

Modeling Drug Molecule Solubility to Identify Optimal Solvent Systems for Crystallization

Prashant B. Kokitkar* and Ed Plocharczyk

Eli Lilly and Company, Indianapolis, Indiana, U.S.A.

Chau-Chyun Chen

Aspen Technology, Inc., Cambridge, Massachusetts, U.S.A.

Abstract:

The utility of the recently developed nonrandom two-liquid segment activity coefficient (NRTL-SAC) model has been reported for solvent selection in support of industrial crystallization process design. In this paper, we present a recent successful application with NRTL-SAC to screen solvents for a crystallization medium with the goal to maximize API solubility and to minimize solvent usage. The NRTL-SAC model parameters for the molecule in development are first identified from a minimal set of solubility experiments in selected solvents. We then perform numerous *in silico* virtual experiments to explore the solubility behavior of the molecule in other pure solvents and mixed solvents. The modeling results suggested optimal solvent systems for the crystallization medium which are validated in physical laboratories and chosen for process scale-up. This study demonstrated the effectiveness of the NRTL-SAC model and supports its use as a tool in drug development.

Introduction

The chemical synthesis of a drug molecule often involves 3–10 synthetic steps in series, each of which can be composed of several unit operations. The selection of solvent, solvent mixture, operating temperature, etc. can be a daunting task in a complex synthesis. Obtaining physical properties measurements for unique chemical entities in many solvents is costly and time-consuming. The absence of experimental data for new chemical entities and their precursors coupled with the lack of robust predictive models for solubilities of complex pharmaceutical molecules have hampered solvent selection and the subsequent process design and optimization.

Chen and Song¹ recently developed a nonrandom two-liquid segment activity coefficient model (NRTL-SAC) which provides a simple and practical molecular thermodynamic framework to correlate and predict drug solubility in pure and mixed solvents. The NRTL-SAC modeling approach requires scientists and engineers to probe and characterize the molecular interaction characteristics of a drug molecule by taking solubility measurements in a few selected reference solvents which are then used to identify dimensionless hydrophobicity, hydrophilicity, polar-

ity, and solvation parameters for the molecule. Given these molecular parameters, one can then perform robust qualitative prediction on the drug solubility in hundreds of commonly used pure solvents and their solvent mixtures for known polymorphs. The solubility model can further be used in process simulators in support of modeling, simulation, design and optimization of pharmaceutical manufacturing processes.

The utility of the model for efficient development and fundamental understanding of pharmaceutical crystallization processes has been reported in the industry.^{2,3} The applicability of model has also been extended to organic electrolytes.⁴ In this paper, we report a recent successful application with NRTL-SAC to screen solvents for a crystallization medium with the goal to maximize API solubility and to minimize solvent usage.

The drug candidate under evaluation in this case had extremely low solubility in initial solvent screens; the highest solubility obtained was approximately 8 mg/mL. While low solubility is desirable for product recovery, it is undesirable for impurity rejection, solvent usage, cycle time, and processing volume required. Crystallization typically involves moving a system from one with high solubility to low solubility in a controlled fashion; the process development effort focuses on trying to find what “controlled fashion” will provide the purity, crystal form, and crystal habit required. With such low solubility, a practical “starting point” for crystallization development was not apparent. The rapid prediction of the solubility of the drug candidate in many solvents, solvent mixtures, and temperatures provides viable conditions for crystallization which facilitate focused development efforts on solvent systems and operating conditions that were required to obtain sufficient solubility.

Solubility Modeling and NRTL Segment Activity Coefficient Model. The solubility of a solid organic solute can be approximated by the van't Hoff equation:^{1,2}

- (2) Tung, H.-H.; Viankaval, N.; Tabora, J.; Bakken, D.; Chen, C.-C. Prediction of Pharmaceuticals Solubility via NRTL-SAC and COSMO-SAC. *J. Pharm. Sci.* **2000**, September 4. Published online.
- (3) Crafts, P. A. The Role of Solubility Modelling and Crystallization in the Design of Active Pharmaceutical Ingredients. In *Chemical Product Design: Toward a Perspective through Case Studies*; Ng, K. M., Gani, R., Dam-Johansen, K., Eds.; Elsevier: Dordrecht, The Netherlands, 2007; Chapter 2.
- (4) Chen, C.-C.; Song, Y. Extension of Nonrandom Two-Liquid Segment Activity Coefficient Model for Electrolytes. *Ind. Eng. Chem. Res.* **2005**, *44*, 8909–8921.

* Author for correspondence. E-mail: pbk@lilly.com.

(1) Chen, C.-C.; Song, Y. Solubility Modeling with Nonrandom Two-Liquid Segment Activity Coefficient Model. *Ind. Eng. Chem. Res.* **2004**, *43*, 8354–8362.

$$\ln x_I^{\text{SAT}} = \frac{\Delta H_{\text{fus}}}{RT_m} \left(1 - \frac{T_m}{T} \right) - \ln \gamma_I^{\text{SAT}} \quad (1)$$

where x_I^{SAT} is the mole fraction of solute I dissolved in the solvent phase at saturation, ΔH_{fus} is the enthalpy of fusion of the solute, R is the ideal gas constant, T is the temperature (K), T_m is the melting point of the solute, and γ_I^{SAT} is the activity coefficient of the solute in solution at saturation. ΔH_{fus} and T_m vary with polymorphic forms of the solute. Given a polymorph and a temperature, the solute solubility is only a function of its activity coefficient in solution. In other words, the activity coefficient of the solute in solution determines the solute's solubility as the solvent composition changes. "Ideal solubility" corresponds to the saturation condition when activity coefficient is unity. For low to medium pressure conditions, pressure has relatively weak influence on the solubility, and it is excluded from this study.

Numerous thermodynamic models have been proposed in the literature to correlate or predict activity coefficients.¹ Popular semiempirical correlative models, such as van Laar, Wilson, NRTL, or UNIQUAC, require identification of binary interaction parameters from phase equilibrium data for each of the solute–solvent, solute–solute, and solvent–solvent binary mixtures. Such phase equilibrium data are rarely available for API molecules; and consequently, these correlative models find very limited use in pharmaceutical process design.

The predictive UNIFAC model requires only chemical structure information for the solutes and solvents. Unfortunately, UNIFAC fails for API molecules for which either the UNIFAC functional groups are undefined or the functional group additivity rule becomes invalid.¹ Recent developments in computational chemistry yielded COSMO-RS⁵ and COSMO-SAC⁶ predictive models that represent promising alternatives to UNIFAC. However, the predictive powers of UNIFAC and COSMO-based models remain inadequate.^{1,2}

The Hansen model has been the most widely used solubility model in the pharmaceutical industry. Incorporating the "like dissolves like" concept, the model is used as a guide to help chemists and engineers explain API solubility behavior based on molecule-specific solubility parameters. However, due to its oversimplistic assumptions the model has had very limited practical use in the prediction of drug molecule solubility.

The NRTL-SAC model is a modification of the original NRTL (non-random two-liquid) and polymer NRTL models for systems with oligomers and polymers. The combinatorial term is calculated from the Flory–Huggins approximation for combinatorial entropy of mixing, while the residual term is set equal to the sum of the local composition (lc) interaction contribution for each segment as shown in eqs 2 and 3. This segment term is further calculated from the segmental activity coefficient, as shown in eqs 4 and 5.

$$\ln \gamma_I = \ln \gamma_I^C + \ln \gamma_I^R \quad (2)$$

$$\ln \gamma_I^R = \ln \gamma_I^{lc} = \sum_m r_{m,I} [\ln \Gamma_m^{lc} - \Gamma_m^{lc,I}] \quad (3)$$

$$\ln \Gamma_m^{lc} = \frac{\sum_j x_j G_{jm} \tau_{jm}}{\sum_k x_k G_{km}} + \sum_l \frac{x_l G_{ml}}{\sum_k x_k G_{kl}} \left(\tau_{ml} - \frac{\sum_k x_k G_{kl} \tau_{kl}}{\sum_k x_k G_{kl}} \right) \quad (4)$$

$$\ln \Gamma_m^{lc,I} = \frac{\sum_j x_{j,I} G_{jm} \tau_{jm}}{\sum_k x_{k,I} G_{km}} + \sum_l \frac{x_{l,I} G_{ml}}{\sum_k x_{k,I} G_{kl}} \left(\tau_{ml} - \frac{\sum_k x_{k,I} G_{kl} \tau_{kl}}{\sum_k x_{k,I} G_{kl}} \right) \quad (5)$$

with

$$x_j = \frac{\sum_I x_I r_{j,I}}{\sum_I \sum_i x_I r_{i,I}} \quad (6)$$

$$x_{j,I} = \frac{r_{j,I}}{\sum_i r_{i,I}} \quad (7)$$

where I is the component index, i, j, k, l, m are the segment species index, x_I is the mole fraction of component I , x_j is the segment-based mole fraction of segment species j , $r_{m,I}$ is the number of segment species m contained only in component I , Γ_m^{lc} is the activity coefficient of segment species m , and $\Gamma_m^{lc,I}$ is the activity coefficient of segment species m contained only in component I . G and τ in eqs 3 and 4 are local binary quantities related to each other by the NRTL nonrandomness factor parameter α :

$$G_{jm} = \exp(-\alpha \tau_{jm}) \quad (8)$$

In essence, the NRTL-SAC model characterizes the molecules in terms of four predefined conceptual segments, i.e., $r_{m,I}$ in eq 3, with predetermined nonrandomness factor, α , and predetermined segment–segment interaction energies, τ . Specifically, for each solute and solvent molecule, NRTL-SAC describes their effective surface interactions in terms of four types of conceptual segments: hydrophobic segment, electrostatic solvation segment, electrostatic polar segment, and hydrophilic segment. The segment numbers for each molecule are measures of the effective surface areas of the molecule that exhibit surface interaction characteristics of hydrophobicity (X), solvation (Y^-), polarity (Y^+), and hydrophilicity (Z). Here the hydrophilic segment simulates polar molecular surfaces that are "hydrogen bond donor or acceptor". The hydrophobic segment simulates molecular surfaces that show aversion to forming hydrogen bonds. The electrostatic segments (Y^- and Y^+) simulate molecular surfaces that are electron pair donor or

(5) Eckert, F.; Klamt, A. A Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE J.* **2002**, *48*, 369–385.

(6) Lin, S.-T.; Sandler, S. I. A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res.* **2002**, *41*, 899–913.

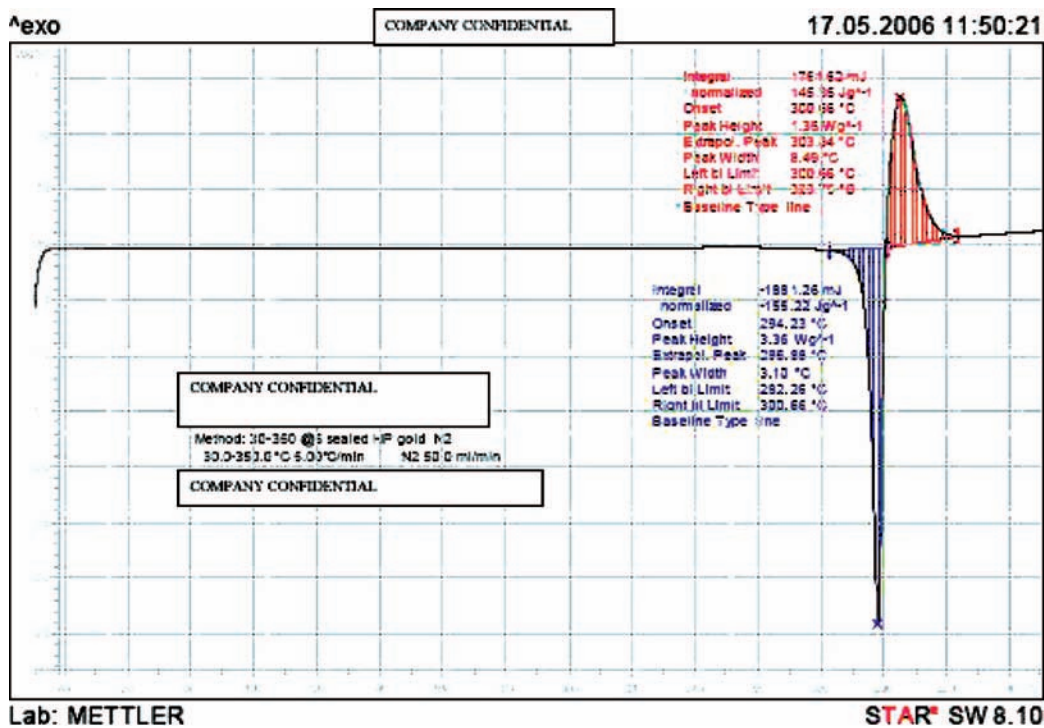


Figure 1. DSC trace of the compound showing decomposition during melting.

acceptor. The electrostatic solvation segment is attractive to the hydrophilic segment while the electrostatic polar segment is repulsive to the hydrophilic segment.

To determine the segment numbers of a solute molecule, solubility data in at least four reference solvents are needed. The parametrization is improved if a range of hydrophilic solvents, polar solvents, solvation solvents, and hydrophobic solvents are used. Once the segment numbers of the solute molecule are determined from the data, the NRTL-SAC model can then be used to predict the solute solubility in other solvents or solvent mixtures.

Model Development. Solubility data for the compound were generated in the R&D laboratory in the course of process development and were used to regress NRTL-SAC model parameters for the compound. The compound modeled was the hydrochloride salt of a secondary amine and included an aromatic amide and ether linkages. This compound had a molecular weight of 410 and melting temperature of 299 °C. A total of 10 data points in six solvents at four different temperatures were available and were used to develop the model. It was assumed that the same polymorph was used in all solubility measurements. The heat of fusion for molecular organic crystals (molecules in the crystal structure held together by noncovalent interactions such as van der Waals forces or hydrogen bonding) is readily determined by DSC. Melting of molecular organic crystals is a reversible event; such crystals usually melt at fairly low temperatures, sufficient to break the weaker forces holding the molecules together in the crystal lattice. In this case, the compound modeled was not a molecular organic crystal, but rather primarily an ionic solid (ions held together by electrostatic attraction). The electrostatic forces hold the molecular ions in the crystal lattice tightly; so tightly that the temperature required for breaking the ionic bonds holding

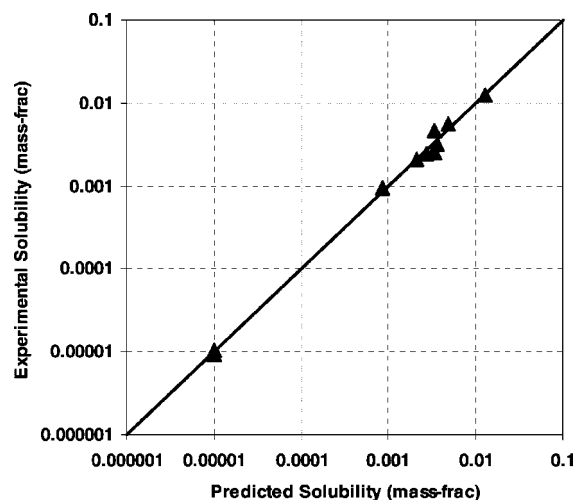


Figure 2. Depiction of model fit obtained during regression.

the crystal lattice together is high enough to actually break the covalent bonds between the atoms of the organic molecule. Ionic solids such as sodium chloride do actually melt reversibly, but the individual ions formed are atoms and not molecules; thus, ionic bonds between the ions are broken at lower temperatures than those required to break individual atoms (temperatures sufficient to initiate thermonuclear reactions). The ionic nature of the crystal lattice is evident in the DSC trace obtained during normal hazards screening for the molecule; the trace in Figure 1 indicates an endotherm at approximately 300 °C (the melting point) immediately followed by an exotherm (indicating thermal decomposition of the molecule). Since the molecule decomposes upon melting, the event is no longer reversible.

Visual observation by the analyst in an OptiMelt melting point apparatus confirmed that there was a visible melt occurring simultaneously with off-gassing and decomposition of the material. The DSC measures net heat flow into or out of the

sample, and cannot differentiate between simultaneous endothermic activity due to melting and exothermic activity due to decomposition; the DSC only measures the net of both combined effects. Since the heat of fusion was not a physical property which could be directly measured for this crystal, the heat of fusion was used as an adjustable parameter to improve model regression as discussed in the Experimental Verification section in this paper. The input solubility data for the model, obtained using HPLC, are given in Table 1 below. Solubility in hexane and heptane was below detection limit and was entered as 0.001 mg/mL for the regression. The heat of fusion estimated from a subsequent DSC run was 70.51 J/g and was used to develop the model.

Table 1. Compound solubility data used for NRTL-SAC regression

solvent name	measured solubility	temperature (°C)
water	3.4	25
water	3.36	50
DMSO	3.0	25
DMSO	5.36	50
DMSO	13.9	70
ethylene glycol	0.93	25
ethylene glycol	4.00	25
methanol	1.68	10
hexane	0.001	25
heptane	0.001	25

The model parameters were regressed using the data set in Table 1. Figure 2 shows the overall fit, which was very reasonable, with the sum of squared error (SSE) of 173.

$$SSE = \sum_{i=1}^k \sum_{j=1}^m \left(\frac{Z_{ij} - Z_{ij}^m}{\sigma_{ij}} \right)^2 \quad (9)$$

Where i is data point number, j is measured variable for a data point, k is total number of points in a data set, m is number of measured variables for a data set, Z is calculated value, Z^m is measured (experimental) value, and σ is the standard deviation. In this study, the standard deviation was set at 5% for the measured solubility data (based on recommendations of the laboratory chemist who performed solubility measurements). Table 2 lists the NRTL-SAC parameters obtained for the compound.

Table 2. Regressed NRTL-SAC parameters for the compound

parameter	value
X	0.0823
Y-	0.5261
Y+	0.5287
Z	0.2771

Once the model parameters were regressed, the model was used to conduct several different solubility modeling studies with the same polymorph. AspenTech has determined NRTL-SAC parameters for 117 solvents from available data in open literature. Some of these NRTL-SAC parameters have recently

been reported by Chen and Crafts.⁷ The solubility of the compound was calculated in 117 pure solvents, thousands of binary mixtures, and a few select ternary solvent mixtures. Each binary solvent mixture was studied at 18 different compositions varying from 0% to 100% of one component. The solubility studies were also conducted at different temperatures (at sufficiently high pressures to prevent boiling) for pure solvents and binary mixtures.

In addition to predicting solubility, the NRTL-SAC model as implemented in an Aspen Properties software package was also used to predict the formation of two liquid layers at process conditions. This information was used along with the NRTL-SAC solubility estimates in recommending promising systems for further study in the laboratory.

Results

Preliminary Studies. The top 10 promising pure solvents identified from the solubility prediction studies conducted at room temperature are given in Table 3.

Table 3. Top 10 promising pure solvents based on predicted solubility at 25 °C

solvent name	predicted solubility (mg/mL)
diethyl amine	7.74
<i>n</i> -methylacetamide	6.55
ethanol	3.80
1,2-propanediol	3.63
isopropyl alcohol	3.47
pyridine	3.41
<i>tert</i> -butyl alcohol	3.17
methanol	3.17
formamide	2.94
formic acid	2.93

Once it was clear that the highest predicted solubility at room temperature was still too low for scale-up purposes, it was decided to try higher temperatures and binary combinations in an attempt to find a combination that would give the desired solubility of 20 mg/mL or higher. Figure 3 shows the effect of temperature on the predicted compound solubility in pure DMSO with NRTL-SAC and IDEAL models. The available experimental solubility data in pure DMSO is also depicted in Figure 3 for comparison purposes. Accounting for nonideality enables the NRTL-SAC model to provide a better prediction of the solubility than the IDEAL model. As seen in the figure, a fairly high temperature of about 85 °C will be required to obtain a solubility of greater than 20 mg/mL in pure DMSO.

Due to the appreciable compound solubility in water, it was desired to study aqueous alcohol mixtures in an attempt to choose an environmentally benign solvent system for process scale-up and commercialization. Figure 4 shows the predicted solubilities of the compound in aqueous mixtures of ethanol, *n*-propanol (NPA), and isopropanol (IPA) at 90 °C. The solubility goes through a maximum with composition, and the

(7) Chen, C.-C.; Crafts, P. 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*, 4816–4824.

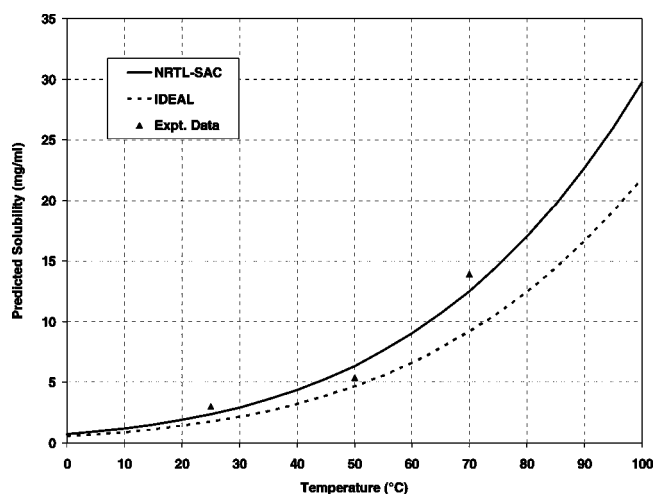


Figure 3. Compound solubilities increase with temperature, predicted with the NRTL-SAC and ideal solubility models and compared with experimental data for DMSO.

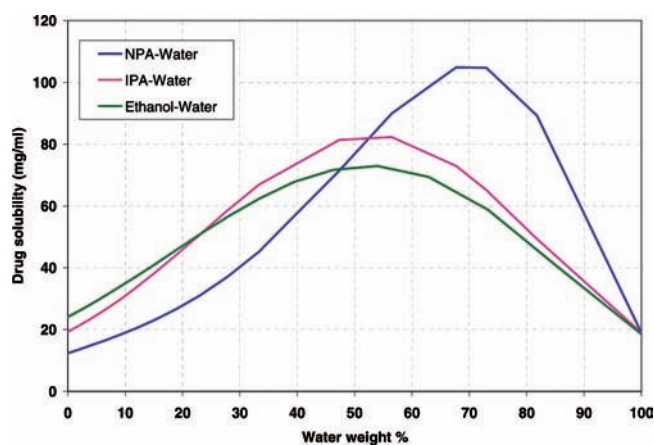


Figure 4. Predicted compound solubilities in ethanol–water mixture at 90 °C.

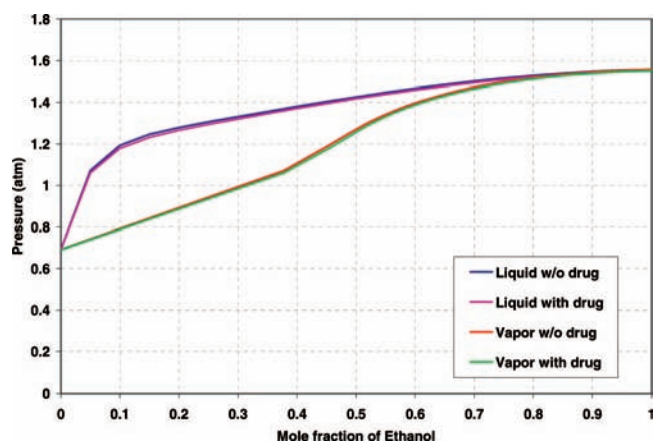


Figure 5. P_{xy} diagram for ethanol–water mixture at 90 °C.

peak solubility is obtained at about 50–70 wt % water, depending on the system.

The pressure required to use the high temperatures needed to obtain high solubility can also be easily estimated using the software. For example, the pressure profile for the ethanol–water mixture at 90 °C is obtained from a P_{xy} diagram, which is depicted in Figure 5. As seen in the diagram, the maximum pressure required to prevent boiling off is about 1.6 atm or 23.5

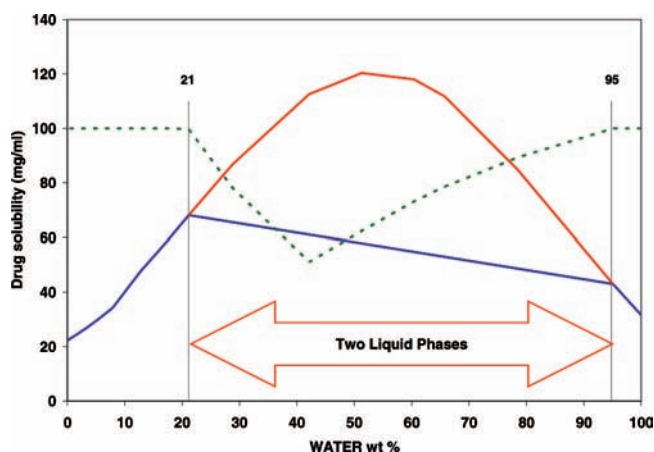


Figure 6. Predicted compound solubility in 1-butanol–water mixture at 110 °C. The dotted line represents $\beta \times 100$, which indicates liquid immiscibility.

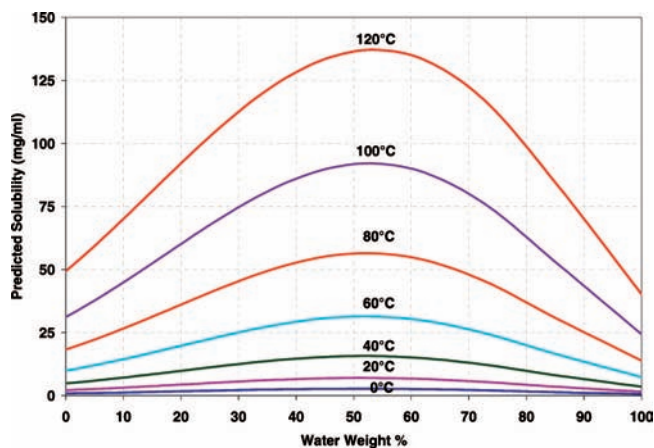


Figure 7. Effect of temperature on predicted compound solubilities in ethanol–water mixtures.

psia, which is well within the capabilities of typical processing equipment used in an R&D pilot plant. The pressure required for higher-boiling alcohols such as NPA or IPA would be even lower.

The results depicted in Figure 6 are obtained for 1-butanol–water binary mixture at elevated temperature and pressure. The dotted line in Figure 6 represents $\beta \times 100$, where β is the ratio of moles of the first liquid phase to the total moles of the liquid phase. Two phases are predicted when parameter $\beta < 1$ (i.e., $\beta \times 100 < 100$). Figure 6 shows the presence of two liquid layers over most of the composition range, and the predicted peak solubility falls in the two-phase region. The actual solubility in the two layers will of course be that given by the intersection of the solubility curve with the vertical lines marking the two-phase region. Thus, the 1-butanol–water system will not provide high compound solubility due to separation into two liquid phases.

Since the model was predicting the ethanol–water system to be promising, the solubilities of the compound in ethanol–water mixtures were further studied at different temperatures. During the calculations, the system was maintained at a sufficiently high pressure to prevent boiling off the solvents. The results are summarized in Figure 7 and show the dramatic increase in solubility with temperature, especially in the middle of the composition range.

Table 4. Top 10 binary systems with water for obtaining maximum drug solubility at 120 °C^a

Solvent1 name (R&D solvent tier)	Solvent1 wt %	Solvent2 name	Solvent2 wt %	max solubility (mg/mL)	LLE?
<i>n</i> -propyl alcohol (–)	32.27	water	67.73	172.2	TRUE
ethanol (2)	46.25	water	53.75	137.1	FALSE
tetrahydrofuran (3)	57.16	water	42.84	124.8	FALSE
pyridine (4)	84.95	water	15.05	118.5	FALSE
diethylamine (–)	64.85	water	35.15	107.3	FALSE
1,4-dioxane (4)	68.97	water	31.03	105.7	FALSE
acetone (2)	59.44	water	40.56	100.7	FALSE
<i>n</i> -methylacetamide (–)	83.91	water	16.09	98.56	FALSE
ethyl formate (–)	37.00	water	63.00	90.80	FALSE
methanol (3)	52.62	water	48.38	82.90	FALSE

^a Number in parenthesis after Solvent1 name indicates that solvent's tier in R&D's Solvent Selection Guide where available.

Automated High-Throughput Studies. A separate program was then written in Microsoft Excel VBA to automate screening of binary solvent systems. The program automatically creates all possible binary combinations from a given list of solvents, 117 in this case. The list of solvents for which NRTL-SAC parameters are available has since grown to 130. It determines compound solubilities at 18 different compositions at a given temperature, and then lists the composition giving the highest solubility for each binary mixture studied. The program also determines if two liquid phases (LLE) are predicted, at calculation temperature and pressure, over the composition range of the binary mixtures. Finally, the program sorts the list first by the presence of LLE and then by predicted solubility. The 117 pure solvents form 6786 binary combinations; 18 compositions for each binary mixture give a total of 122,148 binary mixtures to study. This set of experiments took about 5–6 h of computer time on a high-powered computer in the modeling laboratory. The machine has dual 3.2 GHz Xeon processors with 2 GB RAM, and runs 32-bit Windows XP professional.

A high-throughput screening was conducted at elevated temperatures to determine the upper range of the solubility for the compound. A temperature of 120 °C and a pressure of 7 atm were chosen for the study. Although these extreme conditions were not practical for scale-up, they provided an idea of the increase in solubility possible for the top binary systems. Also, for many systems, the pressure required to keep them from boiling off was significantly less than 7 atm and was within the capabilities of the pilot plant in R&D. The top ten binary systems for the highest compound solubility at 120 °C are given in Table 4. The number in parenthesis after Solvent1 name in the first column indicates that solvent's tier in R&D's Solvent Selection Guide. The Solvent Selection Guide provides a rough ranking of common R&D solvents from an environmental, health, and safety standpoint and is somewhat similar to the ICH classification. Tier 1 contains the most benign solvents, and tier 5 has the least desired or banned solvents. The last column in Table 4 ("LLE?") indicates the presence of two liquid phases for a given binary system. Although a few other binary systems were predicted to provide high solubility as well, they were not included in this list due to their negative environmental and/or health effects (e.g., a solubility of 92 mg/mL was predicted for diethylamine and chlorobenzene mixture).

Table 4 lists those systems that give the highest solubility and that do not form two liquid phases over the entire

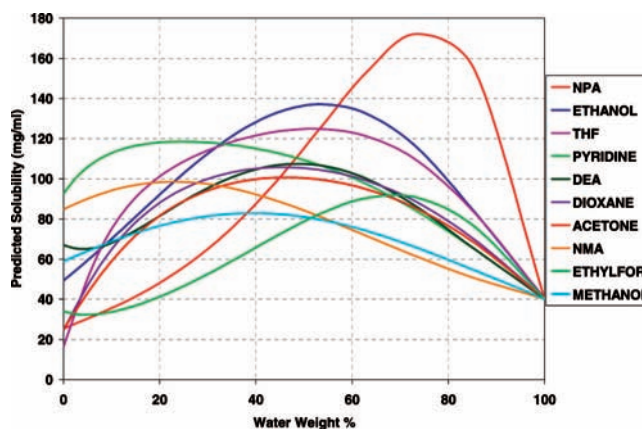


Figure 8. Predicted compound solubility in various binary mixtures with water at 120 °C.

composition range, with the exception of the first entry. For the NPA–water system listed at the beginning of the table, the NRTL-SAC model predicted the formation of two phases over a small composition range. In reality, this binary mixture is not known to form two phases and is included here due to R&D's interest in the aqueous alcohol systems for scale-up. The incorrect model prediction of two phases reflects the qualitative nature of NRTL-SAC in predicting phase behavior of solvent–solvent and solvent–solute mixtures. While NRTL-SAC correctly predicts a transition from a water-miscible alcohol (ethanol) to a water-immiscible alcohol (*n*-butanol), the model is not accurate enough to predict full miscibility of *n*-propanol in water.

The solubility over the entire composition range for these binary mixtures is depicted in Figure 8. Note that the solubility is quite low in pure THF or acetone, but it increases rapidly with the addition of water. All solubility profiles show a maximum in the range; however, the exact location of the maximum varies significantly for each system. Also, the range of solubility values over the composition range for a single binary mixture varies much from one system to another; methanol–water shows the smallest range, whereas NPA–water shows the highest range of values at the temperature studied.

Experimental Verification. Laboratory experiments were conducted to determine experimental solubility for several aqueous solvent mixtures at 50 °C and to evaluate selected model-recommended systems. Figures 9 and 10 depict the comparisons between model predictions and experimental data for the compound solubilities in six aqueous binary mixtures

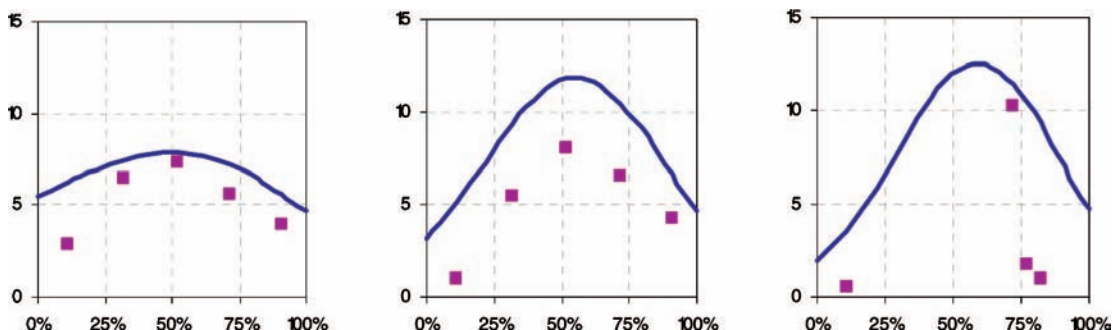


Figure 9. Comparison of experimental solubility data with NRTL-SAC model predictions at 50 °C. Starting from left, graphs are shown for DMSO–water, IPA–water, and DMF–water mixtures. Solubility is expressed in mg/g for all graphs, and the solvent composition is wt % water.

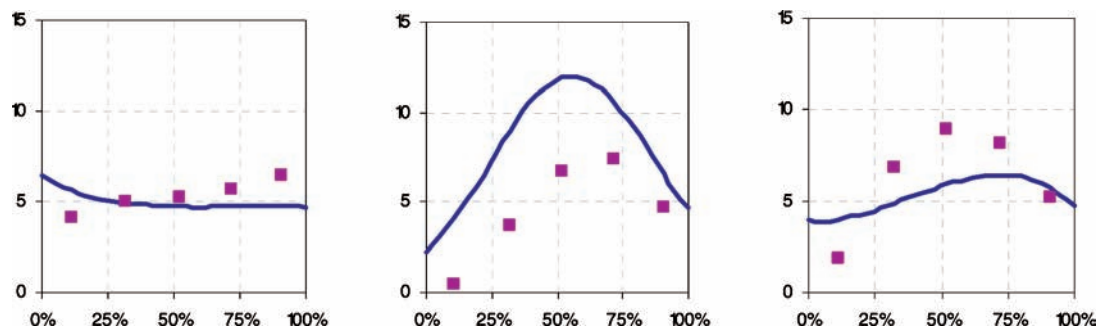


Figure 10. Comparisons of experimental solubility data with NRTL-SAC model predictions at 50 °C. Starting from left, graphs are shown for methanol–water, ethanol–water, and NPA–water mixtures. Solubility is expressed in mg/g for all graphs, and the solvent composition is wt % water.

at 50 °C. The experimental data compared favorably with NRTL-SAC model predictions and was within the expected range of 50–200% in all cases. The fact that NRTL-SAC does identify the solvent composition for maximum solubility reflects the predictive power of NRTL-SAC on solute activity coefficients as affected by the changing solvent composition. The laboratory experiments confirmed that the aqueous alcohol mixtures did provide high solubility as predicted by the model and offered a relatively green solvent system for the scale-up. Ultimately, a 70 wt % water and 30 wt % NPA mixture at 70 °C was chosen for the scale-up of the crystallization process in the R&D pilot plant.

A comparison of all available experimental solubility values for the compound in binary mixtures with those predicted by NRTL-SAC is shown in Figure 11. The data consist of solubilities in nine binary mixtures with water at temperatures varying from 7 °C to 100 °C. As seen in Figure 11, many of the points lie close to the diagonal line, which indicates a very good fit between model prediction and experimental data. This is well within the expected range of 50–200% of the experimental data. Figure 11 also shows a few outliers within the data set. All of these outliers are near the low end of the range of the values and could be due to the impact of measurement uncertainty on experimental determination of low solubility values that are closer to the lower limit of detection.

It can also be observed from Figure 11 that the model seems to overpredict the solubility for the compound in most cases. Most of the data shown are at temperatures other than ambient temperature, and the heat of fusion is the key model parameter for temperature effects on solubility. As seen in Figure 1, the compound undergoes significant decomposition while melting,

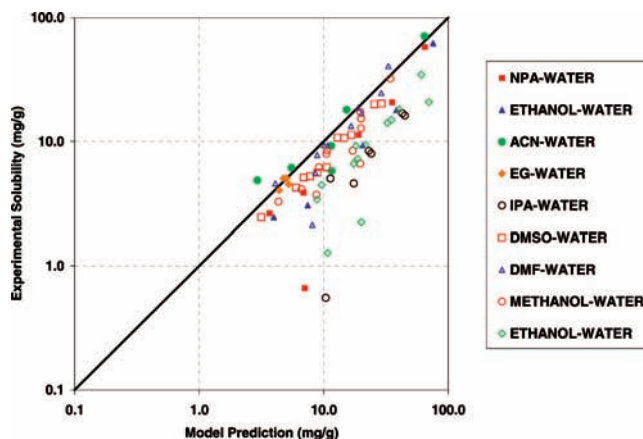


Figure 11. Comparison of overall model prediction with experimental data in binary mixtures. The heat of fusion obtained from DSC was used for regression of NRTL-SAC parameters.

which makes it difficult to experimentally decouple the energies of melting and decomposition and obtain an accurate value for the heat of fusion.

Since the experimental value of ΔH_{fus} was imprecise due to the combined effects of melting and decomposition, a series of calculations were made using different corrections to the ΔH_{fus} used in regression. The resulting data sets of model predictions and experimental data were analyzed using JMP software by SAS Institute, Inc. The data were transformed as ratios of the model predictions to the experimental values, and these transformed data were fitted to a straight line with a zero intercept. In this analysis, a perfect model prediction would give a slope of 1. The distribution of the transformed data around the straight line can be evaluated with the coefficient of variation

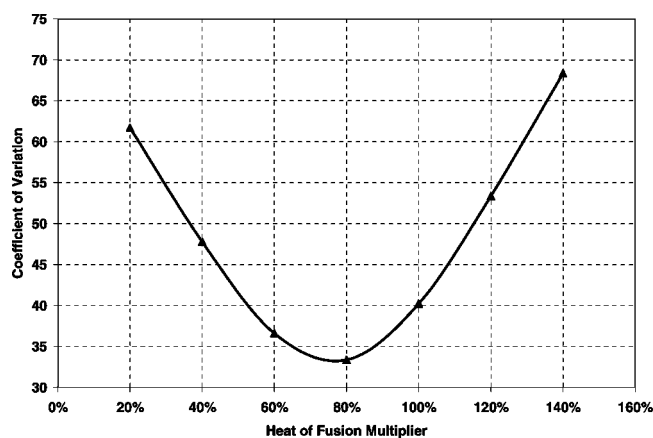


Figure 12. Variation of CV between experimental data and model predictions with different values of the heat of fusion.

Table 5. NRTL-SAC parameters obtained with corrected heat of fusion

parameter	value
X	0.0936
$Y-$	0.4767
$Y+$	0.3746
Z	0.5794

(CV), which is obtained by dividing the standard deviation by the mean. Figure 12 shows the plot of the CV versus the heat of fusion multiplier used in regression. In Figure 11, 100% denotes the experimental value of ΔH_{fus} , and other percentages correspond with the factor used to modify the value for determination of NRTL-SAC parameters using regression. As seen in Figure 12, the CV is lowest for a value of about 80% of the ΔH_{fus} determined from DSC, indicating that this lower ΔH_{fus} value gives the best model fit to the data.

The NRTL-SAC model was then adjusted by using this 80% of the value for ΔH_{fus} during regression. The NRTL-SAC parameters obtained with the corrected value of the heat of fusion are shown in Table 5. In addition to the corrected value of heat of fusion for the compound, the NRTL-SAC parameters for the solvents themselves were slightly updated by AspenTech during this time. The new solvent parameters were obtained by regressing a greater number of data sets, and also included NRTL-SAC parameters for several new solvents. The parameters are qualitatively similar to those obtained earlier, with the largest difference being in the value of Z , which accounts for the hydrophilic contribution.

Figure 13 shows the comparison of model predictions with experimental data using the corrected value of the heat of fusion for parameter regression. It is clear that the overall fit between predicted and experimental solubility values was improved after the correction to the heat of fusion. The optimum value of the

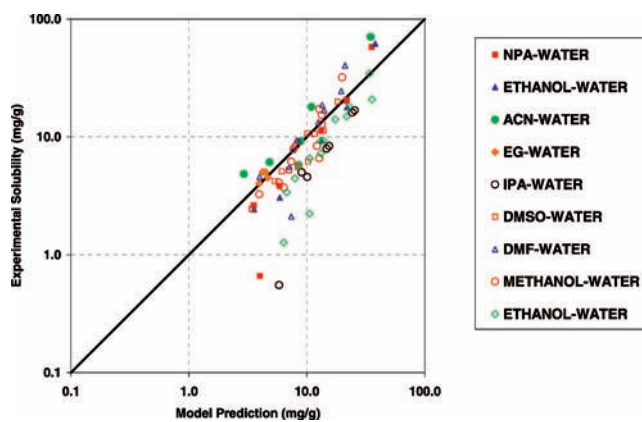


Figure 13. Comparison of overall model prediction with experimental data in binary mixtures. A corrected value of heat of fusion was used for regression of NRTL-SAC parameters.

heat of fusion for the best model fit is lower than the experimental value by 20%. This is consistent with the DSC plot in Figure 1, which shows that decomposition may have contributed to the heat of fusion measurement.

Conclusions. The NRTL-SAC model was used to identify potential solvent mixtures and process conditions for further laboratory study during the synthetic process development. Upon laboratory verification, a model-recommended binary solvent mixture and temperature were chosen for process scale-up. Model predictions were within the expected accuracy of 50–200% when compared with laboratory data later obtained. A correction was applied to the experimental heat of fusion to account for overlap with decomposition on DSC curve, and further analysis was used to show that the corrected value was reasonable and provided improved accuracy of predictions.

Also, a separate high-throughput screening tool based on Excel has been developed, which can be used to quickly screen over 120,000 binary solvent mixtures and identify promising candidates for laboratory studies. The technique can complement green chemistry efforts by searching a much larger set of solvent systems and conditions to identify environmentally benign solvents that are suitable for the process.

The use of solubility modeling to extend a few key solubility measurements to many systems has proven quite useful in targeting laboratory research efforts to systems with the most promise. The technique enables one to search a significantly larger solvent space for an optimal solvent than would be practical in a laboratory, and can complement green chemistry efforts by potentially leading to an environmentally benign solvent.

Received for review June 13, 2007.

OP700130Z