

# Algorithmic Framework for Quality Assessment of Phase Equilibrium Data

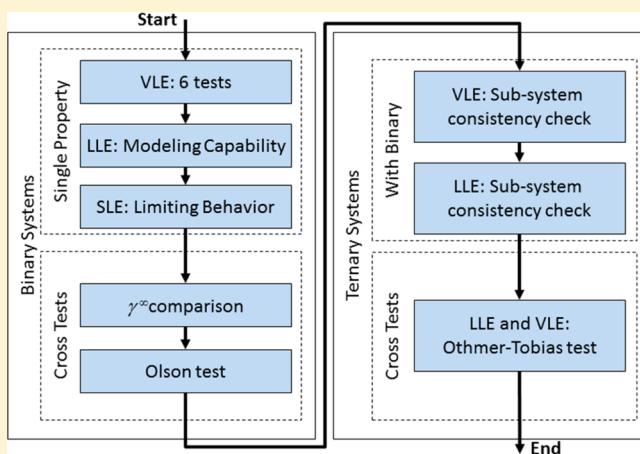
Jeong Won Kang,<sup>†,§</sup> Vladimir Diky,<sup>‡</sup> Robert D. Chirico,<sup>‡</sup> Joseph W. Magee,<sup>‡</sup> Chris D. Muzny,<sup>‡</sup> Andrei F. Kazakov,<sup>‡</sup> Kenneth Kroenlein,<sup>‡</sup> and Michael Frenkel\*,<sup>‡</sup>

<sup>†</sup>Department of Chemical and Biological Engineering, Korea University, 5-1 Anam-Dong, Sungbuk-Ku, Seoul 136-701, South Korea

<sup>‡</sup>Applied Chemicals and Materials Division, National Institute of Standards and Technology, Boulder, Colorado 80305-3337, United States

## Supporting Information

**ABSTRACT:** Data quality assessment procedures for various types of phase equilibrium data are reviewed and analyzed. Experimental data for vapor–liquid equilibrium (VLE), liquid–liquid equilibrium (LLE), solid–liquid equilibrium (SLE), infinite dilution activity coefficients (ACT), and excess enthalpy (HE) are used for interproperty consistency checks, together with relevant properties of pure compounds. These procedures are essential for validation of experimental data and development of high quality prediction and correlation models. A single framework involving consistency tests for all types of phase equilibria data is proposed, together with an algorithm implementation.



## 1. INTRODUCTION

Previously, we described a quality-assessment algorithm for vapor–liquid equilibria (VLE) data.<sup>1</sup> That algorithm was subsequently implemented in the NIST ThermoData Engine software<sup>2,3</sup> and was used in the development of the technology for automated production of UNIFAC<sup>4–6</sup> parameters, resulting in establishment of the NIST-KT-UNIFAC prediction method.<sup>7</sup> The NIST ThermoData Engine software (TDE) is designed to implement the concept of dynamic data evaluation for thermophysical properties.<sup>8</sup> It has been described extensively in the literature.<sup>2,9–16</sup> Development of evaluation tests for phase-equilibrium data remains of high interest, as evidenced by recent work by Hajipour et al.,<sup>17</sup> who proposed a high pressure VLE test, and Cunico et al.,<sup>18</sup> who discussed consistency tests for solid–liquid equilibria.

Activity coefficient (AC) models are widely used in industry and academia for representation of phase-equilibrium data.<sup>19</sup> Parameterization of these models is based upon experimental data for binary chemical systems. Since commonly used thermodynamic models, such as NRTL (nonrandom two liquids),<sup>20</sup> UNIQUAC (universal quasichemical),<sup>21</sup> and UNIFAC (universal functional-group activity coefficients),<sup>4</sup> are used in modeling for all types of phase equilibrium, including VLE, as well as LLE (liquid–liquid equilibria), SLE (solid–liquid equilibria), and enthalpies of mixing processes, it follows that experimental data for all of these data types should be used together in a single framework to evaluate the model parameters. Table 1 shows the

**Table 1. Types of Experimentally Measured Phase Equilibria Properties and Information Derived<sup>a</sup>**

data type	abbreviation	information derived
vapor–liquid equilibrium	VLE	$\gamma = f(x)$
liquid–liquid equilibrium	LLE	$\gamma = f(x, T)$ , partially miscible system
solid–liquid equilibrium	SLE	$\gamma = f(x)$ , at low $T$
activity coefficient at infinite dilution	ACT	$\gamma = f(x; 0)$ , limiting condition
excess enthalpy excess heat capacity	HE CPE	$\gamma = f(T)$ , primarily at ambient condition

<sup>a</sup> $\gamma$  is activity coefficient,  $T$  is temperature,  $x$  is the mole fraction of one component of the mixture in the liquid phase.

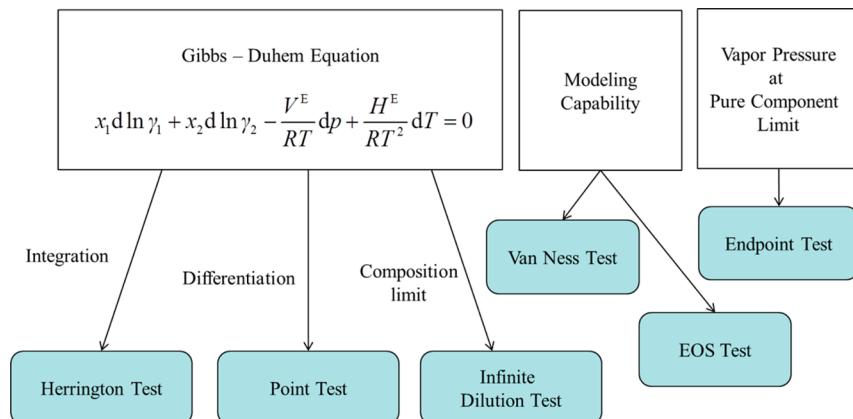
relationship between phase-equilibrium measurements common in the literature and the information that can be gleaned from these concerning the dependence of the activity coefficients on system variables, particularly temperature and composition. This table is similar to that given by Gmehling et al.<sup>22</sup> Currently, there is no single model capable of covering all temperatures and compositions, and model parameters are applicable only to limited variable ranges or limited properties.

Development of comprehensive models capable of predicting all phase-equilibria properties over broad variable ranges is

Received: April 9, 2014

Accepted: May 19, 2014

Published: June 3, 2014



**Figure 1.** Outline of the VLE data quality assessment algorithm.

hindered significantly by the requirement that such a development be based on fitting all available experimental data sets, since it is well-known that many those data sets are thermodynamically inconsistent. These inconsistencies exist for many reasons, such as poor calibration of experimental apparatus, insufficiently characterized samples, poor or absent uncertainty assessment, low-quality data analysis, errors in data reduction, errors in the publication process, including article preparation, technical editing, and typesetting. After publication, additional errors can occur during propagation of the data from the original source to secondary sources, such as property databases and software applications. These data-transfer errors can be caused by ambiguous or incomplete definition of the experimental data in the original report, as well as through transcription errors and misinterpretations during the transfer process. Indeed, the results of a study undertaken by the U.S. National Institute of Standards and Technology (NIST), in cooperation with major journals in the field, indicate that approximately one-third of articles reporting experimental thermophysical properties contained erroneous (or incomplete) numerical and metadata information, even after peer review.<sup>23</sup>

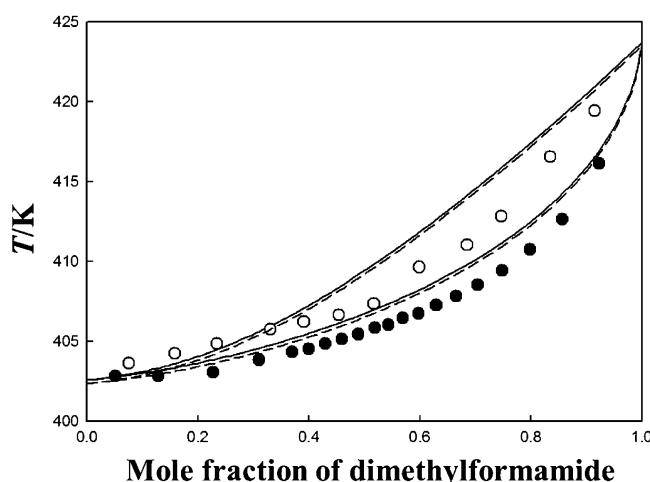
In this article, we describe an algorithmic approach to deploy a sequence of phase-equilibrium data tests within a unified framework involving all relevant properties and their combinations, in order to automatically select reliable data sets and determine appropriate weighting factors for regression analysis. In the development of this framework, we do not suggest new data quality tests, but rather analyze previously described tests in the large data set environment and propose their structural deployment encompassing properties of pure compounds, binary mixtures and ternary mixtures. In the process of the selection of appropriate tests as the components for the framework, we tried to focus on tests that are applicable to a significant number of chemical systems characterized with a sufficient amount of experimental data. Therefore, not all currently known tests are included in the developed system. For example, a previously described test for consistency between excess Gibbs energy, liquid–liquid equilibrium properties, and solid–liquid equilibrium properties<sup>24</sup> relevant for a small number of systems and requiring experimental data for LLE and SLE over a broad temperature range is not included in the proposed framework. However, because of the modular nature of the proposed framework, this test, as well as some other additional tests currently not included, could be incorporated in the future. The developed framework is useful not only for assessment of data quality but is also critical in evaluation of parameters for

predictive models, without the need for labor-intensive manual prescreening. The data assessment framework described here was used to evaluate group parameters for the NIST-Modified UNIFAC prediction method, which is based on the mathematical formulation of Modified UNIFAC,<sup>25–29</sup> together with the extensive collection of experimental data stored in the NIST/TRC Data Archival System.<sup>30</sup> A complete description of the development of NIST-Modified UNIFAC will be published separately.<sup>31</sup>

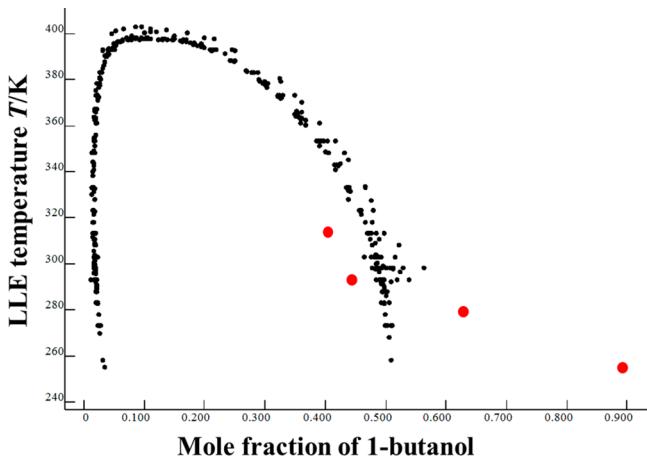
## 2. QUALITY ASSESSMENT PROCEDURES

### 2.1. Consistency Tests for Binary Vapor–Liquid Equilibria Data.

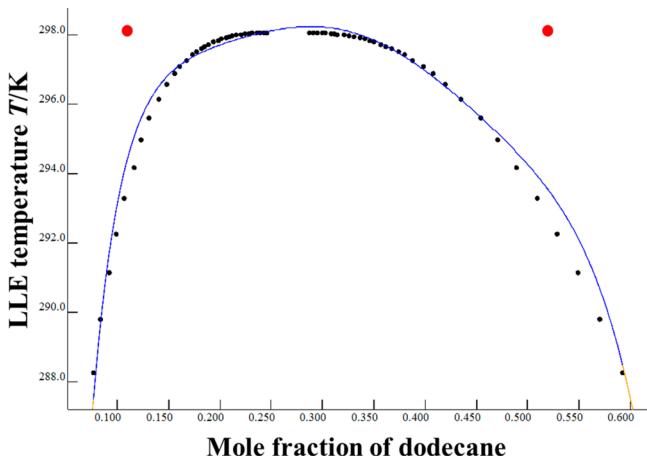
The quality-assessment algorithm reported by us previously<sup>1</sup> for vapor–liquid equilibrium data is a normalized combination of four tests based on the Gibbs–Duhem relationship, plus a test of consistency between vapor pressures (or boiling temperatures) of the components and the binary data at composition extremes. A high-pressure equation of state (EOS) test was subsequently added to expand the range of application.<sup>2</sup> The procedure was essential in automated parameter regressions for the UNIFAC model.<sup>7</sup> An outline of the procedure is shown in Figure 1, with full details given in refs 1 and 2.



**Figure 2.** Comparison of experimental VLE data for the system (dimethylformamide + chlorobenzene) with NIST-Modified UNIFAC<sup>31</sup> prediction. Experimental data: Black circles, Vittal Prasad et al.<sup>32</sup> at pressure  $p = 95.5$  kPa; empty circles, Ramadevi et al.<sup>33</sup> at  $p = 94.9$  kPa; solid line, NIST-Modified UNIFAC<sup>31</sup> at  $p = 95.5$  kPa; dashed line, NIST-Modified UNIFAC at  $p = 94.9$  kPa.



**Figure 3.** Graphical representation of LLE data available in SOURCE Data Archival System<sup>30</sup> for the system (1-butanol + water) for pressure  $p$  near 0.1 MPa. One inconsistent data set is shown in red.<sup>34</sup> The plot was generated with the TDE software.<sup>3</sup> Sources of experimental data (black circles) are listed in the Supporting Information for this article.

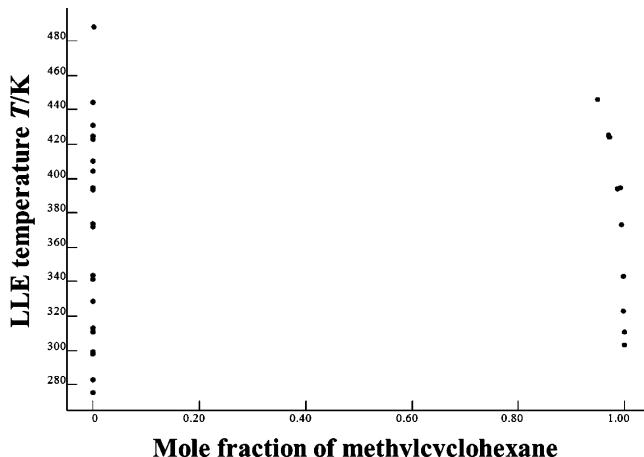


**Figure 4.** Rejection of an inconsistent LLE data set<sup>36</sup> (red circles in the upper left and right of the plot) for the system (dodecane + dimethyl carbonate) with the methods described in the text. The black circles represent the experimental data of ref 37. The plot was generated with the TDE software.<sup>3</sup>

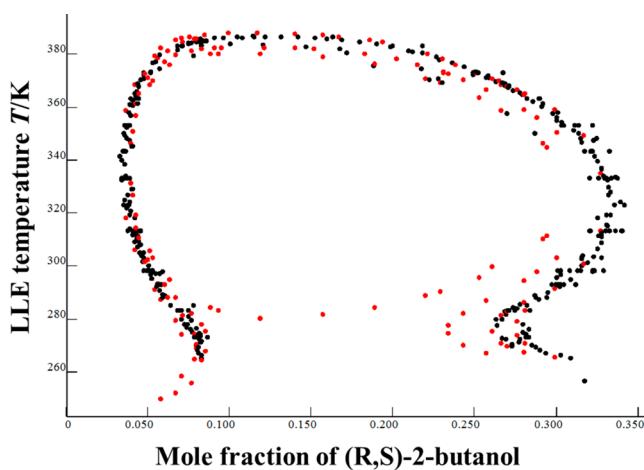
Even though this procedure is useful and applicable for the large majority of cases, limitations arise where experimental data are incomplete and disagree outside the range of their estimated uncertainties. Such a situation is illustrated in Figure 2 for the binary system (*N,N*-dimethylformamide (DMF) + chlorobenzene), for which there are only two published data sets.<sup>32,33</sup> Though pressure conditions for these data sets are very similar ( $p/\text{kPa} = 95.5$  and 94.9), measured boiling temperatures for a given liquid composition are quite different. Both data sets were produced with essentially the same method (a recirculating still), and similar sample purities were claimed. For these data sets, the VLE data quality assessment algorithm is not applicable, due to a lack of vapor-composition measurements. As a result, parameters derived based on these data for the NIST-Modified UNIFAC<sup>31</sup> regression lead to a prediction of boiling temperatures that are simply an average of the two sets. This example shows how the automated analysis can be used to identify systems for which additional measurements are needed.

## 2.2. Consistency Tests for Binary Liquid–Liquid Equilibria Data.

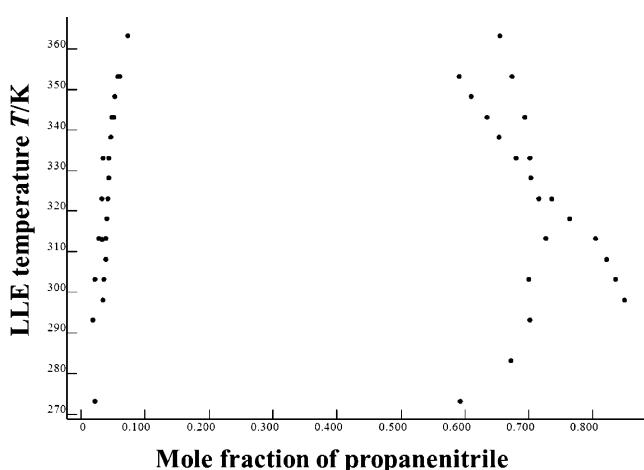
The Gibbs–Duhem equation can be applied



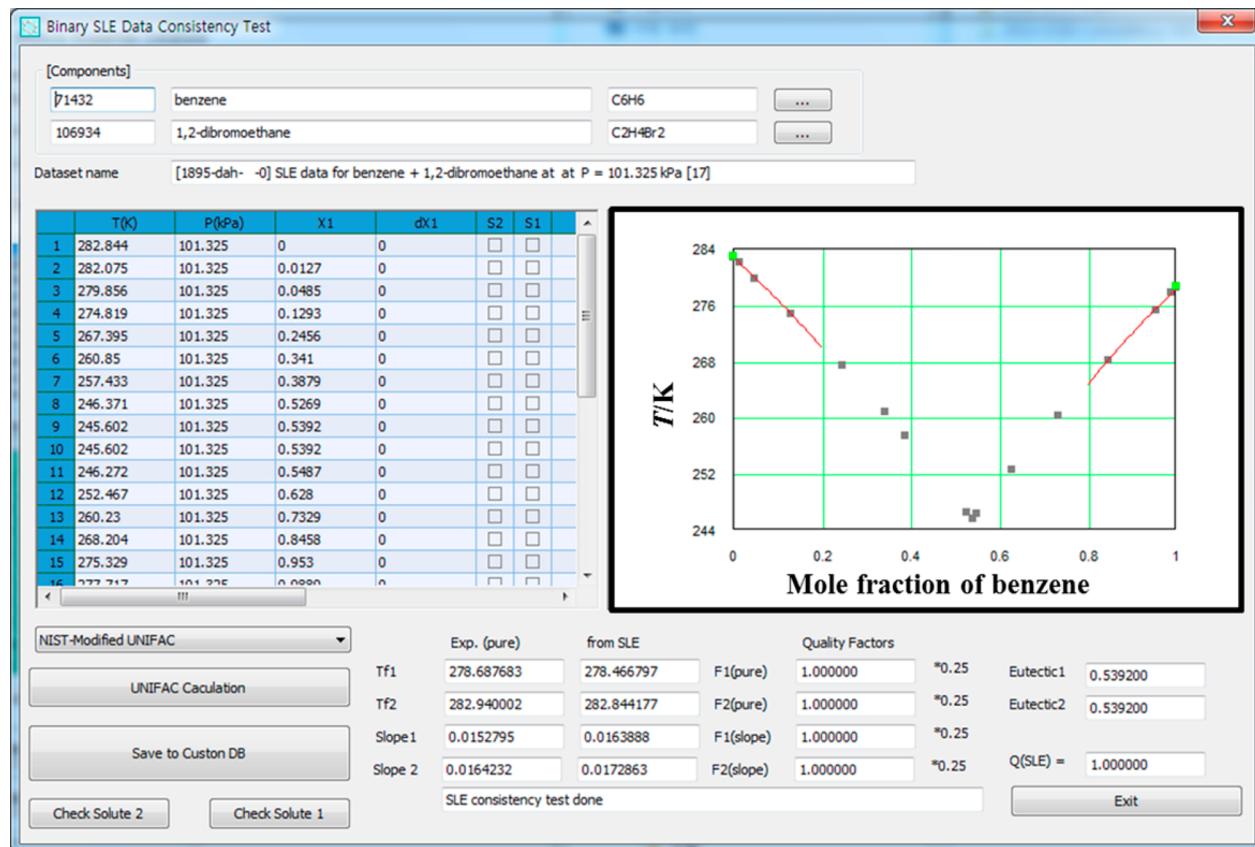
**Figure 5.** Well separated experimental LLE compositions for the system (methylcyclohexane + water). The plot was generated with the TDE software.<sup>3</sup> Sources of experimental data (black circles) are listed in the Supporting Information for this article.



**Figure 6.** Complex LLE for the system (2-butanol + water). Values reported in ref 38 (red circles) are inconsistent with the other reported experimental values. The plot was generated with the TDE software.<sup>3</sup> Sources of experimental data (black circles) are listed in the Supporting Information for this article.



**Figure 7.** LLE data<sup>39–41</sup> for the system (propanenitrile + water) that requires expert review based on the algorithm described in the text. The plot was generated with the TDE software.<sup>3</sup>



**Figure 8.** Demonstration of consistency with the methods described in the text for SLE temperatures for the system (benzene +1,2-dibromoethane) reported in the literature.<sup>42</sup>

to binary liquid–liquid phase-equilibrium data for each phase at constant pressure:

$$x_1^I d \ln \gamma_1^I + x_2^I d \ln \gamma_2^I + \frac{H^{E,I}}{RT^2} dT = 0 \quad (2.1)$$

$$x_1^{II} d \ln \gamma_1^{II} + x_2^{II} d \ln \gamma_2^{II} + \frac{H^{E,II}}{RT^2} dT = 0 \quad (2.2)$$

where  $x$  and  $\gamma$  are mole fractions and activity coefficients, respectively, of components 1 and 2 in liquid phases I and II,  $H^E$  is the excess enthalpy,  $R$  is the gas constant, and  $T$  is temperature.

Because there are two liquids in the mixture and binary LLE data are a function of temperature, excess enthalpy data should be taken into account. Unfortunately, for most systems with reported LLE data, there are no corresponding excess enthalpy data in both liquids, making application of eqs 2.1 and 2.2 impossible. Moreover, these equations do not “bind” (i.e., relate mathematically) compositions of the two liquid phases in equilibrium. However, the dependence of liquid phase composition on temperature at constant, near atmospheric, pressure should be continuous over both liquid phases, and therefore, this requirement can be used in quality assessment. Figure 3 shows experimental LLE data from numerous sources for (1-butanol + water). Clearly, one data set<sup>34</sup> (shown in red) is inconsistent with the rest.

Depending on the particular data scenario (i.e., the extent of experimental data and the degree of consistency), an analysis like that illustrated in Figure 3 leads to either automatic rejection of the inconsistent set or to a programmed warning indicating the

lack of consistency and the need for further expert consideration. The analytical representation used for LLE data was described previously,<sup>35</sup> as was the general algorithm for automatic rejection of inconsistent data sets.<sup>9</sup> As this approach is included in the present larger framework, the algorithm is summarized briefly here.

The rejection procedure is similar to conventional 3-sigma rejection with consideration of local data quality for each value. A value is rejected if it deviates from the model by more than the criterion value  $\Gamma$ , which is calculated for the  $i$ th value as

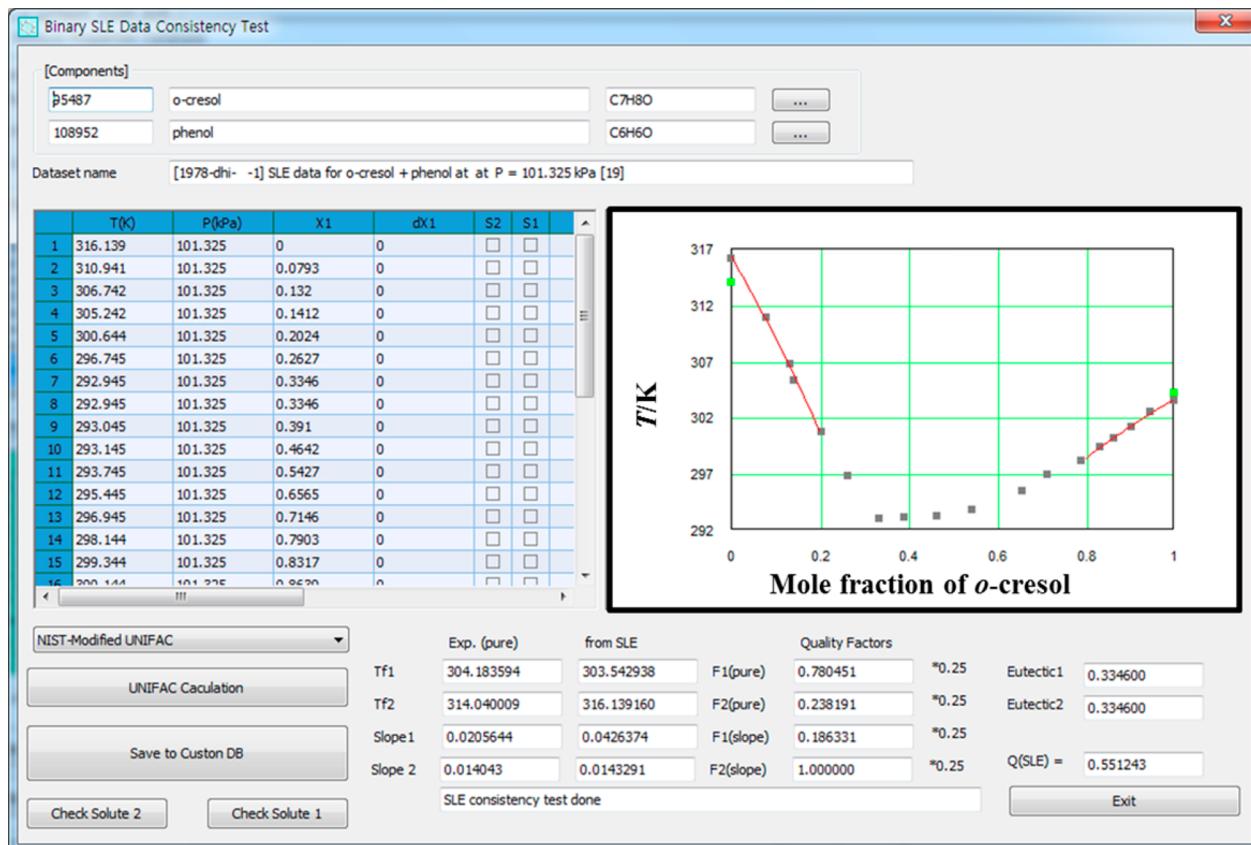
$$\Gamma_i = f \sum_j w_j \Delta_j \exp(-|X_j - X_i|/k) \quad (2.3)$$

where  $f$  is the rejection factor (usually 3),  $w_j$  is the statistical weight of the  $j$ th value with curve deviation  $\Delta_j$ ,  $X_j$  is the state variable, and  $k$  is the damping distance, which is a characteristic length at which the influence of a particular value decreases by a factor of approximately 2.7 (i.e.,  $e$ ). For compositions, the damping distance is, typically, near 0.1 to 0.2 mole fraction. Application of this approach is demonstrated in Figure 4 for a rejected<sup>36</sup> and accepted data set.<sup>37</sup>

In three LLE data scenarios, the “3-sigma” rejection procedure is unsuccessful, and in two of these, the judgment of an expert is required. These cases are characterized by a high value (usually larger than three) of the “inadequacy”  $A$ ,<sup>10</sup> which is defined as

$$A = \frac{\sum_i (\Delta_i / u_i)^2}{n} \quad (2.4)$$

where  $n$  is the number of data points and  $\Delta_i$  and  $u_i$  are the curve deviation and estimated uncertainty of the  $i$ th data point. (A model is said to be inadequate when curve deviations are large relative to the estimated uncertainties of the experimental data.) In the first



**Figure 9.** Demonstration of consistency with the methods described in the text for SLE temperatures for the system (*o*-cresol + phenol) reported in the literature.<sup>43</sup>

case, two liquids are nearly immiscible (Figure 5), and a wide composition gap allows for automated identification of phases and checking of data consistency for each phase separately. Two other cases require the judgment of an expert. These are (1) complex LLE curve shape (see Figure 6, which includes inconsistent values<sup>38</sup> shown in red) or (2) highly inconsistent data<sup>39–41</sup> for which no dominant trend can be derived (see Figure 7 for mole fractions near 0.7). In Figure 7, the results of two studies<sup>40,41</sup> disagree far beyond the estimated range of the estimated uncertainties, making automated data evaluation unreliable. Often, only new reliable measurements can resolve the disagreement.

**2.3. Consistency Tests for Binary Solid–Liquid Equilibria Data.** The large majority of binary organic mixtures form eutectics rather than solid solutions. Solid solutions are found only when the crystal structures of the components are very similar. For eutectic solid–liquid equilibria, the following equation can be used to calculate the solubility of a solid in a liquid with the assumption that heat capacities of the solid and liquid are nearly constant.

$$\ln\left(\frac{1}{\gamma_2 x_2}\right) = \frac{\Delta_c^1 H_m^\circ}{RT_m} \left(\frac{T_m}{T} - 1\right) - \frac{\Delta C_p}{R} \left(\frac{T_m}{T} - 1\right) + \frac{\Delta C_p}{R} \ln\left(\frac{T_m}{T}\right) \quad (2.5)$$

where  $x$  and  $\gamma$  are the mole fraction and activity coefficient of component 2 in the liquid phase,  $T_m$  and  $\Delta_c^1 H_m^\circ$  are the temperature and molar enthalpy of melting,  $\Delta C_p$  is heat capacity of the liquid minus that of the solid at the melting temperature,  $R$  is

the gas constant, and  $T$  is the SLE temperature. Based on eq 2.5, the limiting behavior of the liquidus line should meet the following criteria:

$$T \rightarrow T_{m,1} \text{ as } x_2 \rightarrow 0 \quad (2.6)$$

and

$$\lim_{x_2 \rightarrow 0} \left( \frac{dx_2}{dT} \right)_{eq2.5} = \lim_{x_2 \rightarrow 0} \left( \frac{dx_2}{dT} \right)_{expt} \quad (2.7)$$

The first criterion (eq 2.6) implies that the limiting intercept of the liquidus line should match the melting temperature of the pure component  $T_{m,1}$ , and the second criterion (eq 2.7) requires that the experimental (expt) limiting slope be in accord with the solubility equation (eq 2.5). The experimental limiting slope is estimated with SLE-temperature values for the mole fraction ranges 0 to 0.2 and 0.8 to 1. The following empirical solubility equation was used:

$$\ln x_2 = A + BT + C \ln T \quad (2.8)$$

where  $A$ ,  $B$ , and  $C$  are fitted parameters.

Another consistency requirement for SLE arises from the cryoscopic relation:

$$\left( \frac{dT}{dx_1} \right)_{SLE x_1=1} = - \frac{RT^2}{\Delta_c^1 H_m^\circ} \quad (2.9)$$

which is based on the SLE condition

$$\Delta_c^1 G_{m,1}^\circ = 0 \quad (2.10)$$

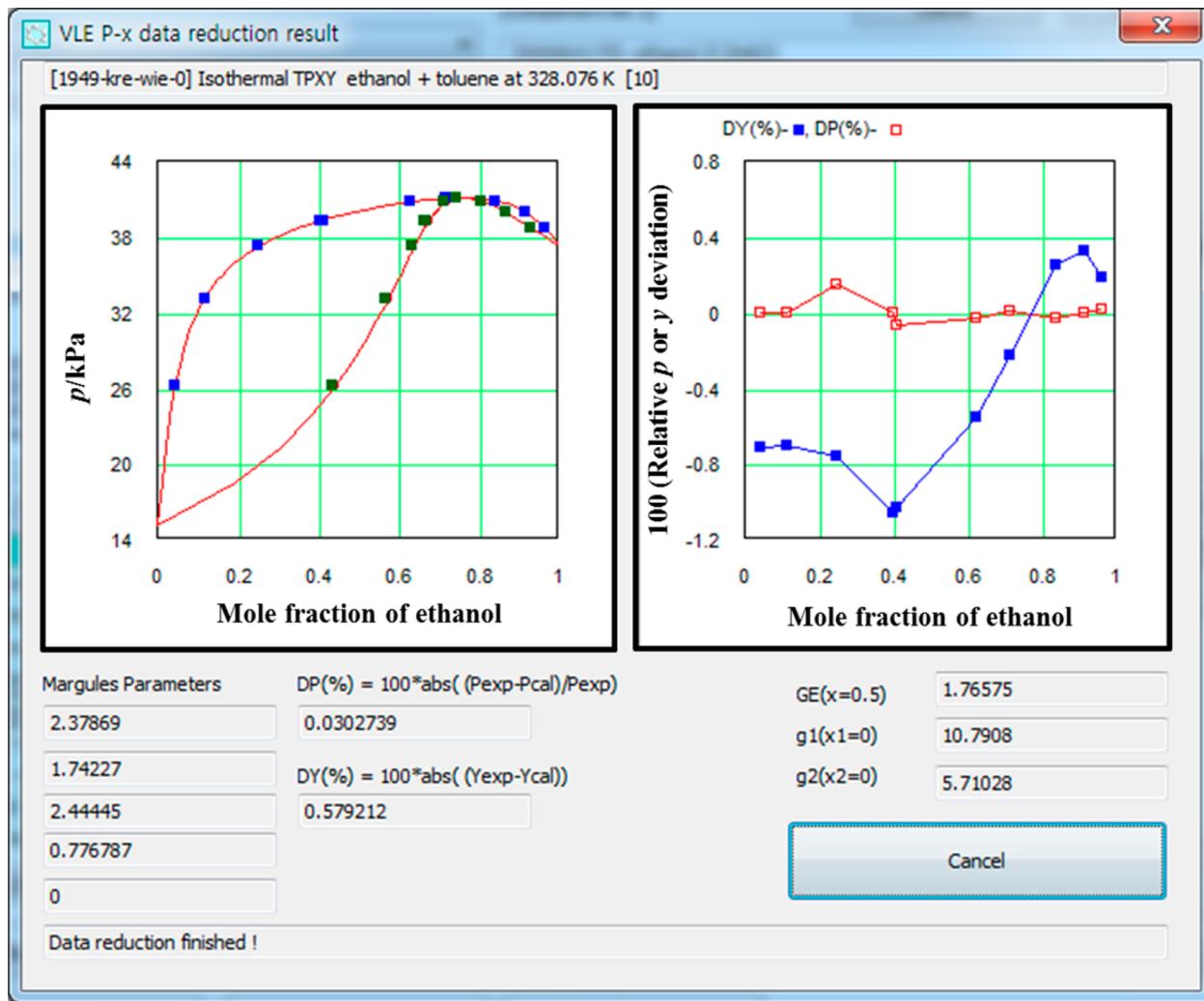


Figure 10. Results of VLE data reduction for the system (ethanol + methylbenzene), using the 4-parameter Margules equation.<sup>45</sup> Sources of the experimental data shown in the plots are listed in the Supporting Information for this article.

and its differential:

$$d\frac{\Delta_c^1 G_{m,1}^\circ}{RT} = \frac{\Delta_c^1 H_{m,1}^\circ}{RT^2} dT + \left( \frac{\partial \ln(\gamma_1 x_1)}{\partial x_1} \right)_T dx_1 \quad (2.11)$$

where  $\Delta_c^1 G_{m,1}^\circ$  is the Gibbs energy of transfer of component 1 from the solid to liquid state at temperature  $T$  and mole fraction  $x_1$  of component 1;  $\Delta_c^1 H_{m,1}^\circ$  is the enthalpy of transfer of component 1 from the solid to liquid state (i.e., the enthalpy of melting, if  $x_1 = 1$ ); and  $\gamma_1$  is the activity coefficient of component 1 ( $\gamma_1 = 1$  for  $x_1 = 1$ ), using the implicit assumption of common mixture models (e.g., UNIFAC,<sup>4</sup> UNIQUAC,<sup>21</sup> and NRTL<sup>20</sup>):

$$\left( \frac{\partial \gamma_1}{\partial x_1} \right)_T = 0 \text{ at } x_1 = 1 \quad (2.12)$$

A “quality factor” for SLE,  $Q_{\text{SLE}}$  (in analogy to those previously introduced for VLE data sets<sup>1</sup>), can be computed based on differences between values calculated from solubility data and the pure component properties.

At the first step for the assessment of SLE data, measured liquidus temperatures are split into two sets separated at the eutectic composition, and data near the composition limits are fitted with the solubility equation (eq 2.8). Values calculated with

eq 2.8 (based on the binary SLE data) and eq 2.5 (based on pure-component properties) are compared, and the quality factor  $Q_{\text{SLE}}$  is calculated.

$$Q_{\text{SLE}} = 0.25F_{1,T_m} + 0.25F_{2,T_m} + 0.25F_{1,\text{slope}} + 0.25F_{2,\text{slope}} \quad (2.13)$$

where 1 and 2 indicate the component number, and  $F$  is a function of two quantities;

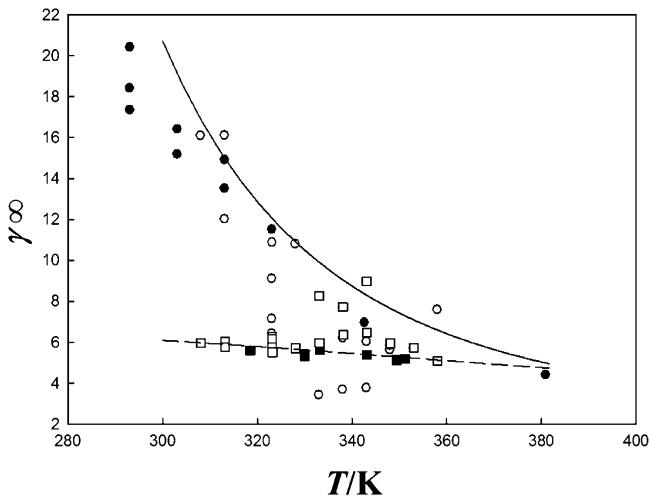
$$\Delta T_{m,i} = |T_{m,\text{expt}} - T_{m,\text{eq2.8}}| \quad (2.14)$$

$$\Delta \text{slope}_i = \left| \lim_{x_i \rightarrow 0} \left( \frac{dx_i}{dT} \right)_{\text{eq2.5}} - \lim_{x_i \rightarrow 0} \left( \frac{dx_i}{dT} \right)_{\text{expt}} \right| \quad (2.15)$$

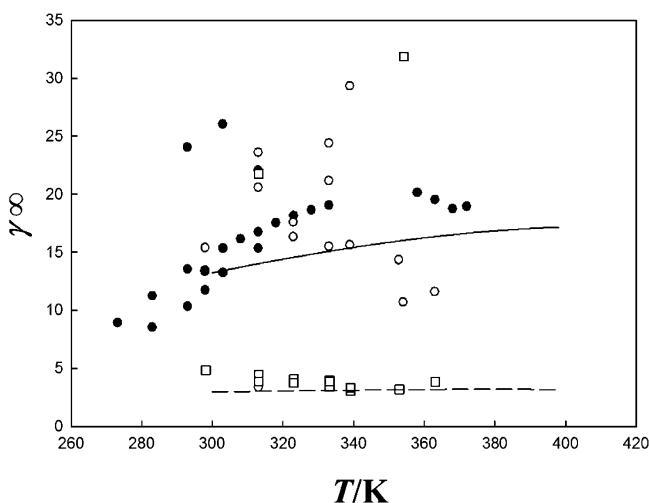
where  $i$  is the component number.  $T_{m,\text{eq2.8}}$  is calculated by extrapolation of eq 2.8 to zero mole fraction.  $T_{m,\text{expt}}$  is the critically evaluated normal melting temperature, based on pure-component data alone. The four quantities  $F_{1,T_m}$ ,  $F_{2,T_m}$ ,  $F_{1,\text{slope}}$ , and  $F_{2,\text{slope}}$  are evaluated as follows.

$$F_{1,T_m} = 1 \text{ when } \Delta T_{m,1} \leq 0.5 \text{ K, and}$$

$$F_{1,T_m} = 0.5 / \Delta T_{m,1} \text{ when } \Delta T_{m,1} > 0.5 \text{ K} \quad (2.16)$$



**Figure 11.** Activity coefficients at infinite dilution  $\gamma^\infty$  for the system [ethanol (1) + methylbenzene (2)]. Unfilled circles, values derived with VLE data for component 1; unfilled squares, values derived with VLE data for component 2; black circles, direct measurements of  $\gamma^\infty$  for component 1; black squares, direct measurements of  $\gamma^\infty$  for component 2. Lines represent values predicted with NIST-Modified UNIFAC<sup>31</sup> (solid line—component 1, dashed line - component 2). Sources of experimental data are listed in the Supporting Information for this article.



**Figure 12.** Infinite dilution activity coefficients  $\gamma^\infty$  for the system [propan-1-ol (1) + water (2)]. Unfilled circles, values derived with VLE data for component 1; unfilled squares, values derived with VLE data for component 2; black circles, direct measurements of  $\gamma^\infty$  for component 1. Lines represent values predicted with NIST-Modified UNIFAC<sup>31</sup> (solid line—component 1, dashed line - component 2). Sources of experimental data are listed in the Supporting Information for this article.

$$F_{2,T_m} = 1 \text{ when } \Delta T_{m,2} \leq 0.5 \text{ K, and}$$

$$F_{2,T_m} = 0.5/\Delta T_{m,2} \text{ when } \Delta T_{m,2} > 0.5 \text{ K} \quad (2.17)$$

$$\begin{aligned} F_{1,\text{slope}} &= 1 \text{ when } \Delta \text{slope}_1 \\ &\leq 0.2, \text{ and } F_{1,\text{slope}} \\ &= 0.2/\Delta \text{slope}_1 \text{ when } \Delta \text{slope}_1 \\ &> 0.2 \end{aligned} \quad (2.18)$$

$$\begin{aligned} F_{2,\text{slope}} &= 1 \text{ when } \Delta \text{slope}_2 \\ &\leq 0.2, \text{ and } F_{2,\text{slope}} \\ &= 0.2/\Delta \text{slope}_2 \text{ when } \Delta \text{slope}_2 \\ &> 0.2 \end{aligned} \quad (2.19)$$

If the number of data points near the composition extremes (i.e., within 0.2 mole fraction of the pure components) is less than four, the quality factors for slope ( $F_{i,\text{slope}}$ ) are set to 0.5. Figure 8 shows the results of the SLE consistency test for the mixture (benzene + 1,2-dibromoethane).<sup>42</sup> (Figures 8, 9, 10, 13, and 14 were generated with an in-house tool developed to implement the algorithm described in this article.) Good agreement is demonstrated between the binary SLE and pure-component data, and the calculated quality factor  $Q_{\text{SLE}}$  is 1. Figure 9 shows a less consistent data set for (*o*-cresol + phenol),<sup>43</sup> where the normal melting temperature of phenol derived from the SLE data ( $T_m = 316.1$  K) deviates significantly from the value based on pure-component data ( $T_m = 314.0$  K). Here, the quality factor  $Q_{\text{SLE}}$  was calculated to be near 0.55.

**2.4. Cross-Check between ACT and VLE data.** Activity coefficients at infinite dilution (ACT) are auxiliary information in support of VLE data, and they are of particular importance in calculation of mutual solubility, as well as evaluation of UNIFAC parameters.<sup>25–29</sup> Van Ness et al.<sup>44</sup> proposed methods for VLE data reduction and consistency assessment. Those methods provide guidance for extrapolation of the excess Gibbs energy at a defined temperature to the infinite dilution limit. We described implementation of a similar procedure previously.<sup>1</sup> If experimentally determined activity coefficients at infinite dilution are available, these can be compared with extrapolated values based on VLE measurements. The results of such a comparison for the system (ethanol + methylbenzene) are shown in Figure 10. The 4-parameter Margules equation<sup>45</sup> was used to model the VLE data.

A comparison of experimentally measured and extrapolated activity coefficients at infinite dilution is illustrated for (ethanol + methylbenzene) in Figure 11, where agreement between the data sources is seen to be good. The results can be used for automatic detection of highly inconsistent data; however, in many instances, uncertainties for the extrapolated values are large, making such an analysis difficult. This situation is illustrated for the (propan-1-ol + water) system in Figure 12, where agreement between the two methods is only qualitative, and large scatter is seen in all of the reported values. As demonstrated in Figure 12, the predictive capability of the UNIFAC method for activity coefficients at infinite dilution is limited. In summary, consistency assessment based on comparisons between directly measured activity coefficients at infinite dilution and those based on extrapolated VLE data is often inconclusive due to large uncertainties in the extrapolations.

**2.5. Cross-Check between HE and VLE data.** Olson<sup>46</sup> proposed a procedure to compare the temperature dependence of the excess Gibbs energy with that of excess enthalpy data for use in cross checking HE and VLE data. Based on the second law of thermodynamics

$$G^E/RT = H^E/RT - S^E/R \quad (2.20)$$

where  $G^E$ ,  $H^E$ , and  $S^E$  are the excess Gibbs energy, excess enthalpy, and excess entropy. Equation 2.20 indicates that the temperature dependence of the excess Gibbs energy should be the same as that of the excess enthalpy at a given temperature.

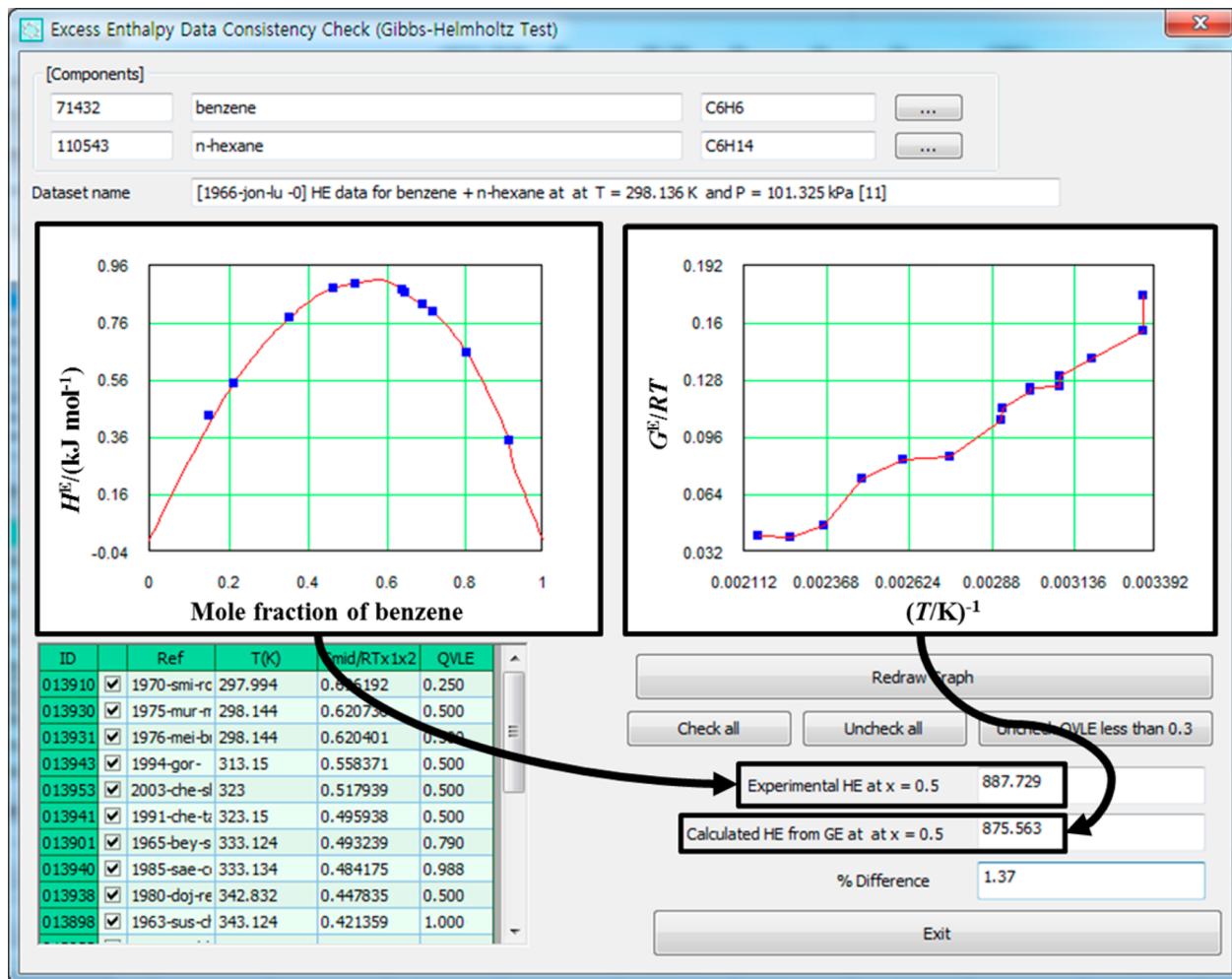


Figure 13. Example of successful application of the Olson<sup>46</sup> Gibbs–Helmholtz consistency test for excess enthalpy and VLE data for the system (benzene + hexane). Sources of experimental data are listed in the Supporting Information for this article.

For a binary mixture, the excess Gibbs energy is related to the activity coefficient by the following equation.

$$G^E/RT = x_1 \ln \gamma_1 + x_2 \ln \gamma_2 \quad (2.21)$$

where mole fractions  $x$  and activity coefficients  $\gamma$  are for components 1 and 2.

Olson<sup>46</sup> suggested that the slope of excess Gibbs energy plotted against temperature at the composition midpoint ( $x_1 = 0.5$ ) could be interpreted as the excess enthalpy for that composition and mean temperature.

A test criterion for testing of consistency between experimentally determined excess enthalpies  $H_{\text{expt}}^E$ , and those determined from VLE data,  $H_{\text{calc(VLE)}}^E$ , was established as follows:

$$0.5H_{\text{expt}}^E \leq H_{\text{calc(VLE)}}^E \leq 1.5H_{\text{expt}}^E \quad (2.22)$$

The value of the excess Gibbs energy at the midcomposition can be obtained with the method of Van Ness et al.,<sup>44</sup> and the results are demonstrated in Figure 13 for the system (benzene + hexane). The excess enthalpy and excess Gibbs energy were fit with a spline function, and the excess enthalpy data were interpolated to obtain the value at mole fraction 0.5. The resulting values of  $H_{\text{expt}}^E = 888 \text{ J}\cdot\text{mol}^{-1}$  and  $H_{\text{calc(VLE)}}^E = 876 \text{ J}\cdot\text{mol}^{-1}$ , for temperature  $T/K = 298.15$ , are in very good agreement.

Unfortunately, applicability of this test is limited, particularly when the absolute value of the excess Gibbs energy is small (i.e., when the solution is near ideal). In such cases, the test can become unreliable because variations in the excess Gibbs energy can be significantly larger than its absolute value. An example is shown Figure 14, where  $H_{\text{expt}}^E$  for (ethanol + propan-1-ol) is more than 40 times smaller than that for the system represented in Figure 13 (benzene + hexane), and the slope of the excess Gibbs energy with temperature (right side of Figure 14) cannot be evaluated with any reliability.

## 2.6. Othmer–Tobias Test for Ternary Mixtures.

Consistency of LLE data for ternary mixtures can be tested by applying the Othmer and Tobias procedure.<sup>47</sup> This procedure consists of establishing tie lines for a binodal curve of ternary LLE through use of partial pressures of the components over a solution. If tie lines are also determined by analysis of the phases, this method can be used to test consistency between vapor pressures, vapor composition, and composition of the liquid phases at LLE. Although useful, the number of chemical systems to which it can be applied is limited because of the diverse and extensive data required.

**2.7. Cross-Check between Ternary and Binary Phase-Equilibrium Data (LLE and VLE).** Another test for LLE data involves cross-checking between ternary LLE and data available for the binary subsystems. The approach is illustrated in

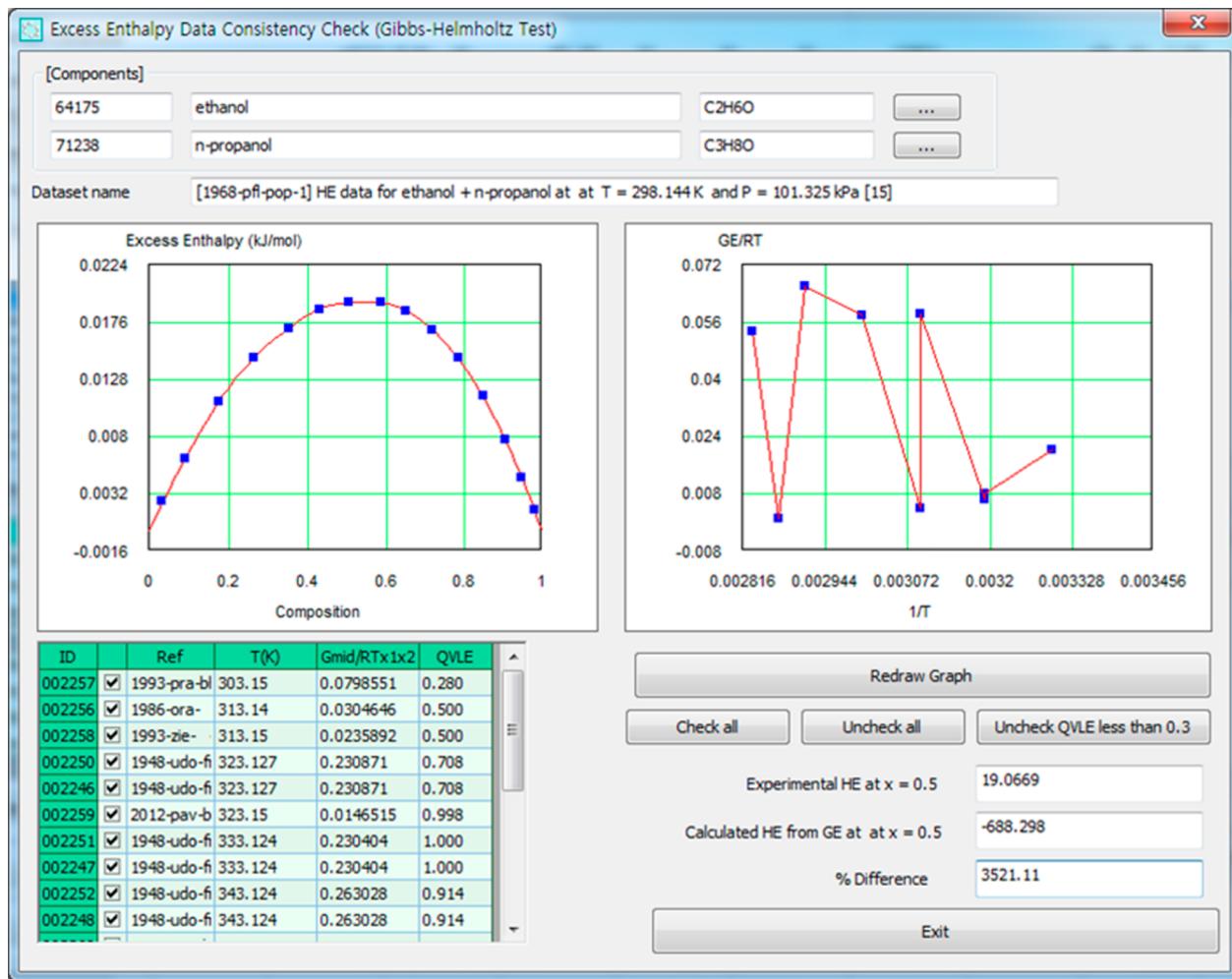


Figure 14. Example of unsuccessful application of the Olson<sup>46</sup> Gibbs–Helmholtz consistency test for excess enthalpy and VLE data for the system (ethanol + propan-1-ol). Sources of experimental data are listed in the Supporting Information for this article.

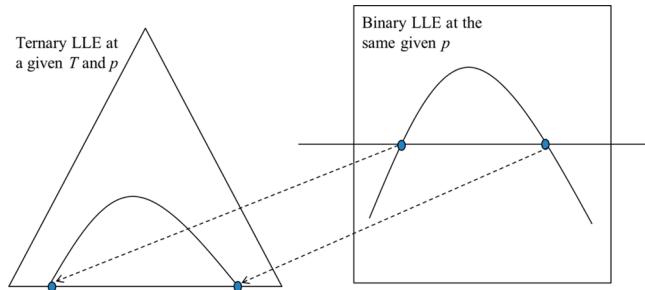


Figure 15. Schematic representation of the cross-checking procedure for ternary and binary LLE data. The triangle represents a three-component phase diagram at a given pressure  $p$  (typically,  $p \approx 0.1$  MPa) with pure components located at the corners. The figure at the right represents a plot of LLE temperature against composition for the binary mixture represented at the base of the 3-component triangle for the same given pressure. The horizontal line in the figure at the right represents the temperature of the 3-component equilibrium plotted on the left. The dashed lines connect values that are common to both plots.

Figure 15. A similar procedure can be applied to ternary VLE data, as well. In addition, procedures implemented in TDE<sup>3</sup> check end points for consistency with pure-component property values and overall consistency for properties described by an

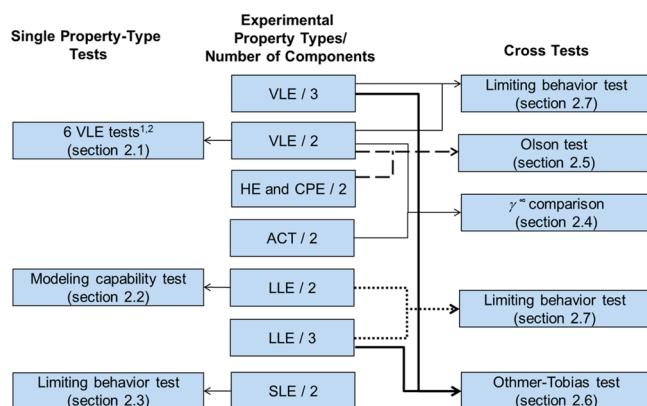
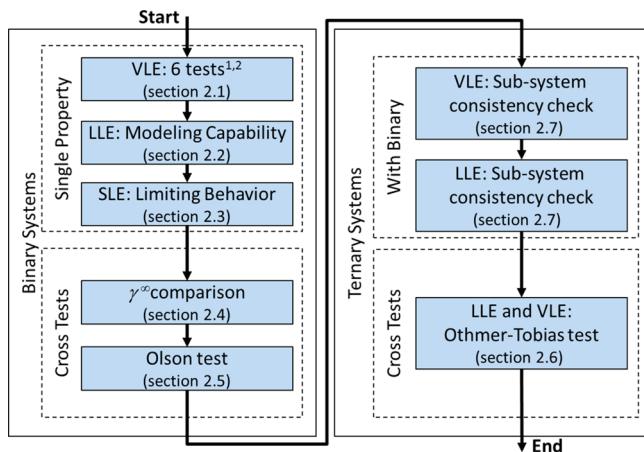


Figure 16. Experimental property types assessed are shown in the center column of the figure. The arrows connect the property types to the related quality-assessment and consistency tests. The various line types (dashed, bold, etc.) are provided as an aid to the eye. “Cross tests” involve more than one property type. The “section” number indicates the location in the text where the test is discussed.

EOS or AC model.<sup>11</sup> For an AC model, apparent inconsistency, as reflected by large deviations of the experimental data from equations, may be caused by erroneous data or an inadequate model.



**Figure 17.** Algorithmic implementation of the phase equilibria test framework. The “section” number indicates the location in the text where the test is discussed.

**2.8. Phase-Equilibrium Consistency Test Framework.** A summary of the various phase equilibrium data-quality assessment procedures described in the text is given in Figure 16. The sequence for application of the tests is shown in Figure 17. This sequence provides the basis for algorithmic implementation of the framework.

### 3. CONCLUSIONS

Assessment procedures for various types of phase-equilibrium data were reviewed and analyzed. These procedures are critical for independent validation of experimental data and development of high quality prediction and correlation models. A consistency test framework involving all types of phase-equilibria data was developed, and an algorithmic implementation is proposed. This approach was used for automated determination of weighting factors, data set rejection, and identification of chemical systems requiring expert review for thousands of phase-equilibrium experimental data sets in generation of NIST-Modified UNIFAC<sup>31</sup> parameters.

### ■ ASSOCIATED CONTENT

#### S Supporting Information

Sources for the data used in various figures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

### ■ AUTHOR INFORMATION

#### Corresponding Author

\*E-mail: frenkel@boulder.nist.gov.

#### Notes

The authors declare no competing financial interest.

<sup>§</sup>Guest Researcher (2009–2010) at the National Institute of Standards and Technology, Boulder, Colorado 80305–3337, USA.

### ■ REFERENCES

- (1) Kang, J. W.; Diky, V.; Chirico, R. D.; Magee, J. W.; Muzny, C. D.; Abdulagatov, I.; Kazakov, A. F.; Frenkel, M. Quality Assessment Algorithm for Vapor-Liquid Equilibrium Data. *J. Chem. Eng. Data* **2010**, *55*, 3631–3640.
- (2) Diky, V.; Chirico, R. D.; Muzny, C. D.; Kazakov, A. F.; Kroenlein, K.; Magee, J. W.; Abdulagatov, I.; Jeong Won Kang, I.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 7. Ternary Mixtures. *J. Chem. Inf. Model.* **2012**, *S2*, 260–276.

(3) Frenkel, M.; Chirico, R. D.; Diky, V.; Kroenlein, K.; Muzny, C. D.; Kazakov, A. F.; Magee, J. W.; Abdulagatov, I. M.; Lemmon, E. W. NIST ThermoData Engine, NIST Standard Reference Database 103b-Pure Compounds, Binary Mixtures, and Chemical Reactions, version 8.0; Standard Reference Data Program; National Institute of Standards and Technology: Gaithersburg, MD, 2013.

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

(5) Fredenslund, A.; Gmehling, J.; Rasmussen, P. *Vapor Liquid Equilibrium Using UNIFAC*; Elsevier: Amsterdam, 1977.

(6) Hansen, H. K.; Rasmussen, P.; Fredenslund, A.; Schiller, M.; Gmehling, J. *Ind. Eng. Chem. Res.* **1991**, *30*, 2352–2355.

(7) Kang, J. W.; Diky, V.; Chirico, R. D.; Magee, J. W.; Muzny, C. D.; Abdulagatov, I.; Kazakov, A. F.; Frenkel, M. A New Method for Evaluation of UNIFAC Interaction Parameters. *Fluid Phase Equilib.* **2011**, *309*, 68–75.

(8) (a) Wilhoit, R. C.; Marsh, K. N. Future Directions for Data Compilation. *Int. J. Thermophys.* **1999**, *20*, 247–255. (b) Frenkel, M. Dynamic Compilation: A Key Concept for Future Thermophysical Data Evaluation. In: Forum 2000: Fluid Properties for New Technologies—Connecting Virtual Design with Physical Reality; Rainwater, J. C., Friend, D. G., Hanley, M., Harvey, A. H., Holcomb, C. D., Laesecke, A., Magee, J. W., Muzny, C. D., Eds.; *NIST Special Publication 975*; National Institute of Standards and Technology: Gaithersburg, MD, 2001; pp 83–84.

(9) Frenkel, M.; Chirico, R. D.; Diky, V.; Xinjian, Y.; Qian, D.; Muzny, C. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. *J. Chem. Inf. Model.* **2005**, *45*, 816–838.

(10) Diky, V.; Muzny, C. D.; Lemmon, E. W.; Chirico, R. D.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 2: Equations of State on Demand and Dynamic Updates over the Web. *J. Chem. Inf. Model.* **2007**, *47*, 1713–1725.

(11) Diky, V.; Chirico, R. D.; Kazakov, A. F.; Muzny, C. D.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 3: Binary Mixtures. *J. Chem. Inf. Model.* **2009**, *49*, 503–517.

(12) Diky, V.; Chirico, R. D.; Kazakov, A. F.; Muzny, C. D.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 4: Chemical Reactions. *J. Chem. Inf. Model.* **2009**, *49*, 2883–2896.

(13) Diky, V.; Chirico, R. D.; Kazakov, A. F.; Muzny, C. D.; Magee, J. W.; Abdulagatov, I.; Kang, J. W.; Kroenlein, K.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 5: Experiment Planning and Product Design. *J. Chem. Inf. Model.* **2011**, *51*, 181–194.

(14) Kroenlein, K.; Muzny, C. D.; Diky, V.; Kazakov, A. F.; Chirico, R. D.; Magee, J. W.; Abdulagatov, I.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 6: Dynamic Web-Based Data Dissemination through the NIST Web Thermo Tables. *J. Chem. Inf. Model.* **2011**, *51*, 1506–1512.

(15) Diky, V.; Chirico, R. D.; Muzny, C. D.; Kazakov, A. F.; Kroenlein, K.; Magee, J. W.; Abdulagatov, I.; Kang, J. W.; Gani, R.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 8: Properties of Material Streams and Solvent Design. *J. Chem. Inf. Model.* **2013**, *53*, 249–266.

(16) Diky, V.; Chirico, R. D.; Muzny, C. D.; Kazakov, A. F.; Kroenlein, K.; Magee, J. W.; Abdulagatov, I.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 9: Extensible Thermodynamic Constraints for Pure Compounds and New Model Developments. *J. Chem. Inf. Model.* **2013**, *53*, 3418–3430.

(17) Hajipour, S.; Satyro, M. A.; Foley, M. W. Uncertainty Analysis Applied to Thermodynamic Models and Process Design. 2. Binary Mixtures. *Fluid Phase Equilib.* **2014**, *364*, 15–30.

- (18) Cunico, L. P.; Ceriani, R.; Sarup, B.; O'Connell, J. P.; Gani, R. Data, Analysis and Modeling of Physical Properties for Process Design of Systems Involving Lipids. *Fluid Phase Equilib.* **2014**, *362*, 318–327.
- (19) Poling, B. E.; Prausnitz, J. M.; O'Connell, J. P. The Properties of Gases and Liquids, 5th ed.; McGraw-Hill: New York, 2001; Chapter 8.1.
- (20) Renon, H.; Prausnitz, J. M. Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures. *AIChE J.* **1968**, *14*, 135–144.
- (21) Abrams, D. S.; Prausnitz, J. M. Statistical Thermodynamics of Liquid Mixtures: A New Expression for the Excess Gibbs Energy of Partly or Completely Miscible Systems. *AIChE J.* **1975**, *21*, 116–128.
- (22) Gmehling, J.; Wittig, R.; Lohmann, J.; Joh, R. A Modified UNIFAC (Dortmund) Model. 4. Revision and Extension. *Ind. Eng. Chem. Res.* **2002**, *41*, 1678–1688.
- (23) Chirico, R. D.; Frenkel, M.; Magee, J. W.; Diky, V.; Muzny, C. D.; Kazakov, A. F.; Kroenlein, R.; Abdulagatov, I.; Hardin, G. R.; Acree, W. E., Jr.; Brenneke, J. F.; Brown, P. L.; Cummings, P. T.; De Loos, T. W.; Friend, D. G.; Goodwin, A. R. H.; Hansen, L. D.; Haynes, W. M.; Koga, N.; Mandelis, A.; Marsh, K. N.; Mathias, P. M.; McCabe, C.; O'Connell, J. P.; Pádua, A.; Rives, V.; Schick, C.; Trusler, J. P. M.; Vyazovkin, S.; Weir, R. D.; Wu, J. Improvement of Quality in Publication of Experimental Thermophysical Property Data: Challenges, Assessment Tools, Global Implementation, and Online Support. *J. Chem. Eng. Data* **2013**, *58*, 2699–2716.
- (24) Snow, R. L.; Ott, J. B.; Goates, J. R.; Marsh, K. N.; O'Shea, S.; Stokes, R. H. (Solid + liquid) and (vapor + liquid) phase equilibria and excess enthalpies for (benzene + n-tetradecane), (benzene + n-hexadecane), (cyclohexane + n-tetradecane), and (cyclohexane + n-hexadecane) at 293.15, 298.15, and 308.15 K. Comparison of  $G_m^E$  calculated from (vapor + liquid) and (solid + liquid) equilibria. *J. Chem. Thermodyn.* **1986**, *18*, 107–130.
- (25) Weidlich, U.; Gmehling, J. A Modified UNIFAC Model. 1. Prediction of VLE,  $h^E$ , and  $\gamma^\infty$ . *Ind. Eng. Chem. Res.* **1987**, *26*, 1372–1381.
- (26) Gmehling, J.; Li, J.; Schiller, M. A Modified UNIFAC Model. 2. Present Parameter Matrix and Results for Different Thermodynamic Properties. *Ind. Eng. Chem. Res.* **1993**, *32*, 178–193.
- (27) Gmehling, J.; Lohmann, J.; Jakob, A.; Li, J.; Joh, R. A Modified UNIFAC (Dortmund) Model. 3. Revision and Extension. *Ind. Eng. Chem. Res.* **1998**, *37*, 4876–4882.
- (28) Gmehling, J.; Wittig, R.; Lohmann, J.; Joh, R. A Modified UNIFAC (Dortmund) Model. 4. Revision and Extension. *Ind. Eng. Chem. Res.* **2002**, *41*, 1678–1688.
- (29) Jakob, A.; Gremseemann, H.; Lohmann, J.; Gmehling, J. Further Development of Modified UNIFAC (Dortmund): Revision and Extension 5. *Ind. Eng. Chem. Res.* **2006**, *45*, 7924–7933.
- (30) Kazakov, A.; Muzny, C. D.; Kroenlein, K.; Diky, V.; Chirico, R. D.; Magee, J. W.; Abdulagatov, I. M.; Frenkel, M. NIST/TRC SOURCE Data Archival System: The Next-Generation Data Model for Storage of Thermophysical Properties. *Int. J. Thermophys.* **2012**, *33*, 22–33.
- (31) Kang, J. W.; Diky, V.; Kroenlein, K.; Magee, J. W.; Muzny, C. D.; Abdulagatov, I.; Kazakov, A. F.; Frenkel, M. *Fluid Phase Equilib.* to be submitted.
- (32) Vittal Prasad, T. E.; Adi Sankara Reddy, A.; Kailash, S.; Prasad, D. H. L. Activity Coefficients and Excess Gibbs Energy of Binary Mixtures of N,N-Dimethyl Formamide with Selected Compounds at 95.5 kPa. *Fluid Phase Equilib.* **2008**, *273*, 52–58.
- (33) Ramadevi, R. S.; Venkatesu, P.; Prabhakara Rao, M. V.; Ramakrishna, M. *Thermochim. Acta* **1996**, *255*, 133–144.
- (34) Berkengem, T. I. Application of the Fischer Method for the Determination of the Solubility of Water in Organic Compounds and in their Mixtures at Various Temperatures. *Zavod. Lab.* **1941**, *10*, 592.
- (35) Chirico, R. D.; Diky, V.; Magee, J. W.; Frenkel, M.; Marsh, K. N. Thermodynamic and Thermophysical Properties of the Reference Ionic Liquid: 1-Hexyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]-amide (Including Mixtures). Part 2. Critical Evaluation and Recommended Property Values (IUPAC Technical Report). *Pure Appl. Chem.* **2009**, *81*, 791–828.
- (36) Regueiro, J. M. P. Propiedades Termodinámicas de Mezclas Binarias: Carbonato de Dialquilo + Hidrocarburo. Ph.D. Thesis; University of Vigo: Ourense, Spain, 1999.
- (37) Huang, M.; Chen, Z.; Yin, T.; An, X.; Shen, W. Critical Behavior of Binary Mixtures of Dimethyl Carbonate + Nonane and Dimethyl Carbonate + Dodecane: Measurements of the Coexistence Curves. *J. Chem. Eng. Data* **2011**, *56*, 2349–2355.
- (38) Dolgolenko, W. The lower critical solubility temperature of two liquids. *Z. Phys. Chem., Stoechiom. Verwandschaftsl.* **1908**, *62*, 499.
- (39) Van Arkel, A. E.; Vles, S. E. Solubility of Organic Compounds in Water. *Recl. Trav. Chim. Pays-Bas* **1936**, *55*, 407–411.
- (40) Stephenson, R. M. Mutual Solubility of Water and Nitriles. *J. Chem. Eng. Data* **1994**, *39*, 225–227.
- (41) Grande, M. D. C.; Marschoff, C. M. Liquid-liquid Equilibria of Water + Benzonitrile + Acetonitrile or + Propionitrile. *J. Chem. Eng. Data* **2000**, *45*, 686–688.
- (42) Dahms, A. Über die Gefrierpunkte einiger binärer Gemenge heteromorpher Substanzen (Sur les points de congélation de quelques mélanges binaires de substances hétéromorphes). *Wied. Ann.* **1895**, *54*, 486.
- (43) Dhillon, M. S. Solid + liquid phase diagrams of mixtures containing phenol, cresols and dihydroxyphenols. *Z. Phys. Chem. (Leipzig)* **1978**, *259*, 243–248.
- (44) Van Ness, H. C.; Byer, S. M.; Gibbs, R. E. Vapor-liquid equilibrium: Part I. An Appraisal of Data Reduction Methods. *AIChE J.* **1973**, *19*, 238–244.
- (45) Margules, M. Über die Zusammensetzung der gesättigten Dämpfe von Misschungen. *Sitzungsber. Akad. Wiss. Wien, Math.-Naturwiss. Kl.* **1895**, *104*, 1243–1278.
- (46) Olson, J. D. Thermodynamic Consistency Testing of PTx-data via the Gibbs-Helmholtz Equation. *Fluid Phase Equilib.* **1983**, *14*, 383–392.
- (47) Othmer, D. F.; Tobias, P. E. Partial Pressures of Ternary Liquid Systems and the Prediction of Tie Lines. *Ind. Eng. Chem.* **1942**, *34*, 696–700.