Group Contribution Prediction of Surface Charge Density Distribution of Molecules for COSMO-SAC

Tiancheng Mu
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.11933

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

Introduction

COSMO-RS[1,2] methods was developed by incorporating “COSMO” (the conductor-like screening model[3]) with statistical thermodynamics to perform calculations in chemical thermodynamics. COSMO is a variant of continuum solvation models and was implemented in Gaussian[4,5], TURBOMOLE[6], and other software packages. Further variants of COSMO-RS including COSMO-SAC (COSMO Segment Activity Coefficient)[7] and COSMO-RS (Ol)[8] were developed recently.

The surface charge density distribution (σ profile) for each component in the mixture is a prerequisite to perform a calculation in all COSMO-RS type models. Until now only one database of s profiles is available free of charge for 1423 compounds[9]. The σ profile calculation is time-consuming, especially in case of large molecules. A group contribution method for the prediction of averaged σ profiles (GC-COSMO-RS(Ol)) has been published[10] as a shortcut to generate the required basic compound information for COSMO-RS(Ol). In this article, the GC-COSMO-SAC method for the prediction of σ profiles for COSMO-SAC is presented.

Theory and Model

The basic idea of GC-COSMO is to generate the molecule σ profiles by summing up group contributions. The database used for regression of group contribution parameters comes from quantum chemistry calculations (Gaussion 03 package with B3LYP DFT functional). The molecular structures of the compounds are directly taken from chemical structure database (ChemDB) of the DDB[11] and further optimized using a high level of theory. No conformation effects were taken into account in the σ profiles calculation. The detailed theory and model of COSMO-SAC and GC-COSMO were given by references 7 and 10, respectively. In COSMO-SAC, a σ profile contains 50 segments, each 0.001 e/Å^2 wide, in the range -0.025 to 0.025 e/Å^2. The data base used to regress the new σ profiles of compounds was enlarged from 1363 to 1776 since the development of GC-COSMO-RS(Ol). Structural groups and correction groups are identical with GC-COSMO-RS(Ol).[8]

Since σ profiles based on Gaussian 03[12] B3LYP (G03 B3LYP) provided the best results when performing COSMO-SAC calculations, these were chosen for the regression.[13]

Results and Discussion

Figure 1 presents the calculated and estimated s profile of 2-heptanone as an example for a good agreement, while Figure 2 shows the example of 1-buten-3-yne where less perfect agreement was observed. As described below, strong conjugation of p bonds is one of the reasons for the discrepancy.

The shapes of the calculated and estimated σ profiles are in good agreement for most compounds. To check the reliability of GC-COSMO-SAC, predictions by COSMO-SAC using the two different profiles were compared with the experimental results for a large amount of and VLE data.

The activity coefficients at infinite dilution and the vapor-liquid equilibria data used to test COSMO-SAC using σ profiles from G03 B3LYP and GC-SOSMO-SAC were taken from the Dortmund Data Bank. The selection criteria for the

---

Correspondence concerning this article should be addressed to T. Mu at tcmu@chem.ruc.edu.cn or J. Gmehling at gmehling@tech.chem.uni-oldenburg.de

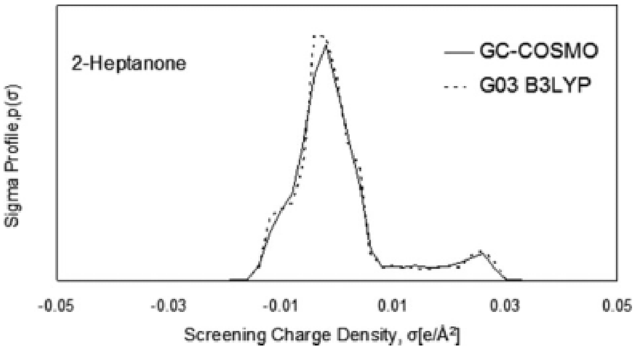

Figure 1. The σ profile of 2-heptanone calculated by GC-COSMO-SAC and G03 B3LYP.

data used in this comparison are described in reference 10 in detail. In total, 17778 -values and 1623 binary VLE data sets were used for the comparison. VLE calculations were performed assuming ideal vapor phase behavior except in case of carboxylic acids, where the chemical theory was applied.

The relative average deviations (RAD) in activity coefficients, the absolute deviations in vapor phase composition, as well as the relative deviation in system pressure and temperature were calculated using Eqs. 1–4 to check the prediction quality of the COSMO-SAC method for activity coefficients and VLE data. The total deviations of these properties are reported in Figure 3 and Table 1.

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

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

$$\Delta P_{\mathrm{rel}}\% = \frac{1}{n}\left(\sum \left|\frac{P_k - P_{k,\mathrm{calc}}}{P_k}\right|\right) \times 100 \tag{3}$$

$$\Delta T_{\mathrm{abs}} = \frac{1}{n}\left(\sum \left|T_k - T_{k,\mathrm{calc}}\right|\right) \tag{4}$$

where n is the number of data points, the subscript of calc and exp denote the corresponding calculated and experimental data. The RAD results of activity coefficients at infinite dilution predicted by UNIFAC[14] and modified UNIFAC (Do)[15,16] are also given in Figure 3 for comparison.

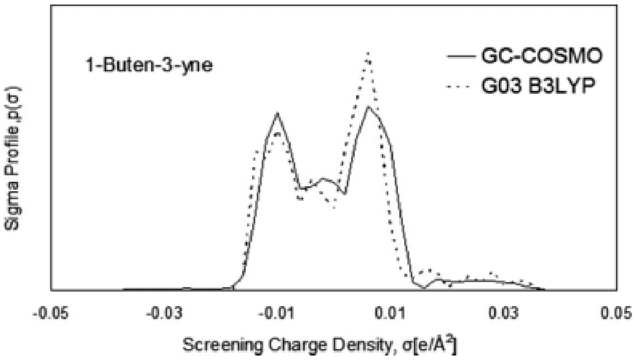

Figure 2. The σ profile of 1-buten-3-yne calculated by GC-COSMO-SAC and G03 B3LYP.

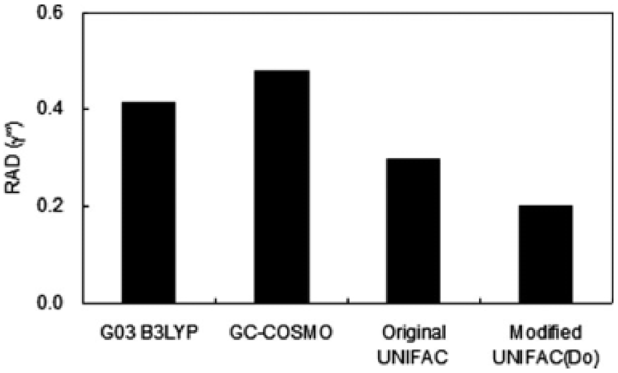

Figure 3. RAD between experimental and calculated γ^{∞} based on different models (12,776 data points).

Both Figure 3 and Table 1 show that the highly trained mod. UNIFAC(Do) method provides the best results while the results from GC-COSMO are only slightly worse than the results of COSMO-SAC with calculated σ profiles based on quantum chemistry.

As in the previous work, deviations for combinations of different compound classes are also checked. Since the results are similar to the results published earlier,[10] only a brief description of the results is given here. In case of nonpolar compounds, σ profiles from GC-COSMO lead to similar results as profiles from quantum chemistry. In case of mixtures of nonpolar and polar compounds, estimated σ profiles lead to larger deviations. In mixtures of polar compounds, the results are similar to nonpolar and polar compounds.

The predicted cavity volume results are in good agreement with the values calculated using quantum mechanics and predefined atomic radius of the atoms.[10]

GC-COSMO-SAC is based on the assumption of additivity of the group contribution profiles. Discrepancies are therefore observed for chlorosilanes, thiols, sulfides and some highly conjugated compounds where this assumption is only partially valid. It is recommended that GC-COSMO-SAC can be applied only to large molecules, for which the quantum chemical calculations are very time consuming and prefer original profiles whenever economically feasible. GC-COSMO-SAC should be considered as complimentary method rather than a replacement of the quantum chemical calculations. The fundamental parameters used in the COSMO-SAC were not optimized using the σ profiles generated by GC-COSMO-SAC method.

Both GC-COSMO-SAC and UNIFAC (including modified UNIFAC) contain a large number of parameters. However, no experimental data are required to obtain the parameters of GC-COSMO-SAC, while these are necessary for the group interaction parameters of the UNIFAC methods.

Conclusions

A group contribution model (GC-COSMO-SAC) to generate the s profiles for COSMO-SAC was developed. The predicted thermophysical properties were compared to those calculated from s profiles calculated by G03 B3LYP. It was shown that GC-COSMO-SAC can generate relatively reliable σ profiles for most compounds. Generally, for alkanes,

Table 1. Comparison of the Mean Deviation of the Activity Coefficients and Vapor-Liquid Equilibria

$\Delta\gamma_{rel}\%$		Δy_{abs}		$\Delta P_{rel}\%$		ΔT_{abs}	
1623 Data sets		1623 Data sets		837 Data sets		786 Data sets	
G03 B3LYP	GC-COSMO	G03 B3LYP	GC-COSMO	G03 B3LYP	GC-COSMO	G03 B3LYP	GC-COSMO
9.38	11.15	0.022	0.027	6.44	7.75	1.66	2.05

aromatics, alkenes, esters, ethers, ketones, epoxides, anhydrides, alcohols, carboxylic acids, the predicted results are good. Larger discrepancies are obtained for chlorosilanes, thiols, sulfides and some highly conjugated compounds. If the predicted σ profiles are poor, the respective parameters are not given in the complementary material.

Acknowledgements

The authors thank the DDBST GmbH for providing the Dortmund Data Bank for the model comparison and DFG (Deutsche Forschungsgemeinschaft SPP-1155) for financial support of the research project. Mu also thanks for the support of State Key Laboratory of Heavy Oil Processing, China University of Petroleum.

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;99:2224–2235.
2. Klamt A, Eckert F. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilibr*. 2000;172:43–72.
3. 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.
4. 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.
5. 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.
6. 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.
7. Lin ST, Sandler SI. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind Eng Chem Res*. 2002;41: 899–913.
8. 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.
9. Mullins E, Oldland R, Liu YA, Wang S, Sandler SI, Chen CC, Zwolak M, Seavey KC. Sigma-profile database for using COSMO-based thermodynamic methods. *Ind Eng Chem Res*. 2006;45:4389–4415.
10. Mu TC, Rarey J, Gmehling J. Group contribution prediction of surface charge density profiles for COSMO-RS(OI). *AIChE J*. 2007;53: 3231–3240.
11. DortmundData Bank and DDB Software Package, DDBST *GmbH*, Oldenburg, Germany, 2008 (HYPERLINK “http://www.ddbst.de”).
12. Frisch MJ, Trucks GW TG, 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.
13. 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.
14. Fredenslund A, Jones RL, Prausnitz JM. Group contribution estimation of activity coeficients in nonideal liquid mixtures. *AIChE J*. 1975;21:1086–1098.
15. Weidlich U, Gmehling J. A modified UNIFAC model. 1. Prediction of VLE, h^E, and Gamma Infinite. *Ind Eng Chem Res*. 1987;26:1372–1381.
16. Gmehling J, Li JD, Schiller M. A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Ind Eng Chem Res*. 1993;32:178–193.

Manuscript received Feb. 6, 2009, and revision received Mar. 19, 2009.