Prediction of Vapor Pressures and Enthalpies of Vaporization Using a COSMO Solvation Model

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

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

Received: March 17, 2004; In Final Form: June 24, 2004

We have developed a general predictive method for vapor pressures and enthalpies of vaporization based on the calculation of the solvation free energy that consists of three components; the electrostatic, dispersion, and cavity formation contributions. The electrostatic contribution is determined using the quantum mechanical COSMO solvation model. Thermodynamic perturbation theory for hard-core molecules is used for the cavity term, and the dispersion term is modeled using a mean field term proportional to the density and molecular surface area. The proposed model uses one set of van der Waals atomic radii to describe molecular shape, two universal interaction parameters for the electrostatic interaction, one set of atom-specific dispersion coefficients, one universal parameter to scale the atomic exposed surface area, and a single universal parameter for the ratio of the hard-core to atomic radii. The model parameters have been determined using 371 pure substances of varying molecular structure, functionality, and size. The average accuracy of the model for vapor pressures and enthalpies of vaporization at the normal boiling temperature is found to be 76% and 4.81 kJ/mol, respectively, with temperature-independent parameters. The average error in the normal boiling temperature is found to be 16 K for species whose boiling points range from 191 to 610 K.

Introduction

Vapor pressure of a pure substance is an important thermodynamic property that is essential in chemical process design and fate analysis of environmental pollutants. Experimental data for the vapor pressures of small molecules are abundant in the literature but are scarce or of low accuracy for larger molecules of low volatility. Thus, there have been many attempts to develop a predictive method for the vapor pressure of compounds of arbitrary molecular structure and functionality. The most successful estimation methods have been based on empirical extensions of the Clausius−Claperyon equation or corresponding states theory.[1] However, such correlation methods usually require other data, such as critical properties, boiling temperature, and acentric factor, which makes their use somewhat limited. Predictive group contribution methods[2,3] have been developed using semiempirical equations for the vapor pressure, but such methods may not be applicable to compounds of very different chemical structure. Another empirical approach is the quantitative structure−property relations (QSPR)[4−7] in which vapor pressures are correlated with descriptors calculated from molecular structure. A theoretically more appealing approach is to predict vapor pressures from solvation free energies. Such an approach includes the SM5 solvation models of Winget et al.[8] and the COSMO-RS (conductor-like-screening model for real solvent) model of Klamt and co-workers.[9,10] There has been considerable interest in the latter due to its novel way of treating the solvent medium and also the relative simplicity when applying the method to mixtures.

In COSMO-RS, the molecules are regarded as collection of surface segments and the chemical potential of each segment is self-consistently determined from a statistical mechanical relation. The difference in the segment chemical potential between the mixture and the pure liquid gives the segment activity coefficient, and the activity coefficient of a molecule is then obtained from the summation over the segment activity coefficients. In this way, thermodynamic properties of liquids can be predicted in an a priori manner, and the COSMO-RS method has been used to predict phase equilibria of mixtures.[10,11] On the basis of the same physical picture, Lin and Sandler[11] invoked the necessary thermodynamic consistency criteria and developed a variation, the COSMO-SAC model. The work here is an extension of the COSMO-SAC model.

Despite the reasonably successful implementation of these COSMO-based models for mixture phase equilibrium calculations, a question that remains is whether such models give the accurate solvation free energy of pure substances, the free energy required to transfer a molecule from the vapor phase to its own liquid. This information is needed to calculate vapor pressures and related properties. In fact, the free energy difference between the vapor and liquid phases is more difficult to estimate than the difference in solvation free energy of a molecule between two liquid phases (related to the species activity coefficient). Owing to the large difference in density, the solvation free energy difference for a molecule between the vapor and liquid phases is much larger than that between two liquid phases, and there is no cancellation of errors, as occurs in the latter case. Thus, the test of a solvation model for the vapor pressure of pure substances is more stringent than is the case for activity coefficients and vapor−liquid equilibria in mixtures.

* Author to whom correspondence should be addressed.
[†] California Institute of Technology.
[‡] University of Delaware.

10.1021/jp048813n CCC: $27.50
Published on Web 08/12/2004

In this work, we develop an accurate solvation model for pure substances by using theoretically based models for the cavity and dispersion terms of the solvation free energy. In doing this, we use thermodynamic perturbation theory (TPT)[12−15] for the cavity term; a recent computer simulation study for hard diatomic fluids shows that the TPT theory gives more accurate estimates for the cavity free energy than the scaled particle theory.[16] For dispersion interactions, we use a mean field theory taking into account the accessible surface area of the molecules. Since the dispersion interaction is significant for both polar and nonpolar compounds, we use an atom-specific dispersion model that can account for the chemical bonding environment of an atom, as covalently bonded neighbor atoms affect the induced dipole fluctuation and the dispersion interaction. This is a significant improvement over the simple dispersion model of the previous COSMO models[9,10,17,18] that used only a single dispersion parameter for each atomic species.

Theory

The determination of chemical potential is the key to the prediction of vapor pressures. For pure substances, the chemical potential is

$$\mu = \left(\frac{\partial G}{\partial N}\right)_{T,p} = \left(\frac{\partial A}{\partial N}\right)_{T,V} \tag{1}$$

The vapor pressure P^{vap} can be determined by the equality of chemical potentials in both the vapor and the liquid phases at some temperature, T

$$\mu_{i/V}(T,P_i^{\text{vap}}) = \mu_{i/L}(T,P_i^{\text{vap}}) \tag{2}$$

To solve eq 2 for P^{vap}, we consider the expression of μ in solvation theory[19]

$$\mu_{i/S}(T,P) = \mu^*_{i/S}(T,P) + kT \ln \rho_{i/S}\Lambda_i{}^3 \tag{3}$$

where $\mu^*_{i/S}$, the pseudochemical potential, is the chemical potential of molecule i at a fixed position in solution S; Λ_i is the de Broglie wavelength of i; and $\rho_{i/S}$ is the number density. The solvation free energy, defined as the work required to move a molecule from a fixed position in an ideal gas phase to a fixed position in solution S, is

$$\Delta g^{*\text{sol}}_{i/S} = \mu^*_{i/S} - \mu^*_{i/IG} \tag{4}$$

The vapor pressure, P^{vap}, can then be expressed in terms of the solvation free energy by first subtracting the ideal gas chemical potential from both sides of eq 2

$$\mu_{i/V}(T,P_i^{\text{vap}}) - \mu_{i/IG}(T,P_i^{\text{vap}}) = \mu_{i/L}(T,P_i^{\text{vap}}) - \mu_{i/IG}(T,P_i^{\text{vap}}) = \Delta g^{*\text{sol}}_{i/L} + kT \ln \frac{\rho_{i/L}}{\rho_{i/IG}} \tag{5}$$

where eqs 3 and 4 are used for the last identity. Replacing the departure of the gas-phase chemical potential from that of an ideal gas with the fugacity coefficient, $\mu_{i/V}(T,P_i^{\text{vap}}) - \mu_{i/IG}(T,P_i^{\text{vap}}) = kT \ln f(T,P_i^{\text{vap}})/P_i^{\text{vap}}$, and the ideal vapor phase density with the pressure, $P_i^{\text{vap}} = \rho_{i/IG}kT$, we have

$$\ln P_i^{\text{vap}} = \frac{\Delta g^{*\text{sol}}_{i/i}}{kT} + \ln kT\rho_{i/L} - \ln \frac{f(T,P_i^{\text{vap}})}{P_i^{\text{vap}}} \tag{6}$$

or in a molar basis,

$$\ln P_i^{\text{vap}} = \frac{\Delta \underline{G}^{*\text{sol}}_{i/i}}{RT} + \ln \frac{RT}{\underline{V}_{i/L}} - \ln \frac{f(T,P_i^{\text{vap}})}{P_i^{\text{vap}}} \tag{7}$$

where $\underline{V}_{i/L}$ is the molar volume of the liquid phase. It should be noted that the 3rd term on the RHS of eq 7 accounts for the vapor phase nonideality, and the first two terms give the vapor pressure when the vapor phase is assumed to be ideal.

The solvation free energy consists of a van der Waals term, which includes cavity and dispersion interactions, as well as an electrostatic term, which contains the electrostatic interactions and polarization between the solute and solvent

$$\Delta \underline{G}^{*\text{sol}}_{i/i} = \Delta \underline{G}^{*\text{vdw}}_{i/i} + \Delta \underline{G}^{*\text{el}}_{i/i} \tag{8}$$

The electrostatic contribution to the free energy is represented by the sum of the ideal solvation (is), charge averaging correction (cc), restoring free energy (res, including hydrogen bond corrections), i.e.,

$$\Delta \underline{G}^{*\text{el}}_{i/i} = \Delta \underline{G}^{*\text{is}}_{i} + \Delta \underline{G}^{*\text{cc}}_{i} + \Delta \underline{G}^{*\text{res}}_{i/i} \tag{9}$$

The ideal solvation term is the difference in energy between the ideal gas and ideal (or perfect) conductor state

$$\Delta \underline{G}^{*\text{is}}_{i} = E^{\text{COSMO}} - E^{\text{IG}} = E(q^*) + E_{\text{diel}}(q^*) - E(0) \tag{10}$$

in which $E(q)$ is the total electronic energy of the solute in the external field arising from the charges, q, located at the molecular surface [note that $E^{\text{IG}} = E(0)$]; $E_{\text{diel}}(q^*) = 1/2 \sum_\upsilon \phi_\upsilon q^*_\upsilon$ is the dielectric energy,[20] where q^*_υ is the screening charge at position υ and ϕ_υ is the electrostatic potential due to the solute at position υ. The sum of the first and the third terms in the second identity of eq 10 is the solute electronic polarization energy; the second term is the solute−solvent electrostatic interaction.

The other electrostatic contributions are determined by treating the interactions in the condensed phase as interacting surface segments, similar to the COSMO-RS and the COSMO-SAC models, but with some modifications. The charge-averaging correction considers the shift of the energy state from the ideal screening charges, q^*_υ, determined in the COSMO calculation to some "averaged" charges, q_υ, that are later used in the determination of the electrostatic interactions (i.e., the restoring and hydrogen bond interactions). The charge averaging is used to ensure that segment pairs form independent pairs in solution, a fundamental assumption in the COSMO-SAC model. An empirical charge-averaging scheme was suggested by Klamt et al.;[10] however, a different radius than the one used in the later restoring free energy calculation is needed. Here, we propose a semi-theoretical approach that allows the use of a single averaging radius throughout our model, and we will show later that it also gives excellent internal consistency in our model. The new approach is based on the conservation of overall dielectric energy before and after averaging the charges within a circular region of radius of $\sqrt{4\pi/a_{\text{eff}}}$ centered on a segment υ

$$E_{\text{diel}}(q^*) = \sum_{v}^{N} \frac{a_{\text{cosmo}}}{2} \sqrt{\frac{4\pi}{a_v}}\, q_v^{*2} + \sum_{v=1}^{N} \sum_{u=v+1}^{N} \frac{q_u^* q_v^*}{r_{uv}} = \frac{a_{\text{cosmo}}}{2} \sqrt{\frac{4\pi}{a_{\text{eff}}}}\, q_v^{\ 2} + \sum_{v=m+1}^{N} \frac{a_{\text{cosmo}}}{2} \sqrt{\frac{4\pi}{a_v}}\, q_v^{*2} + \sum_{v=m+1}^{N} \sum_{u=v+1}^{N} \frac{q_u^* q_v^*}{r_{uv}} \quad (11)$$

which leads to the screening charge on surface segment v averaged over an area of a_{eff}

$$q_v^{\ 2} = \frac{\sum_{u=1}^{N} w(r_{uv}) \frac{a_{\text{cosmo}}}{2} \sqrt{\frac{4\pi}{a_u}}\, q_u^{*2} + \frac{1}{2} \sum_{u=1}^{N} w(r_{uv}) \sum_{k=1,k\neq u}^{N} \frac{q_u^* q_k^*}{r_{uk}}}{\frac{a_{\text{cosmo}}}{2} \sqrt{\frac{4\pi}{a_{\text{eff}}}}} \quad (12)$$

where $a_{\text{cosmo}} = 1.07$ is a constant in the COSMO theory;[21] N is the total number of surface segments on the molecule; a_v is the surface area of segment v; r_{uv} is the distance between segments u and v; $w(r_{uv}) = 1$ if $r_{uv} \leq \sqrt{a_{\text{eff}}/\pi}$ and $w(r_{uv}) = 0$ if $r_{uv} > \sqrt{a_{\text{eff}}/\pi}$. Equation 12 only determines the magnitude of q_v; the sign of the charge on segment v is set to be the same as the original net charge over the averaging area a_{eff}. The correction for the free energy shift due to the charge averaging process is[10]

$$\Delta \underline{G}_i^{*\text{cc}} = f_{\text{pol}}^{\ 1/2} [E_{\text{diel}}(q) - E_{\text{diel}}(q^*)] \quad (13)$$

The polarization factor f_{pol} in eq 13 takes into account the solute electron polarization energy, normally ~20% of the total polarization energy.[9,10] This value of f_{pol} can be determined by regressing $E^{\text{COSMO}} - E^{\text{IG}} = f_{\text{pol}}^{1/2} E_{\text{diel}}(q^*)$ for a set of compounds and was determined to be 0.6916, as is discussed later.

The restoring term accounts for changing from the corrected screening state to the actual solvent, and is obtained as a summation over the segment activity coefficients[11]

$$\frac{\Delta \underline{G}_{i/i}^{*\text{res}}}{RT} = n \sum_{\sigma_m} p(\sigma_m) \ln \Gamma_i(\sigma_m) \quad (14)$$

where n is the number of segments in the molecule, which is the ratio of the surface area to the area of the standard surface segment ($= A_i/a_{\text{eff}}$) [not to be confused with the total number of segments, N, used in eqs 11 and 12 that is determined in the COSMO calculation]; $p(\sigma)$, called the σ profile, is the probability of finding a segment of charge density $\sigma = q/a$; and $\Gamma(\sigma)$ is the segment activity coefficient obtained from

$$\ln \Gamma_S(\sigma_m) = -\ln\left\{ \sum_{\sigma_n} p_S(\sigma_n) \Gamma_S(\sigma_n) \exp\left[\frac{-\Delta W(\sigma_m,\sigma_n)}{RT}\right]\right\} \quad (15)$$

where $\Delta W(\sigma_m,\sigma_n)$ is the segment exchange energy calculated from

$$\Delta W(\sigma_m,\sigma_n) = f_{\text{pol}} \frac{0.3 a_{\text{eff}}^{\ 3/2}}{2\epsilon_0} (\sigma_m + \sigma_n)^2 \quad (16)$$

ϵ_0 is the permittivity of vacuum, and the polarization factor f_{pol} allows for the reduction of the misfit energy due to the electronic polarization of the molecules.

For hydrogen-bonding species, a further correction is introduced to account for strong directional attractions between segments with especially large charge density differences. In the original COSMO-RS and COSMO-SAC models,[10,11] to correct for hydrogen-bonding (H-bond) effects, a simple expression based on the use of a cutoff σ_{hb} value was used. Here, we instead propose the use of a second σ profile, $p_{\text{hb}}(\sigma)$, only for hydrogen-bonding atoms, which are defined to be oxygen, nitrogen, fluorine, and the hydrogen atoms connected to these atoms. The use of a separate σ profile for H-bond atoms not only allows for a more specific identification of segment pairs that are likely to form H-bonds but also eliminates one parameter (σ_{hb}) used in the previous models. Furthermore, one could, in principle, include in $p_{\text{hb}}(\sigma)$ only segments at certain positions with respect to the H-bonding atoms in order to consider the angular dependence of the H-bond effects, though this is not used in this work. Thus, the total σ profile has two components: one from hydrogen-bonding atoms and the other from non-hydrogen-bonding atoms, i.e., $p(\sigma) = p_{\text{nhb}}(\sigma) + p_{\text{hb}}(\sigma)$. The segment activity coefficient (eq 15) becomes

$$\ln \Gamma^t(\sigma_m^t) = -\ln\left\{ \sum_{s}^{\text{nhb,hb}} \sum_{\sigma_n} p^s(\sigma_n^s) \Gamma^s(\sigma_n^s) \exp\left[\frac{-\Delta W(\sigma_m^t,\sigma_n^s)}{RT}\right]\right\} \quad (17)$$

where the superscript t and s can be either hb or nhb, representing the property for a hydrogen-bonding or non-hydrogen-bonding segment. The segment interaction (eq 16) takes the following form

$$\Delta W(\sigma_m^t,\sigma_n^s) = f_{\text{pol}} \frac{0.3 a_{\text{eff}}^{\ 3/2}}{2\epsilon_0} (\sigma_m^t + \sigma_n^s)^2 - c_{hb}(\sigma_m^t,\sigma_n^s)(\sigma_m^t - \sigma_n^s)^2 \quad (18)$$

where $c_{\text{hb}}(\sigma_m^t,\sigma_n^s)$ is the following temperature-independent parameter

$$c_{\text{hb}}(\sigma_m^t,\sigma_n^s) = \begin{cases} c_{\text{hb}} & \text{if } s = t = \text{hb, and } \sigma_m^t \cdot \sigma_n^s < 0 \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

and the restoring free energy becomes

$$\frac{\Delta \underline{G}_{i/i}^{*\text{res}}}{RT} = n \sum_{s}^{\text{nhb,hb}} \sum_{\sigma_m} p(\sigma_m^s) \ln \Gamma_i^s(\sigma_m^s) \quad (20)$$

The van der Waals contribution to the solvation free energy consists of the dispersion and cavity formation (or hard core) terms and is expressed as the sum of the Helmholtz energy plus the PV term

$$\frac{\Delta \underline{G}_{i/i}^{*\text{vdw}}}{RT} = \frac{\Delta \underline{A}_{i/i}^{*\text{vdw}}}{RT} + Z_{i/L} - Z_{i/IG} = \frac{\Delta \underline{A}_{i/i}^{*\text{disp}}}{RT} + \frac{\Delta \underline{A}_{i/i}^{*\text{cav}}}{RT} - 1 \quad (21)$$

where $Z = \underline{PV}/RT$ is the compressibility factor. Here, $Z_{i/IG}$ is the compressibility factor for the ideal gas, and $Z_{i/L}$ for most

liquids is close to zero and, therefore, has been neglected here. The dispersion term is written as a first-order mean field term accounting for attractions between all possible pairs of atoms, each in different molecules, and is given by

$$\frac{\Delta \underline{A}_{i/i}^{*\mathrm{disp}}}{RT} = \frac{\sum_j \sum_k \epsilon_{jk} m_j^i m_k^i}{RT\underline{V}_{i/L}} \tag{22}$$

where $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$ is the energy parameter between atoms of species of i and j with units of energy × volume. The quantity m_i is an effective number of atoms of species i in a molecule that is calculated from

$$m_i = \sum_{a \in i} \left(\frac{S_a}{S_{a0}} \right)^q \tag{23}$$

where S_a is the exposed surface area (or solvent excluded surface area) of atom a of species i, S_{a0} is the surface area of the bare atom, and q is an adjustable parameter. The exposed surface area is determined using the same set of atomic radii as in the COSMO calculation for the electrostatic contributions. Although the exponent q in eq 23 has been empirically introduced, it has some physical significance. When q is zero, m_i is the actual number of atoms of species i, and the dispersion model is a mean-field model disregarding the detailed molecular structure. On the other hand, when q is unity, the dispersion interactions depend strongly on the surface areas. It is presumed that the real situation is intermediate between the two extreme cases, and this parameter is determined such that it gives an optimum fit including branched and cyclic molecules.

The cavity free energy is approximated using thermodynamic perturbation theory[14,15] for nonspherical molecules

$$\frac{\Delta \underline{A}_{i/i}^{*\mathrm{cav}}}{RT} = (2\alpha - 1)\frac{\eta(4-3\eta)}{(1-\eta)^2} - (2\alpha - 2)\ln\left[\frac{1-\eta/2}{(1-\eta)^3}\right] \tag{24}$$

where α is a sphericity parameter and η is the packing fraction. The packing fraction is the ratio of the hard-core volume and the molar volume

$$\eta = V_h/V = N_{av}V_h/\underline{V} \tag{25}$$

where N_{av} is Avogadro's number and $\underline{V}$ is molar volume. The sphericity parameter is a measure of the asymmetry of the molecule as defined by

$$\alpha = \frac{R_h S_h}{3V_h} \tag{26}$$

where S_h and V_h are the surface area and volume of the hard core of molecule, and R_h is the mean radius of curvature defined as the average over all directions (solid angles) of the distance from the center of mass of the molecule to the plane that is tangent to the molecular surface.[22] The values of these parameters are determined from the equilibrium geometry of a dominant conformation of the molecule using the specified value of hard-core radius of each atom. Rather than determining these geometrical properties from a Monte Carlo procedure as used by others,[22] which takes minutes for medium sized molecules and longer for large molecules, we use a much more efficient approach detailed below that takes less than a second.

For compounds that associate in the vapor phase, a correction for the vapor-phase nonideality can be made by computing the compressibility factor and the fugacity coefficient from an equation of state, e.g., the virial equation of state truncated at the 2nd order term.[23] In this work, since we are interested in the vapor pressure only at the normal boiling point (1 atm), we assume that the gas phase is ideal and the fugacity coefficient is unity and exclude species that strongly associate in the vapor phase, such as HF, as second virial coefficients are unavailable for this species.

TABLE 1: Parameters Used in the Model

Universal Parameters

parameter	value
a_{eff} (Å^2)	9.24
c_{hb} (kJ/mol Å^4/e^2)	28 476.21
q	0.272
$R_{\mathrm{hc}}/R_{\mathrm{el}}$	0.611

Atom Specific Parameters

atom type	R_{el} (Å)	ϵ_i/R (K Å^3)
H	1.57	638.69
C	1.90	12773.35
N	1.81	8088.86
O	1.70	6571.79
F	1.71	4062.58
Cl	1.98	27355.53

Finally, the enthalpy of vaporization is obtained analytically from the derivative of the vapor pressure with respect to temperature

$$\frac{\Delta H_i^{\mathrm{vap}}}{RT} = \frac{(\underline{V}_{i/V} - \underline{V}_{i/L})}{R}\frac{dP_i^{\mathrm{vap}}}{dT} = \left[T - \frac{P_i^{\mathrm{vap}}\underline{V}_{i/L}}{R}\right]\frac{d\ln P_i^{\mathrm{vap}}}{dT} \tag{27}$$

Computational Details

To calculate vapor pressure from eq 7 we need to determine the solvation free energy $\Delta \underline{G}_{i/i}^{*\mathrm{sol}}$, the liquid volume $\underline{V}_{i/L}$, and the vapor-phase fugacity coefficient $f(T,P_i^{\mathrm{vap}})/P_i^{\mathrm{vap}}$. Experimental values for $\underline{V}_{i/L}$ taken from the DIPPR database have been used, and the vapor-phase fugacity coefficient is assumed to be unity through out this work. The calculation of the solvation free energy $\Delta \underline{G}_{i/i}^{*\mathrm{sol}}$ consists of two parts: the quantum mechanical (QM) part and the statistical mechanical (SM) part. The QM part of the calculation determines the ideal solvation free energy $\Delta \underline{G}_i^{*\mathrm{is}}$ and also outputs the molecular area and screening charges that are needed later in the SM part for calculating other components of $\Delta \underline{G}_{i/i}^{*\mathrm{sol}}$ (i.e., $\Delta \underline{G}_{i/i}^{*\mathrm{cc}}$, $\Delta \underline{G}_{i/i}^{*\mathrm{res}}$, $\Delta \underline{G}_{i/i}^{*\mathrm{disp}}$, and $\Delta \underline{G}_{i/i}^{*\mathrm{cav}}$).

The COSMO method[21] implemented in Jaguar 4[24] is used for the QM calculation [Appendix I]. Jaguar is used because it is one of the most efficient QM packages that produce very accurate molecular energies from the density functional theory (DFT). [Note that the terms in the parentheses below in this paragraph are used to refer to run time options in the Jaguar program.] First, the geometry of a molecule is optimized using the density-functional theory (DFT) with the B3LYP functional[25,26] and 6-31+G** basis set (idft = 22111 and basis = 6-31G**+). The same minimum energy structure (nogas = 0 and igeopt = 0) is then used in the COSMO solvation calculation (isolv = 3), which is also performed at the B3LYP/6-31+G** level. The solvation cavity is determined using the Connolly algorithm[27] and is represented by a set of points with a density of 4 points/Å^2 (cosfden = 4). The values of the solute atomic radii (vdw2) are listed in Table 1, and a probe with a radius of 1.57 Å (radprb = 1.57, same as the hydrogen radius) is used. Ultrafine grid and tight cutoffs are selected (iacc = 1).

The escaping charges are corrected using the double shell method by Klamt and Jonas[28] with the outer shell thickness set to 90% of rprobe (cskin = 1.413). The time used in the COSMO calculation ranges from seconds for small solutes (e.g., 14 s for water) to hours for larger solutes (e.g., 1.8 h for 1-octadecanol) on a Linux machine with an Intel P4 Xeon CPU.

The ideal solvation free energy $\Delta\underline{G}_i^{*\mathrm{is}}$ (eq 10) is obtained directly from the output of the Jaguar COSMO calculation (extension jcosmo). Also reported in the output file are the molecular surface area A_i, volume V_i, the screening charges q_υ^*, area a_υ, and position vector $\bar{r}_\upsilon$ of each molecular surface segment.

The magnitude of the "averaged" screening charges of all the segments are determined using to eq 12 with an averaging area $a_{\mathrm{eff}} = 9.24$ Å^2, and the sign of each segment is set to be the same as the sum of original net charge over the averaging area. [q_υ is set to zero in those rare cases in which the RHS of eq 12 is negative, which would lead to an imaginary value of q_υ.] The dielectric energies for the two sets of screening charges are then determined from eq 11, and the correction for the free energy shift due to the charge averaging process $\Delta\underline{G}_{i/i}^{*\mathrm{cc}}$ is calculated from eq 13.

Using the averaged screening charges, q, and charge density $\sigma_\upsilon = q_\upsilon/a_\upsilon$, the area weighted σ profile $A_i\, p(\sigma)$ is obtained by collecting the histogram of segment areas as a function of charge density with an interval of 0.01 e/Å^2. Segments associated with the hydrogen-bonding atoms (oxygen, nitrogen, fluorine, and the hydrogen atoms connected to these atoms) are collected in $p_{\mathrm{hb}}(\sigma)$ and others in $p_{\mathrm{nhb}}(\sigma)$. The segment activity coefficient, $\Gamma(\sigma)$, is then solved from eq 17 by iteration starting from $\Gamma(\sigma) = 1$ until a self-consistent solution is obtained. The segment exchange energy is calculated from eq 18 with the effective segment area (a_{eff}) set to 9.24 Å^2, the same value used in the charge averaging process. For hydrogen-bonding segments, the constant c_{hb} is set to 28476 kJ/mol Å^4/e^2. The restoring free energy $\Delta\underline{G}_{i/i}^{*\mathrm{res}}$ is then obtained by summing $\Gamma(\sigma)$ for all the segments (eq 20).

To determine the dispersion contributions $\Delta\underline{A}_{i/i}^{*\mathrm{disp}}$ (eq 22), we need the energy parameter ϵ_i, the exposed surface area S_a, the surface area of a bare atom S_{a0}, the scaling factor q, and the molar volume $\underline{V}_{i/L}$. The values of ϵ_i for each atom tabulated in Table 1 are used. Using the same atomic radii (R_a) as in the COSMO calculations (Table 1), the bare atom area S_{a0} is simply $4\pi R_a^2$. The exposed surface area of each atom is determined by doing the Connolly surface calculation using a probe radius of 0.01 Å, a point density of 32 points/Å^2, and summing the area of segments associated with that atom. It is important to note that in the calculation of S_a, we assumed that a hydrogen atom does not screen its neighboring covalently bonded atoms. For example, the surface area of carbon is not reduced by the presence of covalent hydrogen atoms but only by other heavy atoms. This is reasonable because interactions between heavier atoms are longer ranged and stronger than those of hydrogen. This way of disregarding hydrogen atoms is similar to the united atom description frequently used in molecular simulations but its role in the solvation model is to remedy a deficiency of the surface-area-dependent dispersion model. The scaling factor $q = 0.272$ was used (listed in Table 1), and the liquid molar volumes were taken from the DIPPR database.[29]

Finally, we need the packing fraction η, the hard core surface area S_{h}, volume V_{h}, and mean radius of curvature R_{h}, which defines the sphericity parameter α, for the cavity formation free energy $\Delta\underline{A}_{i/i}^{*\mathrm{cav}}$ (eq 24). A set of smaller atomic radii is used to determine the hard-core properties. For simplicity, we assume that the ratio of the hard core to electrostatic (and dispersion) radii, i.e., $R_{\mathrm{hc}}/R_{\mathrm{el}}$, is a constant of 0.611. The values of S_{h} and V_{h} were also obtained using a Connolly surface calculation with the hard core radii, a probe radius of 0.01 Å, and a point density of 32 points/Å^2. To determine R_{h}, we first construct a sphere with 162 evenly distributed points and calculate the unit outward normal vectors of each point. For each unit vector direction, we determine the largest distance from the solute center of mass to surface tangential plane on each atom [$\bar{u}\cdot(\bar{r}_a - \bar{r}_{cm}) + R_a$ where $\bar{r}_a$ and $\bar{r}_{cm}$ are the position vectors of atom a and the center of mass of the solute, R_a is the radius of atom a], and choose the largest value among the ones for all the atoms to be the radius of curvature in this direction. Then R_{h} is the average of the radius of curvature over the 162 directions. In Appendix II, we show that this approach gives similar results to the more rigorous MC approach,[22] but the time needed is reduced from minutes to less than one second. The packing fraction is determined as the ratio (eq 25) of V_{h} and the liquid molar volume. The cavity formation contribution can then be determined from eq 24.

To evaluate the heat of vaporization, we need the temperature derivative of $\ln P_i^{\mathrm{vap}}$. This is calculated analytically as expressions are available for the derivatives of each component of the solvation energy with respect to temperature [Appendix III].

Parametrization

In the model proposed above, there are a total of four universal parameters (a_{eff}, c_{hb}, q, $R_{\mathrm{hc}}/R_{\mathrm{el}}$), two atom specific parameters for each atom (R_i and ϵ_i), all of which are not system specific. Thus, there are 16 ($4 + 6 \times 2$) parameters for the set of 371 compounds studied in this work. (A complete list of all the compounds can be found in the Supporting Information for this paper.) The parametrization was carried out to achieve optimum fit for the vapor pressures and the enthalpies of vaporization at the normal boiling point for these compounds using the experimental values in the DIPPR database.[29] The objective function used in the minimization was the sum of RMS errors

$$Obj = \left[\frac{1}{N_{\mathrm{data}}}\sum_i^{N_{\mathrm{data}}}(\ln p_{\mathrm{exp}} - \ln p_{\mathrm{cal}})^2\right]^{1/2} + W\left[\frac{1}{N_{\mathrm{data}}}\sum_i^{N_{\mathrm{data}}}\left(\frac{\Delta H_{\mathrm{exp}} - \Delta H_{\mathrm{cal}}}{\Delta H_{\mathrm{exp}}}\right)^2\right]^{1/2} \quad (28)$$

A weight factor W of 2 was used for the RMS error of the heat of vaporization to achieve an optimum correlation for both properties. The simplex method[30] followed by a simulated annealing[30] minimization was used to ensure good parametrization.

The parametrization proceeded as follows. First, a subset of only C- and H-containing compounds (77 compounds consisting of 37 alkanes, 24 aromatics, 9 alkenes, and 7 alkynes) were used to optimize R_{el} for these two atoms. The optimized atomic radii for C and H were then fixed in the further optimization of the radii for other atoms. The optimization of R_{el} for the oxygen atom then used those 77 compounds plus others that contained only C, H, and O atoms (alcohols, aldehydes, ketones, ethers, and esters). Next, additional compounds containing C, H, O, and N atoms (nitro compounds and nitriles) were used to obtain R_{el} for the nitrogen atom (with R_{el} of C, H, and O fixed at the previously optimized values). The atomic radii of fluorine and chlorine were then found using subsets of the data containing

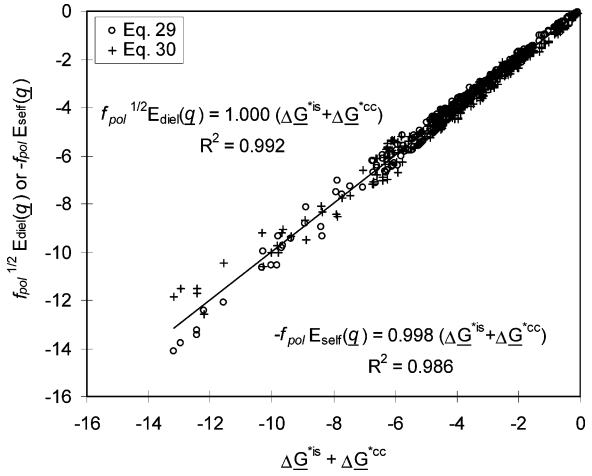


Figure 1. Comparison of the dielectric energy E_{diel} and the self-energy E_{self} to the corrected ideal solvation free energy.

the 77 hydrocarbons plus 5 F-containing and 23 Cl-containing compounds, respectively. During the R_{el} optimization, we fixed the value of a_{eff} at 7.5 Å^2 but allowed all other relevant variables to change. Once the atomic radii for all the atoms were determined, the optimization for the other 10 parameters was performed using the full set of 371 compounds.

The determination of the polarization factor f_{pol} requires some additional discussion. The proportion of solute electron polarization can be estimated from the ratio of the ideal solvation free energy to the dielectric energy $f_{pol}^{1/2} = (E^{COSMO} - E^{IG})/E_{diel}(q^*) = \Delta G^{*is}/E_{diel}(q^*)$. Using eq 13, we can express the energies in terms of the "averaged" screening charges

$$f_{pol}^{1/2}E_{diel}(q) = \Delta G^{*is} + \Delta G^{*cc} \qquad (29)$$

Figure 1 shows that the regression for 419 species using eq 29. [Note that there are an additional 48 compounds in this list but not in the vapor pressure predictions. These are additional compounds for which the experimental vapor pressure or molar volume data were not available in the DIPPR database[29] or for which there is a significant association in the vapor phase, such as HF.] An excellent correlation is obtained ($R^2 = 0.992$) with f_{pol} set to 0.6917. It should be noted that this value of f_{pol} has been determined using only QM data ($E_{diel}(q)$, ΔG^{*is}, and ΔG^{*cc}). Therefore, it is not considered to be a parameter in our model, as it was not fit to experimental data.

Results and Discussion

The accuracy of the proposed model in predicting the vapor pressure and heat of vaporization for the set of 371 compounds at their normal boiling points is listed in Table 2 and shown in Figure 2a and b. (A detailed list of the predicted values for each compound is provided as Supporting Information.) The overall absolute average deviation in vapor pressure is 76%, and the root-mean-square deviation in heat of vaporization is 4.81 kJ/mol. In terms of compound type, the lowest accuracy is obtained for alcohols (90% error in P^{vap} and 8.86 kJ/mol error in ΔH^{vap}) and compounds containing two or more functional groups (average 100% error in P^{vap} and 5.94 kJ/mol error in ΔH^{vap}). Considering the large range of normal boiling points (from 191 to 610 K), the accuracy from the proposed model, which uses only temperature-independent parameters, is quite satisfying. Although the agreement is far from perfect, the present solvation model shows how effectively a theoretically based quantum mechanics model can describe the vapor pressures and heats of vaporization over a considerable temperature range with a single set of parameters.

TABLE 2: Accuracy of the Proposed Model on the Prediction of Vapor Pressure and Heat of Vaporization at the normal Boiling Point. (Numbers in Parentheses Indicate the Maximum Deviation)

compound type	no. of compounds	AD% in P^{vap}	RMSD (kJ/mol) in ΔH^{vap}
alkane	37	56 (194)	1.55 (3.59)
alcohol	31	90 (227)	8.86 (13.75)
aldehyde	8	28 (62)	2.34 (3.80)
ketone	21	67 (222)	2.17 (5.23)
acid	11	33 (102)	6.69 (11.16)
ester	21	54 (111)	2.09 (4.26)
ether	16	73 (243)	2.84 (6.81)
nitro	5	58 (113)	3.14 (4.22)
nitrile	7	70 (270)	3.47 (4.85)
amine	22	86 (319)	3.59 (7.27)
aromatic	24	38 (83)	3.01 (5.60)
alkene	9	47 (124)	1.42 (2.80)
alkyne	7	45 (93)	1.96 (3.01)
F containing	5	53 (127)	2.38 (3.47)
Cl containing	24	75 (593)	2.88 (8.07)
multifunctional	123	100 (851)	5.94 (19.19)
overall	371	76 (851)	4.81 (19.19)

To investigate the predictability of the proposed model at other temperatures, we also compare the prediction of properties at −50, −20, −10, +10, +20, and +50 degrees away from the normal boiling temperature, T_b, for all 371 compounds for which experimental values are available. Figure 3a shows the overall absolute average percentage error in the vapor pressure and the root-mean-square deviation in the heat of vaporization at all seven temperatures (T_b plus the aforementioned six temperatures). [Note that the percentage error is calculated from the exponential of the RMSD in ln P^{vap} (first term on the RHS of eq 28) minus unity.] The RMSD in ΔH^{vap} decreases monotonically from 5.60 kJ/mol at 50 K below T_b to 4.31 kJ/mol at 50 K above T_b. The error in P^{vap} appears to increase rapidly for temperatures below T_b (109% at 50 K below T_b) and remains at about 76% above T_b. The decreased accuracy at temperatures below the normal boiling point is a result of the increasing error in ΔH^{vap}. At temperatures higher than T_b, the improvement in the prediction of ΔH^{vap} helps maintain the accuracy in P^{vap} at 76%.

Figure 4 shows the components of the solvation free energy to the predicted vapor pressure. The contribution from the molar volume (ln(RT/V)) is about 16.8 units and is insensitive to the compound type. At the normal boiling point, the contribution of the ln P^{vap} (=ln(101 325 Pa)) is always 11.5 units. The other solvation free energy components account for the difference between these two values (5.3 units). The (corrected) ideal solvation free energy and the dispersion contributions are, in general, negative and therefore reduce the vapor pressure, whereas the restoring free energy and the cavity term are positive and result in an increase in the vapor pressure. Furthermore, the ideal solvation free energy, the restoring free energy, and the dispersion and the cavity contributions appear to be paired properties as they are of opposite sign and of about the same order of magnitude. Hence, to provide accurate predictions, it is important to have an accurate model for each solvation component.

It is useful to examine the validity of using eq 12 for the charge averaging process. Since the purpose of the charge averaging process is to eliminate the correlations between surface segments, a necessary internal consistency[10] condition for the later calculation of the restoring free energy (eq 20)

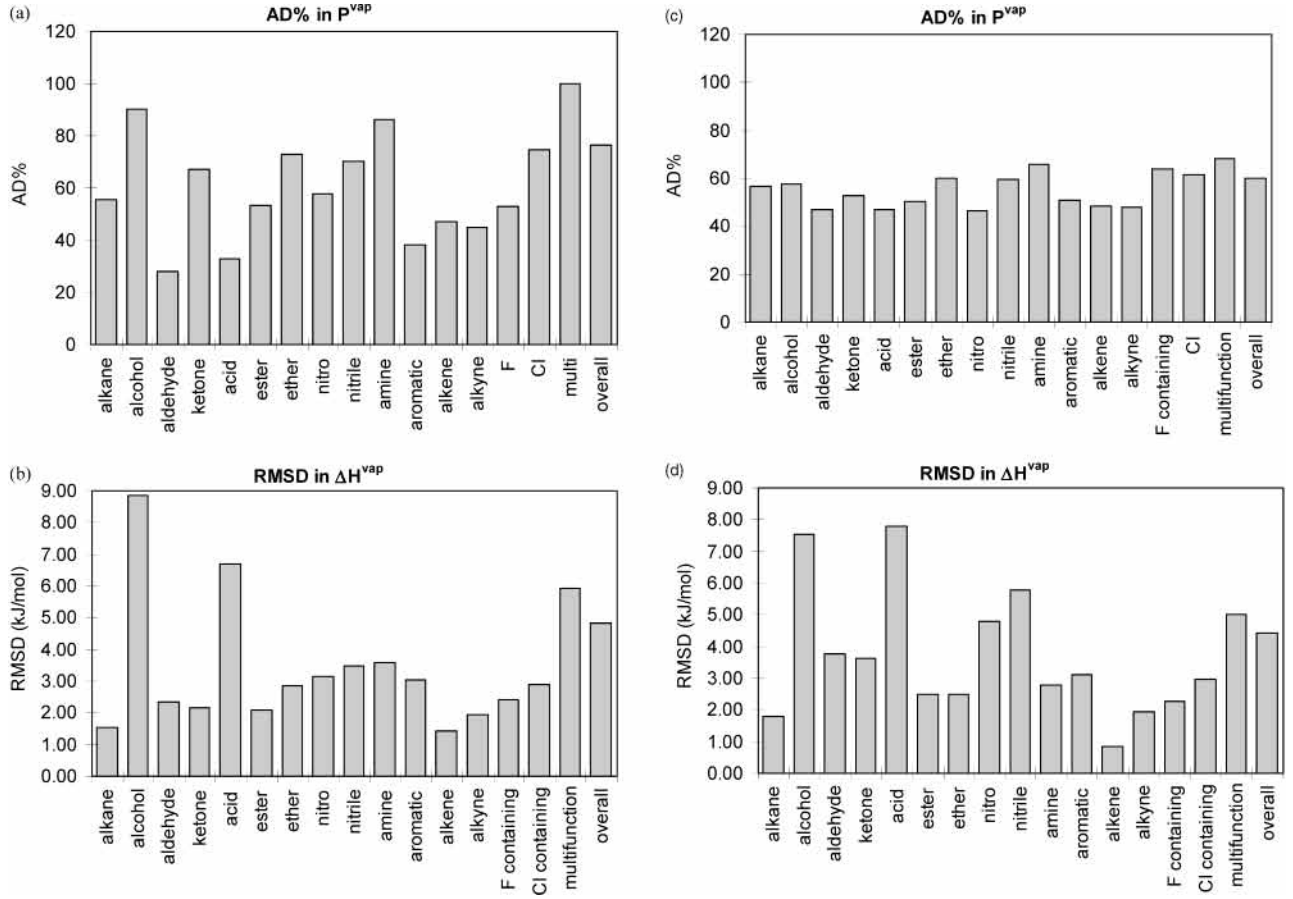


Figure 2. Accuracy of the predicted (a) vapor pressure and (b) heat of vaporization at the normal boiling point. Accuracy of the predicted (c) vapor pressure and (d) heat of vaporization at the normal boiling point. (13 dispersion coefficients model).

requires that the corrected ideal solvation free energy ($\Delta\underline{G}^{*is} + \Delta\underline{G}^{*cc}$) be equivalent to the negative of the energy needed to remove the uncorrelated screening charges, q. This energy is simply the sum of the exchange energy for all segments in contact with a vacuum, i.e., the self-energy $E_{self}(q)$

$$\Delta\underline{G}^{*is} + \Delta\underline{G}^{*cc} = -\sum_m f_{pol} \frac{0.3a_{eff}^{3/2}}{2\epsilon_0}(\sigma_m + 0)^2 = -f_{pol} n \sum_{\sigma_m} p(\sigma_m) \frac{0.3a_{eff}^{3/2}}{2\epsilon_0} \sigma_m^2 = -f_{pol} E_{self}(q) \quad (30)$$

In Figure 1, we also show the correlation between the $\Delta\underline{G}^{*is} + \Delta\underline{G}^{*cc}$ and $-f_{pol}E_{self}(q)$ for 419 compounds with f_{pol} set to the same value of 0.6917. The slope deviates from unity by only 0.2% and has a high correlation coefficient ($R^2 = 0.986$). Thus, the charge averaging procedure provides a good set of uncorrelated averaged screening charges.

It is interesting to examine the variation of the σ profile with the change of the averaging radius, or equivalently, the effective segment surface area a_{eff}. Figure 5 shows the σ profile for six representative compounds using $a_{eff} = 7.5$, 9.24, and 12.5 Å^2, respectively. The charge profiles for the atoms that do not H-bond, $p^{nhb}(\sigma)$, are shown in the figures on the left, and the profiles for the H-bond atoms, $p^{hb}(\sigma)$, are shown on the right. As would be expected, increasing the value of a_{eff} smears out the fine structures in the σ profile. While $p^{nhb}(\sigma)$ is relatively insensitive to changes of a_{eff}, the peaks in $p^{hb}(\sigma)$ broaden and are more separated as a_{eff} decreases. Thus, the value of a_{eff} has a strong influence on the restoring free energy. However, we find that the sum of the restoring free energy and the ideal solvation free energy are often very close to zero, making the prediction of vapor pressures quite insensitive to the choice of a_{eff}. We expect that this parameter can be more accurately determined using VLE or LLE data, where the restoring free energy (and therefore the values of the activity coefficients) dominates the prediction of mixture phase behavior.

Figure 6 shows the segment activity coefficient (ln Γ) determined using the parameters listed in Table 1 at the normal boiling temperature of each of the compounds considered in Figure 5. For compounds that do not contain H-bond atoms (e.g., hexane and chloroform), the segment activity profiles ln Γ^{hb} and ln Γ^{nhb} are very similar and distinctly different from the profiles for the compounds that contain H-bond atoms. The value of ln $\Gamma(\sigma)$ is related to the work needed to remove the screening charge from a segment with a charge density σ.[11] The value of ln Γ^{nhb} approaches large positive numbers as the magnitude of the segment charge density, $|\sigma|$, increases, as more work is needed to remove high-charge-density segments or, equivalently, it is unfavorable to have a high-charge-density segment in the system. However, ln Γ^{hb} becomes negative for negatively charged segments when a compound contains only a hydrogen-bond acceptor (e.g., acetone), and ln Γ^{hb} is negative in value at both ends if a compound contains both the H-bond donor and acceptor (e.g., hexylamine, 1-hexanol, and water). The positive values of ln Γ^{hb} of hexane throughout the whole range of σ indicates hexane will not mix favorably with strongly H-bonding species, such as water (for which $p^{nhb}(\sigma)$ is essentially zero). The negative value of ln Γ^{hb} reflects the fact that additional free energy is released when an H-bond is formed.

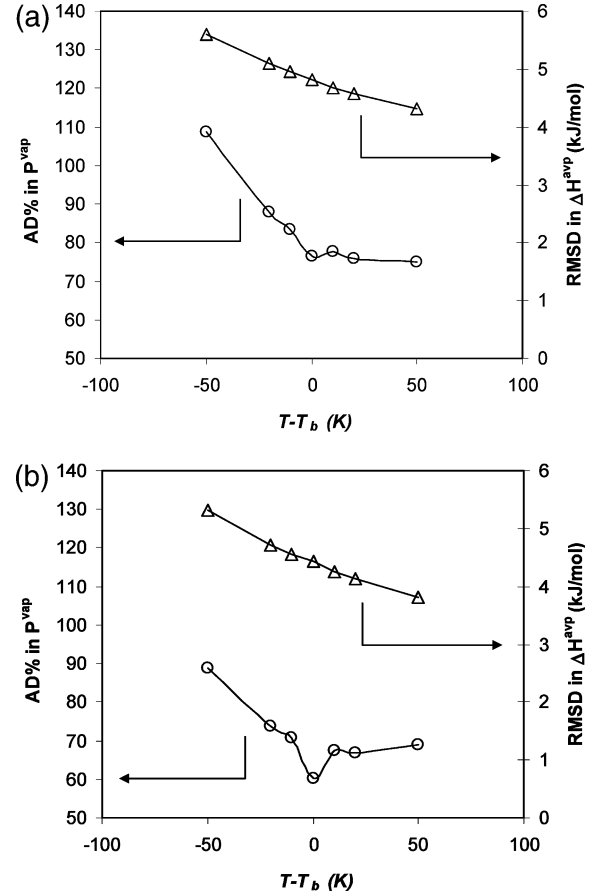


Figure 3. (a) Accuracy of the proposed model in vapor pressure and heat of vaporization at different temperatures. (b) Accuracy of the proposed model in vapor pressure and heat of vaporization at different temperatures (13 dispersion coefficients model).

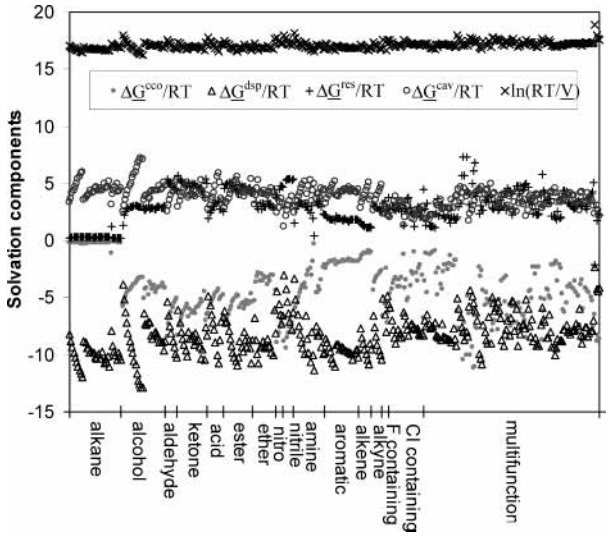


Figure 4. Components of the solvation free energy for all the compounds studied in this work.

The positive ln Γ^{hb} value at large positive σ and negative ln Γ^{hb} value at large negative σ indicate that acetone, being a H-bond acceptor, would favorably mix with H-bond donors, i.e., compounds having a H-bond hydrogen, which provides negatively charged segments. Furthermore, the negative values of ln Γ^{hb} for compounds having both H-bond donor and acceptor atoms make these compounds more stable in the liquid phase, resulting in a higher normal boiling point compared to other similar sized molecules.

TABLE 3: Comparison of the Dispersion Coefficients to the Lennard−Jones Attraction Energy in the Dreiding Force Field and the Optimized Radii to the Lennard−Jones Radii

atom	D_0 (kJ/mol)	$3\epsilon_i/4\pi R_i^3$ (kJ/mol)	R_i (Dreiding)	R_i (this work)
H	0.0635	0.3107	1.60	1.57
C	0.3975	3.4201	1.95	1.90
N	0.3235	2.6125	1.83	1.81
O	0.4000	2.6417	1.70	1.70
F	0.3031	1.5398	1.74	1.71
Cl	1.1842	7.0391	1.98	1.98

TABLE 4: Parameters Used in the Model (13 dispersion coefficients model)

Universal Parameters

parameter	value
a_{eff} (Å^2)	9.24
c_{hb} (kJ/mol Å^4/e^2)	28476.21
q	0.272
R_{hc}/R_{el}	0.611

Atom Bonding Specific Parameters

atom type (bonding specific)	R_{el} (Å)	ϵ_i/R (K Å^3)
H	1.57	1173.51
C_t (sp^3)	1.90	8277.96
C_d (sp^2, ethene, aromatic)	1.90	11 235.27
C_s (sp, ethyne, nitrile)	1.90	12 337.01
N_t (sp^3, NH_3, amine, aniline)	1.81	9039.50
N_d (sp^2, nitro, aromatic)	1.81	11 527.19
N_s (sp, nitrile)	1.81	4057.52
O_tH (sp^3, bonded to H, alcohol, carboxylic)	1.70	9113.10
O_t (sp^3, ether, ester C−O−C, aromatic oxygen)	1.70	8910.89
O_d (sp^2, 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	27 891.21

It is useful to examine the physical significance of other parameters in the proposed model. Interestingly, the solvation radii we obtained (Table 3) are very close in value to the Lennard−Jones radii of the generic Dreiding force field,[31] which is used in atomistic molecular dynamics simulations. Furthermore, as shown in Figure 7, there is a strong correlation (R^2 = 0.949) between the Dreiding force field Lennard−Jones potential well depth, D_0, and the dispersion coefficient divided by the volume of each atom, i.e., $3\epsilon_i/4\pi R_i^3$. The inverse of the slope is about 6.4 (=1/0.156) is a good approximation for the coordination number. Therefore, the parameters obtained in the model reported here are closely related to force field parameters that were derived from a very different basis. This is an intriguing property, as it suggests the theoretical soundness of the model may allow one to estimate the values of parameters in one model from those in the other.

We have also considered the case in which the dispersion coefficients are not only atom-type dependent but also dependent on bonding type; as we expected, this would lead to improved results. For example, instead of one coefficient for each element, three different types of carbon (sp^3 hybrid, sp^2 hybrid and sp hybrid), three different types of nitrogen (sp^3 hybrid, sp^2 hybrid and sp hybrid), four different types of oxygen (sp^3 hybrid bonding with H, sp^3 hybrid, sp^3 hybrid, sp^2 hybrid, and bonding with nitro) are used in the dispersion energy calculation. The total number of parameters therefore increases to 23 (4 universal + 6 R_{el} + 13 ϵ_i, was 16). The optimized values of these parameters are listed in Table 4. The average accuracy of vapor pressures at normal boiling temperatures for the same 371 pure substances is found to improve somewhat to 60% (from 76%)

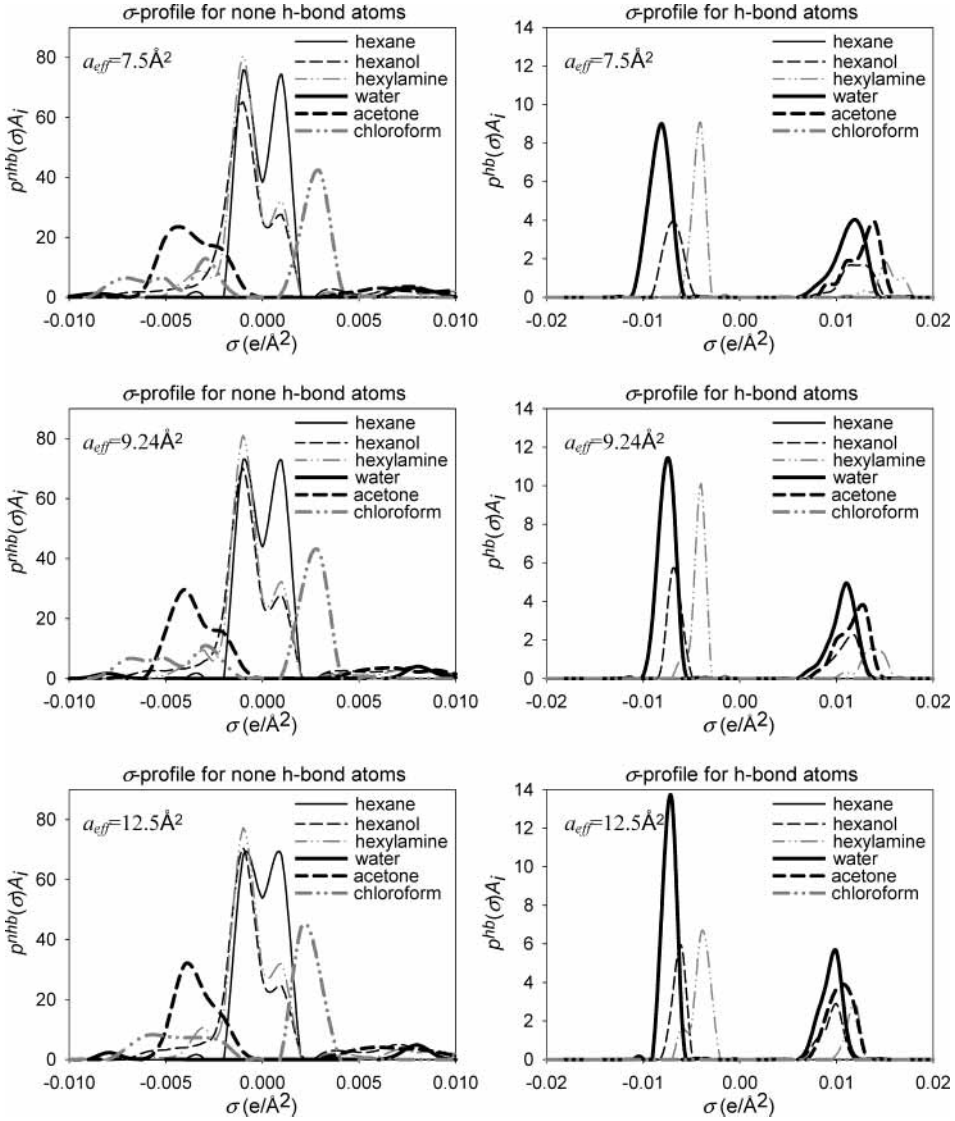


Figure 5. The σ profiles for six selected compounds at different averaging radius.

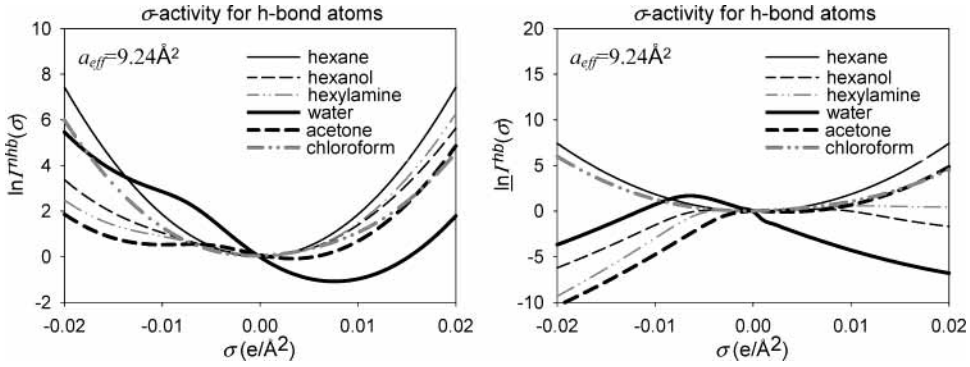


Figure 6. The σ activity for six selected compounds.

and the heat of vaporization to 4.43 kJ/mol (from 4.81). Error analysis for different compound families is given in Table 5 and Figure 2c and d. It is found that the use of bonding-type dependent dispersion coefficient results in better accuracy in the prediction of P^{vap} and also more uniform (consistent) error across different compound families. However, there is little improvement found for ΔH^{vap}. Within temperature range of $T_b - 50$ K and $T_b + 50$ K, the results are improved by nearly 15% in P^{vap} but only 0.42 kJ/mol in ΔH^{vap} (Figure 3b). We have not yet decided whether the magnitude of this improvement justifies the inclusion of the additional parameters in the model.

Conclusion

A new solvation model for pure substances has been developed by incorporating thermodynamic perturbation theory for the cavity free energy and using a surface-area-dependent mean field term for the dispersion interactions. The electrostatic contribution is treated in a manner similar to the previous

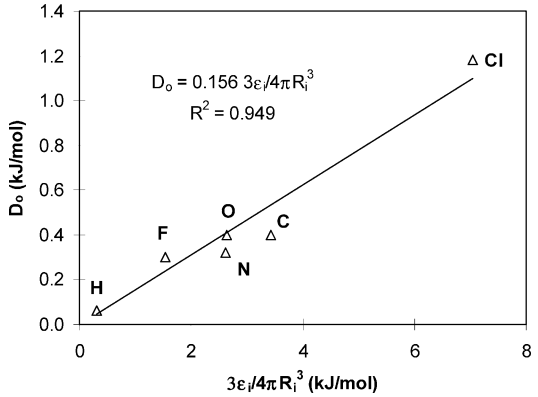


Figure 7. Comparison of the Lennard−Jones potential well, D_0, to the dispersion coefficient.

TABLE 5: Accuracy of the Proposed Model on the Prediction of Vapor Pressure and Heat of Vaporization at the Normal Boiling Point. (13 Dispersion Coefficient Models, Numbers in Parentheses Indicate the Maximum Deviation)

compound type	no. of compounds	AD% in P^{vap}	RMSD (kJ/mol) in ΔH^{vap}
alkane	37	57 (169)	1.76 (4.89)
alcohol	31	58 (90)	7.52 (12.25)
aldehyde	8	47 (52)	3.76 (5.18)
ketone	21	53 (89)	3.64 (8.07)
acid	11	47 (51)	7.77 (11.24)
ester	21	50 (65)	2.51 (4.47)
ether	16	60 (113)	2.51 (6.48)
nitro	5	47 (62)	4.81 (5.81)
nitrile	7	60 (98)	5.77 (6.81)
amine	22	66 (213)	2.76 (5.85)
aromatic	24	51 (66)	3.09 (6.65)
alkene	9	48 (55)	0.84 (1.63)
alkyne	7	48 (61)	1.96 (2.59)
F containing	5	64 (103)	2.26 (3.85)
Cl containing	24	61 (223)	2.97 (7.27)
multifunctional	123	65 (272)	5.02 (16.05)
overall	371	60 (328)	4.43 (16.05)

COSMO-RS and COSMO-SAC models. However, the present solvation model has been parametrized so that it correlates the vapor pressures and the enthalpies of vaporization for 371 organic compounds that contain H, C, N, O, F, and Cl atoms, with average errors of 76% in the vapor pressures and 4.81 kJ/mol in the heats of vaporization. The model accounts for the large temperature variation in the normal boiling points of the components studied while using temperature-independent parameters. The good performance of the model in correlating experimental data is due to its theoretical basis, and further evidence is provided by the fact that the model parameters obtained are closely related to the force field parameters used in atomistic molecular simulations. The only experimental data required to use the model are liquid molar volumes; however, this is not a significant limitation since there are good correlations for this property if experimental data are not available or the liquid density can be estimated from atomistic MD simulations. Therefore, the proposed model is quite general and can be applied to all chemical species composed of the atoms we have considered.

Acknowledgment. Financial support of this research was provided from contract DE-FG02-85ER13436 from the division of Basic Energy Sciences of the U.S. Department of Energy and Grant CTS-0083709 from the U.S. National Science Foundation.

Appendix I: Implementation of COSMO Solvation in Jaguar

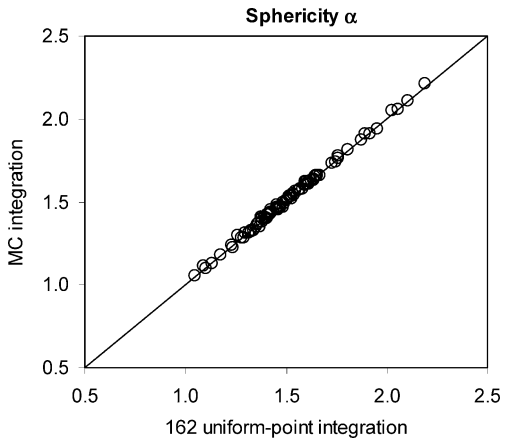
The COSMO calculation in Jaguar is invoked by setting isolv = 3 in the **gen** section. The keywords specifically related to solvation calculations are the probe radius (radprb), the atomic radius (vdw2 in the **atomic** section), the density of points used to represent the solvation cavity (cosfden, in unit of points/Å^2), and the skin thickness (cskin, in unit of Å) for correction of the escaping charge. The last two keywords are specific to the COSMO solvation calculation only.

The COSMO implementation in Jaguar is largely based on the work of Klamt and Schuurmann,[21] and the escaping charge correction is based on the work of Klamt and Jonas.[28] One difference is the way of constructing the molecular surface (or, equivalently, the solvation cavity), which is the trace of contact points as the probe rolls on the surface of the solute.[27] The Connolly surface obtained provides a smooth transition at the junction of two (or more) atoms and, thus, reduces the numerical instability due the cusps between the two atoms. Furthermore, not only the position but also the shape (concave, convex, or saddle), area, outward normal of each surface point, and the atom with which each point is associated are determined. These allow for the calculation of the area and volume of the solute molecule. They also provide an easy way to construct an outer shell for making the escaping charge corrections. The locations of points at the outer shell are simply extended along the surface normal to the user specified distance (sckin). The areas of the segments at the outer shell are scaled to the square of the distances to the atomic center (for convex segments) or to the probe center (for concave and saddle segments).

It should be noted that the area associated with each surface point is important for an accurate evaluation of screening charges at the surface. For this purpose, the actual area of each segment is re-determined from a separate calculation of the Connolly surface using a surface point density 50 times higher than the user request value (cosfden). The actual area of a surface point is then collected from the area of the finer segments located closest to that point. The time needed to construct the segments and evaluate their area is less than 10 s even for the largest molecule (1-octadodecanol) studied in this work.

Appendix II: Fast Determination of the Mean Radius of Curvature

The mean radius of curvature, R_h, is defined as the average distance from the center of mass of a molecule to the tangential plane at each solid angle. One way of obtaining the average is to perform MC integration over the solid angle. We have found that this value quickly converges with only 162 uniformly distributed points over the solid angle. The following figure



shows the excellent agreement between the calculated sphericity factor ($\alpha = R_h S_h / 3V_h$) for 91 compounds using MC and uniform point integration. While MC integration usually takes minutes, the calculation of radius of curvature using uniformly distributed points takes less than a second.

Appendix III: Analytical Expressions for the Temperature Derivatives

Here we present the analytical expressions for the temperature derivative of the solvation free energy needed for the heat of vaporization calculation.

Ideal solvation free energy:

$$\frac{d}{dT}\left(\frac{\Delta \underline{G}_i^{*\mathrm{is}}}{RT}\right) = -\frac{\Delta \underline{G}_i^{*\mathrm{is}}}{RT^2}$$

Free energy shift due to charge averaging:

$$\frac{d}{dT}\left(\frac{\Delta \underline{G}_i^{*\mathrm{cc}}}{RT}\right) = -\frac{\Delta \underline{G}_i^{*\mathrm{cc}}}{RT^2}$$

Restoring free energy:

$$\frac{d}{dT}\left(\frac{\Delta \underline{G}_{i/i}^{*\mathrm{res}}}{RT}\right) = n \sum_{\sigma_m} p(\sigma_m) \frac{d \ln \Gamma_i(\sigma_m)}{dT}$$

where

$$\frac{d \ln \Gamma^t(\sigma_m^t)}{dT} = -\sum_s^{\mathrm{nhb,hb}} \sum_{\sigma_n} \left\{ \left[\frac{d \ln \Gamma^s(\sigma_n^s)}{dT} + \frac{\Delta W(\sigma_m^t, \sigma_n^s)}{kT^2} - \frac{1}{kT} \frac{d \Delta W(\sigma_m^t, \sigma_n^s)}{dT} \right] \times p^s(\sigma_n^s) \exp\left[\frac{-\Delta W(\sigma_m^t, \sigma_n^s)}{kT} + \ln \Gamma^s(\sigma_n^s) + \ln \Gamma^t(\sigma_m^t) \right] \right\}$$

Dispersion term:

$$\frac{d}{dT}\left(\frac{\Delta \underline{A}_{i/i}^{*\mathrm{disp}}}{RT}\right) = -\frac{\sum_j \sum_k \epsilon_{jk} m_j^i m_k^i}{RT^2 \underline{V}_{i/L}} - \frac{\sum_j \sum_k \epsilon_{jk} m_j^i m_k^i}{RT \underline{V}_{i/L}^2} \frac{d \underline{V}_{i/L}}{dT}$$

Cavity term:

$$\frac{d}{dT}\left(\frac{\Delta \underline{A}_{i/i}^{*\mathrm{cav}}}{RT}\right) = -\left[(2\alpha - 1) \frac{(4 - 2\eta)}{(1 - \eta)^3} - (2\alpha - 2) \frac{(5 - 2\eta)}{(1 - \eta)(2 - \eta)} \right] \eta \frac{d \ln V_{i/L}}{dT}$$

Supporting Information Available: Detailed listing of the prediction of vapor pressure (P^{vap}) at the normal boiling point (T_b) for all 371 compounds considered in this study. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Reid, R. C.; Prausnitz, J. M.; Poling, B. E. *The Properties of Gases & Liquids*, 4th ed.; McGraw-Hill: New York, 1987.
(2) Macknick, A. B.; Prausnitz, J. M. *Ind. Eng. Chem. Fundam.* **1979**, *18*, 348.
(3) Edwards, D. R.; Prausnitz, J. M. *Ind. Eng. Chem. Fundam.* **1981**, *20*, 280.
(4) Liang, C. K.; Gallagher, D. A. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 321.
(5) Katritzky, A. R.; Wang, Y. L.; Sild, S.; Tamm, T.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 720.
(6) Basak, S. C.; Mills, D. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 692.
(7) Basak, S. C.; Gute, B. D.; Grunwald, G. D. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 651.
(8) Winget, P.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2000**, *104*, 4726.
(9) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224.
(10) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074.
(11) Lin, S. T.; Sandler, S. I. *Ind. Eng. Chem. Res.* **2002**, *41*, 899.
(12) Wertheim, M. S. *J. Stat. Phys.* **1984**, *35*, 19.
(13) Wertheim, M. S. *J. Chem. Phys.* **1987**, *87*, 7323.
(14) Boublik, T.; Vega, C.; Diazpena, M. *J. Chem. Phys.* **1990**, *93*, 730.
(15) Walsh, J. M.; Gubbins, K. E. *J. Phys. Chem.* **1990**, *94*, 5115.
(16) Ben-Amotz, D.; Omelyan, I. P. *J. Chem. Phys.* **2000**, *113*, 4349.
(17) Eckert, F.; Klamt, A. *AIChE J.* **2002**, *48*, 369.
(18) Eckert, F.; Klamt, A. *Ind. Eng. Chem. Res.* **2001**, *40*, 2371.
(19) Ben-Naim, A. *Solvation Thermodynamics*; Plenum Press: New York, 1987.
(20) Baldridge, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622.
(21) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
(22) Alejandre, J.; Martinezcasas, S. E.; Chapela, G. A. *Mol. Phys.* **1988**, *65*, 1185.
(23) Sandler, S. I. *Chemical and Engineering Thermodynamics*, 3rd ed.; John Wiley & Sons: New York, 1999.
(24) Jaguar, v. 4.2; Schrodinger, Inc.,: Portanld, OR, 1991−2000.
(25) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(26) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(27) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548.
(28) Klamt, A.; Jonas, V. *J. Chem. Phys.* **1996**, *105*, 9972.
(29) *BYU DIPPR801 Thermodynamic Properties Database*, 2002.
(30) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes—The Art of Scientific Computing*; Cambridge: New York, 1986.
(31) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897.