Solubility Prediction of Pharmaceutical and Chemical Compounds in Pure and Mixed Solvents Using Predictive Models

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · JANUARY 2012

Impact Factor: 2.59 · DOI: 10.1021/ie201344k

2 AUTHORS:

Ehsan Sheikholeslamzadeh

4 PUBLICATIONS **19** CITATIONS

SEE PROFILE

Sohrab Rohani

The University of Western Ontario

280 PUBLICATIONS **3,351** CITATIONS

SEE PROFILE

I&EC research
Industrial & Engineering Chemistry Research

ARTICLE

pubs.acs.org/IECR

Solubility Prediction of Pharmaceutical and Chemical Compounds in Pure and Mixed Solvents Using Predictive Models

Ehsan Sheikholeslamzadeh and Sohrab Rohani*

Department of Chemical and Biochemical Engineering, The University of Western Ontario, London, Ontario, Canada, N6A 5B9

ABSTRACT: Thermodynamic models offer a fast, reliable, and cost-effective method to select the best solvent or solvent mixtures for crystallization of solid components. To optimize the performance of the unit operations which produce active pharmaceutical ingredients (APIs), the physical properties of the solute and solvent must be known. Solubility prediction is very crucial in the fine and specialty chemical industries, as the total cost of production is high in most cases. In this study, the solubility of three chemical compounds, 3-pentadecylphenol, lovastatin, and valsartan, in different solvents and solvent mixtures were studied experimentally and theoretically. The thermodynamic models of the UNIFAC and the NRTL-SAC model were used for prediction. The results of the prediction from the two models and their average relative deviation for the three model compounds showed a better performance for the NRTL-SAC model compared to the UNIFAC. For the case of lovastatin and valsartan, the NRTL-SAC model gives the average relative deviation of 0.2401 and 0.3843, respectively. Because of the flexibility of the NRTL-SAC program code that is written for the phase behavior prediction, it can be used for further analysis and optimization of the performance of crystallization processes (i.e., solvent screening and yield of the process). This study shows that the NRTL-SAC model can be used effectively in pharmaceutical industry, especially for solvent screening purposes.

1. INTRODUCTION

One of the main properties of polymorphic compounds is their solubility difference in solvents and solvent mixtures. With this important property, different polymorphs of a drug can be separated by crystallization processes. In the synthesis of pharmaceuticals there are factors that affect the overall performance of unit operations such as solvent selection, operating temperature, and supersaturation.[1] Empirical selection of solvents requires extensive experimentation and high cost.[2] The predictive thermodynamic models can be a good choice in estimation of the phase behavior and solubility of drugs in different solvents and solvent mixtures.[3] There have been many thermodynamic models for the prediction of the phase behavior of liquid−liquid systems, such as the Margules equation,[4] Wilson equation,[5] Van Laar equation,[6] NRTL equation,[7] and UNIQUAC equation.[8] These models can also be used for solid−liquid equilibrium behavior. Generally, the equations of the activity coefficient for solubility prediction can be divided in two categories:

(1) The correlative models, which require many experimental data under different conditions. In some cases, the data on ternary mixtures are also needed, such as for Wilson's model.
(2) The group contribution and predictive models, which only require the chemical structure of the molecule and/or a few experimental data points to predict the phase behavior of the solid in different solvents, such as the UNIFAC model.

From the above two categories, the first is not very useful for solubility prediction and solvent screening purposes.[1] The main reason for this is the lack of experimental data for the binary interaction parameters of the solute−solvent, solute−antisolvent, and solvent−antisolvent systems. Matsuda et al.[9] used Wilson's model to predict the solubility of salicylic acid, benzocaine, acetanilide, and phenacetin in water mixed with different co-solvents. They referred to several literature reports to get the parameters for Wilson's equation. For estimation of the pure parameters, they used the Tassios method[10] followed by using DECHEMA VLE collection.[11] In addition, they had to consider some assumptions (i.e., they used the simplified method to find the interaction parameters, or the molar volumes were estimated using the group contribution method) which led to some errors in prediction.

From the second category, the universal functional activity coefficient (UNIFAC) model is one of the most well-developed methods. This model is basically used for prediction of the phase behavior of nonelectrolytic and nonideal systems, which was first introduced by Fredenslund et al.[12] in 1975. They used the UNIFAC model to predict the solubility of naphthalene, anthracene, and phenanthrene in various solvents and their mixtures. They showed the applicability of the UNIFAC model for prediction of the phase behavior of solutes in solvents. After the introduction of the model, some modifications have been made to the original form of the UNIFAC model to improve the prediction. Sheikhzadeh et al.[13] used the UNIFAC and its modified form to predict the solubility of the two forms of buspirone hydrochloride in 2-propoanol and its mixture with water. They concluded that the modified UNIFAC models are not suitable for prediction of highly soluble drugs. Gracin and co-workers[14] studied the solubility of a wide variety of chemical compounds in water and some organic solvents. Although they obtained some reasonable

Received: June 22, 2011
Accepted: November 11, 2011
Revised: November 10, 2011
Published: November 14, 2011

dx.doi.org/10.1021/ie201344k | *Ind. Eng. Chem. Res.* 2012, 51, 464–473

predictions for a few chemicals, the model was not appropriate for solvent screening purposes. They used some known group parameters in the UNIFAC method instead of complex and unknown groups. In addition, because of the dependency of the property of some functional groups on the rest of the molecule, they concluded that the UNIFAC was not a suitable method for crystallization process design and solvent screening. Kan et al.[15] used the UNIFAC method for solubility prediction of some compounds from alkanes, alkylbenzenes, alkenes, and benzenes. The solubility of 11 out of 14 compounds was best predicted by the UNIFAC model in their study, while the chlorinated alkenes, phthalates, and long-chain alkanes were not predicted well.

The NRTL-SAC model was first introduced by Chen et al.[16] in 2004. This model was proposed to compensate for the weaknesses of the UNIFAC model in predicting the solubility of complex chemical molecules with functional groups for which the UNIFAC parameters had not been determined or where the UNIFAC group addition rule becomes invalid.[3] One of the main advantages of the NRTL-SAC model in comparison to the other predictive methods is its ability to predict organic electrolyte systems.[17] The UNIFAC method identifies the molecule in terms of its functional groups, while the NRTL-SAC model divides the whole surface of the molecule into three segments. Each compound can have three conceptual segments: hydrophobic, hydrophilic, and polar. Hydrophobic segment (X) accounts for the molecular surfaces that do not tend to form a hydrogen bond, such as hexane. Hydrophilic segment (Z) contributes to the part of the molecule that tends to form a hydrogen bond, such as water. The polar segments, (Y−) and (Y+), do not belong to neither hydrophobic nor hydrophilic segment. The polar attractive segment (Y−) shows an attraction toward the hydrophilic segment, while the polar repulsive segment (Y+) has a repulsive characteristic toward the hydrophilic segment.

The three segments are identified in terms of the interaction between the molecules in a solution that is expressed in the phase equilibrium experimental data. Chen et al.[3] selected water as a reference for the hydrophilic segment, acetonitrile as a reference for the polar segment, and hexane as a reference for the hydrophobic segment. To predict the solubility of a chemical in a solvent or a mixture of solvents, the segment numbers of solute and solvent(s) should be known. Chen et al. has studied extensive data of liquid−liquid equilibrium (LLE) and vapor−liquid equilibrium (VLE) for various solvents. On the basis of the reference solvents (water, acetonitrile, and hexane), they found the optimized values of segment numbers for other common solvents. Many solvents have only one or two segment numbers that are significant compared to the other segment numbers. There are 62 common solvents for which segment numbers have been adjusted and tabulated for calculation of phase behavior of solid−liquid systems.[17]

The fundamental relation for calculation of the solid−liquid phase behavior is:[18]

$$\ln x_s = -\frac{\Delta H_{fus}}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right) - \ln\gamma_s \quad (1)$$

where the index s refers to the solid phase, ΔH_{fus} is the heat of fusion, and T_m is the melting temperature of the solute. To predict the solubility of a solid, its physical properties (ΔH_{fus} and T_m can be obtained using thermal methods, such as DSC and TGA) and a proper model that describes the activity coefficient are needed.

The group contribution models divide the contribution, which affect the activity coefficient by the following:

- Combinatorial part: this part includes the contribution of the chemical structure and the size (volume and surface of the molecule) of the compound in the activity coefficient.
- Residual part: this part includes the contribution of the group size and binary interaction between pairs of the functional groups in the activity coefficient.

The equations for calculation of the UNIFAC and NRTL-SAC models were obtained from the literature.[12,16] With the above definition, the total activity of a component in a solution is the sum of two parts:

$$\ln\gamma_i = \ln\gamma_i^{C} + \ln\gamma_i^{R} \quad (2)$$

in which γ_i is the activity coefficient of component i in the solution, γ_i^{C} is the combinatorial part, and γ_i^{R} is the residual part.

According to Chen et al.[16] the NRTL-SAC model is based on the derivation of the original NRTL model for polymers. In the NRTL-SAC model, the combinatorial part is calculated from eq 3:

$$\ln\gamma_i^{C} = \ln\frac{\phi_i}{x_i} + 1 - r_i\sum_j\frac{\phi_j}{x_j} \quad (3)$$

with the definitions:

$$r_i = \sum_j r_{j,i} \quad (4)$$

$$\phi_i = \frac{r_i x_i}{\sum_j r_j x_j} \quad (5)$$

where x_i is the mole fraction of component i, r_i is the total segment number of component i, and ϕ_i is the segment mole fraction in the mixture. The residual term is defined as:

$$\ln\gamma_i^{R} = \ln\gamma_i^{lc} = \sum_m r_{m,i}(\ln\Gamma_m^{lc} - \ln\Gamma_m^{lc,i}) \quad (6)$$

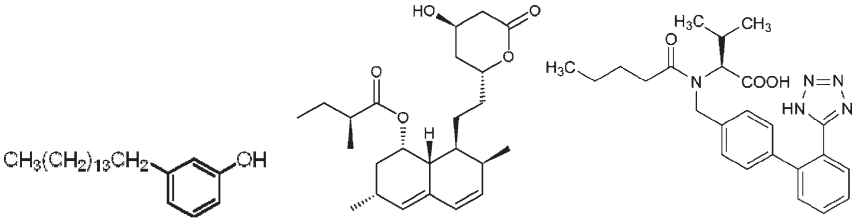
where the terms $\ln\Gamma_m^{lc}$ and $\ln\Gamma_m^{lc,i}$ are the activity coefficients of segment m in solution and component i, respectively, and $r_{m,i}$ is the number of segment m in component i. These terms are found using eqs 7 and 8:

$$\ln\Gamma_m^{lc} = \frac{\sum_j x_j G_{j,m}\tau_{j,m}}{\sum_k x_k G_{k,m}} + \sum_{m'}\frac{x_{m'}G_{m,m'}}{\sum_k x_k G_{k,m'}}\left(\tau_{m,m'} - \frac{\sum_j x_j G_{j,m'}\tau_{j,m'}}{\sum_k x_k G_{k,m'}}\right) \quad (7)$$

$$\ln\Gamma_m^{lc,l} = \frac{\sum_j x_{j,l} G_{j,m}\tau_{j,m}}{\sum_k x_{k,l} G_{k,m}} + \sum_{m'}\frac{x_{m',l}G_{m,m'}}{\sum_k x_{k,l} G_{k,m'}}\left(\tau_{m,m'} - \frac{\sum_j x_{j,l} G_{j,m'}\tau_{j,m'}}{\sum_k x_{k,l} G_{k,m'}}\right) \quad (8)$$

In eqs 7 and 8, l refers to the component and j, k, m, and m′ refer to the segments in each component. $x_{j,l}$ is the segment-based mole fraction of segment species j in component l only.

dx.doi.org/10.1021/ie201344k |Ind. Eng. Chem. Res. 2012, 51, 464–473



(a) (b) (c)

Figure 1. The chemical structure of (a) 3-pentadecylphenol, (b) lovastatin, and (c) valsartan.

The mole fractions of segments in the whole solution and in components are expressed below:

$$x_j = \frac{\sum_l x_l r_{j,l}}{\sum_z \sum_i x_z r_{j,z}} \tag{9}$$

$$x_{j,l} = \frac{r_{j,l}}{\sum_i r_{j,l}} \tag{10}$$

$G_{i,j}$ and $\tau_{i,j}$ are the local binary values between segments i and j, which can be related to each other based on NRTL nonrandom parameters, α_{ij}.[3] In the NRTL-SAC model, after insertion of the values of segments for solvents and the corresponding initial guess values for the solute segments, the written code for NRTL-SAC in Matlab starts solving for the mole fractions at saturation for all of the species in the solution. The only thing here that is different from the UNIFAC model is the parameter estimation loop. This part of the model (which has a separate Matlab code) uses a few experimental data to fit the model output to the experimental data. After optimization of the process and finding the four segment numbers for the solute, the values can be set for that compound and used for further predictions. For the NRTL-SAC model, we used the *lsqnonlin* routine of the Maltab software, which is suitable for solving the nonlinear least-squares problems. The algorithm for solving the least-squares is based on the interior-reflective Newton method.[19] After finding the adjustable parameters for the selected solutes, we used the model equation to predict the solubility of the compound in pure/mixed solvents. For the UNIFAC model, there is no need to find the adjustable parameters. The average relative deviation (ARD) for the whole data is calculated for each case to find which model best describes the system:

$$\mathrm{ARD} = \frac{1}{N} \sum_i \left| \frac{x_i^{\mathrm{model}} - x_i^{\mathrm{exp}}}{x_i^{\mathrm{exp}}} \right| \tag{11}$$

where N is the total number of data points.

2. MATERIALS AND METHODS

2.1. Model Compounds. The model compounds for our study are 3-pentadecylphenol,[20] lovastatin,[21,22] and valsartan.[23,24] 3-Pentadecylphenol with the chemical formula of $C_{21}H_{36}O$ and molecular weight of 304.51 is the product of catalytic hydrogenation of cardanol. Its main use is in the agriculture industry as an emulsifier and coating material. The molecule has a phenolic head and a linear alkyne group, as shown in Figure 1. Because of the long hydrocarbon chain and a near-straight shape of the molecule, one can consider the molecule as if it does not have a benzene ring. This assumption can be reasonable because the benzene ring has a nonpolar characteristic and has a minor effect on the nonideality of the solution. The length of the hydrocarbon chain is about 29.53 Å and that of the benzene ring is 4.73 Å. This means that about 16% of the volume of the molecule is occupied by the benzene ring which is nonpolar while the other part of molecule is mostly the alkyl chain.

Mao et al.[20] studied the solubility of this compound in six pure solvents: ethanol, 1-butanol, toluene, acetone, tetrachloromethane, and ethyl acetate.[20] They used a laser detector to find the point of saturation, and Wilson's equation to get the adjustable parameters for every solute−solvent system. They did not predict the solubility, but only fitted the experimental data to the Wilson's model and derived the values of the binary interactions for the system. We also conducted solubility measurements of this compound in pure and mixed solvents. The material and the experimental procedure will be discussed in the next section. In our study we used the UNIFAC and the NRTL-SAC model to predict the solubility of the compound and compared it with the experimental data.

Lovastatin is a drug that belongs to a group of compounds called statins that lower the lipid levels in the body.[21] There are other statins approved in many countries, such as simvastatin and fluvastatin. The chemical name of lovastatin is butanoic acid 2-methyl-1,2,3,7,8,8 a-hexahydro-3,7-dimethyl-8-[2-(tetrahydro-4-hydroxy-6-oxo-2*H*-pyran-2-yl)ethyl]-1-naphthalenyl ester with the chemical formula of $C_{24}H_{36}O_5$ and molecular weight of 404.54. The chemical structure of lovastatin is represented in Figure 1. There are two main studies that have reported on the solubility of this compound.[21,22] In the study by Nti-Gyabaah et al.,[22] the solubility of lovastatin in a group of acetate solvents was measured. The solvents were ethyl acetate, propyl acetate, isopropyl acetate, butyl acetate, *sec*-butyl acetate, isobutyl acetate, *tert*-butyl acetate, and 2-butanone. All of the solubility experiments were conducted in the range of 285−313 K. In their study, they used the NRTL model to find the binary interaction parameters for solute−solvent mixtures. In another study, the solubility of lovastatin in six alcohols was measured.[21] It was found that in alcohols the solubility of lovastatin increases from ethanol to 1-butanol and then decreases as the carbon chain length increases. This can be rationalized by the solute−solvent interactions.

Valsartan with the chemical formula of $C_{24}H_{29}N_5O_3$ and chemical name (*S*)-*N*-(1-carboxy-2-methylprop-l-yl)-*N*-pentanoyl-*N*-[2′-(1*H*-tetrazol-5-yl)biphenyl-4-ylmethyl]amine is

dx.doi.org/10.1021/ie201344k |*Ind. Eng. Chem. Res.* 2012, 51, 464–473

Table 1. Physical Properties of the Model Compounds Used for Thermodynamic Modeling and Prediction

compound	melting point [K]	heat of fusion [J/mol]	entropy of fusion [J/mol-K]
3-pentadecylphenol[26]	322.35	38092	118.14
lovastatin[23,24]	445.50	43136	96.86
valsartan[25,26]	380.65	31647	84.97

used orally for the treatment of hypertension.[23] The chemical structure of valsartan is illustrated in Figure 1. Valsartan is extracted from the product mixture by ethyl acetate at the end of the production process, because of its high solubility.[24] However, some problems such as poor yield, slow filtration, and long drying time are the results of recrystallization with ethyl acetate. Hexane has been used as an antisolvent to overcome those problems. Liu et al.[24] determined the solubility in ethyl acetate−hexane at different molar ratios over the temperature range of 278.15 K to 313.15 K. The synthetic method was used to obtain the solubility of valsartan in the solvent mixture. In another study, the solubility of valsartan in six different solvents, methyl acetate, *n*-butyl acetate, acetonitrile, *N*,*N*-dimethylformamide, dichloromethane, and chloroform, over the temperature range of 278.15 K to 313.15 K was measured.[24] The melting point temperature, heat of fusion, and entropy change of melting of the three mentioned model compounds are given in Table 1.

2.2. Experimental Setup and Procedure. The solubility of 3-pentadecylphenol in isopropyl alcohol and in a mixture of isopropyl alcohol and water was measured. 3-Pentadecylphenol (90% purity) was purchased from Sigma Aldrich and was recrystallized in ethanol.[20] A specified amount of 3-pentadecylphenol was dissolved in ethanol at 50 °C. After the desired temperature was reached, the solution was filtered (VWR, grade 410, qualitative). The clear filtrate was maintained at 50 °C for 1 h. Then the solution was cooled at a rate of 0.2 °C/min until 20 °C and then maintained at that temperature for 1 h. The precipitated solids were filtered and dried. To identify its purity, we used differential scanning calorimetry (DSC) analysis (DSC, Mettler Toledo, Chicago, IL). The sample was put in a 40 μL aluminum crucible with a hole in the cap to allow venting. The heating rate was 0.5 °C/min in the temperature range of 20 °C to 120 °C. The DSC curve and the obtained melting temperature were compared to those of the purified compound. The results showed good purity of the sample and an effective experimental procedure.

After the purified compound was obtained, it was used to determine the solubility in IPA (isopropyl alcohol) solvent. Gravimetric analysis was used to obtain the solubility data. First, a specified amount of 3-pentadecylphenol was dissolved in 50 mL of IPA solvent at 40 °C. The solution was prepared in a 250 mL Bellco jacketed vessel (Vineland, NJ). A bath circulator (Julabo, Germany) was used for heating and cooling. A Teflon-coated thermocouple was used to read the temperature in the flask. For mixing, a top-mounted electromagnetic stirrer was employed. After complete dissolution of all of the particles in the solvent, the temperature was cooled to a specified temperature to determine the solubility data. The solution was kept at constant temperature for 1 h. A 0.45 μm membrane disk filter (VWR, Mississauga, ON) was used to filter the impurities from the sample solution. The clear saturated solution was kept in closed, weighed vials. The vial with solution was weighed and transferred to the oven (at 60 °C) for one day. The weight of solute and solvent were recorded for every sample. This experimental procedure was repeated three times for each temperature to increase the reliability. The replicates at every temperature were averaged, and the standard deviation was calculated for data analysis. In addition to single solvent experimentation, we used solvent mixtures in our study. Water was selected as an antisolvent, because 3-pentadecylphenol is sparingly soluble in water, but fairly soluble in IPA. The solubility of 3-pentadecylphenol was measured at different solvent volume fractions.

Table 2. Functional Groups That Are Defined in UNIFAC and Their Replication in Two Different Structures of 3-Pentadecylphenol

subgroup	condition I (with benzene ring)	condition II (without benzene ring)
ACOH	1	0
$ACCH_2$	1	0
ACCH	4	0
CH_3	1	1
OH	0	1
CH_2	13	14

2.3. Characterization Methods. *2.3.1. X-ray Powder Diffraction (XRPD).* X-ray powder diffraction was conducted with XRPD equipment (Rigaku, Miniflex) using Cu K_α radiation. The scan angle was 5−40° with the step angle of 0.05°.

2.3.2. Differential Scanning Calorimetry (DSC). Thermal analysis was conducted by differential scanning calorimetry (DSC, Mettler Toledo, Chicago, IL). A sample of 3−5 mg was prepared in a covered 40 μL aluminum crucible with a hole in the lid to allow venting. A heating rate of 1 °C/min was employed, and the temperature range was set from 20 °C to 120 °C. A N_2 flow was used on the crucible at a rate of 100 mL/min. The calibration was performed using indium.

3. RESULTS AND DISCUSSION

3.1. Solubility Prediction of 3-Pentadecylphenol. To model the solubility of 3-pentadecylphenol using the UNIFAC, the chemical structure of the solute was broken down to functional groups. For 3-pentadecylphenol, the alkyl chain length is much longer than the benzene ring. The prediction was based on two functional group arrangements: (1) the structure with a benzene ring and (2) the structure without a benzene ring. The functional groups for the two conditions are shown in Table 2. The results of characterization using XRPD and DSC instruments are shown in Figure 2.

The solubility of 3-pentadecylphenol in the solvents was calculated using the method explained previously. The results of the prediction using the UNIFAC are shown in Figures 3 and 4. For the NRTL-SAC model, the estimation using two solvents is shown in Figure 3 and the prediction is shown for other solvents in Figures 3and 4. Except for butanol and ethanol, in which the two conditions (with and without benzene ring) have nearly the same prediction, the other four solvents exhibit a big difference between the solubility predictions for the two conditions. For acetone with the inclusion of the benzene ring, the prediction is very

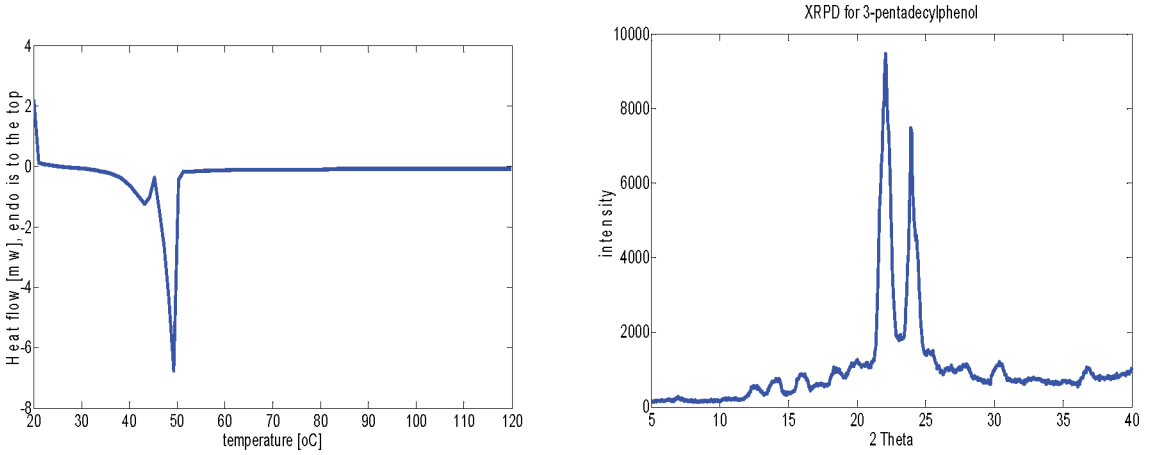


Figure 2. Characterization of recrystallized 3-pentadecylphenol with DSC (left) and XRPD (right).

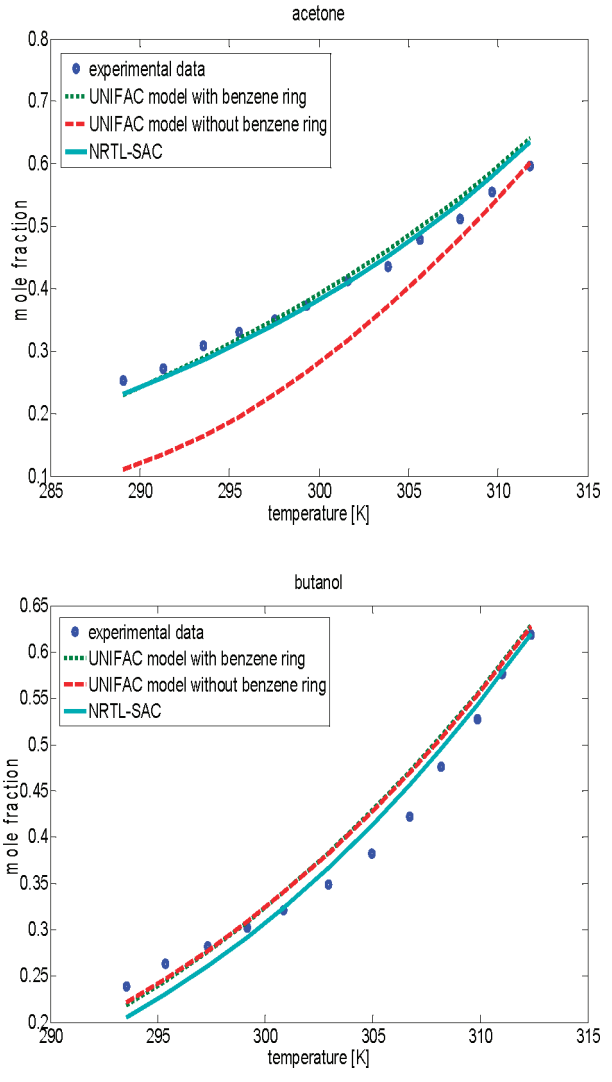


Figure 3. Experimental points of the solubility of 3-pentadecylphenol in two solvents and their curve of estimation using NRTL-SAC model. The curves of UNFAC are also shown here for comparison.

well. For ethyl acetate, tetrachloromethane, and to some extent toluene the assumption of no benzene ring shows better prediction. The reason for the different behavior in the two proposed structures is mainly due to the electron-rich nature of the benzene ring. When the benzene ring is present, the electron cloud on the ring facilitates interaction better with some polar solvents such as acetone. On the other hand, when we assume no benzene ring in the structure, the molecule is a long alkyl chain with an OH group at the terminus, which facilitates interaction with nonpolar or less polar solvents.

For the NRTL-SAC model prediction, we need to find the four segment numbers for the chemical compound and then use the segment values to predict the solubility of the compound in other solvents. We selected 1-butanol and acetone for parameter estimation and optimized the values of segment numbers which are shown in Table 3.

Using the obtained segment numbers for the compound, the NRTL-SAC model predicted solubility in the other solvents (Figure 4). The results show that except for toluene the solubility in the other three solvents were predicted well. For all four solvents, the NRTL-SAC model resulted in better prediction compared to the UNIFAC. As it is evident from Table 3, the polar attractive segment is zero, which implies that the molecule has a repulsive nature toward polar solvent. This is because of the electron-rich part of the benzene ring. As mentioned previously, about 84% of the total volume of the molecule is occupied by the alkyl group and the remaining is the phenolic part, which allows good interaction with other polar solvents. That is why the prediction for toluene has some deviation from the experimental results.

For the system of IPA—water, the addition of water to the pure IPA increases the solubility to a certain value at constant temperature. With more addition of water to the mixture, the solubility decreases and approaches zero as the volume fraction of water goes to unity in the mixture (Table 4). The maximum value of solubility is predicted by the NRTL-SAC model (Figure 5).

3.2. Solubility Prediction of Lovastatin. From the chemical structure of lovastatin, the functional groups in the UNIFAC model are listed in Table 5. The experimental data were obtained from the literature.[21,22] The functional groups are shown schematically in Figure 6. Using the obtained group functions and their replicates in the chemical structure, we modeled the thermodynamic equilibrium of lovastatin in alcohol and acetate group solvents.

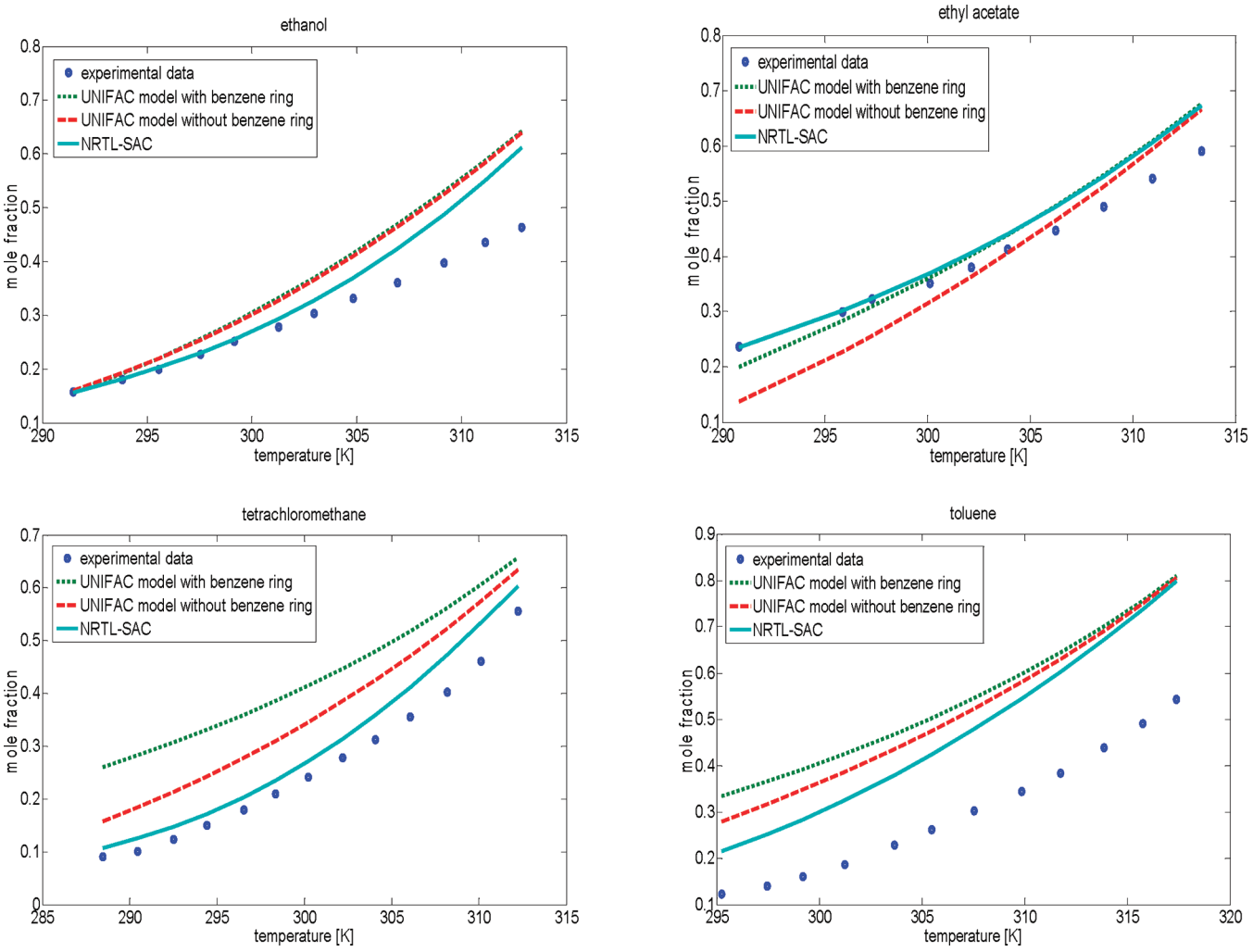


Figure 4. Experimental and predicted solubility for 3-pentadecylphenol in solvents using UNIFAC and NRTL-SAC models.

Table 3. Optimized Segment Numbers for the Model Compounds Using the NRTL-SAC Method

component	X (hydrophobic)	Y− (polar attractive)	Y+ (polar repulsive)	Z (hydrophilic)
3-pentadecylphenol	0.674	0.000	0.571	0.398
lovastatin	1.175	0.000	0.548	0.882
valsartan	0.000	0.946	0.000	0.539

From the acetate group, methyl acetate and ethyl acetate were selected, and from the alcohol group, ethanol and 1-pentanol were chosen for estimation of the NRTL-SAC segment numbers. We modeled the solubility of lovastatin in solvents in three ways:

- Parameter estimation with two solvents from the alcohol group only and solubility prediction for all the remaining solvents.
- Parameter estimation with two solvents from the acetate group only and prediction of solubility for the remaining solvents.
- Parameter estimation with two solvents from alcohol and two solvents from acetate group and solubility prediction for the other remaining solvents.

The summary of the methods are shown in Table 6. For the first estimation (method I), we selected two alcohols from the set of five alcohols.[21] Our selection was based on the alkyl chain to consider the effect of its length on the parameter estimation. Therefore, we used ethanol and 1-pentanol for parameter estimation. With the adjusted value for segment numbers that were generated for the alcohol group, we then predicted the solubility of lovastatin in the remaining solvents.

In Figure 7 the experimental data for four selected solvents with the predicted curves for NRTL-SAC and UNIFAC are shown. From those four curves in each plot, one is for the UNIFAC prediction method and the rest are for the NRTL-SAC model. The curve of "NRTL-SAC" results in the best prediction of the data. The other methods are also shown for comparison. The next possible way of estimation (method II) can be on the acetate group. In this group there are six acetate solvents plus acetone as a basic solvent.[22] Two solvents were selected from the group (methyl and ethyl acetate) for parameter estimation. With the experimental values of these solvents, the error between the experimental and estimated values was minimized. The segment numbers that were generated here were used to predict the solubility of lovastatin in the remaining

dx.doi.org/10.1021/ie201344k |*Ind. Eng. Chem. Res.* 2012, 51, 464–473

Table 4. Experimental Solubility of 3-Pentadecylphenol in Pure IPA and Its Mixtures with Water

	15 °C	20 °C	25 °C	30 °C	35 °C
pure IPA	0.7072 ± 0.0631	0.9436 ± 0.0804	1.4109 ± 0.1233	2.2355 ± 0.1011	3.8502 ± 0.4150
80 vol % IPA	0.7635 ± 0.0659	0.9755 ± 0.1213	1.6026 ± 0.1519	2.3653 ± 0.2608	4.4036 ± 0.2307
50 vol % IPA	0.2302 ± 0.0286	0.3980 ± 0.0419	0.6704 ± 0.0681	1.2739 ± 0.1396	3.1857 ± 0.2867
20 vol % IPA	0.0152 ± 0.0040	0.0188 ± 0.0031	0.0235 ± 0.0045	0.0273 ± 0.0028	0.0329 ± 0.0037

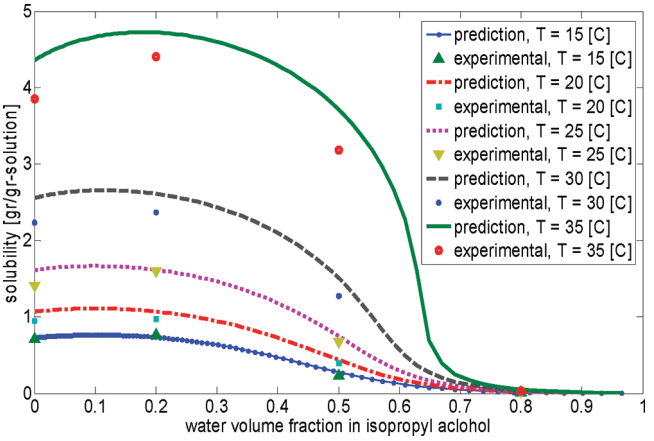


Figure 5. Experimental and predicted values of the solubility data of 3-pentadecylphenol in IPA−water mixture at different temperatures.

solvents. The final part for estimation (method III) deals with four solvents from both the alcohol and acetate groups. For all three cases, the ARD (average relative deviation) was calculated to compare the three methods, qualitatively. The results and summary of the ARD are shown in Table 6. The values of optimized segment numbers for each method are shown in Table 7.

From the four methods of predicting solubility presented in Table 6, method III is the best for describing and predicting the solubility of lovastatin. The next best is method II, which is based on estimation of the two acetate solvents. The UNIFAC method has the highest average deviation from the experimental values. In the second column of Table 6, the ARD is based on only the deviation of predicted solubility from experimental solubility of alcohols. The prediction by the UNIFAC model and method I (which was based on parameter estimation of alcohols only) are nearly the same. In column three of Table 6, the ARD is based on the acetate solvent data. From Table 7, the values of hydrophobic and polar attractive segments are very near to those of method II (acetate estimation). However, this behavior is different for polar repulsive and hydrophilic segments, which are nearly the arithmetic average of methods I and II. Method III was optimum to minimize the total average deviation and hence best describes the solubility of lovastatin in alcohols and acetate solvents. Finally, the results show that the UNIFAC method did not predict the solubility data as well as the NTRL-SAC method.

3.3. Solubility Prediction of Valsartan. For the solubility prediction of valsartan, we used the experimental data of Liu et al.[23,24] to estimate the adjustable parameters for NRTL-SAC. The chemical structure of valsartan is shown in Figure 1. One group contains four nitrogen atoms in a pentagonal ring. To the best of our knowledge and from

Table 5. Functional Groups That Have Been Used in Modeling with the UNIFAC Method

subgroup	number of replicates
CH_3	4
CH_2	6
CH	6
C═C	2
CH−O	1
C−O−C	2
$COCH_3$	2

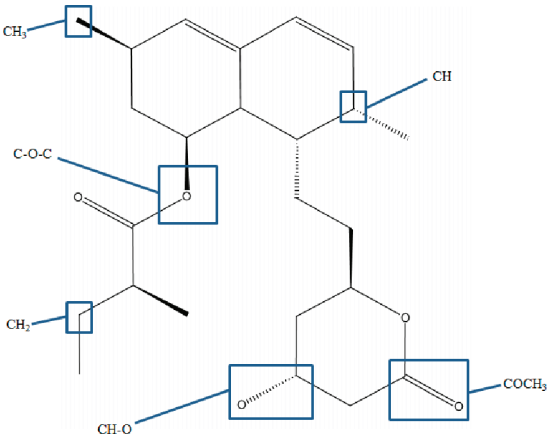


Figure 6. Schematic view of functional groups presented in lovastatin.

Table 6. Average Relative Deviation for UNIFAC and NRTL-SAC in Three Methods

method	ARD based on alcohol solvents	ARD based on acetate solvents	ARD considering both alcohol and acetate solvents
NRTL-SAC (method III)	0.1946	0.1658	0.2401
NRTL-SAC (method II)	0.4980	0.1305	0.3306
NRTL-SAC (method I)	0.4735	0.4841	0.3499
UNIFAC	0.4795	0.7064	0.6153

the thermodynamic references and newly published articles, this group does not have any UNIFAC parameter. This is a weakness of the UNIFAC model in the face of new and complex chemical and pharmaceutical materials. To predict the pure solvent solubility and solvent + antisolvent solubility, three solubility data sets were selected for estimation,

dx.doi.org/10.1021/ie201344k |Ind. Eng. Chem. Res. 2012, 51, 464–473

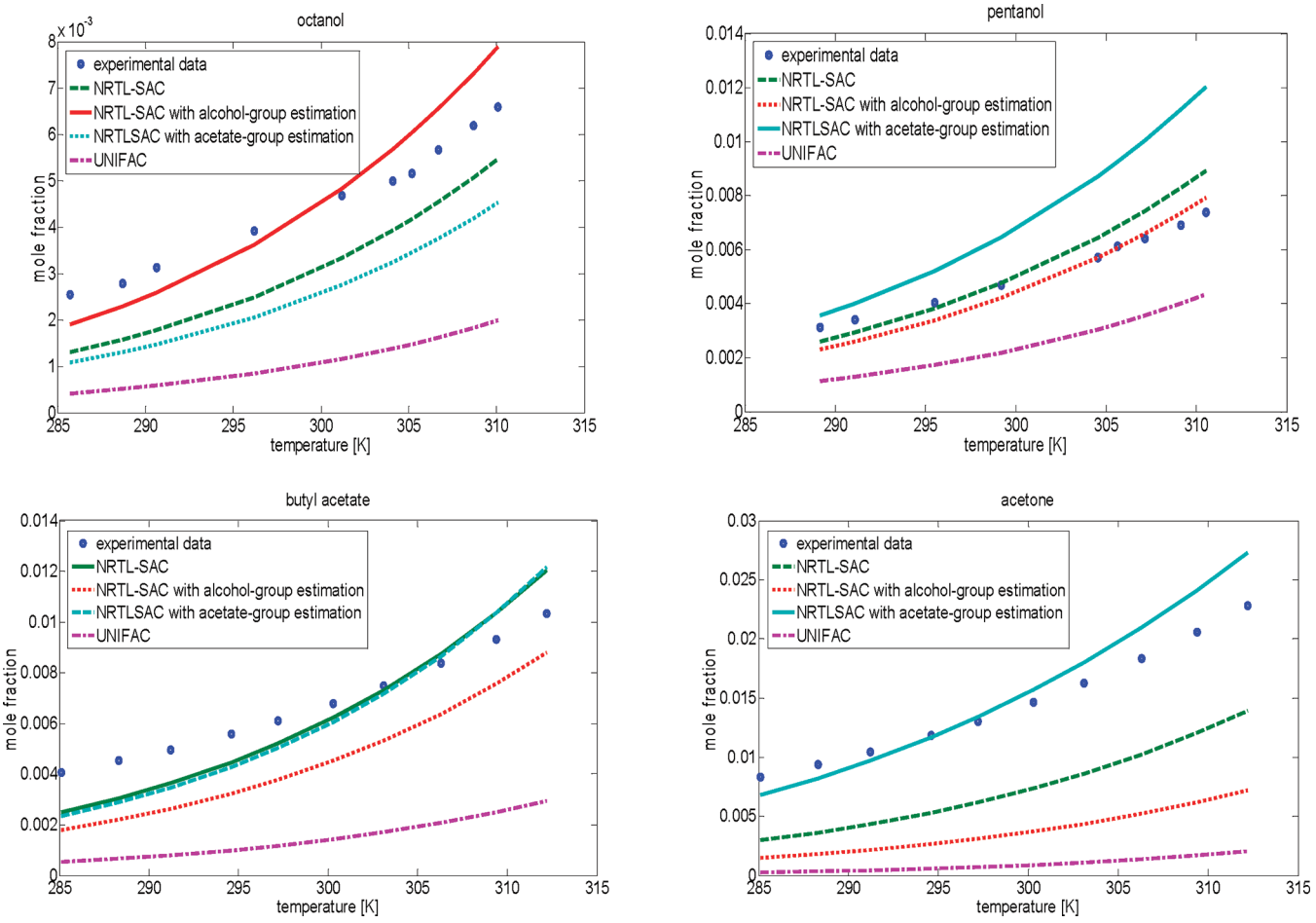


Figure 7. Experimental data with the predictions of solubility of lovastatin in four selected solvents using three methods of the NRTL-SAC and UNIFAC methods.

Table 7. Optimized Segment Numbers for Each Method Used for the NRTL-SAC Model

method	X (hydrophobic)	Y− (polar attractive)	Y+ (polar repulsive)	Z (hydrophilic)
I	0.820	0.366	0.484	0.277
II	1.154	0.001	0.831	1.393
III	1.175	0.000	0.548	0.882

and the remaining data were predicted using the adjusted segment numbers.

The solubility data in methyl acetate, DMF, and dichloromethane were used for parameter estimation. After optimizing the model and obtaining the segment numbers, we predicted the solubility of *n*-butyl acetate and acetonitrile. The ARD for this prediction is 0.3843, which is very promising for prediction purposes. In the two graphs in Figure 8, the square points are the coordination of the point that shows the experimental solubility (*x*-axis) and the modeled data (*y*-axis). For the case of solvent mixture,[24] the optimized segment numbers (Table 3) were used to predict the solubility. Here the ethyl acetate and hexane are used as solvent and antisolvent, respectively. With the assumption that the molar ratio of antisolvent to the solute-free solution is M_w, then the mole fractions of each species in the mixture will be:

$$\begin{cases} x_1 = \text{mole fraction of solute} \\ x_2 = (1 - x_1)(1 - M_w) \\ x_3 = \dfrac{M_w}{(1 - M_w)} x_2 \end{cases}$$

Note that if x_1 is the mole fraction of the solute in the solution, then having the molar ratio (or mass ratio) of antisolvent to solution (solute-free basis) allows the determination of the mole fractions of the two solvents in the system.

Therefore, the mole fractions of all of the three compounds are functions of the solute mole fraction, which is a function of activity coefficient. As a result, for this type of calculation, the trial and error method must be used to converge to a unique value of x_1. Having this included in the modeling code, we started with the trial and error procedure to attain the solubility for each antisolvent to solvent ratio over a range of temperature. The curves of solubility prediction for pure ethyl acetate to the solution of 49.44 mol % and the experimental values are shown in Figure 9.

From the optimized segment numbers for valsartan, we see that this drug does not have the hydrophobic and polar-repulsive parts in its chemical structure. On the other hand, it has polar attractive and hydrophilic segments, which can be rationalized by the carboxyl group (COOH) and also the hydrogen head. Figure 9 shows

dx.doi.org/10.1021/ie201344k |*Ind. Eng. Chem. Res.* 2012, 51, 464–473

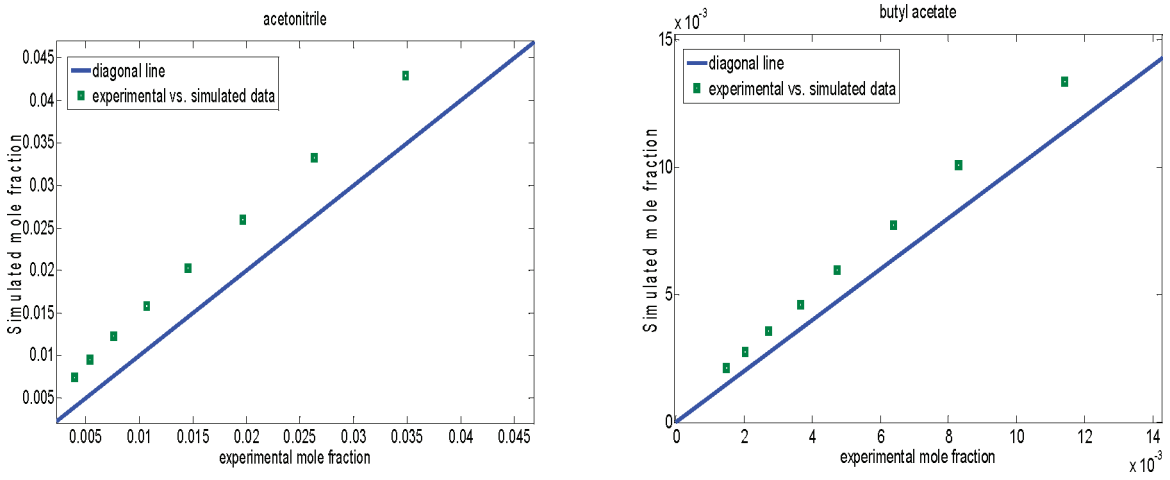


Figure 8. Predicted and experimental values of valsartan in two solvents: left, acetonitrile; right, butyl acetate.

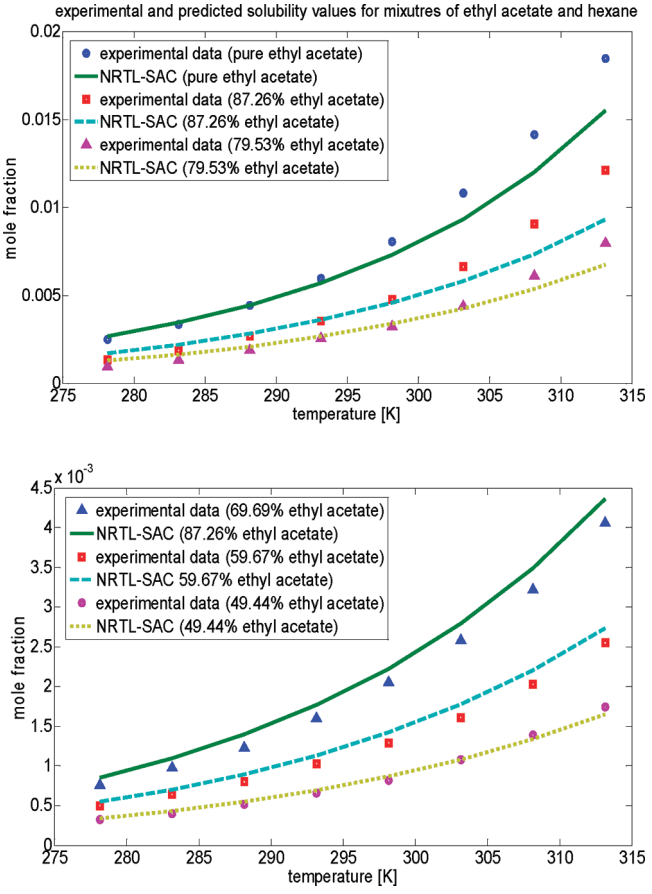


Figure 9. Solubility prediction of valsartan in three ethyl acetate−hexane mixtures for different mole percents of ethyl acetate as a solvent.

satisfactory prediction of the model at different mixture ratios and temperatures. This shows the robustness of the NRTL-SAC model to predict the ternary systems of solute−solvent−antisolvent.

4. CONCLUSIONS

In this study we used two predictive methods for the phase behavior of solid−liquid equilibrium (SLE) systems. There are many thermodynamic models that can describe the solubility of a pharmaceutical or a chemical compound in a solvent or mixture of solvents. However, the correlative models have some restrictions such as the need to have a wide range of experimental data to find the binary interaction parameters (i.e., Wilson). The UNIFAC model described the solubility of three selected solutes in the solvents with a reasonable accuracy. We used some modifications for the optimization of adjustable parameters in the NRTL-SAC model. NRTL-SAC showed very good accuracy for pure solvent systems. Also, for the mixture of solvents, we were not able to use the UNIFAC model. Using a few experimental data points for segment number estimation of NRTL-SAC method allowed us to predict solubility in pure and also mixed solvents, successfully. For the case of solvent−antisolvent solubility prediction, 3-pentadecylphenol phase behavior in the mixture of IPA−water was best predicted by the NRTL-SAC model. The maximum point of solubility was predicted and verified by experimental data points. The NRTL-SAC model predicted the solubility of valsartan in a mixture of hexane and ethyl acetate, successfully. The accuracy and robustness of the model encourages us to optimize the yield of crystallization and obtain the best solvent mixture for each pharmaceutical compound.

AUTHOR INFORMATION

Corresponding Author

*E-mail: srohani@uwo.ca.

ACKNOWLEDGMENT

The authors thank the Natural Sciences and Engineering Research Council (NSERC) of Canada for financial support of the project.

REFERENCES

(1) Kokitkar, P. B.; Plocharczyk, E. Modeling drug molecule solubility to identify optimal solvent systems for crystallization. *Org. Process Res. Dev.* **2008**, *12*, 249.

(2) Liu, W.; Dang, L.; Black, S.; Wei, H. Solubility of Carbamazepine (Form III) in Different Solvents from (275 to 343) K. *J. Chem. Eng. Data* **2008**, *53*, 2204.

(3) Chen, C.; Crafts, P. A. Correlation and prediction of drug molecule solubility in mixed solvent systems with a Nonrandom two-

liquid segment activity coefficient (NRTL-SAC) model. *Ind. Eng. Chem. Res.* **2006**, *45*, 4816.

(4) Margules, M. Uber die Zusammensetzung der gesattigten Dampfe von Mischungen, Sitzungsber Akad. Wiss. *Math.-Naturwiss. Klasse* **1985**, *104*, 1243.

(5) Wilson, G. M. Vapor-liquid equilibrium. XI. A new expression for the excess free energy of mixing. *J. Am. Chem. Soc.* **1964**, *86*, 127.

(6) Peng, D.-Y. Extending the Van Laar Model to Multicomponent Systems. *Open Thermodyn. J.* **2010**, *4*, 129.

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

(8) Abrams, D. S.; Prausnitz, J. M. Statistical thermodynamics of liquid mixutres: A new expression for the excess Gibbs energy of partly or completely miscible systems. *AIChE J.* **1975**, *21*, 116.

(9) Matsuda, H.; Matsumoto, S.; Kaguragi, K.; Kurihara, K.; Tochigi, K.; Tomono, K. Prediction of solubilities of pharmaceutical compounds in water + co-solvent systems using an activity coefficient model. *Fluid Phase Equilib.* **2010**, *290*, 153.

(10) Tassios, D. Presented at the 62nd Annual Meeting of the American Institute of Chemical Engineers, Washington, DC, 1969.

(11) Gmehling, J.; Onken, U. *Vapor-liquid Equilibrium Data Collection*; DECHEMA: Frankfurt, 1977.

(12) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *Am. Inst. Chem. Eng.* **1975**, *21*, 1086.

(13) Sheikhzadeh, M.; Rohani, S.; Taffish, M.; Murad, S. Solubility analysis of buspirone hydrochloride polymorphs: Measurements and prediction. *Int. J. Pharm.* **2007**, *338*, 55.

(14) Gracin, S.; Brinck, T.; Rasmuson, A. C. Prediction of Solid Organic Compounds in Solvents by UNIFAC. *Ind. Eng. Chem. Res.* **2002**, *41*, 5114.

(15) Kan, A. T.; Tomson, M. B. UNIFAC Prediction of Aqueous and Nonaqueous Solubilities of Chemicals with Environmental Interest. *Environ. Sci. Technol.* **1996**, *30*, 1369.

(16) Chen, C. C.; Song, Y. Solubility Modeling with a Nonrandom Two-Liquid Segment Activity Coefficient Model. *Ind. Eng. Chem. Res.* **2004**, *43*, 8354.

(17) Chen, C. C.; Song, Y. Extension of Nonrandom Two-Liquid Segment Activity Coefficient Model for Electrolytes. *Ind. Eng. Chem. Res.* **2005**, *44*, 8909.

(18) Prausnitz, J. M.; Lichtenthaler, R. N.; de Azevedo, E. G. *Molecular Thermodynamics of Fluid-Phase Equilibria*; Prentice Hall Inc.: Upper Saddle River, NJ, 1986.

(19) Matlab software help documents, version 7.6.0.324 (R2008a).

(20) Mao, Z. B.; Luo, T. L.; Cui, T.-B.; Wang, Y.; Liu, Guo-Ji. Solubility of 3-Pentadecylphenol in Ethanol, 1-Butanol, Toluene, Acetone, Tetracholomethane, and Ethyl Acetate. *J. Chem. Eng. Data* **2010**, 543–546.

(21) Nti-Gyabaah, J.; Chmielowski, R.; Chan, V.; Chiew, Y. C. Solubility of lovastatin in a family of six alcohols: Ethanol, 1-propanol, 1-butanol, 1-propanol, 1-hexanol, and 1-octanol. *Int. J. Pharm.* **2008**, *359*, 111.

(22) 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*, 2060.

(23) Liu, Y.; Wang, J.; Wang, X.; Liu, P.; Pang, F. Solubility of Valsartan in different organic solvents and ethanol + water binary mixtures from (278.15 to 313.15) K. *J. Chem. Eng. Data* **2009**, *54*, 986.

(24) Liu, Y.; Wang, J.; Wang, X.; Pang, F. Solubility of Valsartan in Ethyl Acetate + Hexane binary mixtures from (278.15 to 313.15) K. *J. Chem. Eng. Data* **2009**, *54*, 986.

(25) Qu, H.; Kultanen, M. L.; Kallas, J. Solubility and stability of anhydrate/hydrate in solvent mixtures. *Int. J. Pharm.* **2006**, *321*, 101.

(26) Mirmehrabi, M.; Rohani, S.; Murthy, K. S. K.; Radatus, B. Solubility, dissolution rate and phase transition studies of ranitidine hydrochloride tautomeric forms. *Int. J. Pharm.* **2004**, *282*, 73.

(27) Getsoian, A.; Lodaya, R. M.; Blackburn, A. C. One-solvent polymorph screen of carbamazepine. *Int. J. Pharm.* **2008**, *348*, 3.

(28) Trifkovich, M.; Rohani, S. Polymorphic Generation through Solvent Selection: Ranitidine Hydrochloride. *Org. Process Res. Dev.* **2007**, *11*, 138.