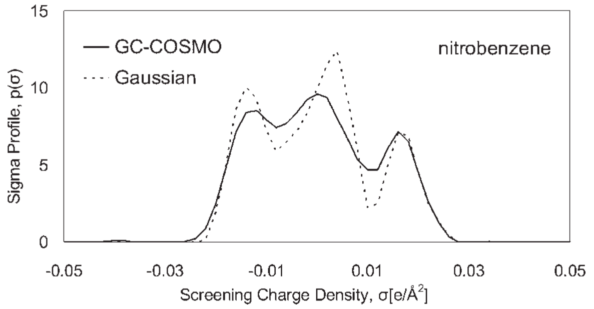

Figure 9. σ **profiles of nitrobenzene based on GC-COSMO and G03 B3LYP.**

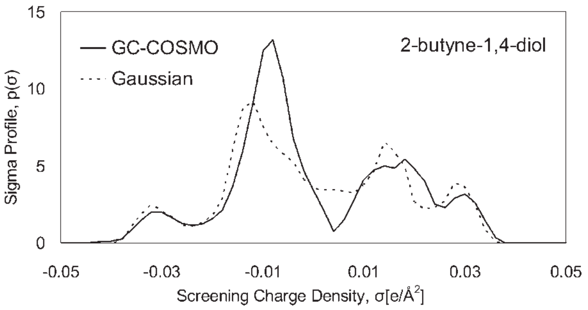

Figure 10. σ profiles of 2-butyne-1,4-diol based on GC-COSMO and G03 B3LYP.

be considered a complimentary method rather than a replacement of the ab initio or DFT methods.

A further extension and improvement of the method is under development. This work is to a significant part aimed at generating profiles for molecules with structural groups, for which up to now no or only limited information is available.

Acknowledgments

The authors thank the Deutsche Forschungsgemeinschaft SPP-1155 for financial support of the ongoing research project. We also thank the DDBST GmbH for providing the latest version of the Dortmund Data Bank for the model comparison.

Literature Cited

1. Katritzky AR, Lobanov VS, Karelson M. QSPR-The correlation and quantitative prediction of chemical and physical properties from structure. *Chem Soc Rev*. 1995;24:279–287.
2. Klamt A. Conductor-like screening model for real solvents—a new approach to the quantitative calculation of solvation phenomena. *J Phys Chem*. 1995;99:2224–2235.
3. Klamt A, Jonas V, Burger T, Lohrenz JCW. Refinement and parameterization of COSMO-RS. *J Phys Chem A*. 1998;102:5074–5085.
4. Klamt A, Eckert F. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib*. 2000;172:43–72.
5. Fredenslund A, Gmehling J, Rasmussen P. *Vapor-Liquid-Equilibia Using UNIFAC*. Amsterdam: Elsevier, 1977.
6. Gmehling J. Present status of group-contribution methods for the synthesis and design of chemical processes. *Fluid Phase Equilib*. 1998;144:37–47.
7. Gmehling J, Li JD, Schiller M. Modified UNIFAC model. II. Present parameter matrix and results for different thermodynamic properties. *Ind Eng Chem Res*. 1993;32:178–193.
8. Amovilli C, Barone V, Cammi R, Cances E, Cossi M, Mennucci B, Pomelli CS, Tomasi J. Recent advances in the description of solvent effects with the polarizable continuum model. *Adv Quantum Chem*. 1999;32:227–261.
9. Klamt A, Schüürmann G. COSMO–a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc-Perkin Trans*. 1993;2:799–805.
10. Truong TN, Stefanovich EV. A new method for incorporating solvent effect into the classical, *ab-initio* molecular-orbital and density-functional theory frameworks for arbitrary shape cavity *Chem Phys Lett*. 1995;240:253–260.
11. Barone V, Cossi M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J Phys Chem A*. 1998;102:1995–2001.
12. Schafer A, Klamt A, Sattel D, Lohrenz JCW, Eckert F. COSMO Implementation in TURBOMOLE: extension of an efficient quantum chemical code towards liquid systems. *Phys Chem Chem Phys*. 2000;2:2187–2193.
13. Andzelm J, Kolmel C, Klamt A. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *J Chem Phys*. 1995;103:9312–9320.
14. Baldridge K, Klamt A. First principles implementation of solvent effects without outlying charge error. *J Chem Phys*. 1997;106:6622–6633.
15. Lin ST, Sandler SI. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind Eng Chem Res*. 2002;41:899–913.
16. Grensemann H, Gmehling J. Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. *Ind Eng Chem Res*. 2005;44:1610–1624.
17. Grensemann H. Weiterentwicklung thermodynamischer Vorhersagemodelle durch Einsatz von quantenchemischen Methode. Thesis, Carl-von-Ossietzky, Universität Oldenburg, 2005.
18. Dortmund Data Bank and DDB Software Package, DDBST GmbH, Oldenburg, Germany, 2006 (Available at: http://www.ddbst.de).
19. Klamt A. *COSMO-RS From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*. Amsterdam: Elsevier, 2005.
20. Mullins E, Oldland R, Liu YA, Wang S, Sandler SI, Chen CC, Zwolak M, Seavey KC. σ Profile database for using COSMO-based thermodynamic methods. *Ind Eng Chem Res*. 2006;45:4389–4415.
21. Hornig M, Klamt A. COSMOfrag: a novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry. *J Chem Info Comp Sci*. 2005;45:1169–1177.
22. Eckert F, Klamt A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J*. 2002;48:369–385.
23. Lee CT, Yang WT, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys Rev B*. 1988;37:785–789.
24. Becke AD. Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys*. 1993;98:5648–5652.
25. Nannoolal Y, Rarey J, Ramjugernath D, Cordes W. Estimation of pure component properties, Part 1: Estimation of the normal boiling point of non-electrolyte organic compounds via group contributions and group interactions. *Fluid Phase Equilib*. 2004;226:45–63.
26. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA. Gaussian 03, Revision C. 02, Gaussian: Wallingford, CT, 2004.
27. Mu TC, Rarey J, Gmehling J. Performance of COSMO-RS with sigma profiles from different model chemistries. *Ind Eng Chem Res*. 2007;46:6612–6629.
28. van Ness HC, Byer SM, Gibbs RE. Vapor-liquid equilibrium. I. Appraisal of data reduction methods. *AIChE J*. 1973;19:238–244.

Manuscript received Apr. 3, 2007, and revision received Sept. 5, 2007.