I&EC research
Industrial & Engineering Chemistry Research

Article

pubs.acs.org/IECR

COSMO-SAC Sigma Profile Generation with Conceptual Segment Concept

Md Rashedul Islam and Chau-Chyun Chen*

Department of Chemical Engineering, Texas Tech University, Lubbock, Texas 79409-3121, United States

ABSTRACT: Solvation thermodynamics-based models for predicting liquid-phase nonidealities and fluid-phase equilibria are gaining attention in modern chemical process and product development. Among this class of thermodynamic models, COSMO-RS and COSMO-SAC are two variants used extensively in industry. A key input to these models is the so-called sigma profile, i.e., a histogram of charge density distribution over the molecular surface. Typically, sigma profiles are generated from quantum mechanical calculations with molecular structure and conformation information as inputs to the calculation. We present an alternative approach for generating sigma profiles from experimental fluid-phase equilibrium data, i.e., solubility. Specifically, we incorporate the conceptual segment concept of NRTL-SAC activity coefficient model into sigma profile generation. We generate "apparent" sigma profiles from linear combination of sigma profiles of conceptual segments represented by reference molecules selected for hydrophobic, polar attractive, polar repulsive, and hydrophilic conceptual segments. Conceptual segment numbers of the molecule of interest are identified from regression of available experimental data. We show applicability of this sigma profile generation approach with solubility modeling for four drug molecules: caffeine, aspirin, paracetamol, and lovastatin.

INTRODUCTION

A priori prediction of liquid-phase nonidealities and fluid-phase equilibria has played a key role in modern chemical process and product development. A number of such predictive thermodynamic models have been widely used with either qualitative or semiquantitative accuracy. Examples include group contribution method, i.e., UNIFAC,[1,2] conceptual segment approach, i.e., NRTL-SAC,[3] and solvation thermodynamics approach, i.e., COSMO-RS[4] and COSMO-SAC.[5]

The group contribution method is one of the earliest of the prediction models. Among the group contribution methods, UNIFAC is the most accurate and widely used.[1] UNIFAC defines chemical compounds and their mixtures in terms of tens of predefined chemical functional groups. Binary interaction parameters which account for intermolecular interactions between functional groups are first optimized from millions of available experimental phase equilibrium data for thousands of molecules structured with the predefined functional groups. They are then employed to predict liquid-phase nonidealities, i.e., activity coefficients, of molecules in mixtures with the predefined functional groups. UNIFAC fails for molecules with functional groups not included in the predefined UNIFAC functional group database, and it is unable to distinguish between isomers as the same set of functional groups is present. Additionally, UNIFAC yields poor predictions for molecules with complex rigid molecular structure as the functional group additivity rule is applicable only to linear molecules.

In contrast, NRTL-SAC[3] defines four conceptual segments each uniquely representing molecular fragments exhibiting hydrophobic, polar attractive, polar repulsive, and hydrophilic nature in molecular interactions. Like UNIFAC, binary interaction parameters for the four conceptual segments are identified from available experimental data of selected reference molecules that exhibit hydrophobicity, polarity, and hydrophilicity. Conceptual segment numbers of the concerned molecules, similar to numbers and types of functional groups in UNIFAC, are the NRTL-SAC model parameters, and they are determined from experimental data of the molecule in the presence of reference solvents. Because the conceptual segment numbers are pure component parameters, NRTL-SAC can then be used to predict phase behavior of the molecule in other solvents and solvent mixtures as long as conceptual segment numbers are known for the solvents.

Solvation thermodynamics-based models have received increased attention in recent years. Among the solvation thermodynamics-based models, conductor-like screening models (COSMO) are the most widely used. There are two different variants of COSMO, i.e., COSMO-RS[4] and COSMO-SAC.[5] Unlike UNIFAC and NRTL-SAC, this method determines the interaction between molecules based on a so-called sigma profile, i.e., a histogram of charge density distribution over the molecular surface based on molecular structure and quantum mechanical calculations. Used together with a statistical thermodynamic expression, the resultant charge density distributions are used to compute chemical potentials of molecules in solution. The solvation thermodynamic models are advantageous over UNIFAC and NRTL-SAC when no experimental data are available. However, the COSMO models require knowledge of molecular structure and conformation to generate sigma profiles from quantum mechanical calculations, and the prediction quality of the COSMO models is qualitative in nature and often considered less reliable than that of UNIFAC and NRTL-SAC. In practice, there is a need to find a way to use the COSMO models without knowledge of molecular structure.[6] Also, empirical

Special Issue: Scott Fogler Festschrift

Received: July 20, 2014
Revised: November 21, 2014
Accepted: November 25, 2014

dx.doi.org/10.1021/ie503829b | *Ind. Eng. Chem. Res.* XXXX, XXX, XXX−XXX

treatments are proposed to correct the difference between the model predictions and the experimental data.[7]

In this work we incorporate the conceptual segment concept of NRTL-SAC into the sigma profile generation of solvation thermodynamic models. Specifically, we select four reference molecules to represent hydrophobic, polar attractive, polar repulsive, and hydrophilic conceptual segments. On the basis of conceptual segment numbers of the molecules we generate sigma profiles of the concerned molecules from linear combination of sigma profiles of the four reference molecules. In practice, conceptual segment numbers of the molecules are identified from fitting available phase equilibrium data involving the molecule and the four reference molecules or their equivalents. This approach allows generation of sigma profiles without knowledge of molecular structure and without use of quantum mechanical calculations. Furthermore, we achieve improved prediction quality with the solvation thermodynamic models because the sigma profiles are fitted against available data. We demonstrate this approach to generate sigma profiles and improved prediction results on solubility in pure solvents and solvent mixtures for four drug molecules: caffeine, aspirin, paracetamol, and lovastatin.

SOLUBILITY THERMODYNAMICS

Solubility of a solid crystal in a solution is dictated by solid−liquid equilibria (SLE). At equilibrium conditions, for a given solute, I, the solid-phase fugacity, f_I^{s}, and liquid-phase fugacity, f_I^{l}, are equal

$$f_I^{\mathrm{s}} = f_I^{\mathrm{l}} \tag{1}$$

At a given temperature, T, and pressure, P, liquid-phase fugacity is the product of saturation concentration, x_I^{sat}, activity coefficient, γ_I^{sat}, and reference state liquid-phase fugacity, $f_I^{0\mathrm{l}}$.

$$f_I^{\mathrm{s}} = x_I^{\mathrm{sat}}\gamma_I^{\mathrm{sat}}f_I^{0\mathrm{l}} \tag{2}$$

The ratio of solid-phase fugacity and reference state liquid-phase fugacity can be approximated as a function of enthalpy of fusion, ΔH_{fus}, and melting temperature, T_{m}, of the solute[8]

$$\ln\frac{f_I^{\mathrm{s}}}{f_I^{0\mathrm{l}}} = \frac{\Delta H_{\mathrm{fus}}}{R}\left(\frac{1}{T_{\mathrm{m}}} - \frac{1}{T}\right) \tag{3}$$

From eqs 2 and 3, solubility of the solute molecule can be expressed as

$$\ln x_I^{\mathrm{sat}} = \frac{\Delta H_{\mathrm{fus}}}{R}\left(\frac{1}{T_{\mathrm{m}}} - \frac{1}{T}\right) - \ln\gamma_I^{\mathrm{sat}} \tag{4}$$

For a given solute with a particular polymorph, enthalpy of fusion, ΔH_{fus}, and melting temperature, T_{m}, are fixed. Equation 4 indicates that at a specific temperature, T, solubility, x_I^{sat}, varies only with activity coefficient, γ_I^{sat}. Equation 4 can be expressed with solubility product constant, K_{sp}.[3] (See eq 5.)

$$\ln x_I^{\mathrm{sat}} = \ln K_{\mathrm{sp}} - \ln\gamma_I^{\mathrm{sat}} \tag{5}$$

From analogy to eq 4, the logarithm of solubility product constant can be expressed as a function of tempearture in eq 6. For a specific polymorph, $A = \Delta H_{\mathrm{fus}}/RT_{\mathrm{m}}$ and $B = -\Delta H_{\mathrm{fus}}/R$.

$$\ln K_{\mathrm{sp}} = A + \frac{B}{T} \tag{6}$$

Solubility modeling requires accurate calculation of solute activity coefficients. As mentioned earlier, a number of activity coefficient models, i.e., UNIFAC, NRTL-SAC, and COSMO-SAC, have been successfully investigated for their use in solubility modeling. An interesting fact for these three models is that they share similar theoretical formulation, as shown in eq 7. Activity coefficients are calculated from a residual term, γ_I^{R}, and a combinatorial term, γ_I^{C}. Different models follow different approaches and assumptions to derive the residual terms and the combinatorial terms. NRTL-SAC and COSMO-SAC are to be presented briefly in the following sections.

$$\ln\gamma_I = \ln\gamma_I^{\mathrm{R}} + \ln\gamma_I^{\mathrm{C}} \tag{7}$$

NRTL-SAC ACTIVITY COEFFICIENT MODEL

The NRTL-SAC model[3] originates from the segment-based concept of the polymer non-random two liquid (NRTL) activity coefficient model.[9,10] To capture the "like dissolves like" phenomenon, Chen and Song represented molecules with four conceptual segments that are selected to reflect major molecular surface characteristics of intermolecular interactions: hydrophobic (X), polar attractive (Y^-), polar repulsive (Y^+), and hydrophilic (Z). Hydrophilic segments act like hydrogen-bond donor, i.e., proton, or acceptor, i.e., lone-pair electrons, whereas hydrophobic segments show strong aversion to hydrogen-bond forming. Polar attractive and polar repulsive segments behave like electron pair donors or acceptors. Both polar segments show weak repulsion with hydrophobic segments; polar attractive segments show certain affinity with hydrophilic segments; and polar repulsive segments show weak aversion with hydrophilic segments. Effective surface interaction characteristics of a molecule are then represented by numbers of conceptual segments of respective nature. The residual activity coefficient is expressed in eq 8 where $r_{m,I}$ is the number of conceptual segment species m contained in component I. Γ_m^{lc} is segment activity coefficient of conceptual segment species m in solution and $\Gamma_{m,I}^{\mathrm{lc}}$ is segment activity coefficient of conceptual segment species m in component I.

$$\ln\gamma_I^{\mathrm{R}} = \ln\gamma_I^{\mathrm{lc}} = \sum_m r_{m,I}[\ln\Gamma_m^{\mathrm{lc}} - \ln\Gamma_{m,I}^{\mathrm{lc}}] \tag{8}$$

where

$$m \in \{X, Y^-, Y^+, Z\}$$

For the combinatorial activity coefficient, γ_I^{C}, the Flory−Huggins equation is adopted (eq 9). Here r_I and φ_I are the total segment number and segment mole fraction of component I, respectively

$$\ln\gamma_I^{\mathrm{C}} = \ln\frac{\varphi_I}{x_I} + 1 - r_I\sum_J\frac{\varphi_J}{r_J} \tag{9}$$

$$r_I = \sum_m r_{m,I} \tag{9a}$$

$$\varphi_I = \frac{r_I x_I}{\sum_J r_J x_J} \tag{9b}$$

A detailed derivation of the NRTL-SAC model is available in the literature.[3] Conceptual segment numbers for common solvent molecules and many drug molecules have been reported through regression of appropriate experimental

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

vapor−liquid equilibrium, liquid−liquid equilibrium, and solid−liquid equilibrium data.[3,11]

COSMO-SAC ACTIVITY COEFFICIENT MODEL

Thermodynamic models based on conductor like screening models are derived from solvation thermodynamics. COSMO-RS[4] and COSMO-SAC[5] are the two main variants. According to solvation thermodynamics, activity coefficient of a solute I is related to the solvation free energy $\Delta G^{*sol}_{I/S}$.[12] Solvation free energy is calculated from the change in energy when a solute molecule I is brought from a fixed position in an ideal gas to a fixed position in a solution S at constant temperature, T, and pressure, P. The solvation process can be described in two steps. First a discharged solute particle is inserted into a cavity of solvent. Energy required for this step is termed cavity formation free energy, ΔG^{*cav}. In the following step, charges are turned on to restore electronic configuration of the solute particle. Energy required for this process is called charging free energy, ΔG^{*chg}.

Cavity formation free energy depends on the shape and size of the solute molecule. Lin and Sandler proposed that ΔG^{*cav} is related to the combinatorial term of eq 7.[13,14] The Staverman−Guggenheim expression is proposed for the combinatorial term,[15,16] i.e., $\gamma^{C}_{I} = \gamma^{SG}_{I/S}$

$$\ln \gamma^{SG}_{I/S} = \ln \frac{\varphi_I}{x_I} + \frac{Z}{2} q_I \ln \frac{\theta_I}{\varphi_I} + l_I - \frac{\varphi_I}{x_I} \sum_J x_J l_J \tag{10}$$

$$l_I = \frac{Z}{2}(r_I - q_I) - (r_I - 1) \tag{10a}$$

$$\theta_I = \frac{x_I q_I}{\sum_J x_J q_J} \tag{10b}$$

$$\varphi_I = \frac{x_I r_I}{\sum_J x_J r_J} \tag{10c}$$

Here, φ_I is the normalized volume fraction, θ_I the normalized surface area fraction, Z the co-ordination number, and x_I the mole fraction of solute I. r_I and q_I are reported as normalized volume and surface area parameters respectively, i.e., $r_I = V_I/r$ and $q_I = A_I/q$. Here, r is the standard volume parameter (66.69 Å^3) and q is the standard surface area parameter (79.53 Å^2).[5]

Lin and Sandler further decomposed the charging free energy into two steps: ideal solvation and restoration of real fluid state. Therefore, ideal solvation free energy, ΔG^{*is}, and restoring free energy, ΔG^{*res}, constitute the charging free energy, ΔG^{*chg}, i.e., $\Delta G^{*chg} = \Delta G^{*is} + \Delta G^{*res}$. Note that ideal solvation free energy is the same for both the solution and pure liquid, i.e., $\Delta G^{*is}_{I/S} = \Delta G^{*is}_{I/I}$. The residual term in activity coefficient can be expressed by eq 11

$$\ln \gamma^{R}_{I/S} = \frac{\Delta G^{*res}_{I/S} - \Delta G^{*res}_{I/I}}{RT} \tag{11}$$

In COSMO-SAC, a molecule is divided into n_I number of segments having fixed surface area, a_{eff} (7.5 Å^2). For a molecule with surface area A_I, n_I will be A_I/a_{eff}. Each segment is characterized by its charge density, σ. If $n_I(\sigma)$ is the total number of segments in a molecule having charge density σ, the probability of finding those segments in pure liquid is

$$p_I(\sigma) = \frac{n_I(\sigma)}{n_I} = \frac{A_I(\sigma)}{A_I} \tag{12}$$

where $A_I(\sigma)$ is the total surface area in a molecule with charge density σ. The histogram of charge density distribution over the molecular surface is called a sigma profile. The sigma profile of a mixture is computed from the weighted average of sigma profiles of molecules in the mixture

$$p_S(\sigma) = \frac{\sum_I x_I n_I p_I(\sigma)}{\sum_I x_I n_I} = \frac{\sum_I x_I A_I p_I(\sigma)}{\sum_I x_I A_I} \tag{13}$$

The sigma profile plays the pivotal role in COSMO calculations. It conveys the electronic properties of the fluid. This histogram in some ways is analogous to the functional group numbers of UNIFAC and the conceptual segment numbers of NRTL-SAC. In COSMO calculations, each segment is considered as an individual entity or ensemble. Segment activity coefficient, $\Gamma(\sigma_m)$, of a pure component or mixture conveys the interaction of segment m with charge density σ_m with all other n segments. Lin and Sandler expressed the restoring free energy as

$$\frac{\Delta G^{*res}_{I/S}}{RT} = n_I \sum_{\sigma_m} p_I(\sigma_m) \ln \Gamma_S(\sigma_m) \tag{14}$$

The segment activity coefficients for pure component and mixture are expressed as

$$\ln \Gamma_I(\sigma_m) = -\ln\left\{ \sum_{\sigma_n} p_I(\sigma_n) \Gamma_I(\sigma_n) \exp\left[\frac{-\Delta W(\sigma_m, \sigma_n)}{RT} \right] \right\} \tag{15a}$$

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

where $\Delta W(\sigma_m, \sigma_n)$ is the exchange energy. This exchange energy is calculated from the following equation:

$$\Delta W(\sigma_m, \sigma_n) = \left(\frac{\alpha'}{2}\right)(\sigma_m + \sigma_n)^2 + c_{hb} \max[0, \sigma_{acc} - \sigma_{hb}] \min[0, \sigma_{don} + \sigma_{hb}] \tag{16}$$

where α' is the misfit energy (16 466 (kcal Å^4)/(mol e^2)), c_{hb} the hydrogen bonding constant (85580 (kcal Å^4)/(mol e^2)), and σ_{hb} the sigma cutoff for hydrogen bonding (0.0084 e/Å^2).[5] The larger and smaller values of σ_m and σ_n are defined as σ_{acc} and σ_{don}, respectively. The functions max and min return the largest and smallest values of their arguments, respectivly. The activity coefficient can be calculated from eq 17. Detailed mathematical derivation and explanation of COSMO-RS and COSMO-SAC are available in the literature.[4,5,17,18]

$$\ln \gamma_{I/S} = n_I \sum_{\sigma_m} p_I(\sigma_m)[\ln \Gamma_S(\sigma_m) - \ln \Gamma_I(\sigma_m)] + \ln \gamma^{SG}_{I/S} \tag{17}$$

SIGMA PROFILE

Sigma profile, $p(\sigma)$, is the probability distribution of surface area having charge density σ. It is observed that ideal screening charge density for most of the molecules are in the range of −0.025 to 0.025 e/Å^2. Therefore, the sigma profile is often reported as a histogram of segment surface over a charge density range of −0.025 to 0.025 e/Å^2. For convenience this

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

interval is further subdivided with the increment of 0.001 e/Å^2, resulting in a vector of 51 elements.[5]

Sigma profile generation requires use of quantum chemistry software packages. Fortunately, there exists an open-source web-based sigma profile database, VT-2005 sigma profile database, for solvents and small molecules (www.design.che.vt.edu).[19] VT-2005 database includes sigma profiles of 1432 common compounds. This database is further supplemented by the VT-2006 database[20] which includes additional 32 solvents and 206 primarily larger pharmaceutical compounds. The reported sigma profiles in the VT databases have been calculated based on density functional theory (DFT)[21] using DMol^3 module[22,23] of Accelrys Materials Studio software.[24]

There are other commercial and open-source quantum chemistry packages in addition to DMol^3. Example includes GAMESS,[25] Gaussian,[26] Jaguar,[27] MOPAC,[28] and TURBOMOLE.[29] GAMESS is an open-source quantum chemistry package. Wang et al. reported a comparison study on phase equilibrium calculations of 45 binary solvents using GAMESS.[30] Additinally, a comparison study of COSMO-RS and COSMO-SAC performance based on sigma profile generated by DMol^3, Gaussian, and TURBOMOLE is reported in the literature.[31] MOPAC uses semiempirical methods to reduce computing time. It is less precise than other packages.[32]

Figure 1 shows sigma profiles of four reference molecules chosen in this study: hexane, dimethyl sulfoxide (DMSO), nitromethane, and water.[19]

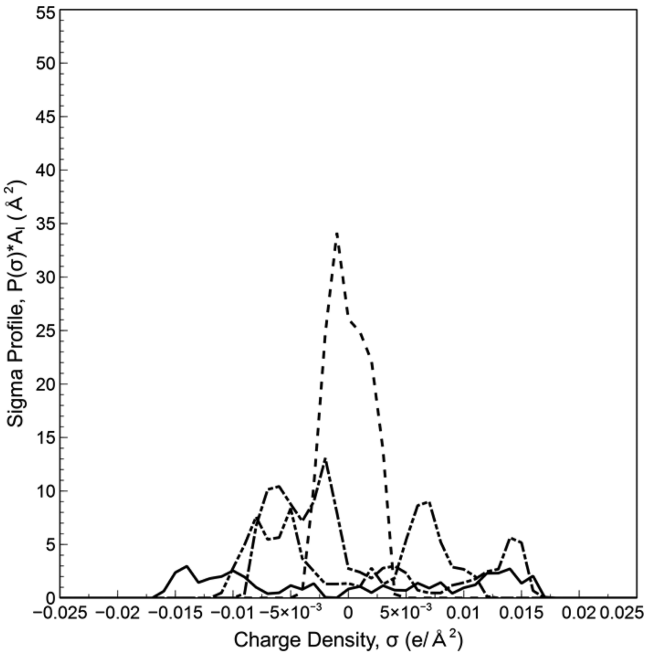


Figure 1. Sigma profiles for hexane (dashed line), DMSO (dash-dotted line), nitromethane (dash-dot-dotted line), water (solid line).

Hexane is a hydrophobic molecule. Its sigma profile is observed to be narrow and inside the sigma cutoff for hydrogen bonding, i.e., $-0.0084 \leq \sigma \leq 0.0084$. In other words, it does not form a hydrogen bond with other molecules.

Water is hydrophilic in nature. Unlike hexane, the sigma profile for water is wide and symmetric. The two polar hydrogen atoms produce a peak at $\sigma = -0.014$. On the other hand, lone-pair electrons on oxygen yield a peak at $\sigma = 0.014$. The region between these two peaks is rather flat and symmetric. Therefore, water demonstrates strong hydrogen bonding interactions with other molecules.

DMSO is a representative polar attractive solvent. It has a highly asymmetric sigma profile. The lone-pair electrons on the sulfur atom form a peak at $\sigma = 0.014$. These electron pair donor electrostatic segments are attractive to hydrophilic segments. In contrast, six hydrogen atoms carry counter charge. The charges on the hydrogen atoms are distributed over a large area on the negative side and form a peak at $\sigma = 0.006$. However, they do not act as an electron pair acceptor.

Nitromethane is a typical polar repulsive solvent. It has a sigma profile almost symmetric around zero charge density, i.e., $\sigma = 0$. The positve charge from the nitrogen atom yields a peak at around the sigma cutoff for hydrogen bonding of $\sigma = -0.0084$. To counterbalance this charge distribition on the negative side, surface charge segments from oxygen are distributed on the positive side with a peak at $\sigma = 0.007$.

PROPOSED SIGMA PROFILE GENERATION METHODOLOGY

We propose an alternative approach for sigma profile generation that builds on the simplicity of NRTL-SAC, the predictive power of COSMO-SAC, and the confidence in actual experimental measurements. No molecular structure information nor quantum mechnical calculations are required. Figure 2 shows the flowchart for the sigma profile generation methodology. Following the NRTL-SAC approach, the methodology generates a sigma profile of a molecule from linear combinations of sigma profiles of four reference molecules, each representing a particular conceptual segment (eq 18).

$$p_I(\sigma)A_I = A_{\text{ref}}\begin{bmatrix} X \\ Y^- \\ Y^+ \\ Z \end{bmatrix} \tag{18}$$

A_{ref} matrix, with dimension of 51 × 4, is generated from the sigma profile vectors of reference molecules. The conceptual segment vector, $[X\ Y^-\ Y^+\ Z]^T$, accounts for the respective contributions from the four conceptual segments. As a simplifying assumption for cavity volume, we consider a spherical cavity having a surface area of A_I which enshrouds the molecule. The radius of the cavity, r_{cav}, is $(A_I/4\pi)^{1/2}$. The cavity volume thus can be calculated as $4\pi r_{\text{cav}}{}^3/3$. For the four conceptual segments, the reference molecules are selected based on the demonstrated nature of hydrophobicity, polarity, and hydrophilicity. In this study, hexane, dimethyl sulfoxide, nitromethane, and water are chosen to represent hydrophobic, polar attractive, polar repulsive, and hydrophilic segments, respectively. Selection of reference molecules may be refined later as our learning evolves.

The resultant sigma profiles and the cavity volumes are used in the COSMO-SAC model (eq 17) to calculate activity coefficients of molecules in the system. The calculated activity coefficients are then passed to the solubility equation (eq 5). An objective function is formulated to minimize the error between the calculated and the experimental values in eq 19.

$$\underset{XY^-Y^+Z}{\text{minimize}}\ \frac{1}{n}\sum_j^n [\ln x_j^{\text{exp}} - \ln x_j^{\text{calc}}]^2$$
$$\text{subject to } p_I(\sigma) \geq 0 \tag{19}$$

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

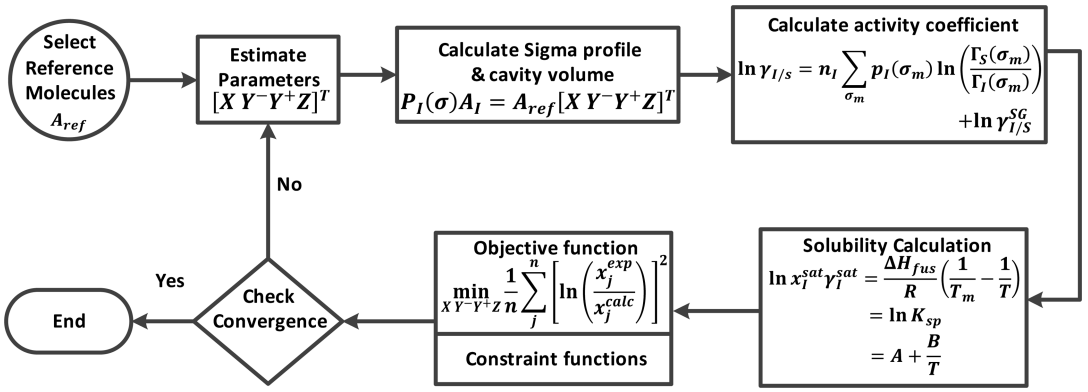
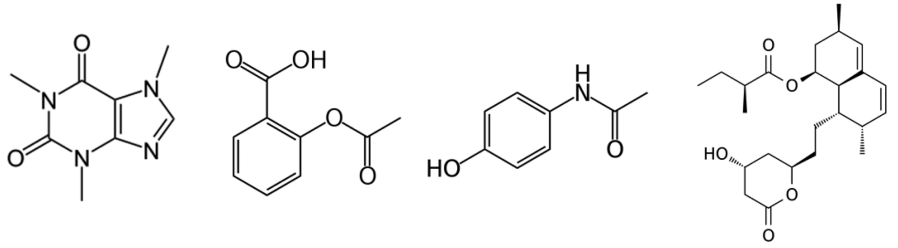


Figure 2. Sigma profile generation methodology.



(a) Caffeine (b) Aspirin (c) Paracetamol (d) Lovastatin

Figure 3. Molecular structure for drug molecules.

Table 1. Experimental and Calculated Solubility of Caffeine

			calculated solubility		
solvent	temp (K)	experimental solubility (mole frac)	VT-2006 sigma profile (mole frac)	apparent sigma profile (mole frac)	data ref
hexane[a]	298.15	3.94×10^{-6}	7.13×10^{-4}	4.10×10^{-6}	45
1,4-dioxane[a]	298.15	8.20×10^{-3}	3.04×10^{-2}	6.48×10^{-3}	45
DMF[a]	298.15	1.26×10^{-2}	4.36×10^{-2}	1.45×10^{-2}	45
water[a]	298.15	2.25×10^{-3}	1.33×10^{-3}	2.24×10^{-3}	45
2-ethoxyethanol	298.15	6.78×10^{-3}	2.04×10^{-2}	1.28×10^{-2}	46
1-octanol	303.15	2.45×10^{-3}	8.69×10^{-3}	1.09×10^{-3}	47
ethanol	298.00	1.70×10^{-3}	1.82×10^{-2}	5.82×10^{-3}	48
ethyl-acetate	298.00	4.09×10^{-3}	2.31×10^{-2}	2.48×10^{-3}	48

[a]Solvents used for conceptual segment number estimation.

where n is number of data points; x_j^{exp} and x_j^{calc} are the experimental and calculated solubility of the molecule in solvent j, respectively. The elements of the conceptual segment vector, i.e., the conceptual segment numbers, are treated as decision variables for the minimization problem.

In the solubility calcuation, enthalpy of fusion, ΔH_{fus}, and melting temperature, T_m, are the input parameters. In case ΔH_{fus} and T_m are not available or unreliable, the solubility product constant, K_{sp}, may be included as one of the decision variables.

For the convenience of referring to sigma profiles and cavity volumes generated from the proposed approach, they are termed "apparent sigma profile" and "apparent cavity volume", respectively.

RESULTS

The proposed sigma profile generation methodology is illustrated with solubility modeling for four drug molecules: caffeine, aspirin, paracetamol, and lovastatin. Figure 3 shows the molecular structure of these drug molecules. These drug molecules have been extensively studied in the literature with NRTL-SAC,[33] UNIFAC,[2] COSMO-SAC,[33] COSMO-RS,[2] PC-SAFT,[34] etc. The purpose of our study is not to show the comparison among these models. Rather, we aim to show the apparent sigma profile, fitted against limited experimental data, offers a practical approach for correlation and prediction with COSMO-SAC and does not distort the model behavior.

To allow for meaningful determination of conceptual segment numbers for the drug molecules, we include solvents that are hydrophobic, polar attractive, polar repulsive, and hydrophilic. From the drug solubility data in the selected solvents we identify the conceptual segment numbers, i.e., $[X\ Y^{-}\ Y^{+}\ Z]^T$, and generate apparent sigma profiles for the drug molecules. The apparent sigma profiles are then used with COSMO-SAC and solubility equations to predict drug solubility in other solvents and solvent mixtures. The calculated solubilities for these drug molecules are compared to their experimental values. Also included in the comparisons are the solubility predictions with drug molecule sigma profiles retrieved from the VT-2006 database.

Caffeine. For caffeine, the melting temperature, T_m, is 512.15 K and the enthalpy of fusion, ΔH_{fus}, is 21 600 kJ/

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

kmol.[35] These thermodynamic data result in $\ln K_{sp} = -3.64$ at 298.15 K. The experimental solubility data for caffeine in pure solvents are collected from the literature and reported in Table 1. Among the eight solvents, hexane, 1,4-dioxane, dimethylformamide, and water are selected to estimate the apparent sigma profile. These four solvents are selected based on their representative molecular surface interaction characteristics. We generate conceptual segment numbers and apparent sigma profile of caffeine through regression of experimental solubility data in these four solvents. From the apparent sigma profile we then compute the caffeine solubility in all eight solvents, and they are presented in Table 1 along with the solubility prediction results from the VT-2006 sigma profile. To show that the conceptual segment numbers are a good measure of the molecular surface characteristics, the conceptual segment numbers are also estimated through regression of experimental solubility data in all eight solvents. Reported in Table 2, the

Table 2. Conceptual Segment Numbers for Caffeine

	segment numbers	
conceptual segment	4 solvents	all solvents
X	0.109	0.017
Y^-	1.057	0.869
Y^+	1.255	1.440
Z	0	0

conceptual segment numbers for caffeine determined with both "4 solvents" and "all solvents" are found to be very similar. Both have significant polar attractive and polar repulsive segments, low hydrophobic segments, and zero hydrophilic segments. The results suggest the conceptual segment numbers and the apparent sigma profile are relatively independent of the number of solvents used in the solubility data regression as long as the solubility data cover solvents of different molecular surface nature.

Figure 4 shows the apparent sigma profile estimated with four solvents along with the VT-2006 sigma profile for caffeine. The apparent sigma profile differs significantly from the VT-2006 sigma profile. The VT-2006 sigma profile for caffeine, similar to that of hexane shown in Figure 1, shows level of hydrophobic segments significantly higher than that of the apparent sigma profile. The apparent cavity volume of the caffeine molecule calculated from the apparent sigma profile is 363.22 Å^3. In comparison, the VT-2006 database reported the cavity volume of the caffeine molecule as 219.23 Å^3. It should be noted that the apparent cavity volume is calculated with the simplifying assumption of a spherical cavity.

Table 3 reports the model errors with the two apparent sigma profiles identified with "4 solvents" and "all solvents" and the VT-2006 sigma profile in terms of root-mean-square error (RMSE) in logarithm of solubility, i.e., $[(1/N)\Sigma_i^N(\ln x_i^{Exp} - \ln x_i^{Calc})^2]^{1/2}$. Figure 5 shows the parity plot comparing the experimental and calculated solubilities. COSMO-SAC with the VT-2006 sigma profile overpredicts caffeine solubility in all solvents except water. This reflects the fact that the VT-2006 sigma profile suggests significant hydrophobic segments similar to that of hexane. However, this is not consistent with the experimental observation as the caffeine solubility data show very low solubility in hexane. In contrast, the apparent sigma profile suggests low hydrophobic segments, zero hydrophilic segments, and signficant polar attractive and polar repulsive segments (Table 2). COSMO-SAC with the apparent sigma profile calculates caffeine solubility in the eight solvents much closer to the experimental data than those predicted with the VT-2006 sigma profile.

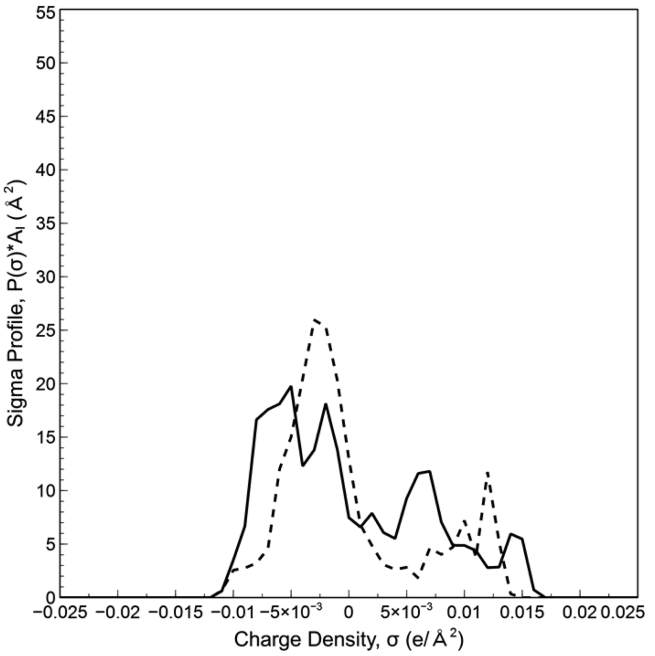


Figure 4. Sigma profile for caffeine: VT-2006 (dashed line), apparent sigma profile (solid line).

Table 3. RMSE in Logarithm of Solubility of Caffeine in Eight Pure Solvents

	apparent sigma profile	
VT-2006 sigma profile	4 solvents	all solvents
2.29	0.60	0.57

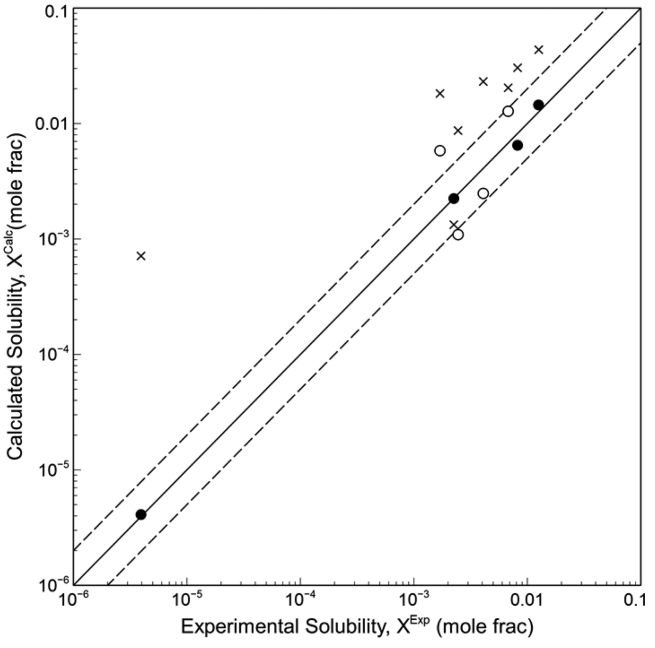


Figure 5. Comparison of experimental solubility of caffeine with model results: prediction with VT-2006 sigma profile (×), prediction with apparent sigma profile (○), data used to identify apparent sigma profile (●), ± 100% error band (dashed lines).

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

Aspirin. For aspirin, the melting temperature and the enthalpy of fusion are 408.15 K and 25 600 kJ/kmol, respectively.[35] ln K_{sp} is calculated as −2.78 at 298.15 K from these thermodynamic data.

Frank et al. reported experimental solubility of aspirin in 23 pure solvents at 298.15 K.[36] (See Table 4). To generate the apparent sigma profile for aspirin, we regress the experimental solubility data of aspirin in four solvents, i.e., methanol, acetone, acetic acid, and cyclohexane. Figure 6 shows the apparent sigma profile of aspirin together with the VT-2006 sigma profile. In contrast to the case of caffeine, the apparent sigma profile for aspirin contains hydrophobic segments significantly higher than that of the VT-2006 sigma profile. To examine the effects of other solvents, the apparent sigma profile is also estimated using 20 solvents with three chlorohydrocarbon solvents excluded. The reason for excluding the three chlorohydrocarbon solvents will be explained later. The conceptual segment numbers for aspirin estimated from two sets of solvents are presented in Table 5. Both show zero or low polar attractive segments and significant hydrophobic, polar repulsive, and hydrophilic segments. The apparent cavity volume of aspirin molecule is calculated to be 328.15 Å^3. In comparison, the VT-2006 database reported the cavity volume of aspirin molecule as 206.19 Å^3.

Solubility calculation results with the apparent sigma profile calculated with four solvents and the VT-2006 sigma profile are reported in Table 4. The VT-2006 sigma profile overpredicts aspirin solubility in all solvents except tetrachloroethylene. The apparent sigma profile yields very good results of aspirin solubility in alcohols, ketones, carboxylic acids, ethers, esters, acetals, and chloroform. However, it is found that the apparent sigma profile grossly under-predicts the aspirin solubility in the three chlorohydrocarbon solvents: tetrachloroethylene, 1,2-dichloroethane, and 1,1,1,-trichloroethane. It is not obvious to us why the model predicts aspirin solubility well for chloroform but rather poorly for the other three chlorohydrocarbon solvents. The poor predictions for the three chlorohydrocarbon solvents remain even if the solubility data in these three solvents are used to generate the conceptual segment numbers and the apparent sigma profile. The model errors in terms of root-mean-square error in logarithm of solubility for the two different apparent sigma profiles identified with "4 solvents" and "all solvents" and the VT-2006 sigma profile are reported in Table 6. Additionally a parity plot of experimental and calculated solubility is presented in Figure 7.

Table 4. Experimental and Calculated Solubility of Aspirin at 298.15 K

solvent	experimental solubility[b] (mole frac)	calculated solubility: VT-2006 sigma profile (mole frac)	calculated solubility: apparent sigma profile (mole frac)
methanol[a]	8.05×10^{-2}	1.88×10^{-1}	6.40×10^{-2}
acetone[a]	1.16×10^{-1}	3.02×10^{-1}	1.14×10^{-1}
acetic acid[a]	4.35×10^{-2}	5.69×10^{-2}	4.86×10^{-2}
cyclohexane[a]	2.34×10^{-5}	8.26×10^{-4}	2.31×10^{-5}
ethanol	6.01×10^{-2}	2.40×10^{-1}	7.92×10^{-2}
1,4-dioxane	1.03×10^{-1}	2.67×10^{-1}	8.40×10^{-2}
methyl ethyl ketone	5.18×10^{-2}	3.08×10^{-1}	1.17×10^{-1}
diacetone alcohol	6.69×10^{-2}	3.88×10^{-1}	1.68×10^{-1}
isopropanol	3.57×10^{-2}	2.53×10^{-1}	8.05×10^{-2}
isoamyl alcohol	5.16×10^{-2}	1.58×10^{-1}	4.37×10^{-2}
2-ethyl hexanol	7.43×10^{-2}	1.65×10^{-1}	4.40×10^{-2}
propylene glycol	4.01×10^{-2}	2.15×10^{-1}	4.64×10^{-2}
chloroform	4.06×10^{-2}	9.68×10^{-2}	6.16×10^{-2}
diethyl ether	2.12×10^{-2}	2.23×10^{-1}	5.05×10^{-2}
methyl benzoate	3.05×10^{-2}	9.56×10^{-2}	4.11×10^{-3}
ethyl butyrate	2.62×10^{-2}	1.59×10^{-1}	1.27×10^{-2}
diethyl maleate	3.83×10^{-2}	1.28×10^{-1}	9.13×10^{-3}
diethyl malonate	3.57×10^{-2}	2.71×10^{-1}	2.59×10^{-2}
acetal	2.66×10^{-2}	1.75×10^{-1}	2.43×10^{-2}
1-octanol	2.19×10^{-2}	1.78×10^{-1}	3.46×10^{-2}
tetrachloroethylene	2.77×10^{-2}	2.15×10^{-3}	5.73×10^{-5}
1,2-dichloroethane	1.67×10^{-2}	3.19×10^{-2}	5.32×10^{-4}
1,1,1-trichloroethane	3.71×10^{-3}	1.10×10^{-2}	3.20×10^{-4}

[a]Solvents used for conceptual segment number estimation. [b]Data from ref 36.

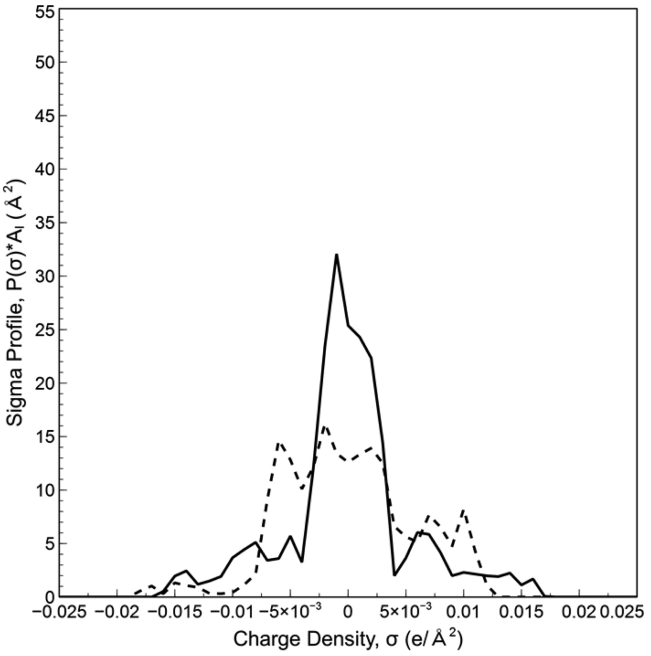


Figure 6. Sigma profile for aspirin: VT-2006 (dashed line), apparent sigma profile (solid line).

Table 5. Conceptual Segment Numbers for Aspirin

conceptual segment	segment number: 4 solvents	segment number: all solvents[a]
X	0.917	0.532
Y^-	0	0.032
Y^+	0.568	1.086
Z	0.823	0.559

[a]With tetrachloroethylene, 1,2-dichloroethane, and 1,1,1-trichloroethane excluded.

Paracetamol. For paracetamol, the melting temperature and the enthalpy of fusion are reported as 441.2 K and 26 000 kJ/kmol, respectively.[37] These reported thermodynamic constants result in ln K_{sp} = −3.40 at 298.15 K and ln K_{sp} = −3.23 at 303.15 K.

Granberg and Rasmuson reported the experimethal solubility of paracetamol in 26 pure solvents at 303.15 K.[38] We estimate the conceptual segment numbers for paracetamol using the experimental data of four solvents, i.e., water, acetone, dimethyl sulfoxide, and toluene. The apparent sigma profile calculated

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

Table 6. RMSE in Logarithm of Solubility of Aspirin in 23 Pure Solvents

	apparent sigma profile	
VT-2006 sigma profile	4 solvents	all solvents
1.68	1.70	1.31
1.68[a]	0.73[a]	0.57[a]

[a]With tetrachloroethylene, 1,2-dichloroethane, and 1,1,1-trichloroethane excluded.

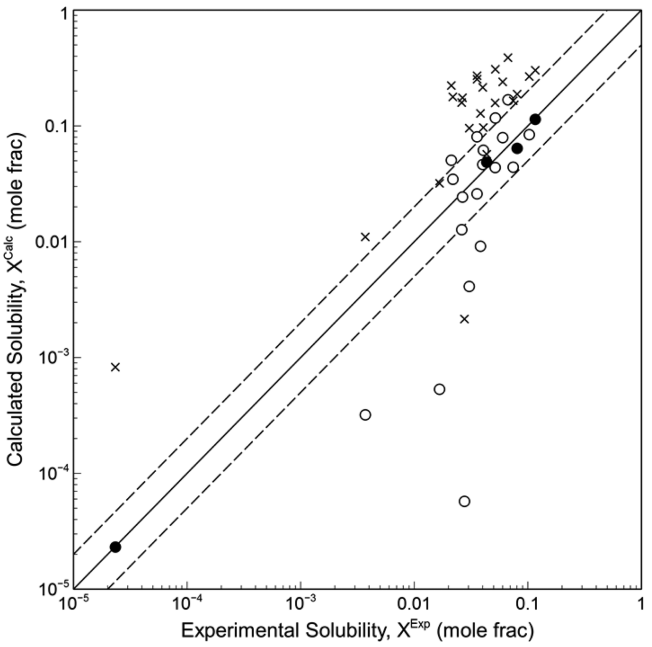


Figure 7. Comparison of experimental solubility of aspirin with model results: prediction with VT-2006 sigma profile (×), prediction with apparent sigma profile (○), data used to identify apparent sigma profile (●), ± 100% error band (dashed lines).

from the conceptual segment numbers is then used to estimate the solubility of paracetamol in all 26 solvents (Table 7). To examine the effects of other solvents, we further estimate the conceptual segment numbers using the entire set of 26 solvents. The conceptual segment numbers for paracetamol estimated from the two sets of solvents are reported in Table 8. The two sets of conceptual segment numbers are found to be consistent, and both suggest substantial hydrophobicity and high hydrophilicity for paracetamol.

The apparent sigma profile, estimated from four solvents, is shown in Figure 8 together with the VT-2006 sigma profile for paracetamol. The apparent cavity volume of paracetamol is calculated as 135.82 Å^3, whereas the VT-2006 cavity volume of paracetamol is 183.80 Å^3.

Table 7 shows the paracetamol solubility in the 26 solvents as calculated with the apparent sigma profile and the VT-2006 sigma profile. Figure 9 shows the parity plot for the experimental and calculated solubility. Table 9 summarizes the model errors for paracetamol for the two apparent sigma profiles and the VT-2006 sigma profile in terms of root-mean-square error in logarithm of solubility. The model error from the VT-2006 sigma profile is significantly larger than those of the apparent sigma profiles. The VT-2006 sigma profile overpredicts paracetamol solubility in all solvents except for diethylamine and carbon tetrachloride. For example, the predictions for alcohol solvents are around two to three times

Table 7. Experimental and Calculated Solubility of Paracetamol at 303.15 K

		calculated solubility	
solvent	experimental solubility[b] (mole frac)	VT-2006 sigma profile (mole frac)	apparent sigma profile (mole frac)
water[a]	2.07×10^{-3}	3.67×10^{-3}	2.10×10^{-3}
acetone[a]	4.11×10^{-2}	2.01×10^{-1}	5.16×10^{-2}
dimethyl sulfoxide[a]	3.69×10^{-1}	4.24×10^{-1}	2.28×10^{-1}
toluene[a]	2.07×10^{-4}	2.19×10^{-4}	1.97×10^{-4}
methanol	7.30×10^{-2}	1.36×10^{-1}	5.83×10^{-2}
ethanol	6.62×10^{-2}	1.60×10^{-1}	6.64×10^{-2}
ethylene glycol	5.59×10^{-2}	6.40×10^{-2}	2.80×10^{-2}
1-propanol	5.01×10^{-2}	1.39×10^{-1}	5.95×10^{-2}
2-propanol	5.09×10^{-2}	1.57×10^{-1}	6.67×10^{-2}
1-butanol	4.39×10^{-2}	1.28×10^{-1}	5.56×10^{-2}
1-pentanol	3.80×10^{-2}	1.08×10^{-1}	4.95×10^{-2}
1-hexanol	3.25×10^{-2}	9.89×10^{-2}	4.66×10^{-2}
1-heptanol	2.80×10^{-2}	8.48×10^{-2}	4.20×10^{-2}
1-octanol	2.31×10^{-2}	7.86×10^{-2}	4.02×10^{-2}
methyl ethyl ketone	3.23×10^{-2}	1.95×10^{-1}	5.53×10^{-2}
methyl isobutyl ketone	1.17×10^{-2}	1.18×10^{-1}	2.97×10^{-2}
tetrahydrofuran	6.90×10^{-2}	2.27×10^{-1}	1.09×10^{-1}
1,4-dioxane	9.86×10^{-3}	1.57×10^{-1}	3.20×10^{-2}
ethyl acetate	6.21×10^{-3}	8.56×10^{-2}	1.33×10^{-2}
acetonitrile	8.84×10^{-3}	6.40×10^{-2}	6.09×10^{-3}
diethylamine	3.89×10^{-1}	2.62×10^{-1}	2.04×10^{-1}
N,*N*-dimethylformamide	3.29×10^{-1}	3.60×10^{-1}	1.84×10^{-1}
acetic acid	3.18×10^{-2}	6.23×10^{-2}	6.85×10^{-2}
dichloromethane	1.80×10^{-4}	4.90×10^{-3}	2.22×10^{-3}
chloroform	1.21×10^{-3}	1.97×10^{-2}	2.15×10^{-2}
carbon tetrachloride	9.05×10^{-4}	2.66×10^{-5}	6.04×10^{-5}

[a]Solvents used for conceptual segment number estimation. [b]Data from ref 38.

Table 8. Conceptual Segment Numbers for Paracetamol

	segment number	
conceptual segment	4 solvents	all solvents
X	0.488	0.338
Y^-	0	0
Y^+	0.153	0
Z	0.868	0.820

higher than the experimental data. The overpredictions for paracetamol solubility in ketone and ester solvents are even more pronounced than those for the alcohol solvents. Furthermore, the VT-2006 sigma profile gives very poor predictions for the three chlorohydrocarbon solvents. It overpredicts the paracetamol solubility in dichloromethane and chloroform by 27 and 16 fold, respectively, and underpredicts the solubililty in carbon tetrachloride by a factor of 34.

Paracetamol solubility model results with the apparent sigma profiles are fairly close to the experimental data for alcohol, ketone, amide, amine, carboxylic acid, and ester solvents. However, like the VT-2006 sigma profile, the apparent sigma profile performs poorly with the paracetamol solubility in the chlorohydrocarbon solvents. Although there is a slight improvement over the VT-2006 sigma profile predictions, the apparent sigma profile overestimates the paracetamol solubility in dichloromethane and chloroform by a factor of 12 and 18,

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

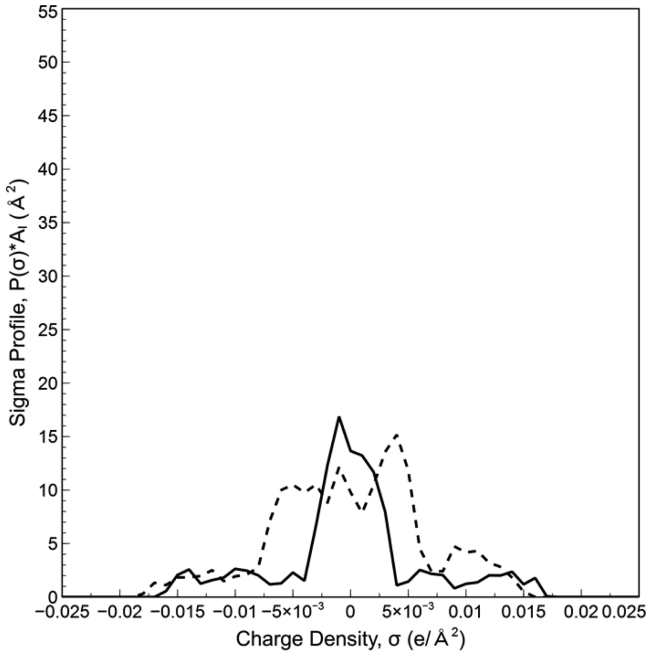


Figure 8. Sigma profile for paracetamol: VT-2006 (dasehd line), apparent sigma profile (solid line).

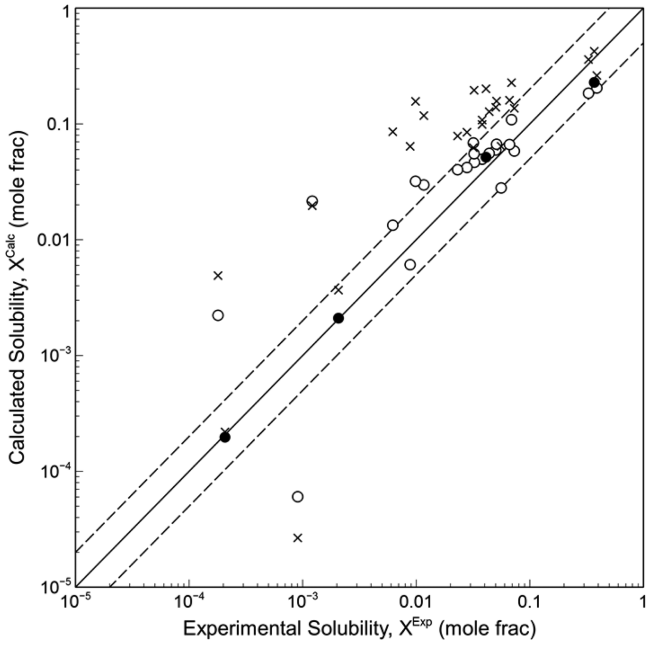


Figure 9. Comparison of experimental solubility of paracetamol with model results: prediction with VT-2006 sigma profile (×), prediction with apparent sigma profile (○), data used to identify apparent sigma profile (●), ± 100% error band (dashed line).

Table 9. RMSE in Logarithm of Solubility of Paracetamol in 26 Pure Solvents

	apparent sigma profile	
VT-2006 sigma profile	4 solvents	all solvents
1.67	1.04	0.95

respectively, and underestimates the solubililty in carbon tetrachloride by a factor of 15. It is worth noting that poor solubility predictions are observed in the chlorohydrocarbon solvents with both aspirin molecule and paracetamol molecule.

We further predict paracetamol solubility in four binary solvents for which solubility predictions have been done previously with NRTL-SAC at 298.15 K.[11] These four binary solvents include methanol−water, acetone−water, acetone−toluene, and methanol−ethyl acetate binaries. Note that, among the solvents, water is hydrophilic, methanol is partly hydrophilic and partly hydrophobic, acetone is polar, toluene is hydrophobic, and ethyl acetate is partly hydrophobic and partly polar.[3] We present the prediction results with the apparent sigma profile determined with four solvents.

Figure 10 shows prediction results for paracetamol solubility in methanol−water binary at 298.15 K with both the apparent

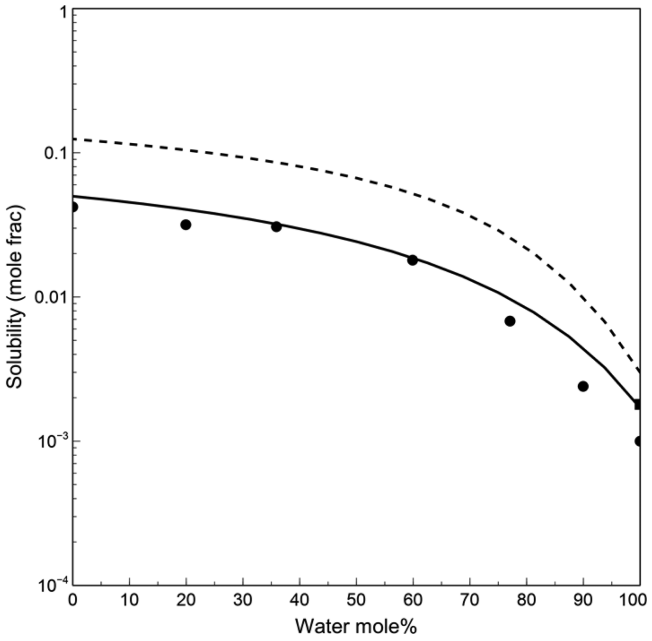


Figure 10. Experimental and predicted solubility of paracetamol in methanol−water binary solvent at 298.15 K: experimental data[43] (●), experimental data[44] (■), prediction with VT-2006 sigma profile (dashed line), prediction with apparent sigma profile (solid line).

sigma profile and the VT-2006 sigma profile. The VT-2006 sigma profile overpredicts the solubility of paracetamol in this binary system. However, the prediction trend seems to be consistent with that of the experimental one. In contrast, the apparent sigma profile predicted paracetamol solubility with excellent accuracy.

Figure 11 shows the model predictions and the experimental data of paracetamol solubility in acetone−water binary at 298.15 K. The VT-2006 sigma profile overpredicts the solubility at low water content and yields reasonable solubilty results when the water content exceeds 50 mol %. However, the VT-2006 sigma profile fails to predict a solubility maximum that is clearly shown by the experimental data. On the other hand, the predictions from the apparent simga profile show a maximal solubility in the binary system that seems to be consistent with the experimental solubility trend.

Figure 12 shows the model predictions and the experimental data of paracetamol solubility in acetone−toluene binary at 298.15 K. Both the apparent sigma profile predictions and the VT-2006 predictions show similar trends. The VT-2006 sigma profile overpredicts the paracetamol solubility throughout the

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

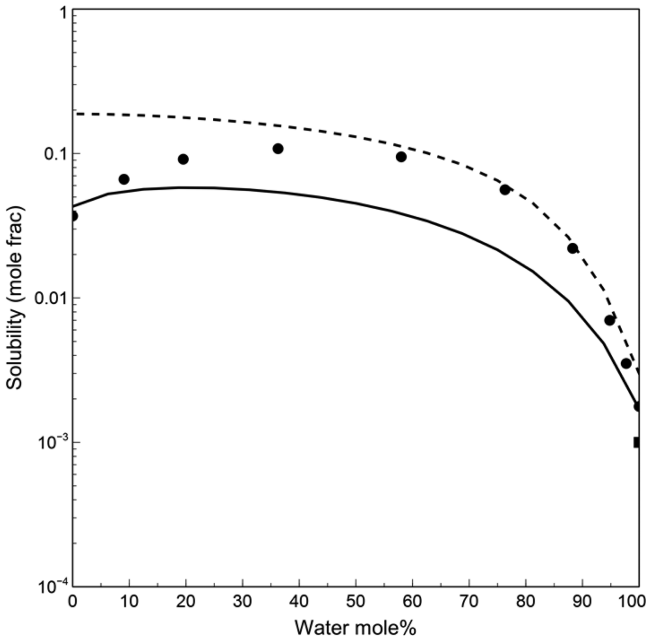


Figure 11. Experimental and predicted solubility of paracetamol in acetone−water binary solvent at 298.15 K: experimental data[44] (●), experimental data[43] (■), prediction with VT-2006 sigma profile (dashed line), prediction with apparent sigma profile (solid line).

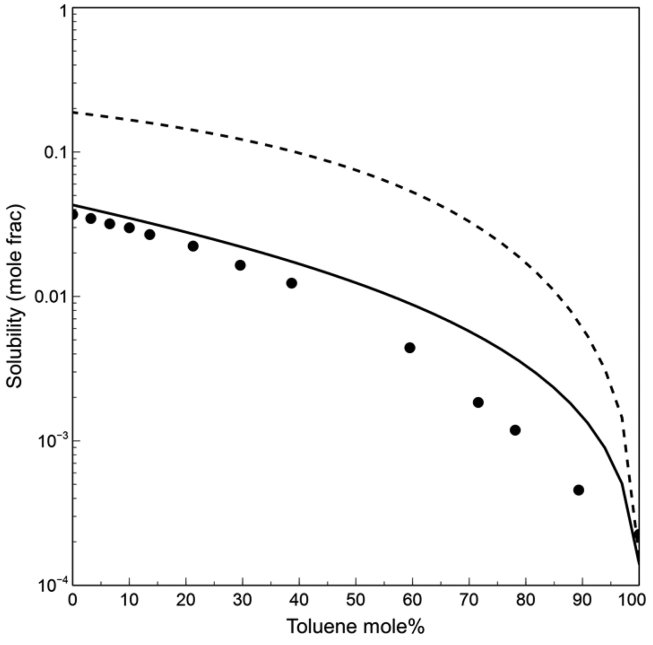


Figure 12. Experimental and predicted solubility of paracetamol in acetone-toluene binary solvent at 298.15 K: experimental data (●),[44] prediction with VT-2006 sigma profile (dashed line), prediction with apparent sigma profile (solid line).

entire concentration range except for pure toluene. The apparent sigma profile predictions are much closer to the data.

Figure 13 shows the model predictions and experimental data of paracetamol solubility in methanol−ethyl acetate binary at 298.15 K. Both the apparent sigma profile predictions and the VT-2006 predictions show similar trends. The VT-2006 sigma profile overpredicts the paracetamol solubility for the entire binary system, whereas the apparent sigma profile predictions follow the experimental data well.

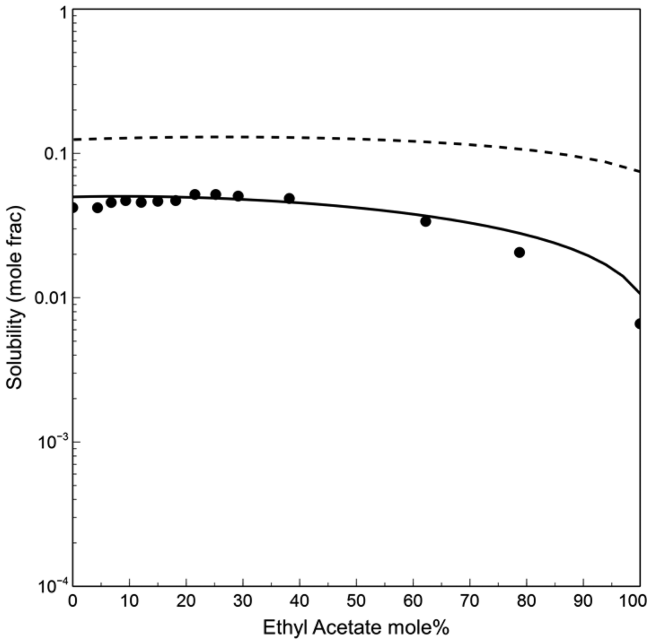


Figure 13. Experimental and predicted solubility of paracetamol in methanol−ethyl acetate binary solvent at 298.15 K: experimental data[43] (●), prediction with VT-2006 sigma profile (dashed line), prediction with apparent sigma profile (solid line).

Lovastatin. For lovastatin, the melting temperature and enthalpy of fusion are reported as 444.25 K and 36 530 kJ/kmol, respectively.[33] From these thermodynamic properties the logarithm of solubility constant, i.e., ln K_{sp}, is calculated to be −4.83 at 298.15 K.

Unlike the other three drug molecules, the sigma profile of lovastatin is not included in VT-2006 database. Therefore, the sigma profile of lovastatin is computed using DMol^3 module of Accelrys Materials Studio software package.[24] We follow the procedure for sigma profile generation outlined in the literature.[19] The computed sigma profile is shown in Figure 14. The most noticeable feature of the DMol^3-generated sigma profile for lovastatin is its very high level of hydrophobic segments.

The available experimental solubility data for lovastatin[33,39−41] in 18 pure solvents are reported in Table 10. To identify the conceptual segment numbers for lovastatin, we regress the solubility data of four solvents, i.e., ethyl acetate, acetone, methanol, and 1-octanol. The regressed conceptual segment numbers are reported in Table 11. The apparent sigma profile for lovastatin is shown in Figure 14 along with the one generated from DMol^3. The estimated conceptual segment numbers indicate relatively high hydrophobic segments and polar attractive segments. The conceptual segment numbers remain similar even when all the solubility data for the 18 solvents are used (Table 11). The apparent sigma profile shows significant hydrophobic segments but much lower than that of the DMol^3-generated sigma profile. The apparent cavity volume is calculated to be 251.45 Å^3. In comparison, the DMol^3 module reported the cavity volume of lovastatin molecule as 521.55 Å^3.

Solubility of lovastatin in the 18 pure solvents is calculated using both the apparent sigma profile and the DMol^3-generated sigma profile. The calculated results are presented in Table 10 together with their experimental values. A parity plot is included in Figure 15 to demonstrate the effectiveness of these

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

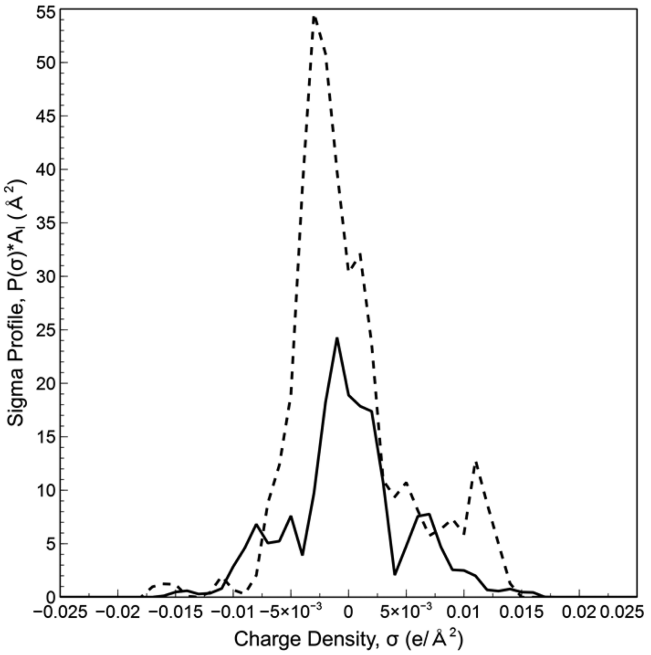


Figure 14. Sigma profile for lovastatin: DMol^3 (dashed line), apparent sigma profile (solid line).

two sigma profiles. The DMol^3-generated sigma profile slightly overpredicts the lovastatin solubility in all solvents except for 1-octanol and *n*-butyl acetate. The apparent sigma profile predicts solubility in all solvents well except for water. The model errors for lovastatin solubility calculation are reported in Table 12 for both the apparent sigma profile and the DMol^3-generated sigma profile. The RMSE for the apparent sigma profile estimated from four solvents assumes the highest value. This is due to the poor prediction of lovastatin solubility in water. Lovastatin is sparingly soluble in water, and its solubility is difficult to ascertain. If the water solubility prediction is excluded from the RMSE calculation, the RMSE for the apparent sigma profile drops drastically from 1.33 to 0.23.

Table 11. Conceptual Segment Numbers for Lovastatin

	segment number	
conceptual segment	4 solvents	all solvents
X	0.670	1.044
Y^-	0.038	0
Y^+	0.840	0.665
Z	0.202	0.212

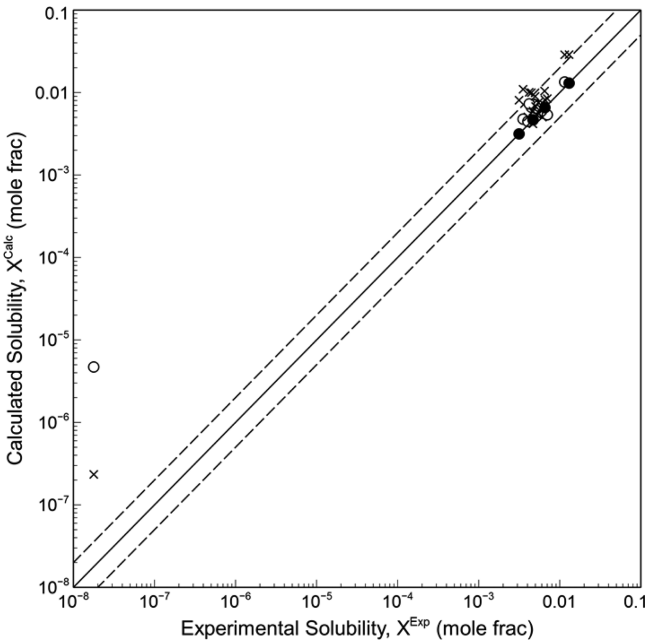


Figure 15. Comparison of experimental solubility of lovastatin with model results: prediction with DMol^3 sigma profile (×), prediction with apparent sigma profile (○), data used to identify apparent sigma profile (●), ± 100% error band (dashed lines).

Table 10. Experimental and Calculated Solubility of Lovastatin

			calculated solubility		
solvent	temp. (K)	experimental solubility (mole frac)	DMol^3 sigma profile (mole frac)	apparent sigma profile (mole frac)	data ref.
ethyl acetate[a]	297.20	6.56×10^{-3}	8.17×10^{-3}	6.56×10^{-3}	41
acetone[a]	297.20	1.30×10^{-2}	2.88×10^{-2}	1.30×10^{-2}	41
methanol[a]	298.35	3.15×10^{-3}	8.07×10^{-3}	3.15×10^{-3}	39
1-octanol[a]	301.20	4.67×10^{-3}	4.17×10^{-3}	4.67×10^{-3}	40
methyl acetate	297.20	4.47×10^{-3}	9.97×10^{-3}	6.63×10^{-3}	41
N-propyl acetate	297.20	5.87×10^{-3}	6.42×10^{-3}	6.11×10^{-3}	41
iso-propyl acetate	297.20	4.97×10^{-3}	8.99×10^{-3}	7.10×10^{-3}	41
N-butyl acetate	297.20	6.11×10^{-3}	5.03×10^{-3}	5.62×10^{-3}	41
isobutyl acetate	297.20	5.04×10^{-3}	6.77×10^{-3}	6.44×10^{-3}	41
sec-butyl acetate	297.20	5.45×10^{-3}	7.55×10^{-3}	6.88×10^{-3}	41
tert-butyl acetate	297.20	4.20×10^{-3}	1.01×10^{-2}	7.28×10^{-3}	41
2-butanone	297.20	1.15×10^{-2}	2.87×10^{-2}	1.34×10^{-2}	41
ethanol	298.15	3.53×10^{-3}	1.10×10^{-2}	4.76×10^{-3}	39
1-propanol	301.65	6.50×10^{-3}	1.04×10^{-2}	5.58×10^{-3}	40
N-butanol	301.20	7.02×10^{-3}	8.44×10^{-3}	5.37×10^{-3}	40
1-pentanol	299.20	4.68×10^{-3}	5.96×10^{-3}	4.64×10^{-3}	40
1-hexanol	298.95	4.04×10^{-3}	5.07×10^{-3}	4.47×10^{-3}	40
water	298.15	1.78×10^{-8}	2.34×10^{-7}	4.70×10^{-6}	33

[a]Solvents used for conceptual segment number estimation.

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

Table 12. RMSE in Logarithm of Solubility of Lovastatin in 18 Pure Solvents

	apparent sigma profile	
DMol^3 sigma profile	4 solvents	all solvents
0.84	1.33	0.70
0.60[a]	0.23[a]	0.54[a]

[a]Water excluded.

We further predict the lovastatin solubility in acetone−water binary solvent. The experimental solubility data in the binary solvent at 318.15 K are available[42] along with the lovastatin solubility in pure acetone.[39] However, we could not find any experimental solubility of lovastatin in water at 318.15 K. Figure 16 shows the prediction results for lovastatin solubility in

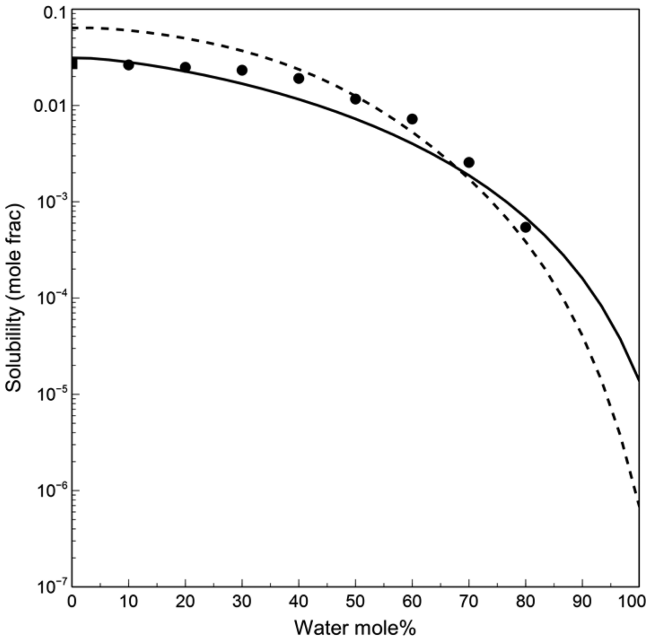


Figure 16. Experimental and predicted solubility of lovastatin in acetone−water binary at 318.15 K: experimental data[39] (■), experimental data[42] (●), prediction with DMol^3-generated sigma profile (dashed line), prediction with apparent sigma profile (solid line).

acetone−water binary with the apparent sigma profile and the DMol^3-generarted sigma profile. The prediction results from both sigma profiles show a similar solubility trend against water content. The apparent sigma profile predicts lovastatin solubility in acetone very close to their experimental value, whereas the DMol^3-generated sigma profile slightly over-predicts lovastatin solubility in the acetone-rich region.

DISCUSSION

To summarize the results for the four drug molecules, Figure 17 shows the parity plot for all the pure solvent solubility data and model results for the four drug molecules. With the 75 solubility data investigated, 29% of the predicted solubilities fall inside the ±100% error band when the sigma profiles from VT-2006 and DMol^3 are used. Most of the predicted solubilities are higher than the experimental values. On the other hand, 73% of the solubilities calculated with the apparent sigma profiles fall inside the ±100% error band. It should be emphasized that the

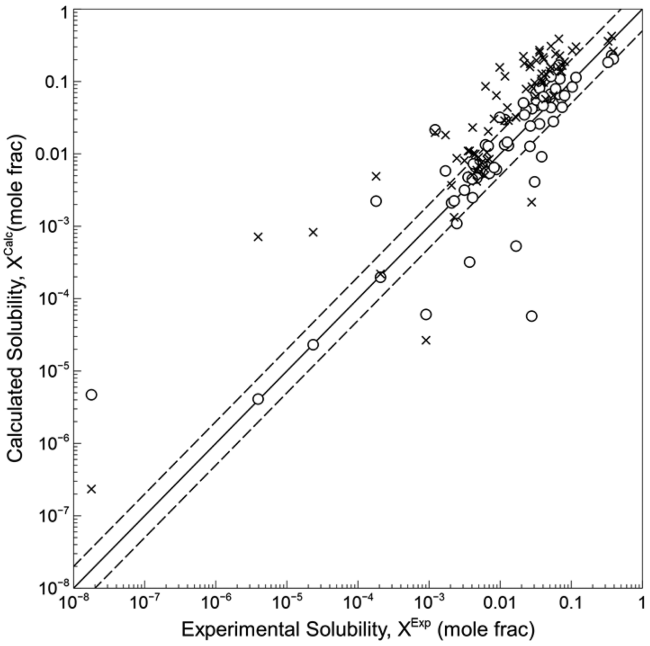


Figure 17. Solubility prediction of four drug molecules: prediction with VT-2006 and DMol^3 sigma profiles (×), model results with apparent sigma profiles (○), ± 100% error band (dashed line).

model results with VT-2006 and DMol^3 sigma profiles represent COSMO-SAC predictions without the use of the experimental data.

The conceptual segment numbers of NRTL-SAC model for the four drug molecules, published in the literature, are reported in Table 13. It is interesting that the conceptual

Table 13. NRTL-SAC Conceptual Segment Numbers for Four Drug Molecules

conceptual segment	caffeine[49]	aspirin[3]	paracetamol[11]	lovastatin[50]
X	−0.105	0.103	0.414	1.175
Y^-	0.282	0	0	0
Y^+	1.211	1.160	0.441	0.548
Z	0.211	1.372	2.462	0.882

segment numbers determined in this study for COSMO-SAC apparent sigma profiles are largely in line with their NRTL-SAC counterparts. The dominant conceptual segments identified by NRTL-SAC for the four drug molecules are also identified by the apparent sigma profiles. Obvisously, the two sets of conceptual segment numbers are not exactly the same because different sets of reference molecules are selected for the two models.

CONCLUSION

The proposed methodology offers a simple and practical approach for generating empirical, apparent sigma profile of any molecule from available experimental solubility data or other relavent phase equilibrium data. The methodology requires no knowledge of molecular structure, no use of DFT calculations, and no quantum chemistry packages. Incorporating the conceptual segment concept, this method generates apparent sigma profiles from sigma profiles for conceptual segment reference molecules, and the necessary conceptual segment numbers are identified from fitting against experimental solubility data. The methodology allows use of valuable

dx.doi.org/10.1021/ie503829b | Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

experimental measurements as input to the COSMO-based models and transforms COSMO-SAC into a correlative model. The apparent sigma profiles represent "best-fit" profiles against experiemntal data and can be used for further predictions. The sigma profile generation methodology should enhance the usability of the COSMO-based thermodynamic models in the predictions of liquid-phase nonideality and fluid-phase equilibria.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +1-806-834-3098. E-mail: chauchyun.chen@ttu.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the generous financial support of the Jack Maddox Distinguished Engineering Chair Professorship in Sustainable Energy sponsored by the J.F. Maddox Foundation.

REFERENCES

(1) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **1975**, *21* (6), 1086−1099.

(2) Hahnenkamp, I.; Graubner, G.; Gmehling, J. Measurement and prediction of solubilities of active pharmaceutical ingredients. *Int. J. Pharm. (Amsterdam, Neth.)* **2010**, *388* (1), 73−81.

(3) Chen, C.-C.; Song, Y. Solubility modeling with a nonrandom two-liquid segment activity coefficient model. *Ind. Eng. Chem. Res.* **2004**, *43* (26), 8354−8362.

(4) Klamt, A. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224−2235.

(5) Lin, S.-T.; Sandler, S. I. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind. Eng. Chem. Res.* **2002**, *41* (5), 899−913.

(6) Ren, D.-B.; Yang, Z.-H.; Liang, Y.-Z.; Ding, Q.; Chen, C.; Ouyang, M.-L. Correlation and prediction of partition coefficient using nonrandom two-liquid segment activity coefficient model for solvent system selection in counter-current chromatography separation. *J. Chromatogr., A* **2013**, *1301*, 10−18.

(7) Loschen, C.; Klamt, A. COSMO quick: A Novel Interface for Fast σ-Profile Composition and Its Application to COSMO-RS Solvent Screening Using Multiple Reference Solvents. *Ind. Eng. Chem. Res.* **2012**, *51* (43), 14303−14308.

(8) Prausnitz, J. M.; Lichtenthaler, R. N.; de Azevedo, E. G. *Molecular thermodynamics of fluid-phase equilibria*; Prentice-Hall, Inc.: Upper Saddle River, NJ, 1998.

(9) Renon, H.; Prausnitz, J. M. Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE J.* **1968**, *14* (1), 135−144.

(10) Chen, C.-C. A segment-based local composition model for the Gibbs energy of polymer solutions. *Fluid Phase Equilib.* **1993**, *83*, 301−312.

(11) Chen, C.-C.; Crafts, A. Correlation and Prediction of Drug Molecule Solubility in Mixed Solvent Systems with the Nonrandom Two-Liquid Segment Activity Coefficient (NRTL-SAC) model. *Ind. Eng. Chem. Res.* **2006**, *45* (13), 4816−4824.

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

(13) Lin, S.-T.; Sandler, S. I. Infinite dilution activity coefficients from ab initio solvation calculations. *AIChE J.* **1999**, *45* (12), 2606−2618.

(14) Lin, S.-T.; Sandler, S. I. Prediction of octanol-water partition coefficients using a group contribution solvation model. *Ind. End. Chem. Res.* **1999**, *38* (10), 4081−4091.

(15) Staverman, A. J. The entropy of high polymer solutions. Generalization of formulae. *Recl. Trav. Chim. Pays-Bas* **1950**, *69* (2), 163−174.

(16) Guggenheim, E. A. *Mixtures: The theory of the equilibrium properties of some simple classes of mixtures, solutions and alloys*; Clarendon Press: Oxford, U.K., 1952.

(17) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. Refinement and parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102* (26), 5074−5085.

(18) Klamt, A.; Eckert, F. COSMO-RS: A novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib.* **2000**, *172* (1), 43−72.

(19) Mullins, E.; Oldland, R.; Liu, Y.; Wang, S.; Sandler, S. I.; Chen, C.-C.; Zwolak, M.; Seavey, K. C. Sigma-profile database for using COSMO-based thermodynamic methods. *Ind. Eng. Chem. Res.* **2006**, *45* (12), 4389−4415.

(20) Mullins, E.; Liu, Y.; Ghaderi, A.; Fast, S. D. Sigma profile database for predicting solid solubility in pure and mixed solvent mixtures for organic pharmacological compounds with COSMO-based thermodynamic methods. *Ind. Eng. Chem. Res.* **2008**, *47* (5), 1707−1725.

(21) Politzer, P.; Seminario, J. M. *Modern density functional theory: A tool for chemistry*; Elsevier: Amsterdam, 1995.

(22) Delley, B. An all-electron numerical method for solving the local density functional for polyatomic molecules. *J. Chem. Phys.* **1990**, *92* (1), 508−517.

(23) Delley, B. Analytic energy derivatives in the numerical local-density-functional approach. *J. Chem. Phys.* **1991**, *94* (11), 7245−7250.

(24) *DMol3, Materials Studio 7.0*. Accelrys Software, Inc.: San Diego, CA.

(25) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14* (11), 1347−1363.

(26) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery, J.; Vreven, T.; Kudin, K.; Burant, J.; et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2008.

(27) *Jaguar*, 4.2. Schrödinger Inc.: Portland, OR, 1991−2000.

(28) Stewart, J. MOPAC program package. *Quantum Chemistry Program Exchange* 1989, No. 455.

(29) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **1989**, *162* (3), 165−169.

(30) Wang, S.; Lin, S.-T.; Watanasiri, S.; Chen, C.-C. Use of GAMESS/COSMO program in support of COSMO-SAC model applications in phase equilibrium prediction calculations. *Fluid Phase Equilib.* **2009**, *276* (1), 37−45.

(31) Mu, T.; Rarey, J.; Gmehling, J. Performance of COSMO-RS with sigma profiles from different model chemistries. *Ind. Eng. Chem. Res.* **2007**, *46* (20), 6612−6629.

(32) Gerber, R. P.; Soares, R. d. P. Prediction of infinite-dilution activity coefficients using UNIFAC and COSMO-SAC variants. *Ind. Eng. Chem. Res.* **2010**, *49* (16), 7488−7496.

(33) Tung, H.-H.; Tabora, J.; Variankaval, N.; Bakken, D.; Chen, C.-C. Prediction of pharmaceutical solubility Via NRTL-SAC and COSMO-SAC. *J. Pharm. Sci.* **2008**, *97* (5), 1813−1820.

(34) Ruether, F.; Sadowski, G. Modeling the Solubility of Pharmaceuticals in Pure Solvents and Solvent Mixtures for Drug Process Design. *J. Pharm. Sci.* **2009**, *98* (11), 4205−4215.

(35) Marrero, J.; Abildskov, J. *Solubility and related properties of large complex chemicals*; DECHEMA: Frankfurt, Ger., 2003; Vol. *15*.

(36) Frank, T. C.; Gupta, S. K.; Downey, J. R. Quickly Screen Solvents for Organic Solids. *Chem. Eng. Prog.* **1999**, *95* (12), 41−61.

(37) Manzo, R. H.; Ahumada, A. A. Effects of solvent medium on solubility. V: enthalpic and entropic contribution to the free energy changes of di-substituted benzene derivatives in ethanol:water and ethanol:cyclohexane mixtures. *J. Pharm. Sci.* **1990**, *79*, 1109.

(38) Granberg, R. A.; Rasmuson, A. C. Solubility of Paracetamol in Pure Solvents. *J. Chem. Eng. Data* **1999**, *44* (6), 1391−1395.

(39) Sun, H.; Gong, J.-b.; Wang, J.-k. Solubility of lovastatin in acetone, methanol, ethanol, ethyl acetate, and butyl acetate between 283 K and 323 K. *J. Chem. Eng. Data* **2005**, *50* (4), 1389−1391.
(40) Chiew, Y. C.; Chmielowski, R.; Chan, V.; Nti-Gyabaah, J. Solubility of lovastatin in a family of six alcohols: Ethanol, 1-propanol, 1-butanol, 1-pentanol, 1-hexanol, and 1-octanol. *Int. J. Pharm. (Amsterdam, Neth.)* **2008**, *359* (1), 111−117.
(41) Nti-Gyabaah, J.; Chiew, Y. C. Solubility of lovastatin in ethyl acetate, propyl acetate, isopropyl acetate, butyl acetate, sec-butyl acetate, isobutyl acetate, tert-butyl acetate, and 2-butanone, between (285 and 313) K. *J. Chem. Eng. Data* **2008**, *53* (9), 2060−2065.
(42) Sun, H.; Wang, J. Solubility of Lovastatin in Acetone+Water Solvent Mixtures. *J. Chem. Eng. Data* **2008**, *53* (6), 1335−1337.
(43) Subrahmanyam, C. V. S.; Reddy, M. S.; Rao, J. V.; Rao, P. G. Irregular solution behaviour of paracetamol in binary solvents. *Int. J. Pharm. (Amsterdam, Neth.)* **1992**, *78* (1), 17−24.
(44) Granberg, R. A.; Rasmuson, A. C. Solubility of paracetamol in binary and ternary mixtures of water + acetone + toluene. *J. Chem. Eng. Data* **2000**, *45* (3), 478−483.
(45) Martin, A.; Paruta, A.; Adjei, A. Extended Hildebrand solubility approach: Methylxanthines in mixed solvents. *J. Pharm. Sci.* **1981**, *70* (10), 1115−1120.
(46) Paruta, A.; Irani, S. Solubility profiles for the xanthines in aqueous alcoholic mixtures I. Ethanol and methanol. *J. Pharm. Sci.* **1966**, *55* (10), 1055−1059.
(47) Yalkowsky, S.; Valvani, S.; Roseman, T. Solubility and partitioning VI: Octanol solubility and octanol-water partition coefficients. *J. Pharm. Sci.* **1983**, *72* (8), 866−870.
(48) Bustamante, P.; Navarro, J.; Romero, S.; Escalera, B. Thermodynamic origin of the solubility profile of drugs showing one or two maxima against the polarity of aqueous and nonaqueous mixtures: Niflumic acid and caffeine. *J. Pharm. Sci.* **2002**, *91* (3), 874−883.
(49) Tanveer, S.; Hao, Y.; Chen, C.-C. Introduction to Solid-Fluid Equilibrium Modeling. *Chem. Eng. Prog.* **2014**, *110* (9), 37−47.
(50) Sheikholeslamzadeh, E.; Rohani, S. Solubility prediction of pharmaceutical and chemical compounds in pure and mixed solvents using predictive models. *Ind. Eng. Chem. Res.* **2012**, *51* (1), 464−473.