

# ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 5. Experiment Planning and Product Design<sup>§</sup>

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

Thermophysical Properties Division, National Institute of Standards and Technology,  
Boulder, Colorado 80305, United States and Department of Chemical and Biological Engineering,  
Korea University, 5-1 Anam-Dong, Sungbuk-Ku, Seoul 136-701, South Korea

Received September 21, 2010

ThermoData Engine (TDE) is the first full-scale software implementation of the dynamic data evaluation concept, as reported recently in this journal. In the present paper, we describe development of an algorithmic approach to assist experiment planning through assessment of the existing body of knowledge, including availability of experimental thermophysical property data, variable ranges studied, associated uncertainties, state of prediction methods, and parameters for deployment of prediction methods and how these parameters can be obtained using targeted measurements, etc., and, indeed, how the intended measurement may address the underlying scientific or engineering problem under consideration. A second new feature described here is the application of the software capabilities for aid in the design of chemical products through identification of chemical systems possessing desired values of thermophysical properties within defined ranges of tolerance. The algorithms and their software implementation to achieve this are described. Finally, implementation of a new data validation and weighting system is described for vapor–liquid equilibrium (VLE) data, and directions for future enhancements are outlined.

## 1. INTRODUCTION

As was discussed in the first four papers of this series,<sup>1–4</sup> the NIST ThermoData Engine (TDE) software represents the first full-scale implementation of the dynamic data evaluation concept for thermophysical properties.<sup>5,6</sup> This concept requires large electronic databases capable of storing essentially all relevant experimental data known to date with detailed descriptions of metadata and uncertainties. The combination of these electronic databases with expert system software, designed to automatically generate recommended property values based on available experimental and predicted data, leads to the ability to produce critically evaluated data dynamically or ‘to order’.

The first version of TDE (version 1.0, released in 2004)<sup>7</sup> was focused on thermophysical properties of pure compounds. The second version (2.0, released in 2006)<sup>8</sup> was expanded to include on-demand generation of equations of state (EOS). Four EOS of different complexity and precision were selected to adequately represent a range of data scenarios.<sup>2</sup> Also, a dynamically updatable database of experimental property values TDE-SOURCE<sup>2</sup> was developed, in part, with the multitier Web–Oracle dissemination system recently established at the National Institute of Standards and Technology (NIST) for several online products.<sup>9,10</sup> As previously described,<sup>2</sup> experimental data used by TDE are extracted from the TRC-SOURCE<sup>11</sup> data

storage system and distributed as a local database (TDE-SOURCE) with the program. TRC-SOURCE is maintained at NIST and is continuously populated with new experimental data by use of Guided Data Capture (GDC) software.<sup>12</sup> In order to more closely adhere to the requirements for dynamic data evaluation, TDE includes a mechanism for dynamically updating the local TDE-SOURCE system based on additions and changes to TRC-SOURCE through a central server.

The third version of TDE (3.0, released in 2008)<sup>13</sup> expanded implementation of dynamic data evaluation to binary mixtures.<sup>3</sup> TDE 3.0 also included initial steps for dynamic data evaluation for chemical reactions. This effort was limited to enthalpies of the formation reaction for compounds composed of carbon, hydrogen, nitrogen, and oxygen.

The fourth paper<sup>4</sup> described the extension of TDE (version 4.0, released in 2009)<sup>14</sup> for dynamic critical evaluation of chemical reactions, including processing for both change-of-state and chemical-equilibrium information. As an essential part of this development, the scope of properties for pure compounds evaluated by TDE was expanded to include enthalpies of formation for sulfur-containing and halogenated compounds, plus the integrated thermodynamic functions (entropy and enthalpy) for all phases.

TDE has proven to be a powerful software tool for thermophysical property data quality assurance,<sup>15</sup> validation of new experimental data,<sup>16</sup> and a variety of engineering applications including chemical process design.<sup>17</sup> It is also a critical component in the implementation of the concept of global information systems in science with application to the field of thermodynamics.<sup>18</sup>

<sup>§</sup> Contribution of the National Institute of Standards and Technology. Products or companies are named solely for descriptive clarity and neither constitute nor imply endorsement by NIST or by the U.S. government.

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

<sup>†</sup> National Institute of Standards and Technology.

<sup>‡</sup> Korea University.

Wakeham et al.<sup>19</sup> envisioned that the future of experimental thermophysical property measurement science will represent a transformation “from accuracy to fitness for purpose.” Indeed, they argued that, in a great many instances, the desire for high accuracy in measurements, which was a driving force in the field for many years, did not lead to a better understanding of natural phenomena nor did it help in the further advancement of theory and simulation. Consequently, one could legitimately conclude that the effort and the resources associated with these measurements were, to a significant degree, inefficiently used, making the need for improved planning for experiments apparent.

In the present paper, we describe the first steps in implementing an algorithmic approach to assist the process of experiment planning with assessment of the entire body of knowledge, including availability of experimental thermophysical property data, variable ranges studied, associated uncertainties, the state of prediction methods, availability of parameters for deployment of prediction methods, and how these parameters can be obtained using targeted measurements, etc., including how the intended measurement can address the underlying scientific or engineering problem under consideration.

The tools described here for experiment planning are also expected to be valuable additions to those developed for prereview of submitted journal manuscripts reporting new experimental thermochemical and thermophysical properties. This is a joint project between NIST and five major journals in this field: *Journal of Chemical and Engineering Data*, *Fluid phase Equilibria*, *The Journal of Chemical Thermodynamics*, *International Journal of Thermophysics*, and *Thermochimica Acta*. The prereview process is described in the Joint Statement of Editors of Journals Publishing Thermophysical Property Data,<sup>16</sup> which was published in 2009 in each of the five journals. It was noted in the statement that, “Often, reviewers cannot make informed decisions regarding the manuscript because the authors have made only a minimal literature review and comparisons. It is then an unacceptable burden to require reviewers to research previously published literature data to ensure a proper comparison has been made, and hence, determine the ultimate worth of the manuscript.”<sup>16</sup> In support of the NIST–journal cooperation,<sup>16</sup> the new tools described here for experiment planning are expected to find application in review, as well as planning processes, particularly with regard to the status of the body of knowledge in the field.

Since TDE operates, in principle, on the premise of analysis of all knowledge available to date, it is clear that it might be used as a powerful instrument to aid in the design of chemical products, where a goal is to solve the “inverse engineering problem” of determining chemical systems possessing desired values of thermophysical properties within defined ranges of tolerance.<sup>20</sup> The algorithms and their software implementation to achieve this goal are described in this paper. Finally, the current paper describes implementation of a recently developed data quality analysis algorithm for vapor–liquid equilibrium data<sup>21</sup> and its implementation in dynamic data evaluation for binary mixtures.

In summary, the present paper describes new extensions of TDE (version 5.0, released in 2010)<sup>22</sup> in three areas: (1) experiment planning, (2) product design, and (3) new data

validation and weighting system for vapor–liquid equilibrium (VLE) data.

## 2. SCOPE OF CHEMICAL SYSTEMS CONSIDERED

While key industrially important inorganic compounds (e.g., ammonia, water, sulfur dioxide, etc.) are within the scope of TDE, the focus of the program is molecular organic compounds composed of the elements C, H, N, O, F, Cl, Br, I, S, and P. This is unchanged from previous versions.

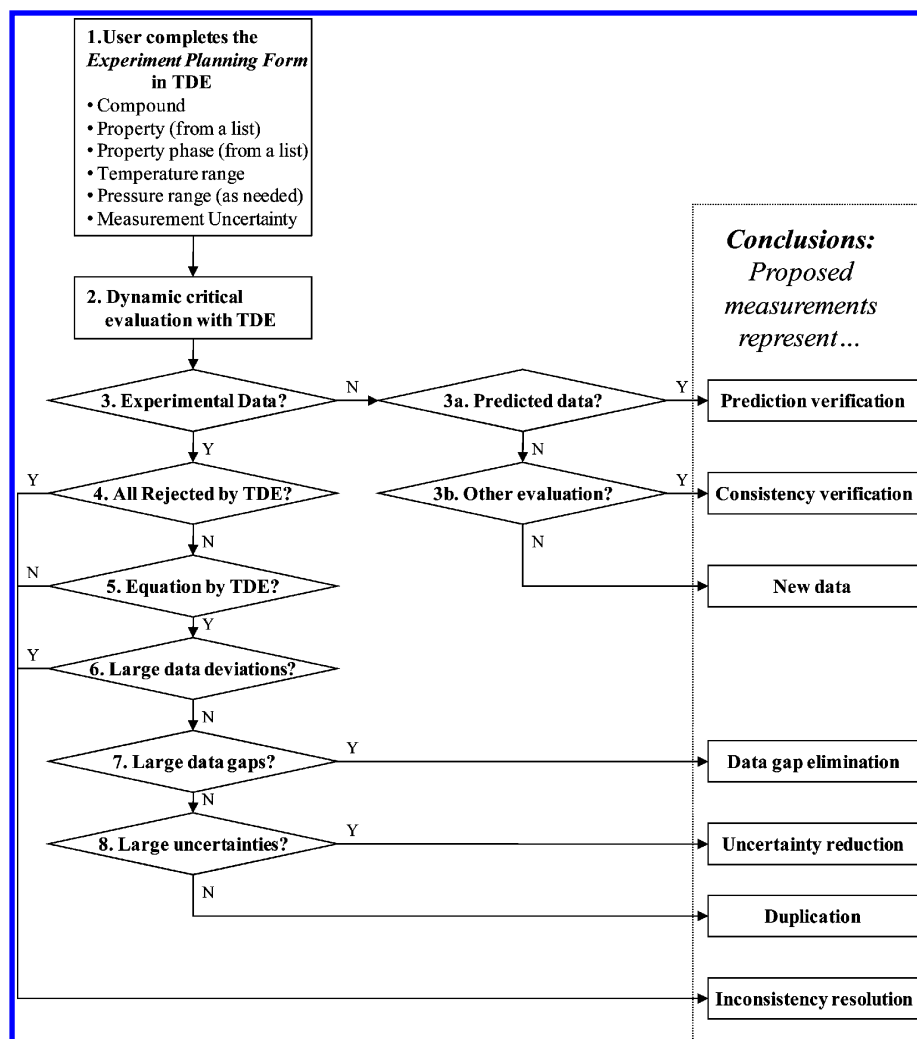
## 3. EXPERIMENT PLANNING

**Background.** Experiment planning for the measurement of thermodynamic and transport properties of chemicals and chemical mixtures involves a series of steps, including selection of chemicals, properties to measure, variable ranges (temperature, pressure, composition), and targeted uncertainty (which is often related to experimental complexity and apparatus capabilities). The extensive collection of experimental property data contained in TRC-SOURCE combined with the data analysis and prediction tools of TDE provides the basis for the development of algorithms to aid the decision-making processes at the experiment planning stage for those steps that are based strongly on the extent and quality of available property values.

As was described previously,<sup>1</sup> combined expanded uncertainties are assigned by the TDE software for all evaluated properties. The essence of the present work is the development of algorithms that use the quantitative information stored in TRC-SOURCE and evaluated with the TDE software to provide qualitative assessments needed for experiment planning. Examples of qualitative assessments that may be returned by TDE are: available experimental data are inconsistent, large gaps exist in the available data, extensive experimental data are available, etc.

**Scope.** It is clear that human information processing related to experiment planning is very complex and multifaceted. Experiment planning is a critical element in the overall scientific discovery process, which will always have technological and societal drivers together with an intuitive component. Nonetheless, we believe that some important aspects of experiment planning can be addressed with the algorithmic approach discussed here.

In TDE 5.0, four experiment planning scenarios are addressed: (1) evaluate a proposed measurement for a pure compound, (2) evaluate a proposed measurement for a binary mixture, (3) recommend measurements for a pure compound, and (4) recommend binary chemical mixtures for extension of a UNIFAC model.<sup>23,24</sup> UNIFAC group parameters in TDE 5.0 are those given in Poling et al.<sup>25</sup> UNIFAC is a group contribution-type prediction method for vapor–liquid equilibrium data, and the approach described here is generally applicable to any prediction method of this class. In the following paragraphs, implementation of each scenario is described together with the algorithm for generation of recommendations and the special features of the program interface. We must emphasize that many practical aspects related to experiment planning, such as availability and cost of samples, complexities and safety of laboratory operations, and stability of compounds during proposed measurement over wide ranges of temperature and pressure, etc., are clearly outside of the current scope of TDE.



**Figure 1.** Algorithm for evaluation of a proposed measurement for a pure compound; experiment planning, Implementation I.

**Implementation I: Evaluate a Proposed Measurement (Pure Compound).** The algorithm for evaluation of a proposed measurement for a pure compound is shown in Figure 1. The initial step is specification by the user of the proposed measurement. Specification involves identification of the compound, property (from a list), property phase (from a list), condition (property and phase-dependent choices: single phase, saturated or near atmospheric pressure, and equilibrium with gas), temperature range, pressure range (if appropriate), and measurement uncertainty (absolute or percentage). A completed experiment planning form is shown at the top of Figure 2. The condition saturated or near atmospheric pressure is used for analysis of properties, such as densities and viscosities, which are often measured at laboratory pressure and within  $\sim 50$  K of room temperature. The difference between these properties at saturation or at one atmosphere is generally negligible, and the properties are combined for analysis with TDE.

Properties supported within TDE 5.0 for analysis of a proposed measurement for a pure compound are the critical temperature, critical pressure, critical density, vapor pressure, sublimation pressure, density, heat capacity, viscosity, thermal conductivity, surface tension, speed of sound, and enthalpy of vaporization (or sublimation). Step 2 (Figure 1) is dynamic critical evaluation for all supported properties by TDE. This evaluation was fully described for pure

components in the first paper of this series.<sup>1</sup> This process is initiated from the experiment planning form, as shown in the upper portion of Figure 2. Subsequent steps (Steps 3–8 in Figure 1) are performed automatically by the TDE software without further user interaction.

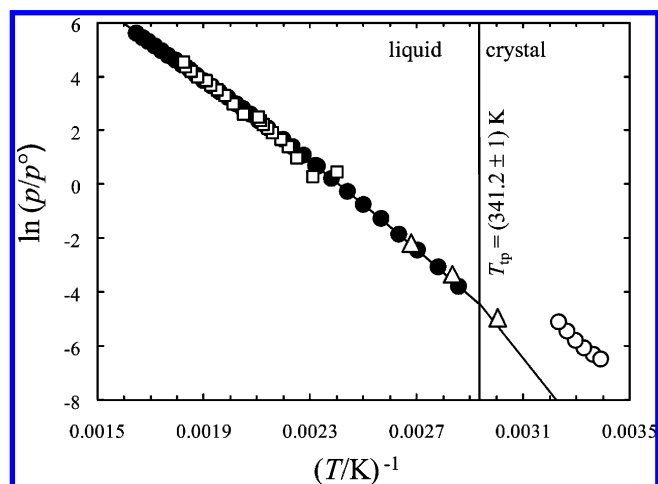
The first step (Step 3 of Figure 1) following critical evaluation of all properties is a check for the existence of literature data that overlap with those proposed. If none are found, then the program checks for the presence of predicted data, i.e., robust predictions with uncertainty estimates made by the TDE software, (Step 3a of Figure 1; Predicted data?) and values extrapolated, interpolated, or calculated indirectly, such as enthalpies of vaporization derived from vapor pressures through the Clapeyron equation (Step 3b of Figure 1; Other evaluation?). Following these checks, one of three conclusions can be made, as shown on the right side of Figure 1: the proposed measurements represent possible (1) prediction verification, (2) consistency verification for the indirect property determination, or (3) new data for the interval defined by the user.

The TDE data analysis algorithms include robust data validation checks to eliminate spurious literature values and possible data storage and processing errors in TRC-SOURCE. Data validation was discussed in the first paper of the series; in particular, see the discussion of Figure 3 of that work.<sup>1</sup> At this step (Step 4 of Figure 1), if all available

**Figure 2.** The experiment planning form for evaluation of a proposed measurement for a pure compound (Experiment Planning, Implementation I). The figures show the form before (upper) and after (lower) analysis by the software.

data for the property in the region of interest were rejected by TDE, then there is a clear data inconsistency, and the program returns the conclusion that the proposed measurements can resolve this (lower right of Figure 1). Step 5 is similar in that a failure by the program to generate a fitting equation for a property implies a gross inconsistency of the available data. An example of such an inconsistency is measured sublimation pressures that cannot be reconciled with a reliable enthalpy of fusion and vapor pressures for the liquid phase. An example of such a data scenario is shown in Figure 3, where sublimation pressures measured by the Knudsen effusion method (unfilled circles)<sup>26</sup> are shown to be inconsistent with available vapor pressures.<sup>27–29</sup> The enthalpy of fusion measured by Steele et al.<sup>27</sup> was used in combination with the experimental vapor pressures to estimate the sublimation curve shown in the lower right of the figure. Difficulties associated with the measurement and data analysis of Knudsen effusion experiments are well documented in the literature.<sup>30</sup>

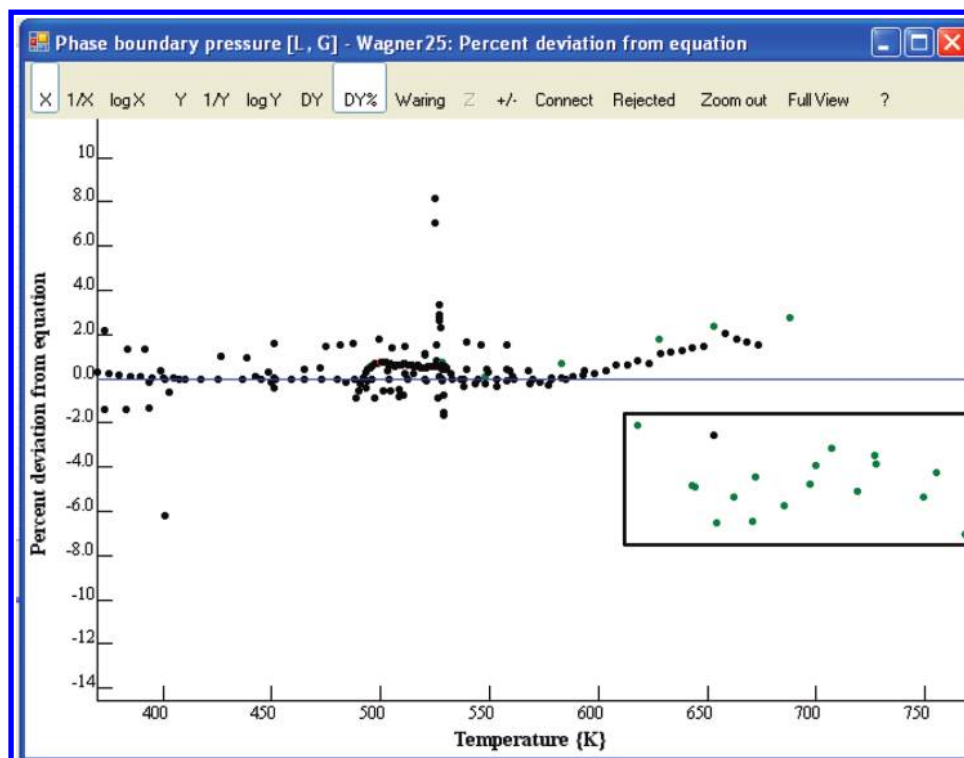
Step 6 (Figure 1) of the TDE analysis involves the cases in which a fitting equation could be generated, but deviations are unexpectedly large. The inconsistency criteria are: (1) the average deviation of the data from the equation exceeds triple the experimental uncertainty, and (2) the average relative deviation exceeds a value between 1% and 20%



**Figure 3.** Experimental and evaluated vapor and sublimation pressures for dimethyl-1,3-benzenedicarboxylate (CASRN: 1459-93-4). The reference pressure  $p^\circ$  is 1 kPa. The vertical line indicates the melting temperature. The curves were critically evaluated by the TDE software. The symbols represent literature experimental values. ●, Steele, et al.;<sup>27</sup> □, Olevskii, et al.;<sup>28</sup> △, Potin-Gautier, et al.;<sup>29</sup> ○, Roux et al.<sup>26</sup>

assigned individually to each property. (For example, 1% for liquid densities near saturation, and 10% for vapor pressures are assigned in TDE 5.0.) The second criterion is





**Figure 4.** Deviation plot for literature data for vapor pressures  $p$  of biphenyl generated by the TDE software. The data available at high temperatures (rectangle) are from a single literature source.<sup>32</sup>

used to reduce the number of false inconsistency detections generated because the assigned uncertainties for the data are too small. As we have discussed,<sup>31</sup> it is common for authors to report repeatabilities rather than combined expanded uncertainties. Consequently, most “uncertainties” reported in the literature are too small. Reported uncertainties are often adjusted before inclusion in TRC-SOURCE (from which TDE-SOURCE is derived), but misassigned uncertainties are inevitable. As shown in Figure 1, if large data deviations are recognized in Step 6, the conclusion is returned that the proposed measurements may resolve the inconsistency.

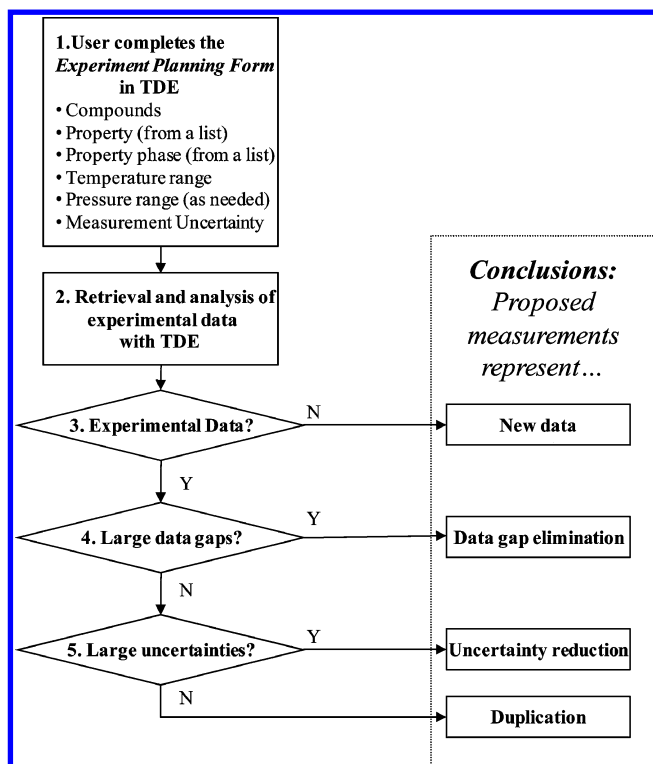
Step 7 (Figure 1) of the TDE analysis involves recognition of large gaps in the experimental data for the variable ranges defined by the user. For properties that are temperature dependent only (e.g., vapor pressure), a gap is said to exist if more than half of the specified temperature range does not include experimental data, provided that the range is not less than 40 K. This implies that a range of less than 20 K cannot be defined as a gap. For properties that are temperature and pressure dependent (e.g., density of a fluid), the variable plane is divided into four quadrants of equal size. The absence of experimental data in one or more quadrants is considered to be a gap, and the conclusion is returned that such a gap could be eliminated by the proposed measurements.

Finally, in Step 8 (Figure 1; large uncertainties?) the uncertainties evaluated in the region of interest are compared with those stated by the user for the proposed measurement. If the uncertainties of the evaluated data are larger, then the conclusion is returned that reduced uncertainties would result from the proposed measurements. If not, then the conclusion is returned that the measurements are redundant with those already available, as shown in the lower right (duplication) of Figure 1.

The lower portion of Figure 2 shows the analysis result for the proposed measurement of vapor pressures for biphenyl. The conclusion is shown in the lower left of the form. The lower right shows the range of the available experimental data together with that proposed by the user (box). Full details of the critical evaluation are immediately available to the user through the “go to property” button in the center right of the form. Figure 4 shows a deviation plot for experimental vapor pressures  $p$  of biphenyl. Only one data set<sup>32</sup> is available at high temperatures, and this set has a repeatability near  $0.05 \cdot p$ , which implies that the standard uncertainty is larger. (See the rectangle in Figure 4.) The uncertainty for the proposed measurements is  $0.005 \cdot p$ . The absence of high-quality data at high temperatures is apparent, in accord with the conclusion shown in the lower screen shot of Figure 2.

**Implementation II: Evaluate a Proposed Measurement (Binary Mixture).** The algorithm for evaluation of a proposed measurement for a binary mixture is shown in Figure 5. The initial step is specification by the user of the proposed measurement. As for the evaluation for a pure compound described above, specification involves identification of the two compounds, property (from a list), property phase (from a list), condition (property and phase-dependent choices: single phase, saturated or near atmospheric pressure, and equilibrium with gas), temperature range, pressure range (if appropriate), and measurement uncertainty (absolute or percentage). A completed experiment planning form for a proposed vapor–liquid equilibrium (VLE) study for the chemical system (cyclohexanone + cyclohexanol) is shown at the top of Figure 6.

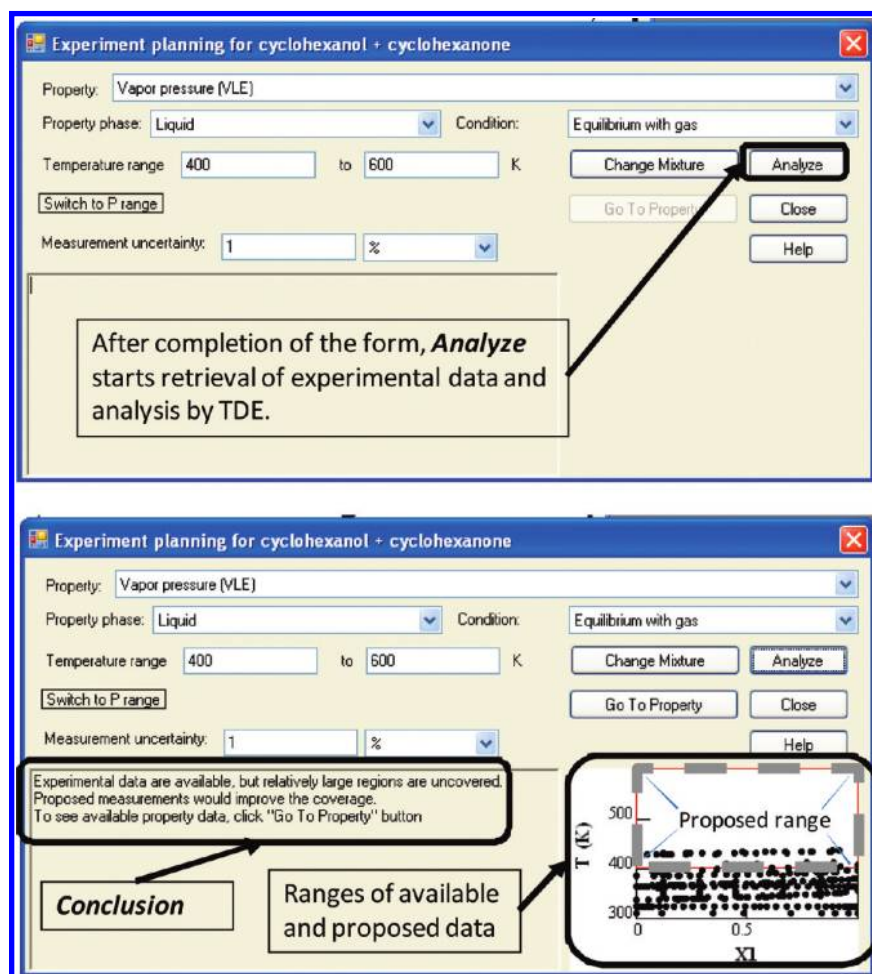
Properties supported within TDE 5.0 for this analysis are the critical temperature, vapor pressure or bubble point temperature (VLE), vapor composition (VLE), liquidus



**Figure 5.** Algorithm for evaluation of a proposed measurement for a binary mixture; experiment planning, Implementation II.

temperature (solid–liquid equilibrium, SLE), activity coefficient (liquid), density, heat capacity, excess enthalpy, viscosity, thermal conductivity, surface tension, and speed of sound. Excess volume is not included explicitly, but such a proposal can be evaluated as a density measurement because of the simple algebraic relationship between the two properties. Liquid–liquid equilibrium (LLE) is not considered at this time because, unlike SLE and VLE, such phase behavior does not exist for many systems; thus, the absence of such data is not an indication of need.

In contrast to the full critical evaluation completed in Step 2 for pure compounds, Step 2 for analysis of the proposed measurements for binary mixtures (Figure 5) is retrieval and analysis of available experimental data by TDE. As was described in the third paper of this series,<sup>3</sup> dynamic critical data evaluation for binary systems is relatively complex and involves substantial user interaction. Consequently, in the present application, the proposed measurements are analyzed with respect to available experimental data only. The criteria for determination of the availability of experimental data for the binary system (Step 3, Figure 5), the presence of significant gaps (Step 4, Figure 5), and comparison of uncertainties between the existing and the proposed measurements are analogous to those described above for pure compounds. Possible conclusions are shown on the right side of Figure 5, in analogy to those shown in Figure 1 for the pure-component analysis.



**Figure 6.** The experiment planning form for evaluation of a proposed measurement for a binary mixture (experiment planning, Implementation II). The figures show the form before (upper) and after (lower) analysis by the software.

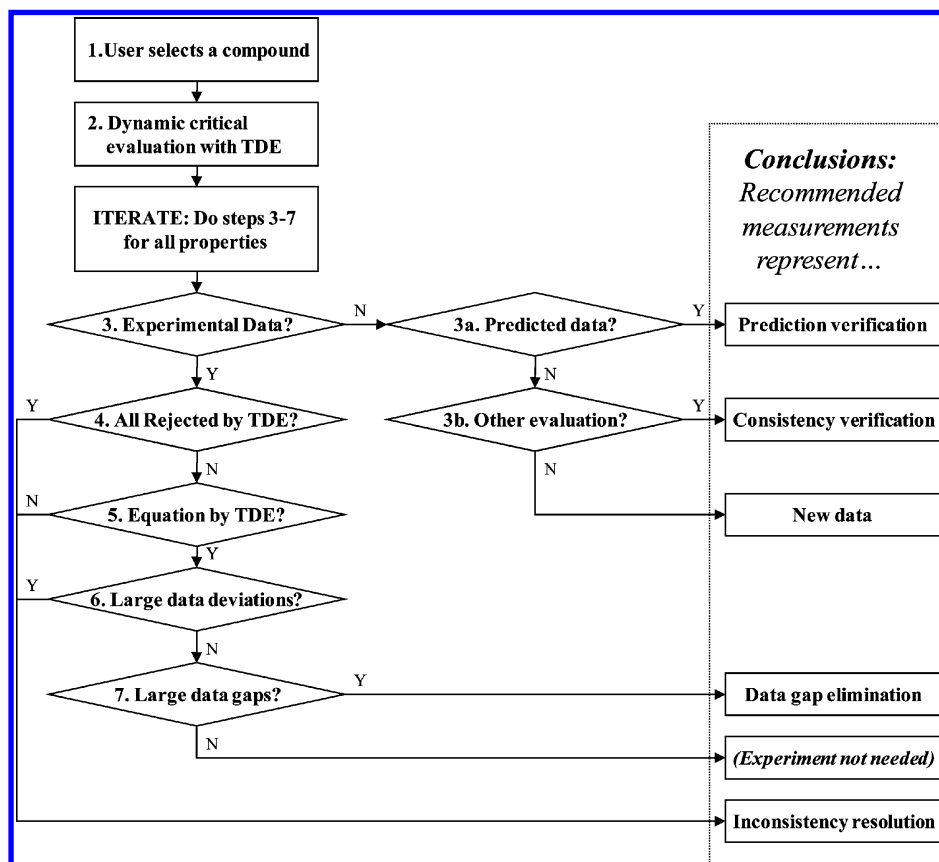


Figure 7. Algorithm for measurement recommendations for a pure compound; experiment planning, Implementation III.

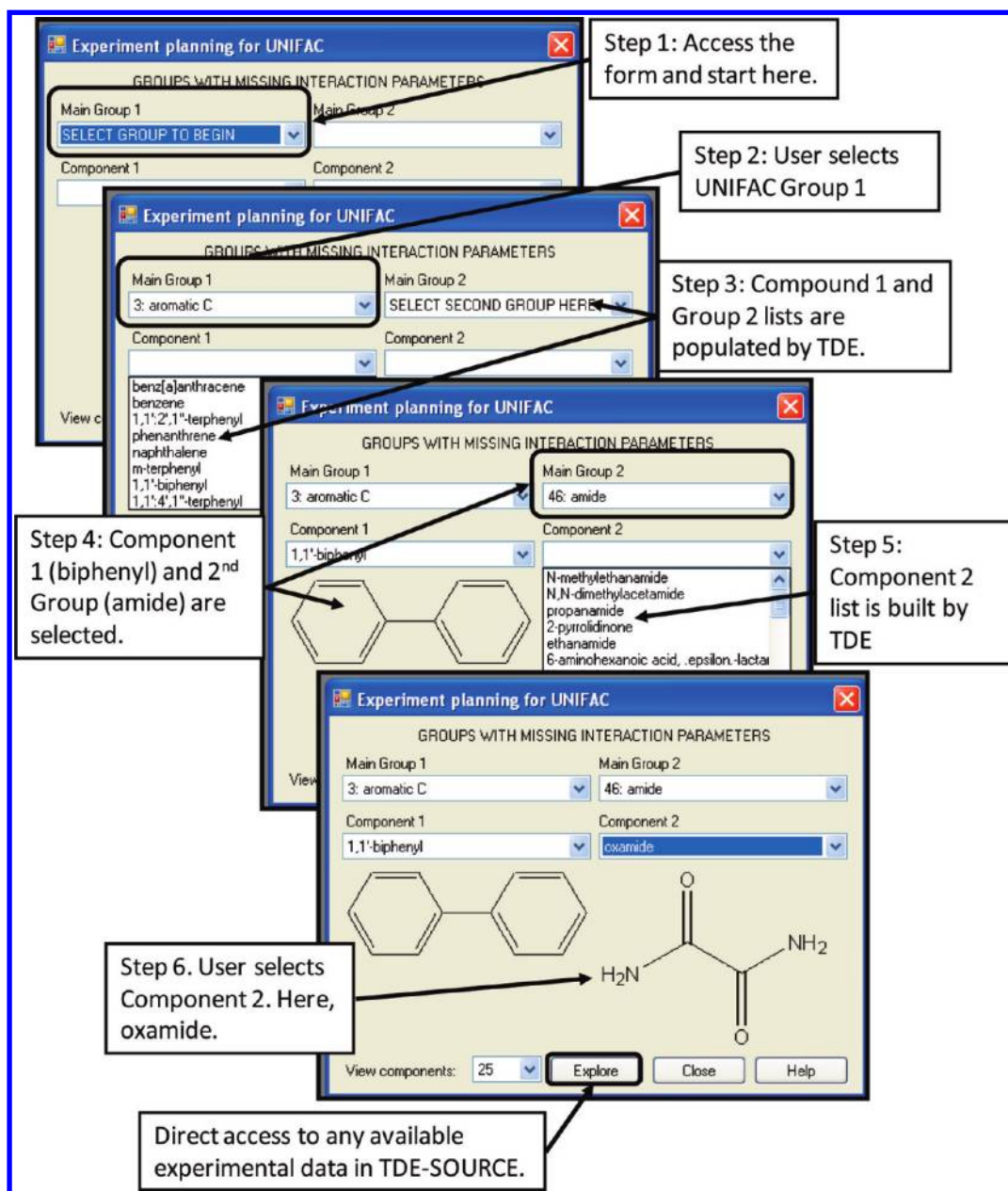
The lower portion of Figure 6 shows the result of the analysis for the proposed measurement of VLE data for the (cyclohexanone + cyclohexanol) system. The conclusion is shown in the lower left of the form, and the basis for this conclusion in graphical form is shown in the lower right. The range for the proposed measurements is shown with a box, while the dots represent the available experimental values. The graph in the lower right of Figure 6 shows clearly that available data for this system are limited to temperatures near  $T = 420$  K and below. Full details of the available experimental data are immediately accessible to the user through the “go to property” button in the center right of the form, as are tools for fitting of activity coefficient models for a more complete analysis of the available data for the proposed system. These were described in the third article of this series.<sup>3</sup>

**Implementation III: Recommend Measurements (Pure Compound).** The third implementation of experiment planning in TDE 5.0 addresses recommendation of a series of measurements for a given pure compound. The approach essentially involves iteration for a series of properties on the algorithm for Implementation I; evaluation of a proposed measurement for a pure compound (Figure 1). Some adjustments to the algorithm and criteria are necessary, and these are described below. The properties considered are the critical properties (temperature, pressure, and density), vapor and sublimation pressure, enthalpy of vaporization and sublimation, density (liquid, gas, and saturated liquid), heat capacity (liquid, gas, crystal, and saturated liquid), thermal conductivity (liquid, gas, and saturated liquid), speed of sound (liquid, gas, and saturated liquid), viscosity (liquid, gas, and saturated liquid), and surface tension.

The algorithm for the recommendation of measurements for a pure compound is shown in Figure 7. Steps 3–6 are analogous to those of Implementation I, as they involve no information (variable ranges and uncertainties) provided by the user. Step 7 of Implementation I involved recognition of data “gaps” that were determined to exist if more than half of the temperature range specified by the user did not include experimental data, provided that the range was not less than 40 K. Similarly, for properties that are temperature and pressure dependent (e.g., density of a fluid), gaps were defined based on ranges provided by the user. Here, the temperature and pressure ranges for definition of gaps are fixed and depend on the phase of the property and are based on approximate practical limits for experimental studies. The temperature range considered in identification of data gaps for properties of the liquid phase is from the melting temperature  $T_m$  to the critical temperature  $T_c$ , while the maximum pressure considered is 100 MPa. Properties of the gas phase are considered to a maximum temperature of  $2 \cdot T_c$  or  $4 \cdot T_m$ , if  $T_c$  is not available through experiment or prediction.

There is no step in Implementation III that is analogous to Step 8 in Implementation I. That step involved comparisons between evaluated uncertainties and those of the proposed measurement.

The program output for Implementation III is a text report that provides one of the six conclusions listed at the right of Figure 7 for each of the properties considered (i.e., those listed at the beginning of this section). In the second article of this series,<sup>2</sup> an algorithm was given (Figure 1 of reference<sup>2</sup>) for deployment of the technical Span–Wagner EOS<sup>33</sup> based on the available experimental data. That



**Figure 8.** Experiment planning steps and interface details for recommendation of binary chemical systems for extension of UNIFAC group parameters (experiment planning, Implementation IV).

algorithm is employed to highlight specific measurements needed for application of this EOS. If all of the required experimental data are available, then the message is given that a high-precision EOS can be deployed. Otherwise, the software provides suggestions of what properties should be measured to enable development of the EOS.

**Implementation IV: Recommend Binary Mixtures for UNIFAC Model Extension.** As described in the first article of this series,<sup>1</sup> structural information for all compounds analyzed by TDE is stored in a connectivity table. Also, a structure drawing tool is available, if the structure is not already included in the TDE-SOURCE collection. The information in the connectivity table can be parsed into various group representations for the application of group-contribution property estimation methods. These features allow the development of experiment planning tools to provide recommended compounds and measurements for the extension of particular estimation methods. The example

given here is for recommendation of binary chemical systems for vapor–liquid equilibrium studies to yield new interaction group parameters for the UNIFAC prediction method. The UNIFAC method<sup>23–25</sup> is used widely in industry, and its implementation in TDE was described previously.<sup>3</sup>

The steps involved to recommend binary chemical systems for model extension are shown in Figure 8 together with details of the program interface. The user initiates the process by opening the experiment planning for UNIFAC form shown in Figure 8. In Step 2, the user selects the first UNIFAC group (here, aromatic C). Immediately (Step 3), TDE populates the Component 1 list of compounds based on the presence of the group chosen and the absence of other less common groups. This is accomplished through a numerical prioritization scheme, where compounds are assigned a value  $\Omega$  based on the following relationship:

$$\Omega = (1 + 0.4/n_{\text{target}}) / \sum n_i w_i, \quad (1)$$



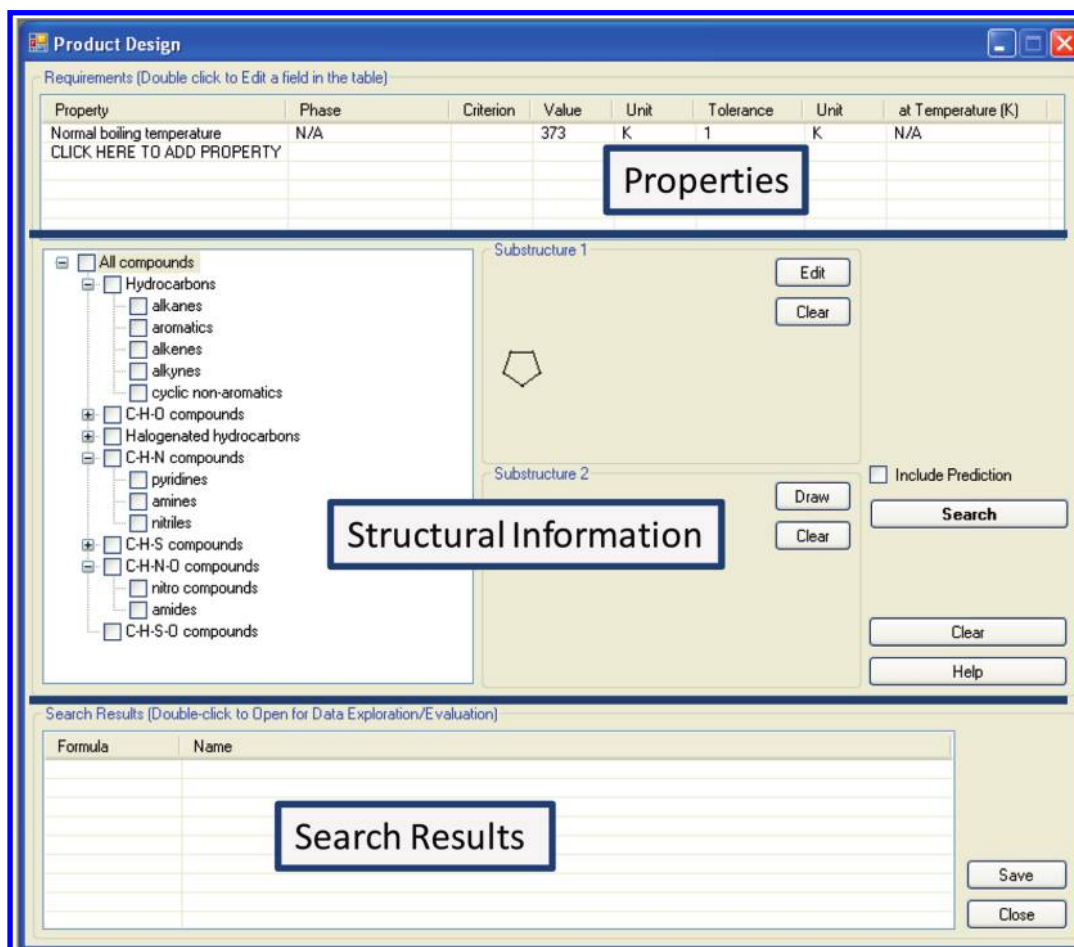


Figure 9. Interface for product design for pure compounds.

where  $n_{\text{target}}$  is the number of target groups in the molecule,  $n_i$  is the number of groups of type  $i$  in the molecule,  $w_i$  is the relative weight for a particular UNIFAC group (0.1 for saturated C–H groups, 0.15 to 0.45 for other groups involving C–H, such as alkenes and aromatics, and 1.0 for all other groups). The summation is over all groups  $i$ . This formulation implies an optimum value of 2 for  $n_{\text{target}}$ . This was chosen based on the practical knowledge that compounds with large values of  $n_{\text{targets}}$  may, in fact, be poor model compounds for the group. For example, if the target group is  $-\text{NO}_2$ , tetranitromethane or hexanitrobenzene would be poor choices for study.

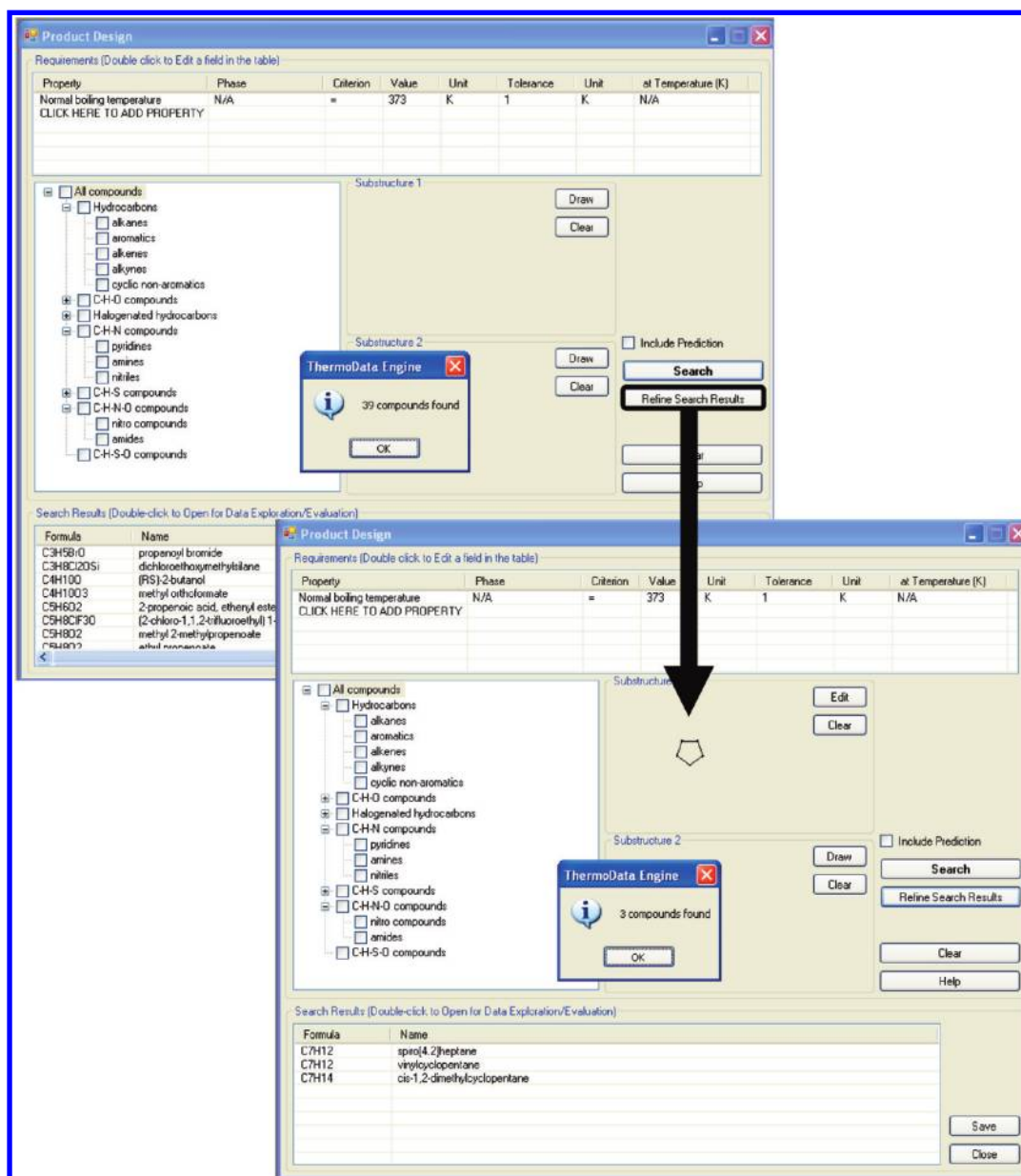
The Group 2 list is populated based on the absence of interaction group parameters between the group and the chosen in Group 1. In Step 4, the user selects Component 1 and Group 2 from the lists. Again, the Component 2 list is populated (Step 5) with the same criteria as those for Component 1, with two additional requirements: (1) that the temperature range of the liquid regions for the two components overlap by a minimum of 80 K, and (2) that there be only one missing group interaction parameter with the selected Group 1. These additions increase the likelihood that a good experimental result can be achieved. In Step 6, the user selects Component 2. At the request of the user, all experimental data for the components and the binary mixture (if any) are loaded from TDE-SOURCE for exploration and processing with the TDE tools described previously for pure components<sup>1,2</sup> and binary mixtures.<sup>3</sup>

#### 4. PRODUCT DESIGN

**Background.** The problem to be addressed is identification of chemical systems possessing desired values of thermo-physical properties within defined ranges of tolerance. Common applications would be selection of solvents, heat exchange fluids, extraction agents, etc. This first implementation in TDE is for pure components only. Existing product design tools, such as those of Molecular Knowledge Systems, Inc.<sup>34</sup> and the Computer Aided Process-Product Engineering Center (CAPEC) of Technical University of Denmark,<sup>35</sup> are based entirely on prediction methods. The distinguishing feature reported here is the extensive use of experimental property data.

**Scope.** This initial implementation of product design in the TDE software covers pure compounds. Properties supported by TDE product design are melting temperature, enthalpy of fusion (mole or mass basis), normal boiling temperature, critical temperature, vapor pressure, density, viscosity, thermal conductivity, enthalpy of vaporization (mole or mass basis), and heat capacity (mole or mass basis). Specific temperatures for temperature-dependent properties are specified by the user. In addition, the user can restrict the search to specific chemical families and substructures. Multiple properties can be specified, each with a specified tolerance or limit.

**Implementation.** In order to increase the speed of the search response, a database of evaluated properties was compiled using the automatic TDE evaluation procedures<sup>1</sup>



**Figure 10.** Example of search refinement with addition of a structural requirement (a cyclopentyl ring) in product design for a pure compound.

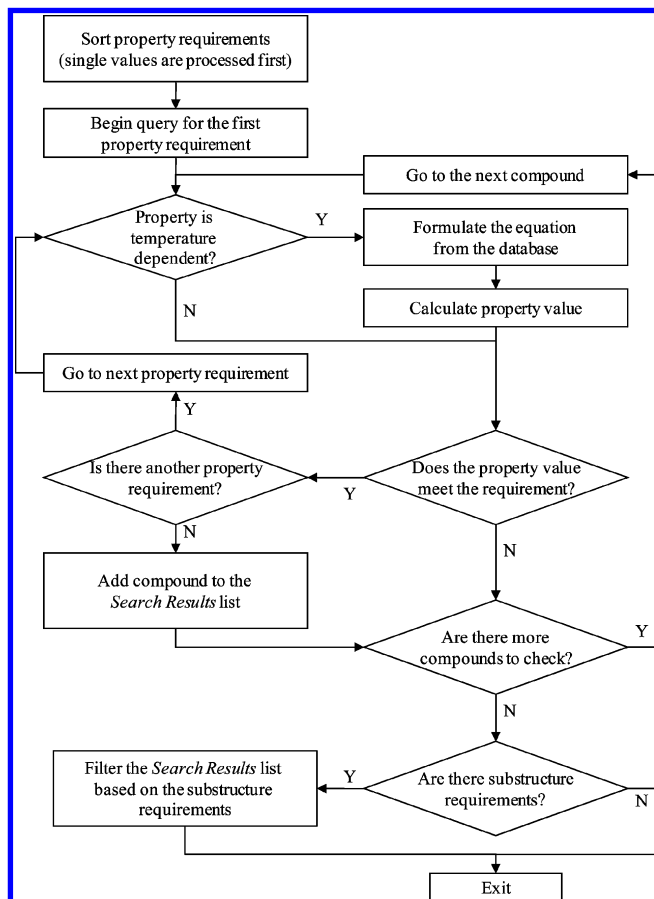
for all pure compounds with any experimental property data in TDE-SOURCE. Single-value properties, such as critical temperature, are stored in separate database fields. Temperature-dependent properties are represented by equations, and the database records include the property, phase, equation name, equation constants and parameters, and the temperature range of validity. A special field identifies each property (value or equation) as being based on some experimental data or based on prediction only. At present, there are roughly 26 000 pure compounds in this collection, and search time durations are typically a few seconds on most modern desktop computers.

The interface for product design is shown in Figure 9. The product design form is divided into three major sections (Figure 9). The upper section is used to define the search requirements: property, property phase, criterion (greater than, less than, or equal), the target property value, property unit, tolerance, tolerance unit (property unit or percentage), and temperature of the target property value, if applicable. The central section of the form provides structural information

used to limit the scope of the search based on compound types (e.g., alkynes, amines, etc.) and structural fragments. A cyclopentyl ring is shown as such a fragment in Figure 9. The lower section provides a list of compounds, if any, that meet the property and structural requirements entered by the user.

The user can refine the search by limiting the compounds searched to those in the search results at the bottom of the form (Figure 9). Figure 10 shows an example of such a two-step search. In the upper figure, the search requirement was a specified normal boiling temperature only, and 39 compounds were found to meet this requirement. The addition of a structural requirement (cyclopentyl ring) reduced the number of resulting compounds to three, as seen at the bottom of the figure.

The algorithm that underlies the compound search in product design is shown in Figure 11. The property requirements are sorted so that single-valued properties are considered first. This speeds the analysis considerably because no equation formulations are involved.



**Figure 11.** Compound identification algorithm for product design for pure compounds.

## 5. OTHER NEW FUNCTIONALITIES IN TDE

### VLE Data Set Validation and Weighting System.

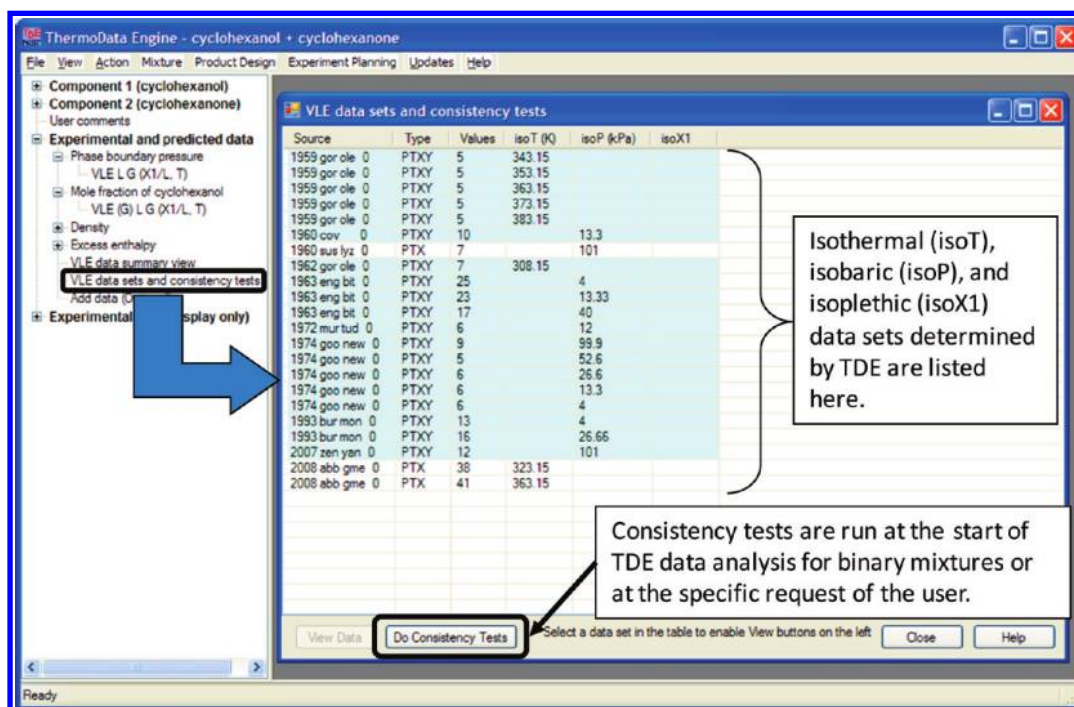
Recently, this research group described a quality assessment algorithm for VLE data.<sup>21</sup> The algorithm combines four widely used tests of VLE consistency based on the require-

ments of the Gibbs–Duhem equation (the Herington or area test, Van Ness modeling test, point or differential test, and infinite dilution test), with a check of consistency between the VLE binary data and the pure compound vapor pressures. Mathematical details of all tests are fully described in the article.<sup>21</sup> A key extension involved calculation of a numerical quality factor  $F_{\text{test},i}$  for each test to replace the simple pass/fail of earlier formulations. The  $F_{\text{test},i}$  values were then combined with a numerical representation  $F_{\text{pure}}$  of the consistency between the end point pressures for the binary mixtures with pure-component vapor pressures assessed with TDE to calculate an overall quality factor  $Q_{\text{VLE}}$ .

$$Q_{\text{VLE}} = F_{\text{pure}}(F_{\text{test},1} + F_{\text{test},2} + F_{\text{test},3} + F_{\text{test},4}); \quad 0.01 \leq Q_{\text{VLE}} \leq 1 \quad (2)$$

$Q_{\text{VLE}}$  and the  $F$  factors were formulated so that if a particular test cannot be performed, the  $F$  factor for that test is set to  $0.5 \cdot F_{\text{test},i, \text{max}}$ , where  $F_{\text{test},i, \text{max}}$  is the maximum value of  $F_{\text{test}}$  for test  $i$ .  $F_{\text{test},i, \text{max}}$  is 0.25 for tests 1–4, and 1 for the pure component consistency test. In the hypothetical case where none of the tests can be applied,  $Q_{\text{VLE}}$  is 0.25. The algorithm can be applied to VLE data sets with at least three state variables reported, pressure, temperature, plus liquid and/or vapor composition, and is applicable to all nonreacting chemical systems at subcritical conditions.

In the present work, the five consistency tests  $F$  and the overall quality factor  $Q_{\text{VLE}}$  were implemented in TDE 5.0 and applied to the fitting activity coefficient models to VLE data. Descriptions of models and general fitting procedures used in TDE for binary mixtures were published in the third article of this series.<sup>3</sup> As VLE data are represented in TDE-SOURCE in the form of pressure  $p(T, x)$  and vapor composition  $y(T, x)$  data sets, where  $T$  is temperature and  $x$  is liquid composition, the overall quality factors  $Q_{\text{VLE}}$  are converted into effective uncertainties of the properties:

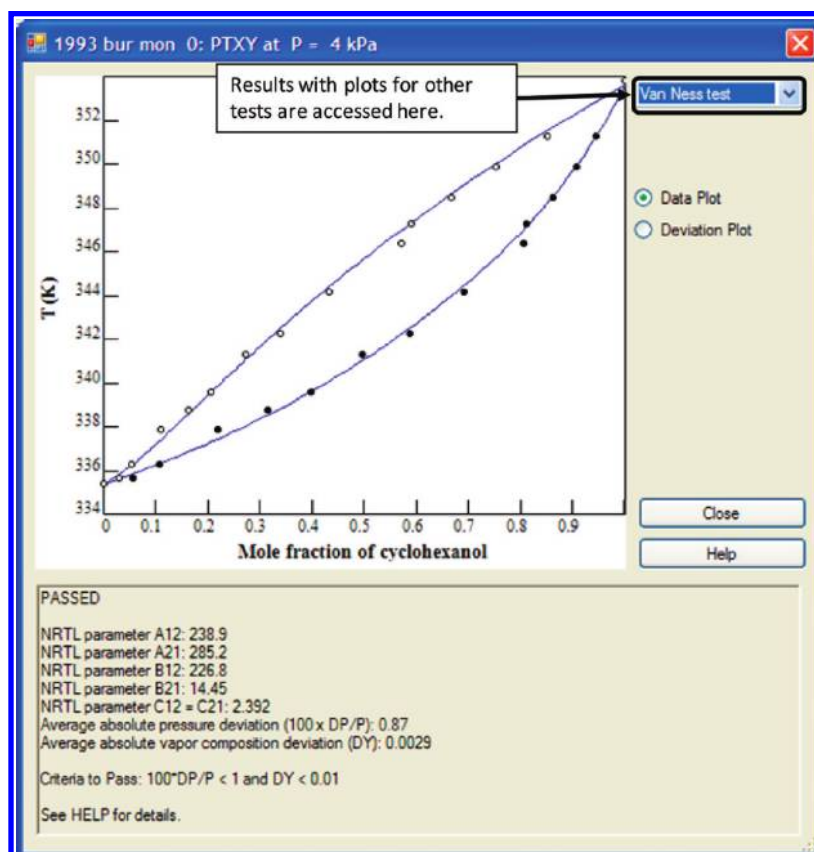


**Figure 12.** Interface for display of VLE data sets and access to results of consistency tests.



Source	Type	Values	iso T (K)	iso P (kPa)	iso X1	Herington	Van Ness	Point Test	Infinite Dilution	Quality Factor
1959 gor ole 0	PTXY	5	343.15			Failed	Failed	Failed	Failed	0.13
1959 gor ole 0	PTXY	5	353.15			Failed	Failed	Failed	Failed	0.19
1959 gor ole 0	PTXY	5	363.15			Failed	Failed	Failed	Failed	0.19
1959 gor ole 0	PTXY	5	373.15			Failed	Passed	Failed	Failed	0.21
1959 gor ole 0	PTXY	5	383.15			Passed	Failed	Failed	Failed	0.31
1960 cov 0	PTXY	10		13.3		Failed	Passed	N/A	Failed	0.22
1960 sus lyz 0	PTX	7		101		N/A	N/A	N/A	N/A	0.22
1962 gor ole 0	PTXY	7	308.15			Failed	Failed	Failed	Failed	0.007
1963 eng bit 0	PTXY	25		4		Failed	Failed	N/A	Failed	0.14
1963 eng bit 0	PTXY	23		13.33		Failed	Failed	N/A	Failed	0.17
1963 eng bit 0	PTXY	17		40		Failed	Failed	N/A	Failed	0.094
1972 murtud 0	PTXY	6		12		Failed	Failed	N/A	Failed	0.12
1974 goo new 0	PTXY	9		99.9		Passed	Passed	N/A	Failed	0.23
1974 goo new 0	PTXY	5		52.6		Failed	Passed	N/A	Failed	0.2
1974 goo new 0	PTXY	6		26.6		Passed	Passed	N/A	Failed	0.24
1974 goo new 0	PTXY	6		13.3		Failed	Failed	N/A	Failed	0.16
1974 goo new 0	PTXY	6		4		Failed	Failed	N/A	Failed	0.14
1993 bur mon 0	PTXY	13		4		Passed	Passed	N/A	Failed	0.72
1993 bur mon 0	PTXY	16		26.66		Passed	Passed	N/A	Failed	0.81
2007 zen yan 0	PTXY	12		101		Failed	Passed	N/A	Failed	0.42
2008 abb gme 0	PTX	38	323.15			N/A	N/A	N/A	N/A	0.06
2008 abb gme 0	PTX	41	363.15			N/A	N/A	N/A	N/A	0.14

**Figure 13.** Interface for display of consistency test results for VLE data sets. The overall quality factor  $Q_{VLE}$  is listed at the far right. Individual test results are seen through selection of a data set and the view test results command.



**Figure 14.** Interface for review of detailed consistency test results. Test details and pass/fail criteria are described separately.<sup>21</sup> The estimated standard uncertainty  $u$  for the pressure  $p$  is  $u(p) = 0.02 \cdot p$ , which when propagated to the observed temperatures  $T$ , yields  $u(T) = 0.5$  K, in accord with the observed data scatter.

$$\sigma(p) = (0.01 \cdot p)(Q_{VLE}^{-1} - 1) \quad \text{and} \quad \sigma(y) = 0.01(Q_{VLE}^{-1} - 1) \quad (3)$$

They are then used for setting relative weights for fitting. Generally, those weights are proportional to the reciprocal square of the uncertainties. If the database uncertainty for a



data point exceeds the effective uncertainty based on the consistency tests, then there is no change to the uncertainty. If the consistency-based uncertainty (eq 3) is greater, then it is used, and the weight of the data point is lowered.

Consistency tests are applied to individual “data sets”, which are either isothermal or isobaric  $p$ – $T$ – $x$ – $y$ ,  $p$ – $T$ – $x$ , or  $p$ – $T$ – $y$  data groups. For application of the tests based on the Gibbs–Duhem equation,  $p$ – $T$ – $x$ – $y$  data are necessary, and all four tests can be applied to isothermal data, while two can be applied to isobaric data sets. In the case of  $p$ – $T$ – $x$  or  $p$ – $T$ – $y$  data, only the pure-component consistency test can be applied. The assignment of data to isothermal and isobaric sets is nontrivial, due to small deviations in reported isothermal or isobaric conditions. Consequently, it was necessary to implement tolerances; 0.01 K for  $T$ ,  $(0.01 \cdot p)$  for  $p$ , and 0.0001 for mole fraction, when composing isovary data sets. The TDE interface for access to all data sets is shown in Figure 12. Application of the consistency tests can be initiated by the user or automatically by the software in preparation for property evaluations for a binary mixture.

A summary of the results for the consistency tests based on the Gibbs–Duhem equation is provided in a single form (Figure 13). Access to details for particular test results is made through this form. An example of detailed test results for the Van Ness test is shown in Figure 14. Results for all consistency tests for the selected data set are accessible through this single form.

The automatic application of consistency tests for VLE data is a necessary precursor to development of an algorithm for reliable dynamic critical evaluation of phase equilibrium data without user intervention. Development of this algorithm is planned for future research.

## 6. CONCLUSIONS AND FUTURE DEVELOPMENT

The software framework of the NIST ThermoData Engine (TDE) for the implementation of the dynamic data evaluation concept has been further expanded to support an algorithmic approach for planning experimental measurements of thermophysical properties for pure compounds and binary mixtures as well as to provide capabilities to assist in chemical product design for pure compounds. Plans for future development consist of implementation of the dynamic data evaluation concept for ternary mixtures as well as expansion for fully unattended evaluation of phase equilibria properties for binary mixtures. Among other venues for development, it is planned to extend the assessment of VLE data to the supercritical region.

## ACKNOWLEDGMENT

The authors express their appreciation to Dr. S. Watanasiri (AspenTech, Burlington, MA) and Dr. M. Satyro (Virtual Materials Group, Calgary, Canada) for their valuable advice with regard to aspects of product design and the VLE data quality assessment algorithm. We also appreciate the encouragement and guidance of Prof. K.N. Marsh of the University of Canterbury (Christchurch, New Zealand) regarding implementation of the VLE data quality assessment algorithm into TDE. The authors thank Dr. A. H. Harvey and Dr. M. L. Huber (both of NIST, Boulder, CO) for their practical suggestions during testing of TDE. In addition, we

thank all staff members of the NIST Thermophysical Properties Division for useful discussions and numerous suggestions related to the implementation of the algorithmic approach for planning experimental measurements of thermophysical properties.

## REFERENCES AND NOTES

- (1) Frenkel, M.; Chirico, R. D.; Diky, V.; Yan, X.; Dong, Q.; Muzny, C. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. *J. Chem. Inf. Model.* **2005**, *45*, 816–838.
- (2) 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.
- (3) Diky, V.; Chirico, R. D.; Kazakov, A. F.; Muzny, C.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 3: Binary Mixtures. *J. Chem. Inf. Model.* **2009**, *49*, 503–517.
- (4) Diky, V.; Chirico, R. D.; Kazakov, A. F.; Muzny, C.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 4: Chemical Reactions. *J. Chem. Inf. Model.* **2009**, *49*, 2883–2896.
- (5) (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, H. J. M., Harvey, A. H., Holcomb, C. D., Laesecke, A., Magee, J. W., Muzny, C., Eds.; NIST: Gaithersburg, MD, 2001; pp 83–84.
- (6) Frenkel, M. Global Communications and Expert Systems in Thermodynamics: Connecting Property Measurement and Chemical Process Design. *Pure Appl. Chem.* **2005**, *77*, 1349–1367.
- (7) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Yan, X.; Dong, Q.; Muzny, C. NIST ThermoData Engine, NIST Standard Reference Database 103, version 1.0. NIST, Standard Reference Data Program: Gaithersburg, MD, 2004.
- (8) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Muzny, C.; Lemmon, E. W.; Yan, X.; Dong, Q. NIST ThermoData Engine, NIST Standard Reference Database 103, version 2.0; NIST, Standard Reference Data Program: Gaithersburg, MD, 2006.
- (9) NIST Ionic Liquids Database, ILThermo, NIST Standard Reference Database 147; NIST, Standard Reference Data Program: Gaithersburg, MD, 2006; <http://ilthermo.boulder.nist.gov/ILThermo/>. Accessed August 1, 2009.
- (10) (a) Web Thermo Tables (Lite Edition) Standard Reference Subscription Database 2; NIST, Standard Reference Data Program: Gaithersburg, MD; <http://www.nist.gov/srd/nistwebsub2.htm>. Accessed August 1, 2009. (b) Web Thermo Tables (Professional Edition) Standard Reference Subscription Database 3; NIST, Standard Reference Data Program: Gaithersburg, MD; <http://www.nist.gov/srd/nistwebsub3.htm>. Accessed August 1, 2009.
- (11) Frenkel, M.; Dong, Q.; Wilhoit, R. C.; Hall, K. R. TRC SOURCE Database: A Unique Tool for Automatic Production of Data Compilations. *Int. J. Thermophys.* **2001**, *22*, 215–226.
- (12) Diky, V. V.; Chirico, R. D.; Wilhoit, R. C.; Dong, Q.; Frenkel, M. Windows-Based Guided Data Capture Software for Mass-Scale Thermophysical and Thermochemical Property Data Collection. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 15–24.
- (13) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Muzny, C.; Kazakov, A. F.; Lemmon, E. W. NIST ThermoData Engine, NIST Standard Reference Database 103b-Pure Compounds and Binary Mixtures, version 3.0; NIST, Standard Reference Data Program: Gaithersburg, MD, 2008.
- (14) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Muzny, C.; Kazakov, A. F. NIST ThermoData Engine, NIST Standard Reference Database 103b-Pure Compounds, Binary Mixtures, and Chemical Reactions, version 4.0; NIST, Standard Reference Data Program: Gaithersburg, MD, 2009.
- (15) Frenkel, M.; Chirico, R. D.; Diky, V.; Muzny, C.; Dong, Q.; Marsh, K. N.; Dymond, J. H.; Wakeham, W. A.; Stein, S. E.; Königsberger, E.; Goodwin, A. R. H.; Magee, J. W.; Thijssen, M.; Haynes, W. M.; Watanasiri, S.; Satyro, M.; Schmidt, M.; Johns, A. I.; Hardin, G. R. New Global Communication Process in Thermodynamics: Impact on Quality of Published Experimental Data. *J. Chem. Inf. Model.* **2006**, *46*, 2487–2493.
- (16) (a) Cummings, P. T.; de Loos, T.; O’Connell, J. P.; Haynes, W. M.; Friend, D. G.; Mandelis, A.; Marsh, K. N.; Brown, P. L.; Chirico, R. D.; Goodwin, A. R. H.; Wu, J.; Weir, R. D.; Trusler, J. P. M.; Pádua, A.; Rives, V.; Schick, C.; Vyazovkin, S.; Hansen, L. D. Joint Statement of Editors of Journals Publishing Thermophysical Property

- Data: Process for Article Submission for The Journal of Chemical Thermodynamics, Fluid Phase Equilibria, International Journal of Thermophysics, Thermochimica Acta, and Journal of Chemical Engineering Data. *Fluid Phase Equilib.* **2009**, *276*, 165–166. (b) *Int. J. Thermophys.* **2009**, *30*, 371–373; (c) *J. Chem. Eng. Data* **2009**, *54*, 2–3; (d) *J. Chem. Thermodyn.* **2009**, *41*, 575–576; (e) *Thermochim. Acta* **2008**, *484*, vii–viii.
- (17) Frenkel, M.; Chirico, R. D.; Diky, V.; Muzny, C. D.; Watanasiri, S.; Chen, C.-C. New Technologies for Product on-Demand Design; 11th International Conference on Properties and Phase Equilibria for Product and Process Design, Hersonissos, Crete, Greece, May 21, 2007; Elsevier: Amsterdam, The Netherlands, 2007. Book of Abstracts, p 99.
  - (18) Frenkel, M. Global Information Systems in Science: Application to the Field of Thermodynamics. *J. Chem. Eng. Data* **2009**, *54*, 2411–2428.
  - (19) Wakeham, W. A.; Assael, M. A.; Atkinson, J. K.; Bilek, J.; Fareleira, J. M. N. A.; Fitt, A. D.; Goodwin, A. R. H.; Oliveira, C. M. B. P. Thermophysical Property Measurements: The Journey from Accuracy to Fitness for Purpose. *Int. J. Thermophys.* **2007**, *28*, 372–416.
  - (20) Gani, R. Chemical Process Design: Challenges and Opportunities. *Comput. Chem. Eng.* **2004**, *28*, 2441–2457.
  - (21) Kang, J. W.; Diky V.; Chirico, R. D.; Magee, J. W.; Muzny, C. D.; Abdulagatov, I.; Kazakov, A. F.; Frenkel, M. *J. Chem. Eng. Data* **2010**, *55*, 3631–3640.
  - (22) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Muzny, C.; Kazakov, A. F.; Magee, J. W.; Abdulagatov, I. M.; Kang, J. W. NIST ThermoData Engine, NIST Standard Reference Database 103b-Pure Compounds, Binary Mixtures, and Chemical Reactions, version 5.0; NIST, Standard Reference Data Program: Gaithersburg, MD, 2010.
  - (23) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **1975**, *21*, 1086–1099.
  - (24) Fredenslund, A.; Gmehling, J.; Rasmussen, P. *Vapor Liquid Equilibrium Using UNIFAC*; Elsevier: Amsterdam, The Netherlands, 1977.
  - (25) B. E.; Prausnitz, J. M.; O'Connell, J. P. *The Properties of Gases and Liquids*, 5th ed; McGraw Hill: New York, 2001, p 8.75.
  - (26) Roux, M. V.; Jimenez, P.; Davalos, J. Z.; Turron, C.; Afeefy, H. Y.; Liebman, J. F. Enthalpies of formation of the methyl benzenecarboxylates. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 887–890.
  - (27) Steele, W. V.; Chirico, R. D.; Knipmeyer, S. E.; Nguyen, A. Vapor Pressure, heat capacity, and density along the saturation line, measurements for dimethyl isophthalate, dimethyl carbonate, 1,3,5-triethylbenzene, pentafluorophenol, 4-tert-butylcatechol, 4-methylstyrene, and N,N'-bis(2-hydroxyethyl)ethylenediamine. *J. Chem. Eng. Data* **1997**, *42*, 1008–1020.
  - (28) Oleviskii, V. M.; Martynenko, R. L.; Pogodina, M. O.; Kuzmina, N. V.; Sokolova, T. K. Boiling temperatures of methyl esters of phthalic acids and their solubilities in organic solvents. *Trudy GIAP (U.S.S.R.)* **1959**, *10*, 93–98.
  - (29) Potin-Gautier, M.; Grenier, P.; Bonastre, J. New analytical application of the method for determining vapor pressures by inert gas saturation. Part II. Case of phthalates. *Anal. Lett.* **1982**, *15*, 1431–1448.
  - (30) Wahlbeck, P. G. Comparison and interrelations for four methods of measurement of equilibrium vapor pressures at high temperatures. *High Temp. Sci.* **1986**, *21*, 189–232.
  - (31) Dong, Q.; Chirico, R. D.; Yan, X.; Hong, X.; Frenkel, M. Uncertainty Reporting for Experimental Thermodynamic Properties. *J. Chem. Eng. Data* **2005**, *50*, 546–550.
  - (32) Glaser, F.; Ruland, H. Vapor Pressure Curves and Critical Data for Several Technically Important Organic Substances. *Chem. Ing. Tech.* **1957**, *29*, 772–775.
  - (33) (a) Span, R.; Wagner, W. Equations of State for Technical Applications. I. Simultaneously Optimized Functional Forms for Nonpolar and Polar Fluids. *Int. J. Thermophys.* **2003**, *24*, 1–39. (b) Span, R.; Wagner, W. Equations of State for Technical Applications. II. Results for Nonpolar Fluids. *Int. J. Thermophys.* **2003**, *24*, 41–109. (c) Span, R.; Wagner, W. Equations of State for Technical Applications. III. Results for Polar Fluids. *Int. J. Thermophys.* **2003**, *24*, 111–161.
  - (34) *Synapse - Chemical Product Design Software*, Molecular Knowledge Systems, Inc.: Bedford, New Hampshire; <http://www.molknow.com/index.htm>. Accessed December 15, 2010.
  - (35) *ICAS - Integrated Computer Aided System*, Technical University of Denmark: Lyngby; <http://www.capec.kit.dtu.dk/>. Accessed December 15, 2010.

CI100373T