27

MOLECULAR THERMODYNAMICS FOR PHARMACEUTICAL PROCESS MODELING AND SIMULATION

CHAU-CHYUN CHEN
Aspen Technology, Inc., Burlington, MA, USA

27.1 INTRODUCTION

In this increasingly demanding pharmaceutical market, chemical engineers face major challenges in maintaining a competitive advantage in process development: (1) driving speed and efficiency into the active pharmaceutical ingredient (API) process development workflow, (2) managing information flow throughout the process development stages to improve collaboration, and (3) enabling improved process design that delivers quality assured product and lower cost of goods in commercial-scale manufacturing operations. In addition, U.S. FDA and other agencies are increasing their emphasis on process understanding. Quality by design (QbD) is an evolving initiative from FDA emphasizing that quality should be built into a product, with a thorough understanding of the product and the process by which it is developed and manufactured, along with a knowledge of the risks involved in manufacturing the product and how best to mitigate those risks. As part of QbD, first-principles modeling, statistically designed experiments, and scale-up correlations will all be considered in determining the approved design space of acceptable manufacturing conditions. To meet these challenges, chemical engineers must strive to apply modern-day first-principles modeling and simulation technology and advance science-based, mechanistic understanding of pharmaceutical manufacturing processes. Without such understanding, pharmaceutical process development would remain "lagging behind potato chip and laundry detergent makers in the use of modern manufacturing systems [1]." However, to date, the use of first-principles modeling and simulation technology in the pharmaceutical industry remains very limited, if any.

27.2 PROCESS SIMULATION AND MOLECULAR THERMODYNAMICS

> Process simulation, which emerged in the 1960s, has become one of the great success stories in the use of computing in the chemical industry. For instance, steady-state simulation has largely replaced experimentation and pilot plant testing in process development for commodity chemicals, except in the case of reactions having new mechanisms or requiring new separation technologies. Tools for steady-state process simulation are nowadays universally available to aid in the decisions for design, operation, and debottlenecking; they are part of every process engineer's toolkit. Their accuracy and predictive ability for decision-making is widely accepted to make routine plant trials and most experimental scale-up obsolete in the commodity chemicals industry [2].

While the foundation of quality by design is the availability of intrinsic (kinetic and mechanistic) process knowledge collected through the use of the various (modeling) tools [1], the scientific foundation of process modeling and simulation technology are the molecular thermodynamic models that provide thermodynamically consistent descriptions of thermophysical properties and

Chemical Engineering in the Pharmaceutical Industry: R&D to Manufacturing Edited by David J. am Ende

phase behavior of chemical systems that are being investigated.

> Indeed, the development of process modeling technologies and software tools follows and evolves around the development of molecular thermodynamics. For example, advances in equations-of-state and activity coefficient models since the 1970s set the stage for widespread applications of process modeling tools in the oil and gas industry and the petrochemical industry. For example, recent advances in molecular thermodynamics have made it possible to develop and apply process modeling technologies for complex chemical systems, i.e., processes involving synthetic fuels, aqueous electrolytes, polymer manufacturing processes, etc. Without these advances in molecular thermodynamics, chemical engineers would not have been able to develop first-principles-based, high-fidelity process models that are instrumental for process and product design and optimization [3].

The key technical barrier to the successful application of process modeling and simulation technology in the pharmaceutical industry has been the lack of accurate and robust molecular thermodynamic models that meet the unique challenges of the pharmaceutical industry. The process modeling and simulation challenges faced by the pharmaceutical industry are not so much with the mathematical formulation or simulation software of various unit operations, batch or continuous. Such mathematical formulation and simulation software have been well advanced and widely applied in the petrochemical, chemical, and specialty chemical industries and they are ready to be used in the pharmaceutical industry. Rather, without robust molecular thermodynamic models that can accurately describe the thermophysical properties and phase behavior of systems with drug molecules, process modeling and simulation technology will not provide the accuracy and predictive capability required to simulate real performance of pharmaceutical manufacturing processes and little process knowledge and few benefits, if any, could ever be derived from it.

27.2.1 Activity Coefficient as Key Thermophysical Property

Models for wide varieties of thermophysical properties and phase behavior are required in process modeling and simulation. Among the various thermophysical properties of concern to pharmaceutical process modeling and simulation, activity coefficient stands out clearly as the single most critical property. Activity coefficient plays the central role in determining the solubility and related phase behavior of pharmaceutical molecules in major pharmaceutical unit operations, for example, crystallization, chromatography, extraction, distillation, reaction, and so on.

As summarized by Lipinski et al. [4], "the knowledge of the thermodynamic solubility of drug candidates is of paramount importance in assisting the discovery, as well as the development, of new drug entities at later stages." Given a solid polymorph and a fixed temperature, equation 27.1 shows that the solubility of drug candidate is only a function of its activity coefficient in solution as the solvent composition changes.

$$\ln x_I^{\text{sat}} \gamma_I^{\text{sat}} = \frac{\Delta H_{\text{fus}}}{R}\left(\frac{1}{T_{\text{m}}} - \frac{1}{T}\right) \tag{27.1}$$

where x_I^{sat} is the mole fraction of solute I at saturation, ΔH_{fus} is the enthalpy of fusion for the solid polymorph, R is the ideal gas constant, T is the temperature, T_{m} is the melting point, and γ_I^{sat} is the activity coefficient of solute I at saturation. ΔH_{fus} and T_{m} vary with polymorphic forms of the solute.

Equation 27.2 shows that the magnitude of solute retention in chromatography under isocratic condition (i.e., constant mobile-phase solvent composition) is also related to solute activity coefficients in the mobile phase and the stationary phase.

$$k_I = \frac{x_{\text{s}} V_{\text{s}}}{x_{\text{m}} V_{\text{m}}} = K_I \Phi = \frac{\gamma_{\text{m}}^{\infty}}{\gamma_{\text{s}}^{\infty}} \Phi \tag{27.2}$$

where k_I is the retention factor for solute I; K_I is the partition coefficient; Φ is the phase ratio, that is, the ratio of the volume of the stationary phase V_{s} to that of the mobile phase V_{m}; x_{s} and x_{m} are the solute concentrations in the stationary phase and the mobile phase; and $\gamma_{\text{s}}^{\infty}$ and $\gamma_{\text{m}}^{\infty}$ are the solute infinite dilution activity coefficients in the stationary phase and the mobile phase, respectively.

Equation 27.3 shows the isoactivity relationship for extraction:

$$x_I^{\text{L1}} \gamma_I^{\text{L1}} = x_I^{\text{L2}} \gamma_I^{\text{L2}} \tag{27.3}$$

where x_I^{L1} and x_I^{L2} are the concentrations for solute I in the first liquid phase and the second liquid phase, respectively. γ_I^{L1} and γ_I^{L2} are the solute activity coefficients in the two liquid phases.

Equation 27.4 shows the isofugacity relationship for distillation:

$$x_I \gamma_I p_I^{\text{o}} = y_I \phi_I^{\text{V}} P \tag{27.4}$$

where x_I and y_I are the concentrations for solute I in the liquid phase and the vapor phase, respectively, γ_I is the solute liquid-phase activity coefficient, ϕ_I^{V} is the vapor-phase fugacity coefficient, p_I^{o} is the solute vapor pressure, and P is the system pressure.

As seen by the inclusion of activity coefficients in the above equations, a prerequisite to executing meaningful process modeling and simulation of any pharmaceutical manufacturing processes is having robust and thermodynamically consistent models that can accurately describe activity

coefficients of various components in the system of interest. Availability of such activity coefficient models is a prerequisite to meaningful first-principles process modeling and simulation in the pharmaceutical industry.

27.2.2 Thermodynamic Activity Coefficient Models

Numerous molecular thermodynamic models have been proposed in the literature to correlate or predict activity coefficients [5]. Many of them have been incorporated into process simulators [6]. Popular semiempirical correlative models, such as NRTL and UNIQUAC, are the gold standard activity coefficient models for process modeling and simulation of the petrochemical and chemical industries. However, these correlative models require identification of binary interaction parameters from phase equilibrium data for each of the solvent–solvent, solvent–solute, and solute–solute binary mixtures. While such solvent–solvent, solvent–solute, and solute–solute binary phase equilibrium data are often available for commodity chemicals, they are rarely available for new chemical entities and reaction intermediates encountered in the pharmaceutical industry. Consequently, these correlative models find very limited use in pharmaceutical process modeling, simulation, and design except in very limited solvent recovery applications.

The predictive, group contribution-based UNIFAC activity coefficient model requires only chemical structure information for the solvents and solutes [7]. Unfortunately, UNIFAC fails for complex pharmaceutical molecules for which either the UNIFAC functional groups are undefined or the functional group additivity rule becomes invalid for rigid molecular structure [8]. Recent developments in computational chemistry yielded the COSMO-RS [9] and COSMO-SAC [10] predictive models that represent promising alternatives to UNIFAC. However, the predictive powers of UNIFAC- and COSMO-based models are still inadequate [11] and their usability has been limited to nonelectrolytes.

The Hansen solubility parameter model has been the most widely used activity coefficient model in the pharmaceutical industry [12]. Incorporating the "like dissolves like" concept, the model is useful as a guide to help chemists and engineers explain API solubility behavior. However, due to its oversimplistic assumptions, the model has very limited practical use in the quantitative estimation of drug molecule solubility [11].

27.2.3 NRTL Segment Activity Coefficient (NRTL-SAC) Model

Designed to overcome the gap between molecular thermodynamics and process modeling and simulation technology for the pharmaceutical industry, the NRTL-SAC model is a very interesting and promising new development [11]. As an extension of the NRTL model [13] and the polymer NRTL [14] model for systems with solvents, solutes, oligomers, and polymers, NRTL-SAC computes activity coefficients from a combinatorial term and a residual term.

$$\ln \gamma_I = \ln \gamma_I^{\mathrm{C}} + \ln \gamma_I^{\mathrm{R}} \tag{27.5}$$

The combinatorial term γ_I^{C} is calculated from the Flory–Huggins approximation for the combinatorial entropy of mixing. The residual term γ_I^{R} is calculated from the local composition (lc) contribution γ_I^{lc} of the polymer NRTL model. Incorporating the segment interaction concept, the equation computes the activity coefficient for component I in solution by summing up contributions to the activity coefficient from all segments that make up component I.

$$\ln \gamma_I^{\mathrm{R}} = \ln \gamma_I^{\mathrm{lc}} = \sum_i r_{i,I} \left[\ln \Gamma_i^{\mathrm{lc}} - \ln \Gamma_{i,I}^{\mathrm{lc}} \right] \tag{27.6}$$

$$\ln \Gamma_i^{\mathrm{lc}} = \frac{\sum_j x_j G_{ji} \tau_{ji}}{\sum_k x_k G_{ki}} + \sum_m \frac{x_m G_{im}}{\sum_k x_k G_{km}} \left(\tau_{im} - \frac{\sum_j x_j G_{jm} \tau_{jm}}{\sum_k x_k G_{km}} \right) \tag{27.7}$$

$$\ln \Gamma_{i,I}^{\mathrm{lc}} = \frac{\sum_j x_{j,I} G_{ji} \tau_{ji}}{\sum_k x_{k,I} G_{ki}} + \sum_m \frac{x_{m,I} G_{im}}{\sum_k x_{k,I} G_{km}} \left(\tau_{im} - \frac{\sum_j x_{j,I} G_{jm} \tau_{jm}}{\sum_k x_{k,I} G_{km}} \right) \tag{27.8}$$

where I is the component index, i,j,k,m are the segment species index, $r_{i,I}$ is the number of segment species i contained only in component I, x_j is the segment-based mole fraction of segment species j, $x_{j,I}$ is the segment-based mole fraction of segment species j in component I, Γ_i^{lc} is the activity coefficient of segment species i, and $\Gamma_{i,I}^{\mathrm{lc}}$ is the activity coefficient of segment species i contained only in component I. G and τ in equations 27.7 and 27.8 are binary quantities related to each other by α $\left(\text{i.e., } G = \exp(-\alpha\tau)\right)$. α and τ are the nonrandomness factor parameter and the segment–segment binary interaction energy parameter, respectively.

Chen and Song identified four unique "conceptual" segments that broadly characterize surface interaction characteristics of molecules, solvents, or solutes [11]. These four conceptual segments, together with their corresponding nonrandomness factor and segment–segment binary interaction parameters (i.e., α and τ), are capable of qualitatively describing the various solvent–solvent, solvent–solute, and solute–solute molecular interactions and the resulting phase behavior of mixtures of solvents and solutes. Specifically, Chen and Song proposed to describe the molecular surface interactions of all solvents and solutes in solution with four types of conceptual segments: hydrophobic segment, electrostatic solvation segment, electrostatic polar segment, and hydrophilic segment. The conceptual segment numbers for each molecule, solvents or solutes, are measures of the

effective molecular surface areas that exhibit surface interaction characteristics of hydrophobicity, solvation, polarity, and hydrophilicity. The hydrophilic segment simulates molecular surfaces that are "hydrogen bond donor or acceptor." The hydrophobic segment simulates molecular surfaces that show aversion to forming a hydrogen bond. The polar and solvation segments simulate molecular surfaces that are "electron pair donor or acceptor." The solvation segment is attractive to the hydrophilic segment while the polar segment is repulsive to the hydrophilic segment. The molecule-specific conceptual segment numbers correspond to $r_{i,I}$ in equation 27.6.

Also proposed are "reference compounds" for the conceptual segments. They are used to identify the segment–-segment nonrandomness factor and binary interaction energy parameters for the conceptual segments from regression of available experimental vapor–liquid and liquid–liquid equilibrium data associated with these reference compounds. Chen and Song further identified the conceptual segment numbers for solvents commonly used in the pharmaceutical industry.

To determine the conceptual segment numbers of a solute molecule, solubility data or equivalent activity coefficient data in at least four solvents of varied surface interaction characteristics are needed. The parameterization 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, the NRTL-SAC model can then provide robust, qualitative prediction for the solute activity coefficient and the corresponding solubility in pure solvents and solvent mixtures.

It is estimated that half of all the drug molecules used in medicinal therapy are administered as salts [15]. This conceptual segment methodology has also been successfully extended for activity coefficient modeling of organic salts [16, 17].

27.2.4 Acetaminophen: An Example

Figure 27.1 shows NRTL-SAC predictions versus experimental data for acetaminophen solubility in 23 pure solvents at 303.15K [18]. As representatives of hydrophilic (water and ethanol), solvation (DMSO), polar (acetone, acetonitrile, and THF), and hydrophobic (chloroform and toluene) solvents, eight pure solvent solubility data points (shown as solid squares) were used to identify the acetaminophen parameters. Empty diamonds represent the predictions for the remaining 15 pure solvents.

Figures 27.2–27.5 show robust predictions (as solid lines) versus experimental data (as solid squares) for acetaminophen solubility in mixed solvents at 298.15K [18]. Figure 27.2 shows the prediction for a mixed solvent of two hydrophilic solvents, for example, ethanol–water binary. The acetaminophen solubility in this binary is nonideal but without significant peak solubility. Figure 27.3 shows the prediction for a mixed solvent of one polar solvent and one hydrophilic solvent, for example, acetone–water binary. The solubility behavior of acetaminophen in this binary is extremely nonideal, with "bell"-shaped solubility behavior as a function of solvent composition and a four- to fivefold solubility increase. Figure 27.4 shows the prediction for

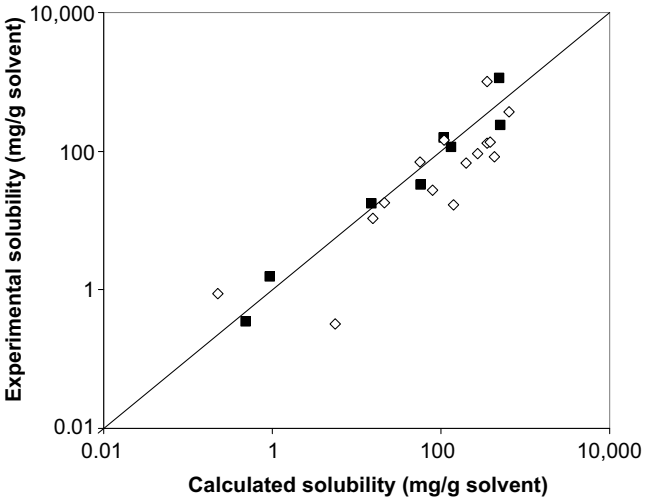

FIGURE 27.1 Model prediction versus experimental data for acetaminophen solubility in pure solvents at 303.15K (solid squares represent the eight pure solvent solubility data points that were used to identify the solute parameters; empty diamonds represent pure solvent solubility data, excluding the eight pure solvents). Reprinted with permission from Ref. 18. Copyright 2006, American Chemical Society.

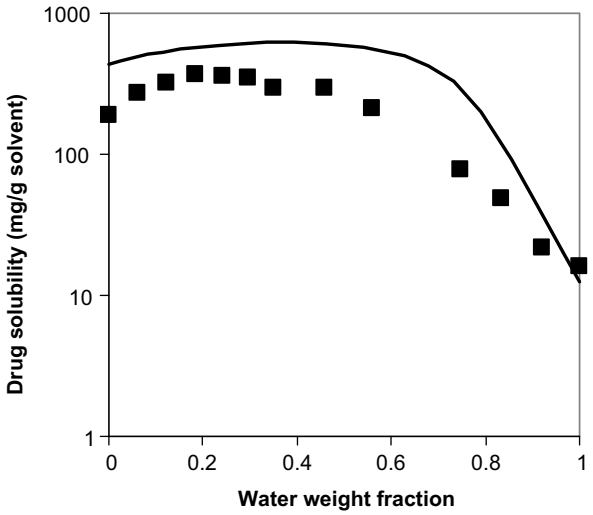

FIGURE 27.2 Model prediction versus experimental data for acetaminophen solubility in ethanol–water binary solvents at 298.15K (solid squares are experimental data and solid line represents model predictions). Reprinted with permission from Ref. 18. Copyright 2006, American Chemical Society.

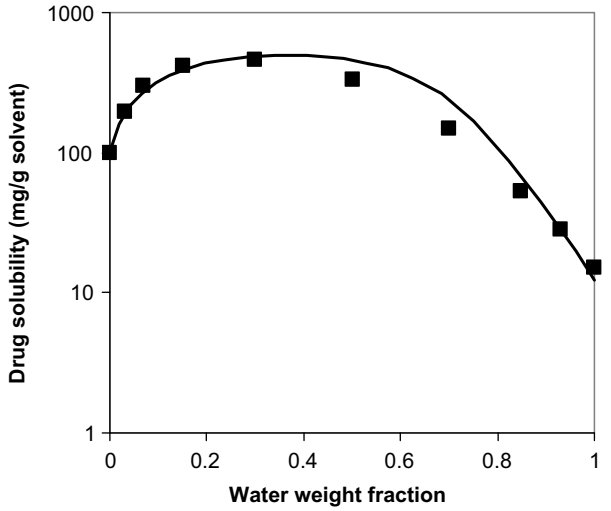

FIGURE 27.3 Model prediction versus experimental data for acetaminophen solubility in acetone–water binary solvents at 298.15K (solid squares are experimental data and solid line represents model predictions). Reprinted with permission from Ref. 18. Copyright 2006, American Chemical Society.

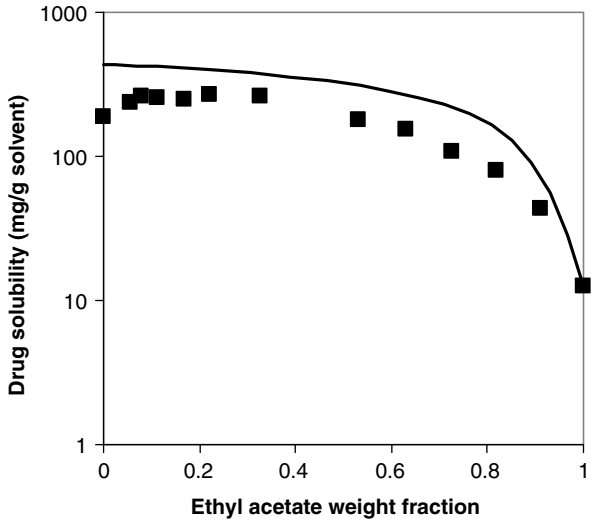

FIGURE 27.5 Model prediction versus experimental data for acetaminophen solubility in ethanol–ethyl acetate binary solvents at 298.15K (solid squares are experimental data and solid line represents model predictions). Reprinted with permission from Ref. 18. Copyright 2006, American Chemical Society.

a mixed solvent of one polar solvent and one hydrophobic solvent, for example, acetone–toluene binary. The acetaminophen solubility in this binary is relatively ideal. Figure 27.5 shows the prediction for a mixed solvent of one hydrophilic solvent and one hydrophobic solvent, for example, ethanol–ethyl acetate binary. Again, the model predicts nonideal acetaminophen solubility in the binary, consistent with the trend exhibited by the experimental data.

While the quality of the NRTL-SAC model predictions depends on the quality of the experimental data used to identify the solute parameters and there is no guarantee that the model will always yield correct, quantitative predictions, the predictive capability of NRTL-SAC has been successfully demonstrated with hundreds of drug molecules. For example, in a study with six Merck compounds [19], Merck researchers showed that NRTL-SAC offers solubility prediction accuracy within the range of $\pm 50\%$ and meets the needs of solvent selection and API process design. In contrast, the prediction accuracies from UNIFAC are in the range of $\pm 500\%$, while the correlation accuracies with the Hansen model are in the range of $\pm 200\%$. In a subsequent study [20], Merck researchers reported application of NRTL-SAC and COSMO-SAC in the solubility estimation of four Merck compounds: lovastatin, simvastatin, rofecoxib, and etoricoxib. They concluded that NRTL-SAC offered superior performance over COSMO-SAC. The maximum average log square error for NRTL-SAC was 0.10 (i.e., prediction accuracy of $\pm 30\%$) while the maximum average log square error for COSMO-SAC was 0.32 (i.e., prediction accuracy of $\pm 100\%$).

A very recent evaluation of NRTL-SAC also found a satisfactory agreement between experimental and calculated values for four drugs: paracetamol, allopurinol, furosemide, and budesonide [21]. The solubility data in pure organic solvents were used to regress the solute model parameters that were used afterward for the prediction of solubility of these compounds in water and in mixed solvent systems. The absolute average deviation was 68% for the correlation in the organic solvents and 38% for the prediction in water. The model was shown to be an appropriate tool to represent and predict the solubility of these compounds.

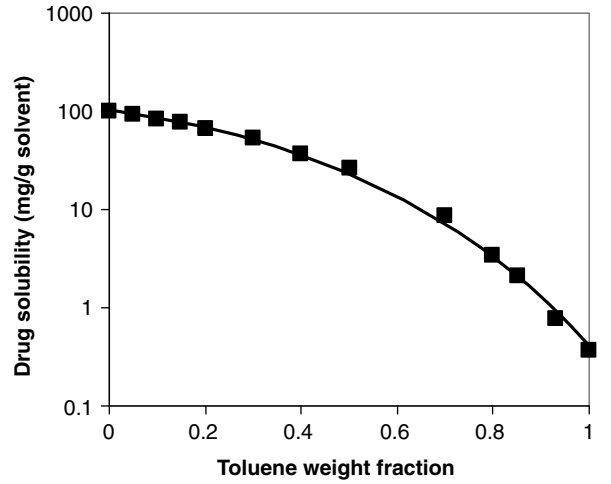

FIGURE 27.4 Model prediction versus experimental data for acetaminophen solubility in acetone–toluene binary solvents at 298.15K (solid squares are experimental data and solid line represents model predictions). Reprinted with permission from Ref. 18. Copyright 2006, American Chemical Society.

27.2.5 Applications

The ability to predict drug molecule activity coefficients and solubility in a reliable and efficient manner is invaluable to the tasks of solvent selection, API process design and optimization, and process modeling and simulation. NRTL-SAC represents a quantum leap forward in the ability to first correlate a limited number of experimental data and then predict drug molecule activity coefficients and resulting phase behavior in pure solvents and mixed solvents. When integrated with Microsoft® Excel and process simulators, NRTL-SAC offers a rigorous and practical thermodynamic framework that provides robust predictions of API activity coefficients across all unit operation models and enables chemical engineers design and optimize pharmaceutical manufacturing processes that deliver required drug purity and yield, minimize solvent usage, reduce hazardous solvent waste, consume less energy, and lower overall cost.

Some successful industrial applications of NRTL-SAC have started to emerge in the public domain. A few examples are summarized here:

Design of crystallization processes for the manufacture of API is a significant technical challenge to process research and development groups. AstraZeneca researchers examined the role of activity coefficient modeling and its application within the crystallization process design framework [22]. NRTL-SAC has been demonstrated, through the case study on cimetidine, to be a valuable aid in solubility data assessment and targeted solvent selection for crystallization process design.

Eli Lilly scientists applied NRTL-SAC to screen solvents for a crystallization medium with the goal to maximize API solubility and to minimize solvent usage [23].

> 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.

Using models, Bristol-Myers Squibb researchers demonstrated an efficient approach to identify optimal solvent compositions during conceptual design of an API process [24].

> A ternary solvent system was considered for a reaction, extraction, distillation, and crystallization sequence. Two thermodynamic models, NRTL-SAC and NRTL, as well as Aspen modeling tools, were employed to predict the liquid-liquid, vapor-liquid, and solid-liquid phase behaviors. We used these modeling tools to identify a solvent composition space for the reaction that allows for reasonable reaction volume while continuously removing a byproduct into a second aqueous phase. This composition also reduces API loss during subsequent aqueous extractions. Furthermore, the composition of the organic phase allows for an efficient azeotropic distillation during solvent exchange, resulting in a shorter cycle time needed to achieve the desired composition for final crystallization. Overall solvent usage for the process is also significantly reduced. This approach was applied retrospectively to a late-stage API process under experimental development and was validated with the production of API of excellent quality at the pilot scale with solvent compositions of the process in agreement with those predicted by the models.

27.2.6 Benefits

At the highest level, and for any R&D centric industry, the sooner improvements to new products (or the quality of decisions surrounding them) can be made, the greater the overall value potential. Value potential here is measured not only by value delivered to a product or a process, but also by "redundant cost avoidance." It is this that hits pharma's "value sweet spot" square on because pharmaceutical research is fundamentally much more "selection" than "instruction" in its nature, and so the biggest value impact may be felt in cost avoidance. This is where process modeling and simulation comes into its own and can yield millions of dollars of accrued value in the course of subsequent R&D and throughout the life cycle of the new drug thereafter. This is illustrated in Figure 27.6: a value plot against timeline to launch a new drug. Here, areas of application for modeling and simulation include lead optimization, API process development, and Drug Product (DP) process development.

Given the recent development of the NRTL-SAC model, and the long timelines for product development required for a new drug, value benefit can only be a qualitative estimation, based on current applications and anticipated capabilities. Notwithstanding the inevitably "estimated" nature of value benefit, any potential benefit should also be seen against the backdrop of apparently ever-increasing R&D costs.

In the 1980s and 1990s, relative R&D spending represented approximately 15–17% of revenue for the average drug company. Today, that average is approaching 20%, and for some companies may exceed that level. Estimates have placed the cost of bringing a new chemical entity (NCE) successfully to market to be anywhere from $700 to $1200 million over the course of 9–12 years of R&D. Some companies estimate that getting as far as completion of lead optimization requires spending some $300 million over the first 4 or 5 years of research. With the discovery and development of high-value medicines becoming harder and harder, and new drug application (NDA) annual submissions on the decline, the time is fast approaching where dramatic

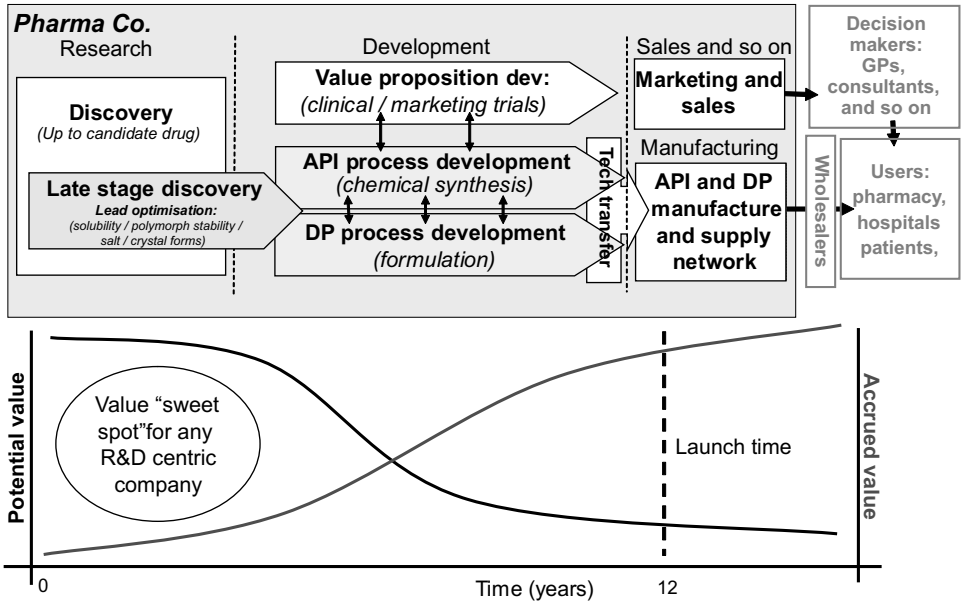

FIGURE 27.6 Value potential and benefit plot against timeline for a new drug.

operational efficiency improvements in R&D will be as much a central plank for competitive advantage in pharmaceutical industry as it already is for manufacturing operations.

With the above in mind, the application of process modeling and simulation should add value in four major ways across any pharmaceutical R&D organization:

1. *Efficiency Improvement*: by driving up the efficiency with which NCE solubility can be fully characterized, and all the potentially advantageous effects that this can confer in the lead optimization space. Literally hundreds of "experimental hours" could be reduced to just a few through the use of modeling and prediction software. This value may manifest as a reduction in cost through headcount reduction or an increase in throughput rate of NCEs in late discovery/early development. The latter is the likelier benefit route for companies with healthy pipelines of NCEs.
2. *Risk Management/Better Decision Making*: by exploiting the predictive power of the model to drive more informed and earlier decisions relating to selecting the candidate drug to best progress with respect to its "processability" downstream in API and DP manufacture. This ultimately enables a more informed and better investment focus. Delaying or dropping candidate drugs exhibiting very significant process challenges could save time, money, and resource or direct attention to solving "knockout" issues first, before devoting more investment. In addition, this should augment an "eyes open" approach to portfolio management of NCEs in early development with respect to their risk profile for manufacturability.
3. *Speed to Market Launch/Continuity of Supply After Launch*: by enabling aspects of process development activity (often delayed owing to insufficient NCE material) to proceed earlier. This can be achieved by using the modeling and predictive power of the software to sidestep this common cause of delay by using prediction to replace what would otherwise be experimentally derived process design data. This may translate into earlier clinical trials, and just possibly faster to market. Further value may manifest by avoiding or reducing the emergence of unforeseen disasters downstream that may severely compromise launch times or continuity of supply after launch. (Late emerging crystal polymorphs are a good example here, in which solubility characteristic of the active drug are permanently changed and can impact dose form stability, and even bioavailability.)
4. *API and Drug Product Manufacturing Process Performance and Cost Profiles*: by enabling the informed design of many aspects of the API and DP manufacturing processes, such that the final developed process is better characterized, optimized, greener, and higher in yield, thereby reducing cost of goods from the outset of launch.

27.3 FUTURE DEVELOPMENTS

NRTL-SAC represents one recent successful molecular thermodynamic model that makes it possible to carry out meaningful first-principles process modeling and simulation for pharmaceutical manufacturing processes. While the development of NRTL-SAC is new, further developments

of NRTL-SAC and other new activity coefficient models will certainly emerge as chemical engineers explore and expand use of process modeling and simulation in the pharmaceutical industry. It should be noted that development of NRTL-SAC and other new models requires extensive, high-quality experimental data sets for model developers to advance physical insights and to validate models. The scarcity and the often questionable quality of public literature data with drug-like molecules, especially regarding exact solid polymorphs, indeed makes such new developments in molecular thermodynamics extremely challenging.

The ability to predict activity coefficients and the corresponding solubility of drug molecules is important not only for API manufacturing processes, but also for formulation design and delivery of a drug to its site of action in the human body [4]. While the development of NRTL-SAC has focused on applications in process modeling and simulation for API manufacturing processes, the conceptual segment methodology of describing molecular surface interaction characteristics is equally applicable to surfactants and polymers that are often used as excipients in drug formulation. Another example of particular interest, although not obvious and under investigation, is solubility modeling and prediction of biologically derived or engineered macromolecules, such as monoclonal antibodies and genetically engineered proteins. This could be particularly exciting if further research shows that the segmentation nature of the NRTL-SAC model lends itself well to more and more complex chemical/biochemical entities. With further experimentation, it may also be possible to characterize human organs, body fluids, and tissues in a similar manner to study drug bioavailability, pharmacokinetics, and toxicity. Application of NRTL-SAC in drug formulation and other pertinent areas should be actively pursued.

27.4 ASPEN TECHNOLOGY'S PROCESS DEVELOPMENT SOLUTION

aspenONE Engineering, AspenTech's model-based process development solution for the pharmaceutical industry, delivers "design for manufacture" (DfM) capability by enabling people to model, simulate, design, and optimize API manufacturing processes. The solution is designed to transform the pharmaceutical process development workflow into a productive and efficient process by

- Enabling development of robust and thermodynamically consistent thermophysical property models including NRTL-SAC
- Providing the modeling and simulation framework for first-principles-based process models, both batch and continuous
- Improving collaboration across the development workflow, from route selection to scale-up to completion of technology transfer
- Facilitating efficient and successful technology transfer to first sites of commercial manufacture
- Enabling the workflow to capture learning and apply improvements iteratively

Underlying AspenTech's process development solution are process modeling and simulation tools that enable engineers to develop first-principles-based process models to achieve mechanistic understanding of the pharmaceutical manufacturing processes. The same process models developed during the design phase can be used to support continuous improvement initiatives in commercial-scale API manufacture—helping engineers eliminate bottlenecks, increase throughput, or achieve better operational efficiency. Table 27.1 shows the *aspenONE Engineering* suite of products comprised of independently deployable components with specific modeling functionalities.

Clearly, one cannot overemphasize the importance of applying exactly the same thermophysical property methods, models, and data across all process modeling and simulation tools and activities to ensure rigor and consistency of results. Therefore, a key common requirement for the various process modeling and simulation products is the library of rigorous, thermodynamically consistent models for thermophysical properties and phase behavior. Aspen

TABLE 27.1 aspenONE Engineering suite of products

Product Name	Description
Aspen Properties	Physical property modeling system with comprehensive chemical database and estimation capability
Aspen Solubility Modeler	Modeling solubility of drug molecules and predicting drug molecule solubilities in solvents and solvent mixtures
Aspen Plus	First-principles rigorous modeling for continuous steady-state processes (including single and multistage separations such as distillation and extraction, reactors, heat exchangers, pumps, compressors, etc.)
Aspen Batch Distillation	First-principles rigorous modeling for batch distillation processes
Aspen Reaction Modeler	First-principles rigorous reaction modeling, including kinetic data fitting
Aspen Custom Modeler	Custom environment for detailed modeling of other unit operations
Aspen Batch Process Developer	Recipe-based process modeling designed for route selection, recipe development, process scale-up, scheduling, and technology transfer

Properties provides such a common technology component and delivers best-in-class thermophysical property methods, models, and data. It includes extensive databases of pure component and phase equilibrium data and libraries of estimation methods. Also included are the activity coefficient models such as NRTL, electrolyte NRTL, UNIQUAC, UNIFAC, Hansen, COSMO-SAC, NRTL-SAC, electrolyte NRTL-SAC, and so on. Ongoing collaboration with the U.S. National Institute of Standards and Technology (NIST) ensures continuing access to the newly available methods, models, and data.

There are three modeling tools in Table 27.1 that deserve special attention due to their particular pertinence to the process modeling and simulation of pharmaceutical manufacturing processes: Aspen Solubility Modeler (ASM), Aspen Batch Distillation, and Aspen Reaction Modeler. Brief summaries are given below.

Aspen Solubility Modeler is a tool with an Excel-based front-end that provides users with the capability to define drug properties and identify NRTL-SAC parameters for drug molecules by regressing user-specified experimental solubility data in pure solvents or mixed solvents. Also available is the NRTL-SAC databank that provides NRTL-SAC parameters for over 150 solvents and excipients commonly used in the pharmaceutical industry. The Data Analysis Excel package allows users to predict phase equilibria (vapor–liquid equilibrium and vapor–liquid–liquid equilibrium) and solubility behavior of drugs under various operating conditions and solvent compositions. Appendix A shows how to set up Aspen Solubility Modeler, how to use ASM to regress solubility data to identify NRTL-SAC parameters for drug molecules (i.e., caffeine), and how to use ASM to perform solubility calculations based on the regressed NRTL-SAC parameters.

Aspen Batch Distillation is a comprehensive simulation tool for conceptual design, analysis, and optimization of batch distillation processes. Key features include the following:

- Intuitive interface designed specifically for simulating batch distillation
- Interoperability of Aspen Batch Distillation models inside the industry-leading Aspen Plus process simulation environment
- Optimization tool enabling rigorous identification of optimum operating steps to minimize cycle time while maintaining operating and performance constraints
- Equation-oriented architecture allowing timely and robust dynamic simulation of complex columns
- Rigorous equipment modeling including flexible, configurable controller models; pressure drop correlations; multiphase, azeotropic, and reactive distillation; and options to start from dry or total reflux conditions

Aspen Reaction Modeler enables users to identify reaction kinetics models using experimental measurements from reaction calorimeters. Key features include the following:

- Easy to use user interface that fully supports the model identification workflow and enables easy copying of experimental data from Excel
- Powerful numerical solvers for finding the best fit to experimental data
- Comprehensive kinetics include power law and Langmuir–Hinshelwood reaction kinetics, reversible reactions, and mass transfer effects
- Kinetic models consistent with other AspenTech products such as Aspen Plus, Aspen Plus Dynamics, and Aspen Batch Distillation, enabling users to apply the fitted parameters directly within these products
- Use of Aspen Properties for estimating physical properties required for the model identification process

27.5 CONCLUSIONS

Modern-day first-principles process modeling and simulation is the enabling technology to advance science-based, mechanistic understanding of pharmaceutical manufacturing processes and to succeed in quality by design. Molecular thermodynamics provides the scientific foundation for process modeling and simulation technology, and the lack of suitable molecular thermodynamic models for systems with complex pharmaceutical molecules has been the primary technical barrier for productive practice of process modeling and simulation in the pharmaceutical industry. The recent development of the NRTL-SAC activity coefficient model brought about a long-awaited breakthrough in molecular thermodynamics and it opened the door for meaningful application of process modeling and simulation technology in the pharmaceutical industry. Incorporating NRTL-SAC and other pertinent molecular thermodynamic models, AspenTech's model-based process development solution for the pharmaceutical industry ensures thermodynamic consistency in modeling of thermophysical properties and phase behavior and enables development of intrinsic process knowledge through the use of first-principles process modeling and simulation technology.

ACKNOWLEDGMENT

C.-C. Chen is grateful for extensive inputs, discussions, and assistance in generating the examples from his colleagues including Phil Norris, Russell Schofield, Catherine Jablonsky, Gary O'Neill, Jack Vinson, David Tremblay, Yuhua Song, Emmanuel Lejeune, Joseph DeVincentis, Jiangchu Liu, and

Zhiming Huang. The Benefits section was published previously in *Pharmaceutical Manufacturing* (October 2009, pp. 30–32) and the author thanks Agnes Shanley of Pharmaceutical Manufacturing for permission to reprint.

APPENDIX A SOLUBILITY MODELING WITH ASPEN SOLUBILITY MODELER

I. Setting Up Aspen Solubility Modeler

Objective

This workshop shows how to set up Aspen Solubility Modeler on your computer.

Description

Task 1: Set up the Aspen Properties Add-Ins in Excel

- Start Excel
- Go to the Tools menu, Add-Ins... (Figure 27.7)
- Click the Browse button (Figure 27.8)
- Navigate to C:\Program Files\AspenTech\Aspen Properties v7.1\Engine\Xeq
- Select the file Aspen Properties.xla, and then click OK (Figure 27.9)

Click OK on any message box that may be displayed. Click the OK button back on the Add-ins window.

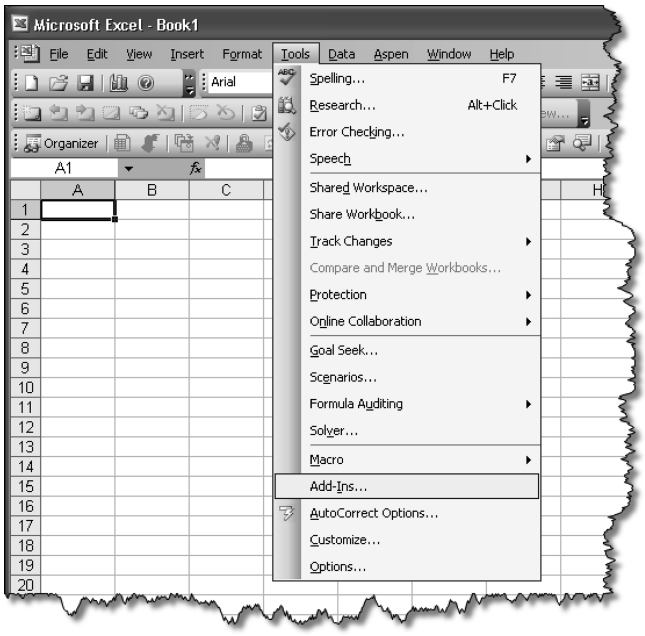

FIGURE 27.7 Tools pull-down menu.

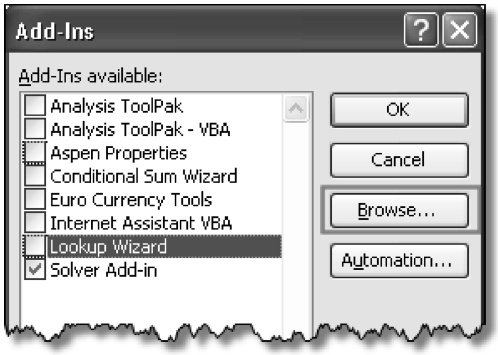

FIGURE 27.8 Add-Ins menu.

Task 2: Allow Macros

- In Tools, Macros, Security, select the "High" security level (or Medium or Low)
- Close Excel
- Go to the Start menu of Windows, Programs, AspenTech, Process Development v7.1, Aspen Solubility Modeler
- Open the spreadsheet "Regression.xls"
- When prompted to allow macros signed by AspenTech Inc., click "Enable macros"
- Close Excel

Note: This procedure needs to be done only once.

II. Caffeine Solubility Data Regression

Objective

This workshop shows how to regress solubility data.

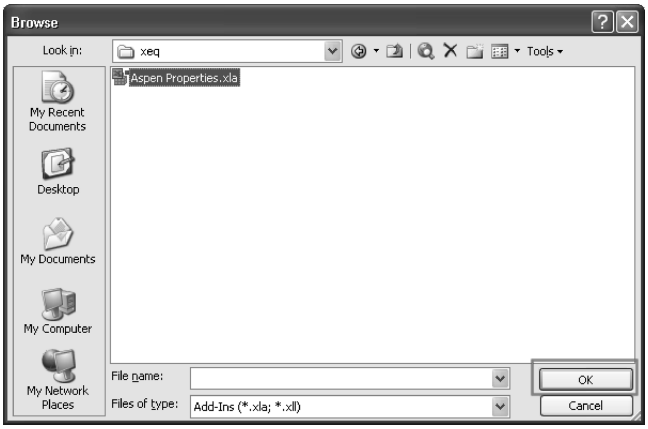

FIGURE 27.9 Browse file.

Description

The solubility of caffeine in different solvents is given in the following table.

Solvent	T (°C)	x (Solubility) (g Drug/g Solvent)	Standard Deviation (%)
N-HEXANE	25	8.8802E–06	40
2-ETHOXYETHANOL	25	1.4707E–02	10
1-OCTANOL	30	3.6693E–03	10
1,4-DIOXANE	25	1.8241E–02	10
1,4-DIOXANE	25	1.8873E–02	10
N,*N*-DIMETHYLFORMAMIDE	25	3.3787E–02	10
WATER	25	2.4274E–02	10
WATER	25	2.4741E–02	10
WATER	25	2.1667E–02	10
WATER	25	2.3316E–02	10
WATER	30	2.0436E–02	10

The caffeine pure properties are given in the following table:

Property	Value
MW	194.19 kg/kmol
Melting point	512.15K
Enthalpy of fusion	21600 kJ/kmol

Source: Ref. 25.

Task 1: Copy the NRTL-SAC Folder

- From the Windows Start menu, go to Programs, AspenTech, Process Development v7.1, Aspen Solubility Modeler
- Select the folder NRTL-SAC, and then copy it into any location that is convenient for you (e.g., the Desktop)

Note: We recommend copying the NRTL-SAC folder instead of working on the original files. This is because some data specific to your project will be stored in the physical property package, that is, the regressed parameters.

Task 2: Specify the Property Package

- Open the file Regression.xls with Excel
- On the worksheet, make sure the option "Pure Solvents" is selected, and then click the OK button

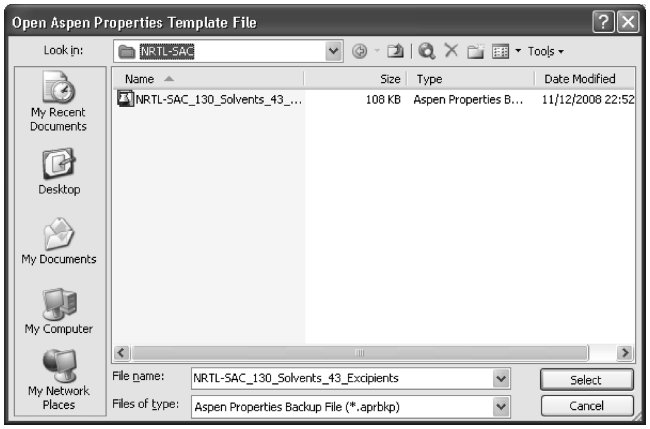

FIGURE 27.11 Open Aspen Properties template file.

- On the "Pure Solvent" sheet, click the "Execute Step 1" button (Figure 27.10)
- Select the file shown, and then click Select (Figure 27.11)

This will launch Aspen Properties as a hidden application, so there will be a delay before you can continue.

Task 3: Define the Drug Parameters

- Change the name to "Caffeine" (this is only for reporting purposes)
- Set the MW to 194.19
- Set the melting point to 512.15K
- Set the enthalpy of fusion to 21600 kJ/kmol
- Clear the entropy of fusion value
- Click the button "Execute Step 2" (Figure 27.12)

The entropy of fusion is now calculated (enthalpy of fusion divided by melting point).

Task 4: Enter NRTL-SAC Model Parameter

- Click the button "Calculate Ksp A & Ksp B" (Figure 27.13)

You should see the values of Ksp A and B are updated. The values of Ksp A and Ksp B can be calculated from the enthalpy and entropy of fusion (Ksp A = entropy of fusion/

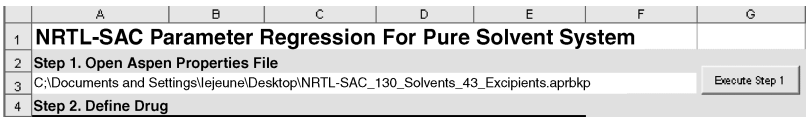

FIGURE 27.10 Properties file.

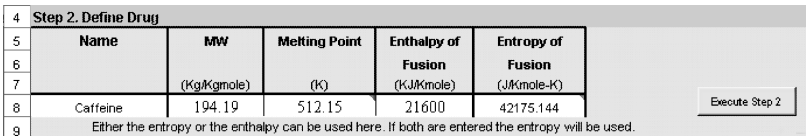

FIGURE 27.12 Enter data for caffeine.

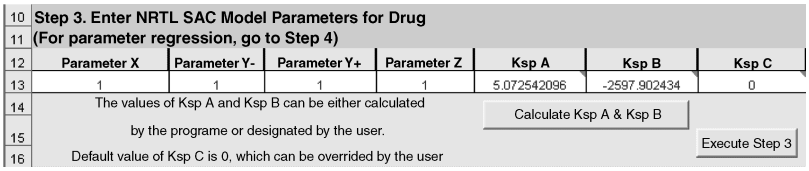

FIGURE 27.13 Select parameters.

gas constant; Ksp B = −enthalpy of fusion/gas constant). The other parameters (X, Y−, Y +, Z) may be left at the value of 1 or specified to a positive value (e.g., $X = 0$ if you believe the component does not manifest any hydrophobic behavior).

- Click the Execute Step 3 button (Figure 27.13)

This will copy the parameters from Excel to the Aspen Properties file.

Task 5: Enter the Experimental Data

- Enter the experimental data
- For each data row, select the solvent using the pull-down list and enter the temperature, the solubility, and the standard deviation
- Delete the data on the rows you are not using

Note: You can copy and paste the data from the spreadsheet "Solubility Data.xls."

- For the parameter Ksp B, select the option "EXCLUDE"

Note: We exclude the parameter Ksp B because the experimental temperature range is not very large.

See Figure 27.14 for how the spreadsheet should look.

Task 6: Run the Regression and Review the Results

- Click the button "Execute Step 4"

The following window will be displayed. Click the OK button (Figure 27.15).The Aspen Properties application will be made visible.

- Click the OK button when this window is displayed (Figure 27.16)

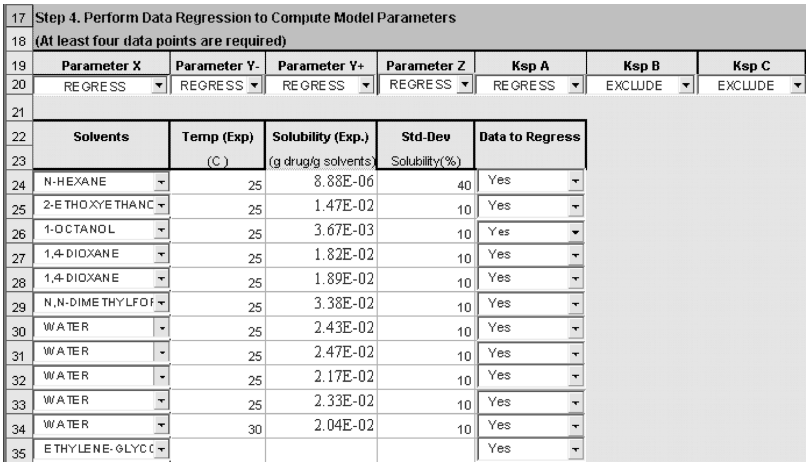

FIGURE 27.14 Perform data degression.

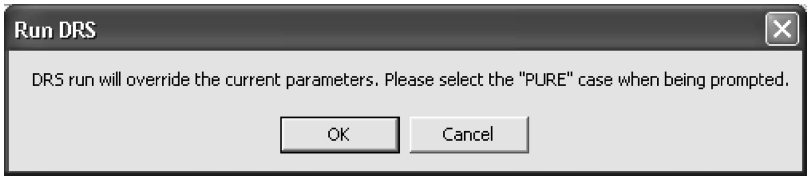

FIGURE 27.15 Dialog box.

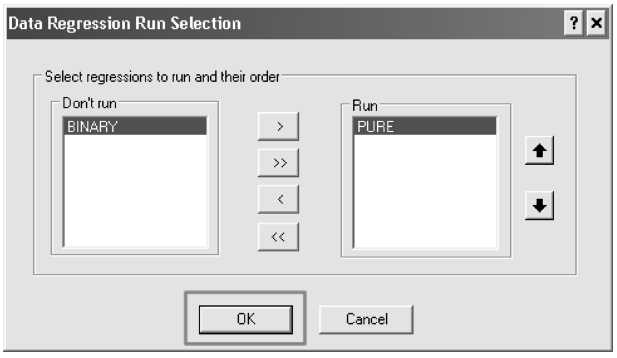

FIGURE 27.16 Dialog box.

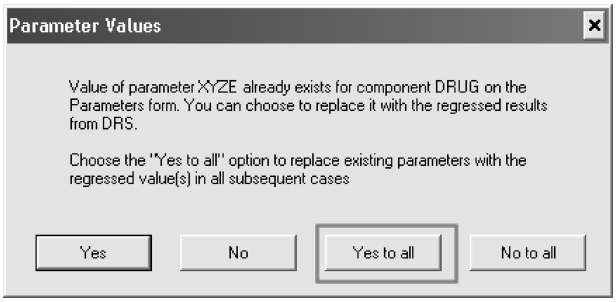

FIGURE 27.17 Dialog box.

Aspen Properties is now processing the data regression. When the calculations are complete, you will see the following window (Figure 27.17).

Click the "Yes to all" button

Back in Excel you can now inspect the regression results.

Solution

The following results are obtained.

	Value	Standard Deviation
Parameter X	0.012959095	0.013825285
Parameter Y−	0.509289197	0.06132528
Parameter Y+	0.920405633	0.060143778
Parameter Z	0.422110828	0.050881132
Ksp A	3.90390191	0.056377518
Ksp B	−2597.902434	0
Ksp C	0	0
SSQ	22.5715669	
R^2	0.915104372	
R^2 (log)	0.988690968	
RMSE	0.0026638	
RMSE (log)	0.104301944	

The regression looks good: the values of the parameters are reasonable, the standard deviations are smaller than the regressed parameter, and the parity plot shows a good correlation.

III. Caffeine Solubility Calculation

Objective

This workshop shows how to use the calculation spreadsheet.

Description

Task 1: Open Calculation Spreadsheet

- Navigate to the folder where the workshop files are stored, in the folder asm-calc-caffeine
- Open the spreadsheet "Calculation.xls"

Note: When Excel displays a warning window about ActiveX controls, click the "Yes" button to allow them. These are required for the ternary diagram plots used on some calculation sheets.

Task 2: Pure Solvent Solubility

- Select the option "Solubility in Solvents"
- On the "Solubility in Solvents" sheet, select the following solvents and enter the experimental solubility (Figures 27.18 and 27.19)

Temperature: 25°C
Pressure: 1 bar

Solvent	Solubility (g/100 g Solvent)	Calculated Solubility (g/100 g Solvent)
N-HEXANE	8.88E−4	
2-ETHOXYETHANOL	1.47	
1-OCTANOL	0.366	
WATER	2.42	

			Input x		Calculated x		Calculated w		Solubility (g solute/100 g solvent)		
Solvent Name	Solute	Solvent ID	Solute	Solvent	Solute	Solvent	Solute	Solvent	Calculated	perimental	% Error
N-HEXANE	DRUG	HEXANE	0.5	0.5	7.62E-06	0.999992	1.72E-05	0.99998284	1.716E-03	8.88E-04	93.28102
2-ETHOXYETHANOL	DRUG	2ETHOXYE	0.5	0.5	0.009147	0.990853	0.019502	0.9804976	1.989E+00	1.47	35.30822
1-OCTANOL	DRUG	OCTANOL	0.5	0.5	0.002148	0.997852	0.0032	0.9967999	3.210E-01	0.366	-12.2849
WATER	DRUG	WATER	0.5	0.5	0.002051	0.997949	0.021675	0.97832495	2.216E+00	2.42	-8.44929

FIGURE 27.18 Output data.

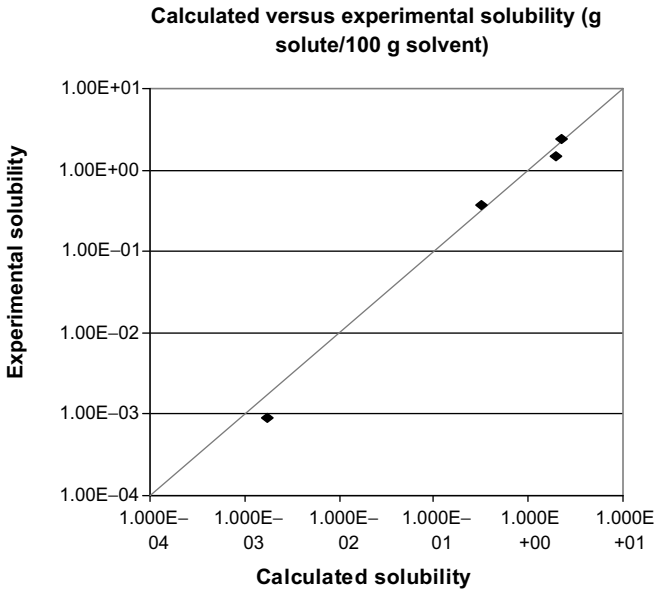

FIGURE 27.19 Comparison of calculated versus experimental solubility.

Task 3: Find the Solubility in Acetonitrile

Using the same sheet, change one solvent to Acetonitrile. This will report the solubility of caffeine in acetonitrile.

Answer: 2.79 g/100 g solvent.

Task 4: Find the Solubility in Binary Mixture

- Click the "Back to Welcome page" button
- Select "Solubility in binary solvent mixture"
- Select ACETONITRILE and WATER as the solvents
- Set the temperature to 25°C

Answer: We can see that the solubility is about five times larger for the mixture 60 wt% acetonitrile/40 wt% water. We can confirm the mixture of solvents is a single phase by checking the value of BETA reported on the spreadsheet (1 = single liquid, <1 = two liquid phases) (Figure 27.20).

Task 5: Use the "High Throughput" Sheet

- Click the "Back to Welcome page" button
- Select "High Throughput Prediction" option
- Select the following solvents:
  - WATER
  - ACETONE
  - ACETONITRILE
  - METHYL-ACETATE
  - 1,4-DIOXANE
  - 1-CHLOROBUTANE
  - TETRAHYDROFURAN
  - ETHYL-ACETATE

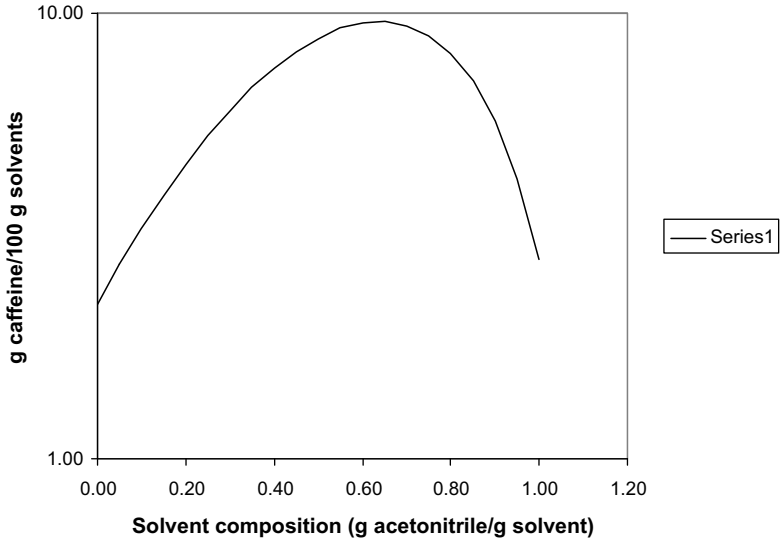

FIGURE 27.20 Graphical answer to Task 4.

- ETHANOL
- METHANOL

Note: When prompted "Do you want to continue," click "No" except when selecting the last solvent. This will prevent Excel recalculating the sheet while you are still setting up the list of solvents.

Is there another binary mixture in which the caffeine solubility becomes larger than in the pure solvents?

Answer: Yes, essentially all binary mixtures with water, especially water/ethanol where the solubility is even higher than that in water/acetonitrile.

REFERENCES

1. McKenzie P, Kiang S, Tom J, Rubin AE, Futran M. *AIChE J*. 2006;52:3990–3994.
2. *Beyond the Molecular Frontier: Challenges for Chemistry and Chemical Engineering*, Committee on Challenges for the Chemical Sciences in the 21st Century, Board on Chemical Sciences and Technology, National Research Council, National Academics Press, Washington, DC, 2003, p. 86.
3. Chen C-C. Fluid Phase Equilib. 2006;241:103–112.
4. Lipinski CA, Lombardo F, Dominy BW, Feency PJ. *Adv. Drug Deliv. Rev.* 1997;23:3–25.
5. Prausnitz JM, Tavares FW. *AIChE J*. 2004;50:739–761.
6. Chen C-C, Mathias PM. *AIChE J*. 2002;48:194–200.
7. Fredenslund A, Jones RL, Prausnitz JM. *AIChE J*. 1975;21: 1086–1099.
8. Gracin S, Brinck T, Rasmuson AC. *Ind. Eng. Chem. Res.* 2002; 41:5114–5124.
9. Eckert F, Klamt A. *AIChE J*. 2002;48:369–385.
10. Lin S-T, Sandler SI. *Ind. Eng. Chem. Res.* 2002;41:899–913.
11. Chen C-C, Song Y. *Ind. Eng. Chem. Res.* 2004;43:8354–8362.
12. Hansen CM. *Hansen Solubility Parameters: A User's Handbook*, CRC Press, 2000.
13. Renon H, Prausnitz JM. *AIChE J*. 1968;14:135–144.
14. Chen C-C. *Fluid Phase Equilib*. 1993;83:301–312.
15. Stahl PH, Wermuth CG, editors. *Handbook of Pharmaceutical Salts: Properties, Selection and Use*, Verlag Helvetica Chimica Acta, Zürich, 2002.
16. Chen C-C, Song Y. *Ind. Eng. Chem. Res.* 2005;44:8909–8921.
17. Song Y, Chen C-C. *Ind. Eng. Chem. Res.* 2009;48:5522–5529.
18. Chen C-C, Crafts PA. *Ind. Eng. Chem. Res.* 2006;45: 4816–4824.
19. Bakken D, et al. Solubility Modeling in Pharmaceutical Process Design, paper presented at AspenTech User Group Meeting, New Orleans, LA, October 5–8, 2003.
20. Tung H-H, Tabora J, Variankaval N, Bakken D, Chen C-C. *J. Pharm. Sci.* 2008;97:1813–1820.
21. Mota FL, Carneiro AP, Queimada AJ, Pinho SP, Macedo EA. *Eur. J. Pharm. Sci.* 2009;37:499–507.
22. Crafts P. The role of solubility modeling and crystallization in the design of active pharmaceutical ingredients. Ng KM, Gani R, Dam-Johanson K, *Chemical Product Design: Toward a Perspective through Case Studies*, Elsevier, 2007.
23. Kokitkar PB, Plocharczyk E, Chen C-C. *Org. Process Res. Dev.* 2008;12:249–256.
24. Hsieh D, Marchut AJ, Wei C-K, Zheng B, Wang SSY, Kiang S. *Org. Process Res. Dev.* 2009;13:690–697.
25. Marrero J, Abildskov J. *Solubility and Related Properties of Large Complex Chemicals, Part 1: Organic Solutes Ranging from C_4 to C_{40}*, Chemistry Data Series, Vol. 15, DECHEMA 2003.