Application of the COSMO−SAC−BP Solvation Model to Predictions of Normal Boiling Temperatures for Environmentally Significant Substances

Shu Wang,[†] Shiang-Tai Lin,[‡] Jaeeon Chang,[†] William A. Goddard III,[‡] and Stanley I. Sandler*,[†]

Center for Molecular and Engineering Thermodynamics, Department of Chemical Engineering,
University of Delaware, Newark, Delaware 19716, and Materials and Process Simulation Center (MSC),
Beckman Institute (139-74), California Institute of Technology, Pasadena, California 91125

We recently reported the COSMO−SAC−BP model for predicting vapor pressure and its temperature derivative, the enthalpy of vaporization. This COSMO−SAC−BP model, which contains no compound specific parameters, is based on determining three major solvation components: (i) an electrostatic contribution, calculated using the quantum mechanical COSMO (conductor-like-screening-model) method with a statistical mechanical correction for solution nonideality (deviation from a perfect conductor); (ii) a dispersion contribution, obtained from a mean field treatment; and (iii) a cavity formation contribution determined from thermodynamic perturbation theory. This COSMO−SAC−BP model was previously validated to successfully correlate normal boiling point temperatures and enthalpies of vaporization for 369 molecules. In this present study, we have extended the COSMO−SAC−BP model to describe large and more-complex molecules, including pollutants, herbicides, insecticides, and drugs,. The average absolute deviation in the predicted boiling points of these complex molecules, which spans the range of 266−708 K is 17.8 K, or 3.7%. This is similar to the value of 3.2% that was obtained for the 369 molecules in the earlier study, indicating that this method can be applied well outside the systems used to train the model. More importantly, we report here the predicted the normal boiling temperatures for 10 pesticides for which no experimental data are available. This illustrates the advantage to the COSMO−SAC−BP model: predicting several properties for a wide variety of molecules simultaneously in a unified framework with few parameters (unlike group contribution methods (or quantitative structure−property relationships).

1. Introduction

The normal boiling temperature of pure components is an essential property for many chemical, biochemical, environmental, and safety studies. A large amount of experimental data for the normal boiling temperatures of common chemicals is available in the literature. However, for large, complex molecules (including some pollutants, insecticides, herbicides, medicinal intermediates, pharmaceuticals, and molecules constructed in computer simulations), there is a paucity of data, partly because the low vapor pressures make experiments difficult to perform.

To estimate this important property, many methods have been developed, as reviewed by Reid et al.[1] Horvath[2] has provided a summary of estimation methods of normal boiling temperatures. Most of these methods are based on the group contribution method (GCM), which assumes that the interactions between groups are the same in different molecules, and that the groups are independent from each other. Joback and Reid[3] used 41 groups in a first-order GCM for correlating the properties of 438 organic compounds, with an average absolute error of 12.9 K and an average absolute percent error of 3.6%. Stein et al.[4] extended that work by increasing the number of groups and using a larger database. They used more than 4000 compounds as the training set for 85 groups, resulting in an average absolute error of 15.5 K, or 3.2%. When they applied their method to more than 6000 other compounds, the errors were 20.4 K, or

4.3%. In addition to these first-order methods, Constantinou and Gani[5] achieved higher accuracy using second-order or nearest-neighbor-bonded group corrections that are based on combinations of first-order UNIFAC-type groups (this adds 43 new second-order groups fitted by comparisons to data on hundreds of compounds in the Design Institute for Physical Property Data (DIPPR) database[6]). They reported an average absolute error of 10.5 K (2.0%) for their training set in the DIPPR database.[6] Other GCMs that were developed by Marrero-Morejon and Pardillo-Fontdevila[7] and Cordes and Rarey[8] are also widely used. Recently, Nannoolal et al.[9] proposed an improved GCM that is based on the method of Cordes et al. in which the experimental data of the normal boiling temperature for ~2800 compounds was regressed to an accuracy of 6.5 K. Limitations of these structural group methods is the restriction to nonpolar molecules, because of the importance of solvation for polar compounds and the empirical aspect of identifying the particular groups to be included in the regression.

Recently, descriptors other than structural groups have been used in QSPR models. For example, Katritzky et al.[10] reported that, with their two-descriptor QSPR model, a standard error of 8 K was obtained for hydrocarbons, 17 K for chlorinated compounds, and 20 K for fluorinated compounds. Ericksen et al.[11] used both structural groups and property descriptors to estimate normal boiling temperatures, resulting in an absolute deviation of 7.75 K (1.9%) for more than 1000 compounds in their training set. Compared to the structural group method, their descriptors have a better physical basis and can be used for polar compounds, although these descriptors do not quantitatively account for the interactions between molecules in the condensed

---

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

[†] University of Delaware.

[‡] California Institute of Technology.

phase. However, the QSPRs and GCMs use different models to correlate different thermodynamic properties.

The COSMO−RS (conductor-like screening model for real solvent) approach of Klamt[12,13] provides a novel way to treat the liquid state for predicting thermodynamic properties. In COSMO theory, energies and charge density distributions for a molecule in a perfect conductor are calculated by ab initio methods. The surface of that molecule then is partitioned into segments, based on a standard segment area, and the segments are assumed to be independent of their neighbors in the perfect conductor phase. The energy for the transition from the conductor state to a real liquid state is then calculated self-consistently from statistical mechanics. This allows the prediction of thermodynamic properties for both mixtures and pure compounds. Lin et al. developed the COSMO−SAC (segment activity coefficients) variation of the COSMO−RS model, to predict the vapor−liquid equilibrium,[14] vapor pressures, and enthalpies of vaporization.[15] A particular advantage of the COSMO-based methods is that they can be used to predict several thermodynamic properties (e.g., normal boiling temperatures and heats of vaporization) simultaneously from the *same model*, which is a significant advantage over QSPRs or GCMs.

The present study extends the COSMO−SAC model for predicting normal boiling points of pure compounds by considering 84 environmentally important chemicals with a variety of functional groups and molecular sizes.

2. Theory

In the COSMO−SAC method, the vapor pressure is obtained from the following expression:

$$\ln P_i^{\text{vap}} = \frac{\Delta \underline{G}_{i/i}^{*\text{sol}}}{RT} + \ln\left(\frac{RT}{\underline{V}_{i/L}}\right) - \ln\left(\frac{f(T,P_i^{\text{vap}})}{P_i^{\text{vap}}}\right) \quad (1)$$

where $\Delta \underline{G}_{i/i}^{*\text{sol}}$ is the Gibbs energy of solvation, $\underline{V}_{i/L}$ is the liquid molar volume of species i, and f is the fugacity coefficient, which accounts for the nonideality of the vapor phase. In this work, because all the calculations are based on the normal boiling point, we assume that the vapor phase is ideal, so that the third term in eq 1 vanishes.

The Gibbs energy of solvation results from the electrostatic and polarization interactions between molecules. It can be written as the sum of a van der Waals (vdW) Gibbs free energy that is due to instantaneous electron correlation and an electrostatic free energy:

$$\Delta \underline{G}_{i/i}^{*\text{sol}} = \Delta \underline{G}_{i/i}^{*\text{vdW}} + \Delta \underline{G}_{i/i}^{*\text{el}} \quad (2)$$

As shown in ref 15, the electrostatic free energy is obtained in this method as follows:

$$\Delta \underline{G}_{i/i}^{*\text{el}} = \Delta \underline{G}_i^{*\text{is}} + \Delta \underline{G}_i^{*\text{cc}} + \Delta \underline{G}_{i/i}^{*\text{res}}$$

where $\Delta \underline{G}_i^{*\text{is}}$ is the absolute energy difference of the molecule between the ideal conductor phase and the ideal-gas phase, and $\Delta \underline{G}_i^{*\text{cc}}$ is the correction to the free energy that is due to the charge-averaging processes in which each segment calculated from quantum mechanics is reorganized into a standard segment. Also, $\Delta \underline{G}_{i/i}^{*\text{res}}$ is the restoring Gibbs energy required to change from the ideal conductor state to the real liquid state, which is calculated from the sigma profile or charge density, using statistical mechanics as described elsewhere.[15]

The van der Waals Gibbs energy has two contributions and is expressed in terms of the Helmholtz energy as follows:

$$\frac{\Delta \underline{G}_{i/i}^{*\text{vdW}}}{RT} = \frac{\Delta \underline{A}_{i/i}^{*\text{vdW}}}{RT} + Z_{i/L} - Z_{i/IG} =$$

$$\frac{\Delta A_{i/i}^{*\text{disp}}}{RT} + \frac{\Delta A_{i/i}^{*\text{cav}}}{RT} - 1 \quad (3)$$

where $Z = P\underline{V}/(RT)$ is the compressibility factor, which is approximately zero in a liquid phase and unity in an ideal gas phase. The Helmholtz energy of cavity formation is modeled using thermodynamic perturbation theory:[16,17]

$$\frac{\Delta \underline{A}_{i/i}^{*\text{cav}}}{RT} = (2\alpha - 1)\frac{\eta(4 - 3\eta)}{(1 - \eta)^2} - (2\alpha - 2)\ln\left[\frac{1 - \eta/2}{(1 - \eta)^3}\right] \quad (4)$$

where α is an sphericity parameter and η is the packing fraction.

The dispersion term was obtained using a first-order mean field approximation to account for all possible pairwise interactions between the atoms in different molecules:

$$\frac{\Delta \underline{A}_{i/i}^{*\text{disp}}}{RT} = \frac{\sum_j \sum_k \epsilon_{jk} m_j^i m_k^i}{RT\underline{V}_{i/L}} \quad (5)$$

where $\epsilon_{jk} = \sqrt{\epsilon_j \epsilon_k}$ is a pair-interaction energy between atom type j and atom type k and has units of energy times volume. The superscript i indicates that only species i is considered, because the liquid phase is pure, and m_j^i is the effective number of atoms of type j within a species i molecule obtained from

$$m_j^i = \sum_{a \in j}\left(\frac{S_a}{S_{a0}}\right)^q \quad (6)$$

where S_a is the exposed surface area or solvent-accessible area of atom a and S_{a0} is the bare surface area, calculated using the atomic radii set (R_{el}), which are used in quantum mechanics (QM) calculations.

This mean field treatment is based on the assumption that the pairwise atom interactions between different molecules is independent of the molecular size and atom position within the molecule so that each atom in one molecule can interact with all the atoms in another molecule, and, thus, contributes to the interaction energy. This is accurate for small molecules. However, because the vdW interaction is short-range, compared to the electrostatic interaction, we revisit this assumption below for the longer molecules considered in this study.

Normal boiling temperatures are calculated by solving eq 1 at a pressure of 1 atm, using Newton's iterative method.

3. Parameters and Testing Sets

The COSMO−SAC model uses 23 (4 + 6 + 13) parameters to describe any compound consisting of C, N, H, O, F and Cl atoms. There are four universal parameters (two for the COSMO calculation (a_{eff} and/or c_{hb}) and two for the dispersion energy calculations (q and $R_{\text{hc}}/R_{\text{el}}$)), one set of radii ($R_{\text{el}}$) for the six atom types, which have been optimized to reproduce boiling points and enthalpies of vaporization for 369 compounds, and one set of dispersion parameters for each of the 13 different bonding types we have used. The values of these parameters are given in Table 1.

Table 1. Parameters Used in the COSMO−SAC−BP Model

Universal Parameters

parameter	value
a_{eff}	9.24 Å²
c_{hb}	6812.49 kcal/mol Å⁴/e²
q	0.272
R_{hc}/R_{el}	0.611

Atomic Bonding Specific Parameters

atom type (bonding specific)	R_{el} (Å)	ϵ_i/R (K Å³)
H	1.57	1173.51
C__t (sp3)	1.90	8277.96
C__d (sp2, ethene, aromatic)	1.90	11235.27
C__s (sp, ethyne, nitrile)	1.90	12337.01
N__t (sp3, NH₃, amine, aniline)	1.81	9039.50
N__d (sp2, nitro, aromatic)	1.81	11527.19
N__s (sp, nitrile)	1.81	4057.52
O__tH (sp3, bonded to H, alcohol, carboxylic)	1.70	9113.10
O__t (sp3, ether, ester C−O−C, aromatic oxygen)	1.70	8910.89
O__d (sp2, carbonyl, ketone, aldehyde, ester, carboxylic)	1.70	4418.22
O__dN (nitro oxygen)	1.70	4890.04
F	1.71	4747.68
Cl	1.98	27891.21

Table 2. Accuracy of the COSMO−SAC−BP Model, in Regard to Prediction of the Normal Boiling Point

		Deviation of Normal Boiling Point (T_b)	
compound type	number of compounds	temperature AAD[a] (K)	AAPD[b] (%)
alkane	37	11.30	3.23
alcohol	31	15.97	3.49
aldehyde	8	7.53	2.12
ketone	21	11.16	2.71
acid	11	9.36	2.05
ester	21	10.63	2.58
ether	16	11.72	3.63
nitro	5	4.80	1.26
nitrile	7	16.24	4.60
amine	22	11.96	3.75
aromatic	24	12.81	2.83
alkene	9	8.14	2.49
alkyne	7	6.95	2.19
F-containing	5	10.34	4.60
Cl-containing	24	11.46	3.24
multifunctional	121	15.68	3.56
overall	369	12.94	3.24

[a] Average absolute deviation. [b] Average absolute percent deviation.

Inputs to this model are (i) sigma profiles generated by the COSMO calculation, representing the unique charge density distribution for each species, and (ii) liquid density obtained from experimental data or by an estimation method. The DIPPR database[6] provides liquid densities at the normal boiling point by extrapolating reported experimental data at different temperatures. The error in density using the DIPPR method for most compounds is 5% or less. For species for which experimental data are not available, the LeBas group contribution method[1] is used to estimate liquid density at the normal boiling point. It is reported[1] that the average error for this method is 4%, and it appears to be applicable to a greater variety of molecules than other methods. Because the liquid phase density does not change much with temperature near the normal boiling point and the errors are relatively small, the liquid densities obtained from the DIPPR database or the LeBas method can be used when experimental data are not available. Also, as we show later, the model predictions are relatively insensitive to liquid density errors of 5%.

There are a total of 84 compounds in the test set. These compounds were obtained from a handbook of the properties of environmentally important chemicals by Mackay et al.[18] We
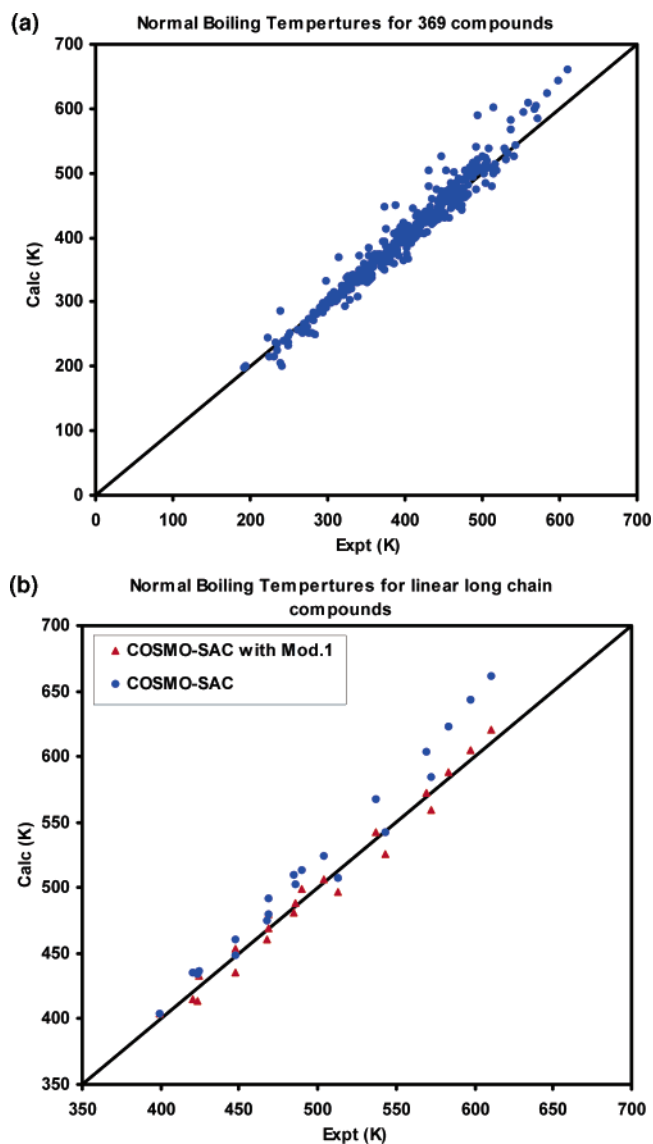

Figure 1. Normal boiling temperatures for several compounds: (a) the 369 compounds in the training set and (b) linear long-chain compounds.

considered only compounds that are stable at their normal boiling point and for which there are experimental data for the normal boiling temperature to compare. The testing set has a variety of species, from polychlorinated biphenyls (PCBs) and polynuclear aromatic hydrocarbons (PAHs) to insecticides.

4. Results and Discussion

For the training set of 369 molecules described elsewhere,[15] the accuracy of the COSMO−SAC−BP model described here is given in Table 2 and Figures 1 and 2. The overall average absolute deviation (AAD) is 12.9 K, with an average absolute percentage deviation (AAPD) of 3.2%. Table 2 shows the large variety of chemical functionalities that can be accurately described with this model. The long-chain alcohols, up to 17 carbon molecules in the training set have high normal boiling temperatures. The largest AAD comes from alcohol, nitrile and compounds that have two or more functional groups. The nitrile compounds have the largest AAD and AAPD, and, therefore, the lowest accuracy, which indicates that care should be taken when using COSMO-based models for compounds that have a nitrile group. Other researchers[19] have also remarked about the inaccuracy of COSMO models for nitrile groups. Considering
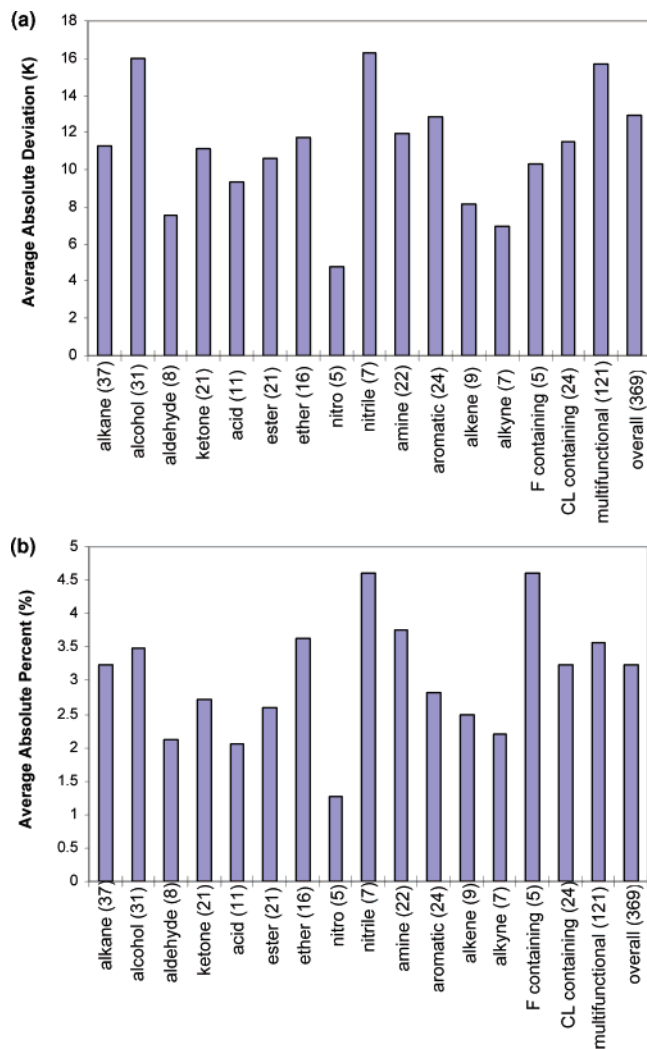
(a)

(b)

Figure 2. (a) Average absolute deviation (AAD) and (b) average absolute percent deviation (AAPD) for the 369 compounds described in the training set.

(a) Normal Boiling Tempertures for 84 environmentally important chemicals

(b) NBT for 84 compounds with mod.1

Figure 3. Normal boiling temperatures for 84 environmentally important chemicals: (a) normal dispersion and (b) dispersion using modification 1.

the large range of functional groups in the molecular structure, the overall AAD is comparable to QSPR methods, which have a much larger number of parameters, compared to only 23 temperature-independent parameters in the COSMO model here.

To test the method, we predicted the normal boiling temperatures of 84 chemicals of interest in environmental studies. This group included alkanes, aliphatic esters, aliphatic amines, aliphatic acids, halogenated hydrocarbons, phthalate esters, and polynuclear aromatic hydrocarbons. These are listed in Table 3 and depicted visually in Figure 3. The overall AAD and AAPD values are 21.9 K and 4.4%, respectively. Except for a few outliers, the boiling temperatures of most of the chemicals are predicted reasonably well. Some of the outliers are larger, long-chain molecules, as shown in Table 4.

Because liquid density is a required input property, we examined the effect of density errors on the predictions. The DIPPR database has up to a 5% uncertainty in experimental liquid density and the Le Bas method results in an error of ~4%. Therefore, we chose to consider the impact of both a 5% increase and decrease in the values of liquid density for the set of 369 training set and 84 environmentally important chemicals. The results show that, with a 5% increase of liquid densities, the overall AAD for 369 compounds in the training set is 13.0 K (3.26% for AAPD); the overall AAD for 84 chemicals is 22.5 K (4.5% for AAPD). However, with a 5% decrease, the
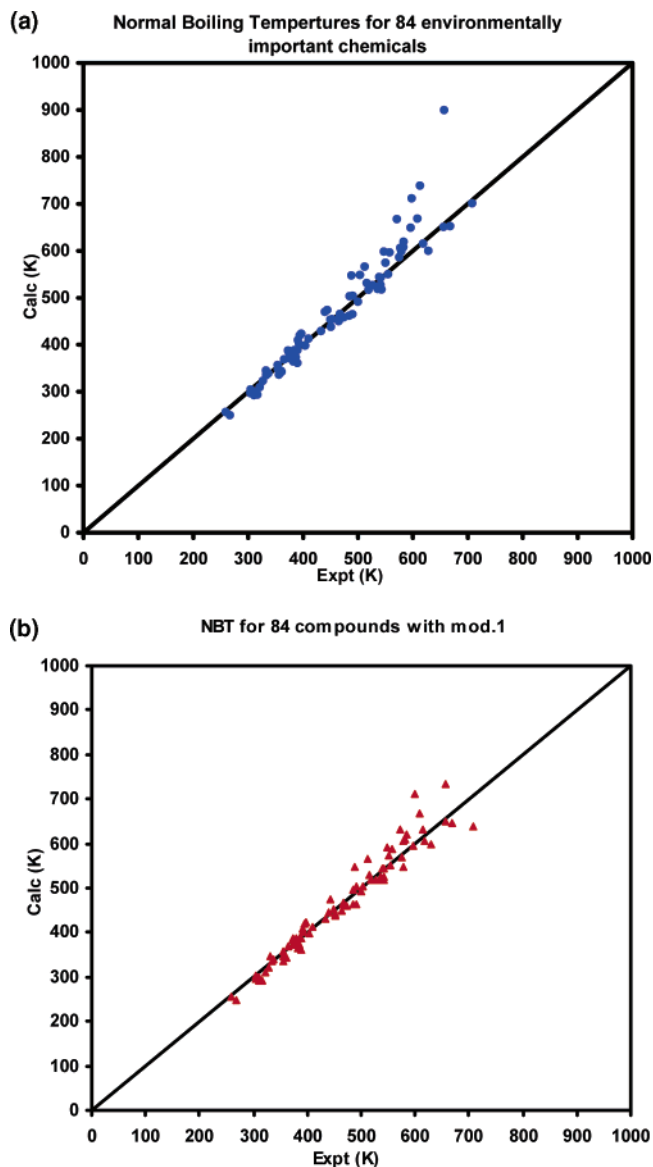
AAD values for the set of 369 compounds and 84 chemicals are 12.97 K (3.25%) and 22.1 K (4.44%), respectively. These results indicate that errors of 5% in the liquid densities lead to average changes of only ~0.3 K in the predicted boiling point. Consequently, we conclude that the uncertainties of liquid densities have little effect on the predictions.

The few outliers in the testing database are all large molecules, such as *bis*-(2-ethylhexyl)-phthalate and dibutylphthalate. One possible model improvement is to revisit the dispersion energy calculation. The dispersion interactions are due to instantaneous electron correlations that decay as $1/R^6$, leading to a limited effective range, compared with the long-range electrostatic forces. Consequently, we will consider the below-given modifications to the mean field treatment of the dispersion interactions.

Here, we consider three modifications from the previous mean field treatment. The first is based only on molecular size. For medium- and large-sized molecules, we assume that atoms in one molecule cannot make their full contribution to the vdW interaction energy. The reason is that the vdW dispersion correlation in the electrons on atom *i* for interacting with atom *j* are not compatible with the vdW dispersion correlation in the

Table 3. List of Normal Boiling Temperatures for Environmentally Important Chemicals

| compound | T_b, experiment (K) | COSMO−SAC−BP | | COSMO−SAC−BP/1 | |
		T_b, calculated (K)	ΔT (K)[a]	T_b, calculated (K)	ΔT (K)[a]
Aliphatic Acids					
formic acid	373.90	384.50	10.60	384.50	10.60
dichloroacetic acid	467.20	465.67	−1.53	465.67	−1.53
Aliphatic Amines					
triethanolamine	608.55	668.41	59.86	668.41	59.86
2,2′-diethanolamine	541.54	542.74	1.20	542.74	1.20
Aliphatic Esters					
ethylformate	327.00	322.60	−4.40	322.60	−4.40
methylacrylate	354.00	356.55	2.55	356.55	2.55
methylmethacrylate	373.50	371.71	−1.79	371.71	−1.79
propylformate	354.10	348.32	−5.78	348.32	−5.78
ethylacrylate	372.75	387.35	14.60	387.35	14.60
Alkanes					
2,2,5-trimethylhexane	397.00	423.59	26.59	423.59	26.59
2-methylheptane	390.70	409.62	18.92	409.62	18.92
Alkenes					
isobutylene	266.70	249.67	−17.03	249.67	−17.03
2-methyl-2-butene	311.00	292.54	−18.46	292.54	−18.46
2-methyl-1-butene	304.30	297.09	−7.21	297.09	−7.21
cis-2-pentene	309.80	300.66	−9.14	300.66	−9.14
trans-2-pentene	309.40	300.05	−9.35	300.05	−9.35
Alkylphenols and Other Substituted Phenols					
2,6-dimethylphenol	474.22	458.19	−16.03	458.19	−16.03
2,4-dimethylphenol	484.13	461.82	−22.31	461.82	−22.31
3,4-dimethylphenol	500.15	491.58	−8.57	491.58	−8.57
Aromatic Acid					
phenylacetic acid	538.70	543.95	5.25	543.95	5.25
phthalic acid	598.00	711.17	113.17	711.17	113.17
Aromatic Amines					
diphenylamine	575.15	585.91	10.76	569.73	−5.42
Aromatic Esters					
ethylbenzoate	485.00	502.94	17.94	497.46	12.46
benzylbenzoate	596.20	649.16	52.96	595.61	−0.59
propylbenzoate	503.70	548.28	44.58	503.58	−0.12
Chlorobenzenes					
hexachlorobenzene	582.55	607.76	25.21	607.76	25.21
pentachlorobenzene	550.20	574.38	24.18	574.38	24.18
1,2,4,5-tetrachlorobenzene	516.20	531.00	14.80	531.00	14.80
Chlorophenols					
pentachlorophenol	583.20	619.30	36.10	619.30	36.10
2,4,6-trichlorophenol	519.20	517.06	−2.14	517.06	−2.14
Cycloalkanes					
1,1,3-trimethylcyclopentane	378.10	386.62	8.52	386.62	8.52
1,1,3-trimethylcyclohexane	410.00	413.16	3.16	413.16	3.16
trans-1,4-dimethylcyclohexane	392.60	401.97	9.37	401.97	9.37
decalin	465.00	450.05	−14.95	450.05	−14.95
propylcyclopentane	404.10	397.74	−6.36	397.74	−6.36
pentylcyclopentane	453.00	454.20	1.20	441.87	−11.13
Cycloalkenes					
cyclopentene	317.00	293.56	−23.44	293.56	−23.44
cyclohexene	356.00	336.01	−19.99	336.01	−19.99
1,4-cyclohexadiene	360.15	344.35	−15.80	344.35	−15.80
cycloheptatriene	389.70	387.92	−1.78	387.92	−1.78
cycloheptene	387.00	373.11	−13.89	373.11	−13.89
1-methylcyclohexene	382.00	364.04	−17.96	364.04	−17.96
1,2,3,4-tetrahydronaphthalene	490.00	465.16	−24.84	465.16	−24.84
D-limonene	450.00	453.41	3.41	453.41	3.41
Dienes					
isoprene	307.00	300.89	−6.11	300.89	−6.11
1-decene	440.00	469.90	29.90	444.81	4.81
Ethers					
methylchloromethylether	332.65	345.22	12.57	345.22	12.57
epichlorohydrin	389.60	360.87	−28.73	360.87	−28.73
tetrahydropyran	361.00	342.40	−18.60	342.40	−18.60
2-methylfuran	337.00	338.96	1.96	338.96	1.96
2-chloroethylvinylether	381.20	373.22	−7.98	373.22	−7.98
1,2-epoxyethylbenzene	467.30	462.07	−5.23	462.07	−5.23
bis-(2-chloroethyl)ether	451.00	437.78	−13.22	437.78	−13.22

Table 3. (Continued)

compound	T_b, experiment (K)	COSMO−SAC−BP		COSMO−SAC−BP/1	
		T_b, calculated (K)	ΔT (K)	T_b, calculated (K)	ΔT (K)
Fungicides					
nitrotrichloromethane	385.00	388.42	3.42	388.42	3.42
Halogenated Hydrocarbons					
1,1-dichloroethylene	304.55	304.73	0.18	304.73	0.18
vinylchloride	259.30	256.67	−2.63	256.67	−2.63
cis-1,2-dichloroethylene	333.00	336.67	3.67	336.67	3.67
trans-1,2-dichloroethylene	321.00	309.57	−11.43	309.57	−11.43
tetrachloroethylene	394.20	419.33	25.13	419.33	25.13
1,1,2,2-tetrachloro-1,2-difluoroethane	365.75	368.90	3.15	368.90	3.15
2-chloro-1,3-butadiene	333.00	335.12	2.12	335.12	2.12
perchlorocyclopentadiene	512.20	565.78	53.58	565.78	53.58
hexachlorobutadiene	488.20	547.26	59.06	547.26	59.06
Heterocyclic Compounds					
2,3-dimethylpyridine	433.00	428.97	−4.03	428.97	−4.03
diphenylenimine	628.20	599.55	−28.65	599.55	−28.65
acridine	618.70	615.17	−3.53	604.43	−14.27
Monoaromatic Compounds					
2-methyl-propylbenzene	444.00	473.56	29.56	473.56	29.56
Nitroaromatic Compounds					
1-nitronaphthalene	577.20	605.94	28.74	605.94	28.74
Nitrophenols					
O-nitrophenol	490.40	504.16	13.76	504.16	13.76
Phthalate Esters					
dimethylphthalate	555.00	550.28	−4.72	550.28	−4.72
diethylphthalate	571.00	667.06	96.06	630.66	59.66
dibutylphthalate	613.20	738.64	125.44	632.42	19.22
bis-(2-ethylhexyl)-phthalate	657.20	899.29	242.09	734.38	77.18
Polychlorinated Biphenyls (PCBs)					
2-chlorobiphenyl (PCB-1)[b]	547.20	597.93	50.73	590.57	43.37
3-chlorobiphenyl (PCB-2)[b]	557.70	596.73	39.03	589.30	31.60
Polynuclear Aromatic Hydrocarbons (PAHs)					
acenaphthylene	543.15	517.23	−25.92	517.23	−25.92
1,4-dimethylnaphthalene	537.00	524.23	−12.77	524.23	−12.77
1,3-dimethylnaphthalene	536.15	518.40	−17.75	518.40	−17.75
2,3-dimethylnaphthalene	541.15	527.15	−14.00	527.15	−14.00
fluoranthene	655.95	650.42	−5.53	649.18	−6.77
pyrene	667.95	652.30	−15.65	646.41	−21.54
2-ethylnaphthalene	527.00	527.04	0.04	518.17	−8.83
benz(a)anthracene	708.15	701.06	−7.09	638.04	−70.11
trans-1,2-diphenylethene	579.15	598.88	19.73	546.39	−32.76

[a] $\Delta T = T_b^{calc} - T_b^{expt}$. [b] As described in ref 17.

electrons on atom i for interacting with atom k. These three-body corrections can be included (the Axelrod−Teller formalism) but are ignored here. The net effect of these three-body corrections is to decrease the net vdW interactions between those atoms in large molecules that are far from each other. To account for this effect, we introduce a size-dependent contribution factor τ. The effective number of atoms of type j within species i is then modified as follows:

$$m_j^i = \sum_{a \in j} \left(\frac{S_a}{S_{a0}} \times \tau\right)^q \quad (7)$$

$$\tau = \begin{cases} 1 & \text{if } l_{effect} \geq l_{max} \\ \dfrac{l_{effect}}{l_{max}} & \text{if } l_{effect} < l_{max} \end{cases} \quad (8)$$

where l_{max} is the maximum molecular length of species i. Here, l_{effect} is an effective length, a value within which the vdW interaction between an atom in one molecule and the atoms in other molecule can be considered to interact fully. We determined, empirically, that an average value of 9 Å is optimum

for l_{effect}. For long linear chain molecules, we found l_{effect} to be dependent linearly on the number of carbons in the molecule as follows: $l_{effect} = 9 + 0.9(C_{num} - 6)$. This expression reduces the overestimate of normal boiling temperatures for long-chain chemicals, while leaving the results for small molecules unchanged.

If the l_{max} value of species i is smaller than the effective length, then $\tau = 1$, so that an atom in one molecule can fully interact with all other atoms. If the l_{max} value of species i is greater than the effective length, an atom makes only a partial contribution to the total interaction energy. That is, with an increase of molecular size, the energy contribution of each atom decreases.

A second possible modification is to make the dispersion interaction position-dependent by considering the local concentration of different atomic types:

$$\Delta \underline{A}_{i/i}^{*disp} = \frac{1}{\underline{V}_{i/L}} \sum_j \sum_k \epsilon_{jk} m_j^i m_k^i P_{jk} \quad (9)$$

where P_{jk} is the likelihood of an interaction between atom types of j and k normalized by the likelihood of interactions between

Table 4. List of Normal Boiling Temperatures for Linear Long-Chain Molecules

		COSMO−SAC−BP		COSMO−SAC−BP/1	
compound	T_b, experiment (K)	T_b, calculated (K)	ΔT (K)[a]	T_b, calculated (K)	ΔT (K)[a]
1-nonene	420.02	435.70	15.68	414.81	−5.21
1-nonyne	423.85	434.79	10.94	413.43	−10.42
octane	398.83	404.11	5.28	404.11	5.28
nonane	423.97	436.05	12.08	432.54	8.57
octanal	447.15	448.07	0.92	435.83	−11.32
nonanal	468.15	474.68	6.53	460.57	−7.58
octylacetate	484.45	509.65	25.20	481.11	−3.34
decane	447.31	460.88	13.57	453.55	6.24
undecane	469.08	491.37	22.29	480.48	11.40
dodecane	489.47	513.11	23.64	498.60	9.13
octanoic acid	512.85	507.04	−5.81	497.30	−15.55
decanoic acid	543.15	542.99	−0.16	525.55	−17.60
lauric acid	571.85	584.54	12.69	559.77	−12.08
1-octanol	468.35	479.57	11.22	469.32	0.97
1-nonanol	486.25	502.35	16.10	488.26	2.01
1-decanol	504.07	524.54	20.47	506.75	2.68
1-dodecanol	536.95	567.81	30.86	542.88	5.93
1-tetradecanol	568.80	604.18	35.38	572.44	3.64
1-pentadecanol	583.40	622.99	39.59	587.69	4.29
1-hexadecanol	597.23	643.22	45.99	604.87	7.64
1-heptadecanol	610.50	661.48	50.98	620.55	10.05

[a] $\Delta T = T_b^{calc} - T_b^{expt}$.

Table 5. Comparison of the Results from Applying COSMO−SAC−BP Model with Modifications

		Training Set				Test Set		
	number of parameters	small[a] (305)	mid−large[b] (43)	long-chain[c] (21)	overall (369)	small[a] (68)	mid−large[b] (16)	overall[c] (84)
COSMO−SAC−BP	23	12.4 K (3.2%)	13.5 K (2.9%)	19.3 K (3.7%)	12.9 K (3.2%)	16.1 K (3.6%)	44.8 K (7.6%)	21.9 K (4.4%)
COSMO−SAC−BP/1	26	12.4 K (3.2%)	11.9 K (2.6%)	7.6 K (1.5%)	12.1 K (3.0%)	16.1 K (3.6%)	24.7 K (4.1%)	17.8 K (3.7%)
COSMO−SAC−BP/2	26	14.3 K (3.7%)	15.9 K (3.5%)	14.4 K (3.0%)	14.3 K (3.6%)	16.9 K (3.8%)	15.0 K (2.4%)	16.7 K (3.5%)
COSMO−SAC−BP/3	39	12.2 K (3.1%)	17.4 K (3.8%)	19.7 K (4.1%)	13.0 K (3.2%)	15.5 K (3.4%)	20.8 K (3.6%)	16.6 K (3.4%)

[a] For small molecules, the maximum molecular length is <9 Å. [b] For mid−large molecules, the maximum molecular length is >9 Å, with nonlinear structure. [c] The long-chain molecules are those with straight linear carbon chains longer than 9 Å.

atom type j and all atom types within a molecule. The Boltzmann distribution is used for this probability function:

$$P_{jk} = \frac{\exp\{-[(\epsilon_j/D_j^3)(\epsilon_k/D_k^3)]^{1/2}/T\}}{(1/\alpha)\sum_\alpha \exp\{-[(\epsilon_j/D_j^3)(\epsilon_k/D_k^3)]^{1/2}/T\}} \quad (10)$$

where ϵ is the dispersion coefficient, D_j the diameter (twice as large as R_{el}) of atom type j; α is the number of all atom types within molecule i. For this calculation, we assumed T to be room temperature (300 K). This choice is somewhat arbitrary; however, other temperature choices lead to additional problems. (Thus, using a different temperature for each compound adds additional complexity and parameters to the model.) If all atom types would have the same dispersion energy, the probability factor reduces to unity. On the other hand, a large dispersion coefficient for a particular atom type increasing the probability for this type of interaction is increased in this modification.

The last modification we considered is to use, in eq 7, a different q value (q_j for each atom type j):

$$m_j^i = \sum_{a \in j} \left(\frac{S_a}{S_{a0}} \times \tau\right)^{q_j} \quad (11)$$

The value of q (eq 7) or q_j (eq 11) was optimized using the 369-compound training set. The range of q is from zero to unity. If q_j is zero, then m_j is the actual number of atoms of species j;

if q_j is unity, then the dispersion interaction is highly dependent on the surface area.

We refer to these three modifications as follows: modification 1 (denoted as COSMO−SAC−BP/1), which involves size dependence (see eqs 7 and 8); modification 2 (denoted as COSMO−SAC−BP/2), which involves size and intramolecular position dependence (eqs 7−10); and modification 3 (denoted as COSMO−SAC−BP/3), in which size, intramolecular position dependence, and multiexponent parameters are the properties of interest (eqs 7−11).

The results for each of these modifications are summarized in Table 5 and compared with the results of the original model. Each of the three cases has its advantages and disadvantages.

(1) COSMO−SAC−BP/1, which includes only molecular size, leads to an AAD for linear long-chain molecules that reduces from 19.3 K to 7.6 K, and the AAD for moderately large molecules in the test set is reduced from 44.8 K to 24.7 K. This modification leads systematically to lower boiling points.

(2) COSMO−SAC−BP/2, which includes both size and intramolecular position, leads to an AAD for linear long-chain molecules that decreases only to 14.4 K, which is twice the value obtained with modification 1. However, for the moderately large molecules in the test set, AAD decreases to 15.0 K, which is only 60% of the error for modification 1. However, modification 2 required reoptimization for all dispersion interaction parameters, whereas this was not required within modification 1.

(3) COSMO−SAC−BP/3, which has multiexponent parameters q_j, leads to an AAD value for small molecules of 12.2 K

Table 6. Predicted Normal Boiling Temperatures for Ten Pesticides

pesticide compound	Predicted Boiling Temperature (K)		
	COSMO−SAC−BP	COSMO−SAC−BP/1	Stein and Brown's group contribution method
	Herbicides		
Atrazine	659.68	622.18	586.20
Barban	727.14	671.55	609.77
Bifenox	848.32	745.70	698.31
MCPB	678.66	621.46	619.26
	Insecticides		
Aminocarb	653.51	612.13	571.05
Bendiocarb	678.78	657.81	588.32
Carbofuran	691.82	675.09	584.58
	Fungicides		
Anilazine	705.55	671.19	638.06
Benomyl	864.52	800.54	735.05
Carbendazim	636.80	600.84	596.08

Table 7. Predicted Sublimation and Vapor Pressures at 298.15 K for Ten Pesticides[a]

pesticide compound	experimental sublimation, $\ln(P^{sub})$	COSMO−SAC−BP		COSMO−SAC−BP/1		Grain-Watson's Method for Liquid/Solid	
		$\ln(P^{vap})$	$\ln(P^{sub})$	$\ln(P^{vap})$	$\ln(P^{sub})$	$\ln(P^{vap})$	$\ln(P^{sub})$
		Herbicides					
Atrazine	−9.21	−1.49	−6.96	−0.17	−5.64	−0.89	−4.46
Barban	−6.62	−4.97	−6.46	−2.97	−4.46	−2.05	−3.08
Bifenox	−8.0	−10.01	−11.81	−6.15	−7.95	−6.54	−7.89
MCPB	−9.21	−3.82	−6.60	−1.67	−4.45	−2.52	−4.20
		Insecticides					
Aminocarb	−6.08	−3.08	n.a.	−1.47	n.a.	−0.15	n.a.
Bendiocarb	−5.38	−4.62	n.a.	−3.78	n.a.	−0.99	n.a.
Carbofuran	−5.91	−5.18	n.a.	−4.51	n.a.	−0.81	n.a.
		Fungicides					
Anilazine	−14.01	−1.40	−5.34	−0.32	−4.27	−3.46	−6.63
Benomyl	−12.22	−13.50	n.a.	−10.83	n.a.	−8.46	n.a.
Carbendazim	−8.80	−2.50	n.a.	−1.10	n.a.	−1.37	n.a.

[a] Sublimation pressures at 298.15 K are estimated from the predicted vapor pressures using the relation $\ln(P^{vap}/P^{sub}) = \ln(f^{vap}/f^{sub}) \approx \Delta H^{fus}(T_m)/(RT)[1 - (T/T_m)]$. All pressures given in pascals. Experimental data of heat of fusion and melting point are available for five compounds that allows sublimation pressure predictions to be made. For the remaining compounds, only predicted vapor pressures of the liquid are reported, and the term "n.a." is used to indicate that a sublimation pressure prediction could not be made.

(3.1%); however, the results for long-chain and mid−large molecules are less satisfactory.

Consequently, we recommend COSMO−SAC−BP/1 for the prediction of normal boiling temperatures, especially for linear long-chain or moderately large molecules.

Both to illustrate the application of the COSMO−SAC−BP model, and to challenge the COSMO−SAC−BP model, Table 6 lists the normal boiling point temperature predictions for four herbicides, three insecticides, and three fungicides for which no experimental normal boiling temperatures are available. For comparison, predictions using Stein and Brown's GCM are listed. The COSMO−SAC−BP predictions are systematically higher, although we cannot yet determine which is more accurate (or whether the chemicals would decompose before the predicted boiling point is reached).

To make comparisons with experimental data in Table 7, we compare experimental data[18] and predictions for the sublimation pressures of these pesticides at 298.15 K. Because these compounds are solids at 298.15 K and COSMO−SAC predicts liquid vapor pressures, we have corrected the predictions from hypothetic vapor pressures to sublimation pressures, using

$$\ln\left(\frac{P^{vap}}{P^{sub}}\right) = \ln\left(\frac{f^{vap}}{f^{sub}}\right) \approx \frac{\Delta \underline{H}^{fus}(T_m)}{RT}\left(1 - \frac{T}{T_m}\right) \qquad (12)$$

for the five compounds for which reported heat of fusion and

melting point data are available.[18] The predicted vapor pressures for the hypothetical liquids are also given in the table. For comparison, predictions using the generally accepted Grain−Watson[20] method also are listed. The measured and predicted sublimation pressures are very low, and the accuracy of the COSMO−SAC−BP method, although not very good, is generally better than the Grain−Watson method. One reason for this may be that the COSMO−SAC−BP method was developed to predict normal boiling points; however, here it is being used for vapor pressure predictions at temperatures that are hundreds of degrees lower. Also, eq 12 neglects terms that are related to the heat capacities of the liquid and solid, which may be important over such temperature ranges, but for which data are not available. Finally, some of experimental data may not be accurate. Anilazine is a case in point; it is a relatively small molecule with a predicted boiling point that is moderate, whereas its reported vapor pressure in unusually low, much lower than the predictions of methods considered here. Also, the two methods considered here for the predictions of sublimation pressures—the Grain−Watson method and the modified COSMO−SAC−BP method—are in reasonable agreement with each other, even though they are based on very different assumptions.

5. Conclusions

Our previous work developed a new method for predicting vapor pressures and enthalpies of vaporization that leads to

satisfactory accuracy for a test data set of 369 species. This method combines COSMO solvation with a mean field treatment and perturbation theory. The present work uses the COSMO−SAC−BP model to predict normal boiling temperatures for 84 environmentally important chemicals that contain C, H, O, N, F, and Cl atoms. For molecules smaller than 9 Å, the normal boiling temperatures are obtained without any modification to the previous model. However, for medium- and large-sized molecules, an effective length is introduced to modify the van der Waals interaction energy contribution. With the introduction of an effective length (COSMO−SAC−BP/1), there is a slight improvement (12.1 K) in predictions for the training set, and a more significant improvement for the testing set and linear long-chain compounds (see Tables 5 and 6 and Figure 1b).

Overall, the predictions of boiling temperatures of the test set of environmentally important chemicals give an average absolute deviation (AAD) of 21.9 K (4.4%), using the original model for the wide range of compound classes considered, from alkanes to polychlorinated biphenyls (PCBs), to fungicides. With COSMO−SAC−BP/1, the AAD values are 17.8 K (3.7%) for the test set and 12.1 K (3.0%) for the training set. Therefore, we recommend the COSMO−SAC−BP/1 model (i.e., the use of eqs 7 and 8) be used to predict the boiling points of large, complex molecules.

We also tested two other modifications of the model in which we included position-dependent weight functions (P_{jk}) and exponential parameters; however, these models with additional parameters did not lead to substantial improvements, compared with COSMO−SAC−BP/1. Finally, we have shown how the quantum mechanics-based model presented here can be used for the predictions of the normal boiling points for 10 pesticides for which no experimental data are available.

Nomenclature

COSMO−SAC−BP = conductor-like screening model-based segment activity coefficient model for boiling-point calculation

P_i^{vap} = vapor pressure for compound i

R = gas constant

i/i = pure compound i in solvent i

i/L = pure compound i in solvent L

i/IG = pure compound i in ideal gas

a_{eff} = standard segment surface area

c_{hb} = hydrogen bonding constant

q = exponential constant

R_{ho}/R_{el} = ratio between hard sphere radii and van der Waals (vdW) atomic radii

Literature Cited

(1) Reid, R. C.; Prausnitz, J. M.; Poling, B. E. *The Properties of Gases and Liquids*, 4th Edition; McGraw−Hill: New York, 1987.

(2) Horvath, A. L. *Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds*; Elsevier: Amsterdam, The Netherlands, 1992.

(3) Joback, K. G.; Reid, R. C. Estimation of Pure-Component Properties from Group-Contributions. *Chem. Eng. Commun.* **1987**, *57*, 233.

(4) Stein, S. E.; Brown, R. L. Estimation of Normal Boiling Points from Group Contributions. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 581.

(5) Constantinou, L.; Gani, R. New Group-Contribution Method for Estimating Properties of Pure Compounds. *AIChE J.* **1994**, *40*, 1697.

(6) Rowley, J. R.; Wilding, W. V.; Oscarson, J. L.; Rowley, R. L. Byu-Dippr Tpl; Brigham Young University, Provo, UT, 2002.

(7) Marrero-Morejon, J.; Pardillo-Fontdevila, E. Estimation of Pure Compound Properties Using Group-Interaction Contributions. *AIChE J.* **1999**, *45*, 615.

(8) Cordes, W.; Rarey, J. A New Method for the Estimation of the Normal Boiling Point of Non-Electrolyte Organic Compounds. *Fluid Phase Equilib.* **2002**, *201*, 409.

(9) Nannoolal, Y.; Rarey, J.; Ramjugernath, D.; Cordes W. Estimation of Pure Component Properties Part 1. *Fluid Phase Equilib.* **2004**, *226*, 45.

(10) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Normal Boiling Points for Organic Compounds: Correlation and Prediction by a Quantitative Structure−Property Relationship. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28.

(11) Ericksen, D.; Wilding, W. V.; Oscarson, J. L.; Rowley, R. L. Use of the Dippr Database for Development of Qspr Correlations: Normal Boiling Point. *J. Chem. Eng. Data* **2002**, *47*, 1293.

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

(13) Klamt, A. Refinement and Parametrization of COSMO−RS. *J. Phys. Chem. A* **1998**, *102*, 5074.

(14) Lin, S.-T.; Sandler, S. I. A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res.* **2002**, *41*, 899.

(15) Lin, S.-T.; Chang, J.; Wang, S.; Goddard, W. A., III; Sandler, S. I. Prediction of Vapor Pressures and Enthalpies of Vaporization Using a Cosmo Solvation Model. *J. Phys. Chem. A* **2004**, *108*, 7429.

(16) Boublik, T.; Vega, C.; Diazpena, M. Equation of State of Chain Molecules. *J. Chem. Phys.* **1990**, *93*, 730.

(17) Walsh, J. M.; Gubbins, K. E. A Modified Thermodynamic Perturbation-Theory Equation for Molecules with Fused Hard-Sphere Cores. *J. Phys. Chem.* **1990**, *94*, 5115.

(18) Mackay, D.; Shiu, W. Y.; Ma, K. C. *Illusrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*; Lewis Publishers: Ann Arbor, MI, 1991; Vols. I−V.

(19) Eckert, F.; Klamt, A. Fast Solvent Screening Via Quantum Chemistry: Cosmo-Rs Approach. *AIChE J.* **2002**, *48*, 369.

(20) Boethling, R. S.; Mackay, D. *Handbook of Property Estimation Methods for Chemicals Environmental and Health Sciences*; CRC Press: Washington, DC, 2000.