

Cranium: Component Software for Physical Property Estimation

Dr. Kevin G. Joback
Molecular Knowledge Systems, Inc.

Introduction

Physical property values are essential for process design, simulation, and optimization. Invariably, for the particular compounds, compositions, temperatures, and pressures the engineer is seeking, data is not available. Fortunately there are numerous computer software programs which provide accurate estimates for many physical properties.

However, with today's widespread use and integration of chemical engineering software, physical property estimation packages must provide more than just good estimates. Physical property estimation software must also:

- Intelligently choose the best estimation technique
- Enable the engineer to add new, possibly proprietary estimation techniques
- Integrate with corporate molecular structure databases
- Integrate with commercial and proprietary simulation packages
- Be very easy and intuitive to use

Cranium is an expert physical property estimation package which addresses each of these issues. Reasoning about a compound's molecular structure or a mixture's components, Cranium identifies the material's chemical family, selects the most accurate estimation technique, automatically dissects molecular structures into group fragments, and then estimates the material's physical property.

Background: Physical Property Estimation

Physical property estimation takes advantage of property-property relationships and property-structure relationships. Figure 1 graphs the critical temperature against the normal boiling point for 535 chemicals. A general trend is evident enabling the engineer to estimate an unknown critical temperature if he or she knows the chemical's boiling point. A statistical regression approximates the data with an averaged absolute error of less than 4%.

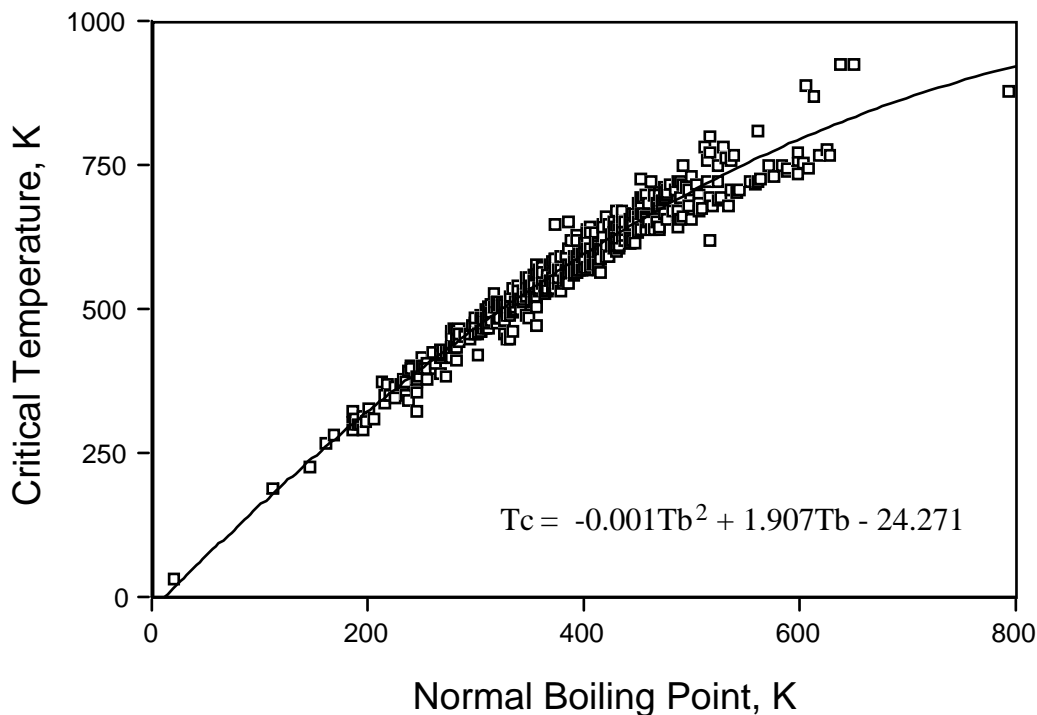


Figure 1: Critical Temperature-Boiling Point Relationship

Estimation techniques that equate a predicted physical property to a function of other, more readily known properties, are called equation oriented techniques. Recursive physical property estimation may be needed if some the required physical properties also must be estimated.

Ultimately a chemical's physical properties are dependent upon the structure of its constituent molecules. Relationships between physical properties and molecular structure have been observed for many years. Figure 2 graphs the critical pressure as a function of carbon number for several normal alkenes.

Transforming the critical pressure to the reciprocal square root uncovers a linear trend. This linear relationship implies that adding a -CH₂- group increases the transformed critical pressure by 9.6E-3. This value is called the -CH₂- group's contribution.

Using statistical regression we can determine the contributions for many groups. Table 1 shows several group contributions used to estimate the critical pressure. To find the critical temperature for a new compound we simply total the contributions for all groups occurring in the compound's structure. These types of estimation techniques are called group contribution techniques.

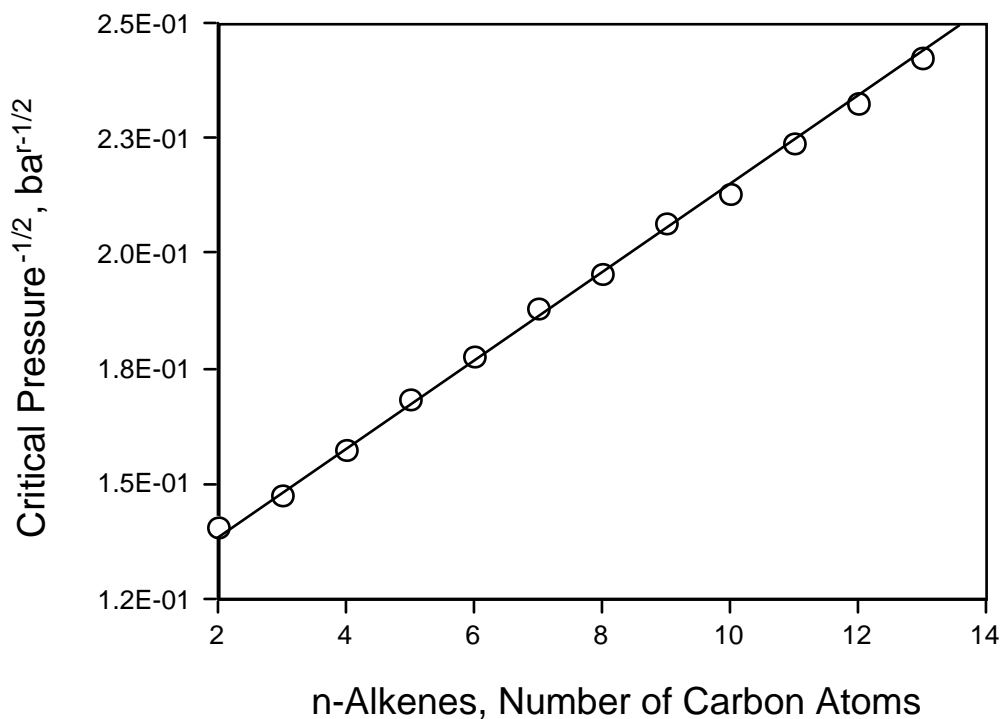


Figure 2: Linear Critical Pressure - Structure Relationship

Hundreds of equation oriented and group contribution techniques exist which estimate many of the physical properties needed by chemical engineers. Some techniques are very accurate yielding estimates within 1 or 2% of experimental data. Others are more approximate, often very simple to use but yielding estimates with errors of 10 or 20%. Some techniques are very general, applicable for many classes of compounds over large ranges of temperature and pressure. Many techniques are very focused, applicable only for certain classes of compounds at specific temperatures and pressures.

Choosing the Best Estimation Technique

Physical property estimation techniques are often applicable to specific types or families of compounds, e.g., polar, hydrocarbons, highly fluorinated, or branched. Before choosing estimation techniques, Cranium identifies the chemical family of each material or mixture. This identification is performed using a set of rules.

Table 1: Critical Pressure Group Contributions

Group	Contribution	Group	Contribution
-CH ₃	0.0140	=CH-	0.0070
-CH ₂ -	0.0096	>C=O	0.0033
>CH-	0.0044	-OH	-0.0048
>C<	-0.0011	-COOH	0.0051
=CH ₂	0.0124	Intercept	0.1130

The majority of rules deal with the occurrence of atoms, bonds, and groups. Figure 3 shows the rule for identifying branched compounds.

Branched Compounds Rule:

A compound is branched if any atom is connected to at least three atoms which are themselves connected to at least two atoms.

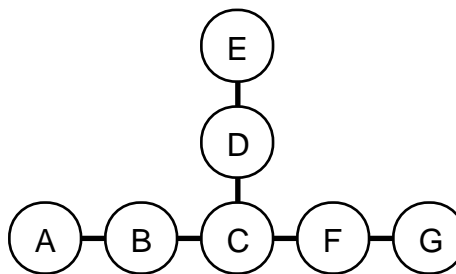


Figure 3: The Branched Compounds Classification Rule

Applying the rule to the structural schematic finds atom C connected to three atoms, B, D, and F, which are themselves connected to two atoms. Atom B is connected to atoms A and C. Atom D is connected to atoms C and E. Atom F is connected to atoms C and G.

Applying the rule to 3-ethyl-pentane identifies the compound as branched. Applying the rule to a normal hydrocarbon identifies the compound as not being branched.

Once the compound's chemical families have been identified, Cranium sorts all techniques by accuracy and then begins trying the most accurate technique first. The preamble of each technique's code uses chemical family, molecular structure, and state variable values to determine the technique's applicability. If the technique is not applicable to branched compounds or high pressure, the preamble code would exit signaling an error. Cranium would record the error and then continue the estimation trying the second most accurate technique.

Depth-First Property Estimation

Cranium executes the ordered techniques in a depth first manner. Figure 4 shows a simplified technique tree for the estimation of vapor pressure. Cranium found two techniques to estimate vapor pressure with Tech 1 more accurate than Tech 2. Cranium tries Tech 1 executing the technique's preamble to determine applicability. Assuming the technique is applicable its core computations are performed. During these computations the values of the critical temperature, normal boiling point, and enthalpy of vaporization are needed. Cranium then recursively repeats this depth-first estimation process with each of these properties. The critical temperature is tried first. All techniques are collected, ordered by accuracy, and then executed in a depth-first manner.

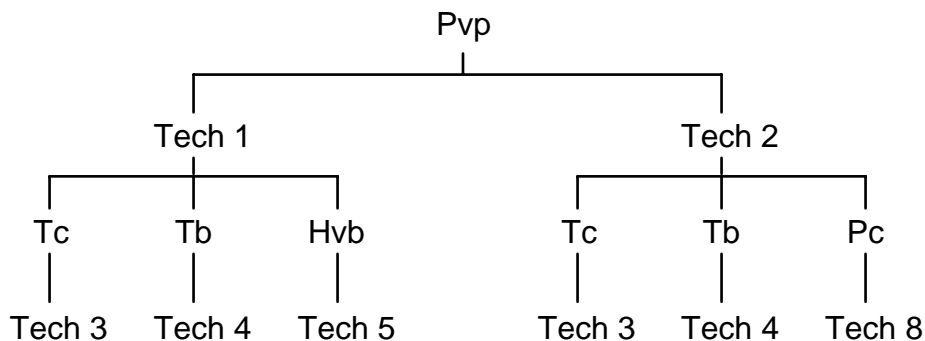


Figure 4: Depth-First Search of Estimation Techniques

In our simple example we assume there is only one technique for the critical temperature. This technique is executed and its value returned to Tech 1. The depth-first traversal continues estimating the normal boiling point and then the enthalpy of vaporization.

Let us assume the enthalpy of vaporization could not be estimated. This will occur if there are no applicable techniques, missing group contributions, or required properties which themselves could not be estimated. In this case an error is signaled which is logged by Cranium for later examination by the engineer.

This error propagates up the tree to the last decision point, the choice of Tech 1 in our example. Tech 1 was tried first because it was the most accurate. However, it could not produce an estimate. Cranium then continues attempting to estimate the vapor pressure using Tech 2. Tech 2 also requires values for the critical temperature and normal boiling point. However, these were already successfully computed during the execution of Tech 1. For efficiency, Cranium stores the property values computed during an estimation. Therefore, during the execution of Tech 2 the critical temperature and normal boiling point values are already available. Finally the critical pressure is estimated and a final estimate for the vapor pressure is obtained.

Because many physical properties are related to one another Cranium keeps track of each technique used to ensure circularities do not occur. A technique can not be executed with the same parameters more than once. Executing an estimation technique with different parameters, e.g., different temperatures, is allowed to enable the iterative solution of implicit physical properties.

Adding Estimation Techniques

With the availability of powerful statistical analysis software many companies are developing their own proprietary physical property estimation techniques. This has the advantage of creating techniques with higher accuracy because development focused only on those compounds of importance to the company.

Adding new estimation techniques to physical property software has traditionally been difficult if not impossible. This is because estimation techniques were treated as code to be entered by the software developer and compiled into an executable program. Unlike these programs, Cranium treats estimation techniques as data.

Cranium's Techniques chapter enables engineers to enter their own physical property estimation techniques. A very simple, C-like language is used. Once entered, Cranium compiles the estimation techniques for rapid execution, and then stores the code in the same way it stores a compound's boiling point. Figure 5 shows Cranium's estimation technique editor displaying code for the Peng-Robinson equation of state.

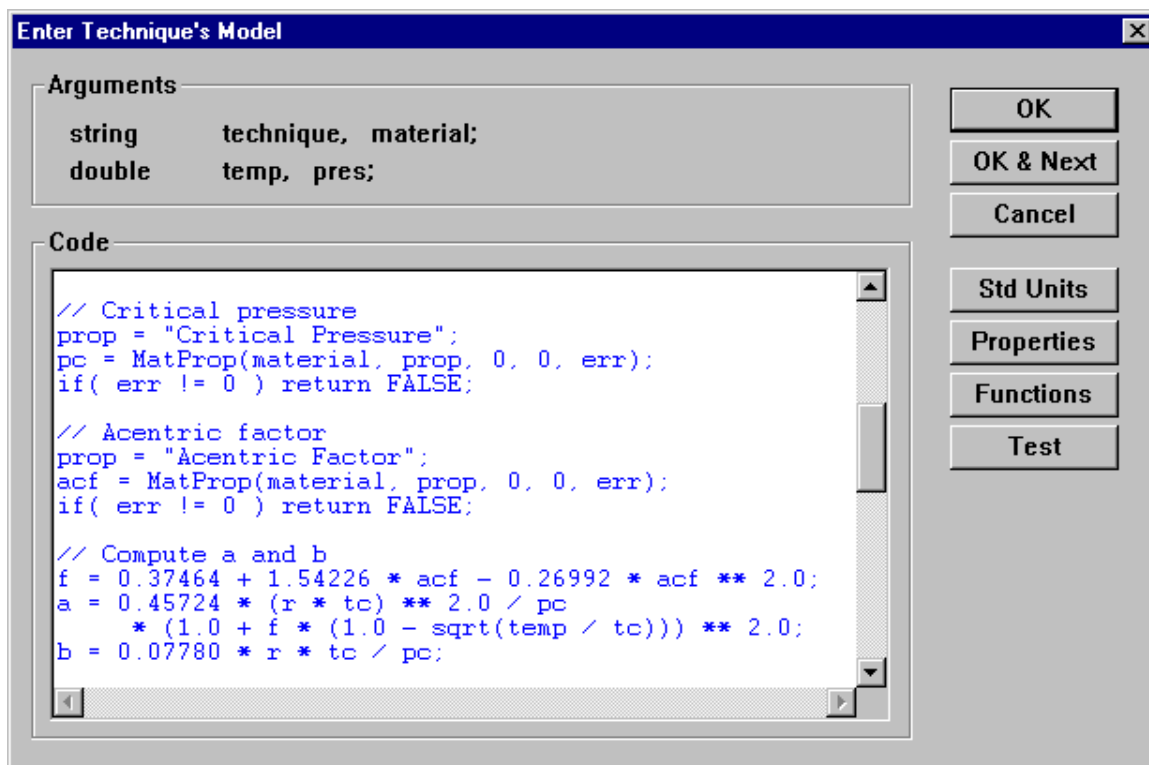


Figure 5: Cranium's Estimation Technique Editor

Molecular Structure Management

Dissecting molecular structures into groups is the most tedious step of using group contribution techniques. This is complicated by the fact that invariably each technique uses its own unique combination of groups. Cranium eliminates this tedious step by providing a graphical structure editing interface which records all the connectivity information needed to automatically decompose a structure into groups. All the engineer needs to do is draw the molecule's structure.

Figure 6 shows Cranium's structure editor. Modeled after a simple drawing program, the editor enables the engineer to quickly place atoms and connect them with bonds to fully specify a compound's molecular structure. The resulting two dimensional structures are sufficient for the majority of estimation techniques.

Cranium uses a network pattern matching algorithm to identify all groups within a structure. Figure 7 shows the example of finding an alcohol group in diacetone alcohol's structure. Each atom in the group is first associated with each matching atom in the molecule. The oxygen atom in the -OH group, O21, is associated with the two oxygen atoms in the structure, O5 and O19. The hydrogen atom in the -OH group, H22, is associated with the twelve hydrogen atoms in the structure. The bonds for each atom in an association are then compared. The oxygen atom in the -OH group is connected by two single bonds. O5 is connected to a double bond and therefore does not continue to match. O19 is connected to two single bonds and therefore does continue to match. Finally, the neighboring atom of each associated atom is checked. The hydrogen atom in the -OH group has an oxygen atom as its neighbor. There is only one hydrogen atom in the structure which has an oxygen neighbor, H20. The other hydrogens all have a carbon neighbor and are thus not matched. The -OH group thus matches atoms O19 and H20 in the structure.

Cranium uses this algorithm to automatically match a molecular structure with a set of groups. The flexibility of the algorithm enables engineers to easily enter new techniques containing new groups.

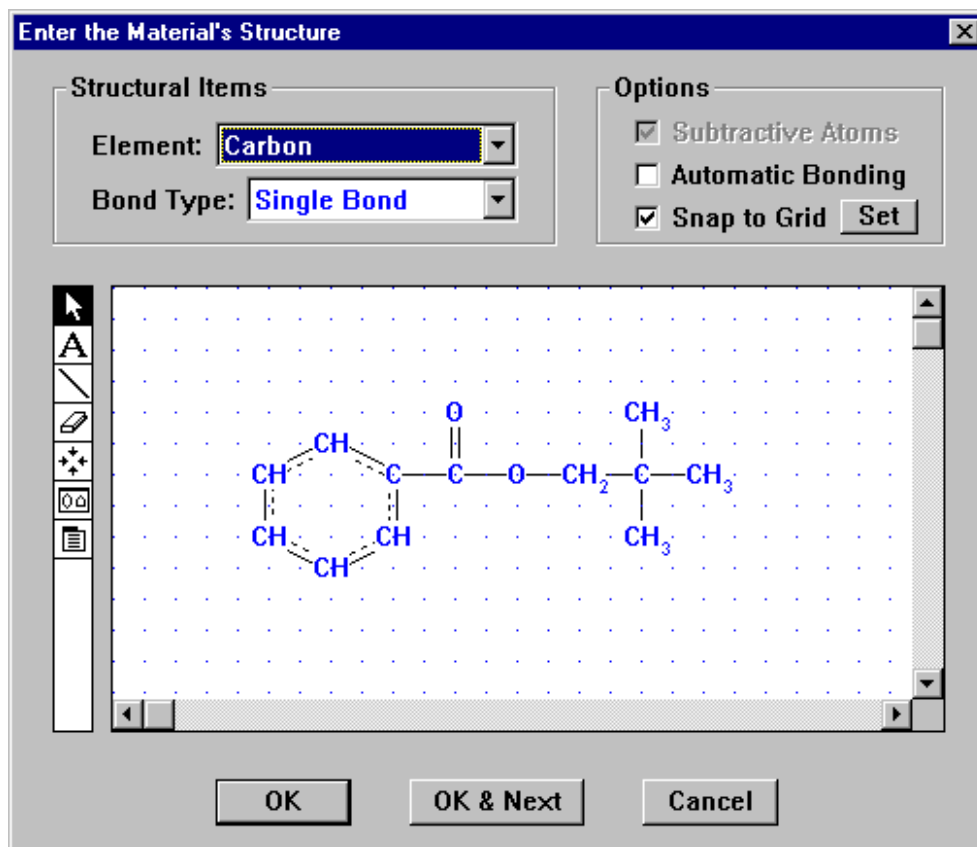


Figure 6: Cranium's Molecular Structure Editor

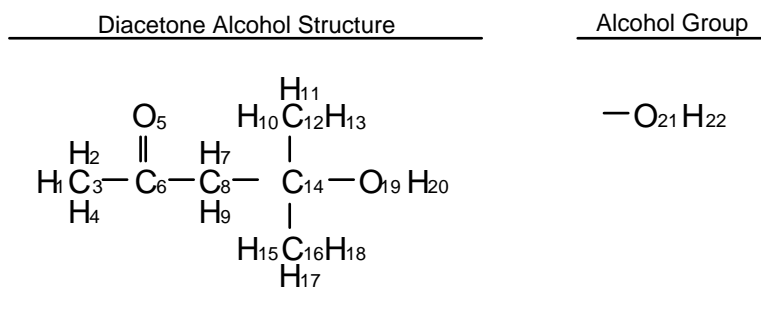
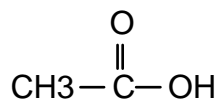


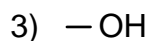
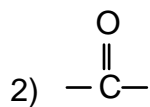
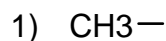
Figure 7: Dissection of Molecular Structure into Groups

Before dissection groups are ordered from most specific to least specific. Cranium defines specificity by the number of atoms and bonds in a group. This ensures multiple interpretations have the most specific interpretation. For example, using the groups presented earlier in Table 1, acetic acid can be dissected in the two ways presented in Figure 8.

Acetic Acid's Structure



Dissection 1



Dissection 2

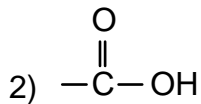
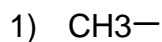


Figure 8: Alternative Dissections of Acetic Acid

Both dissections are correct, but Dissection 1 uses a ketone and an alcohol group to represent the acid group. Dissecting groups with the largest number of atoms and bonds first, the acid group in our example, preserves the acid characteristic of the molecule which will typically yield better estimates.

Integration with Structure Databases

Many chemical companies have large corporate databases containing information on thousands of chemicals. Very often the molecular structure of each chemical is also stored in the database.

Numerous file formats have been developed for storing molecular structure. The most common file format is called a molfile format. Cranium can import a molfile into its molecular structure editor. This enables the engineer to simply access the chemical's database file, import the molecular structure, and estimate its properties.

Integration with Simulation Packages

Cranium can export physical property data to third-party software packages. Currently, export capabilities are limited to copying and pasting data between applications or writing data to text files. Figure 9 shows a dialog window detailing the exporting of physical property data into a format compatible with BPRC's MultiBatchDS batch distillation simulation program.

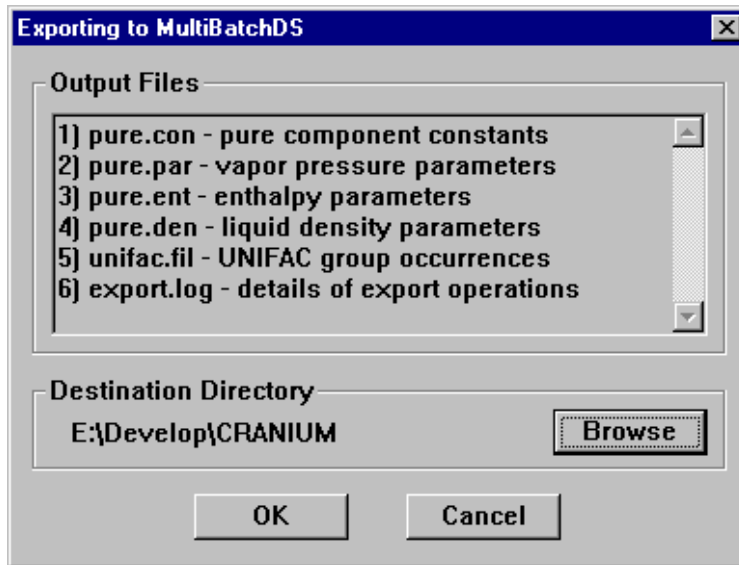


Figure 9: Cranium can Export Data to Simulation Packages

Cranium's export capabilities enable it to be used a central repository of physical property knowledge. Molecular structures, physical property data, and estimation techniques can all be added to Cranium in a single, consistent manner. This data can then be translated into different formats and exported to various commercial or proprietary software programs.

User Interface Concepts

Cranium was developed with the flexibility to accommodate the tremendous amount of data and estimation techniques used in industry. Figure 10 shows the user interface to a Cranium knowledge base. A knowledge base is analogous to an electronic book. Each knowledge base is divided into chapters and pages. The Introduction chapter provides identifying information and on-line documentation. The remaining seven chapters provide data on elements, structures, materials, mixtures, estimation techniques, units of measure, and literature references.

Each chapter is divided into pages. Each page holds data on a single entity. Thus in the elements chapter, each page holds data on a single element. The push buttons on the window's bottom move to the next page or the previous page.

Figure 10 shows a page of Cranium's Mixtures chapter. The top of the window shows the tabs designating each chapter. The central data pane scrolls to display fields for physical properties. The engineer can enter data in the data fields or press the compute button to have Cranium generate estimates for each physical property.

Engineers are able to create their own custom knowledge bases. These would contain their own material data, estimation techniques, molecular structures, etc. These knowledge bases can be freely distributed to other Cranium users or to non-users using a runtime package. Cranium thus enables engineers to quickly generate technical, electronic literature for internal, research, or marketing purposes.

The Future

Physical property estimation has long been a utility task serving design or simulation software. However, we are moving into a new age in which the simulation of products is becoming more important than the simulation of processes. Major challenges in the fields of pharmaceuticals, refrigerants, lubricants, and

polymers are WHAT compounds to make not HOW to make them. Physical property estimation software and its companion molecular design software will continue to be used to provide a competitive edge in the race to market better chemical products.

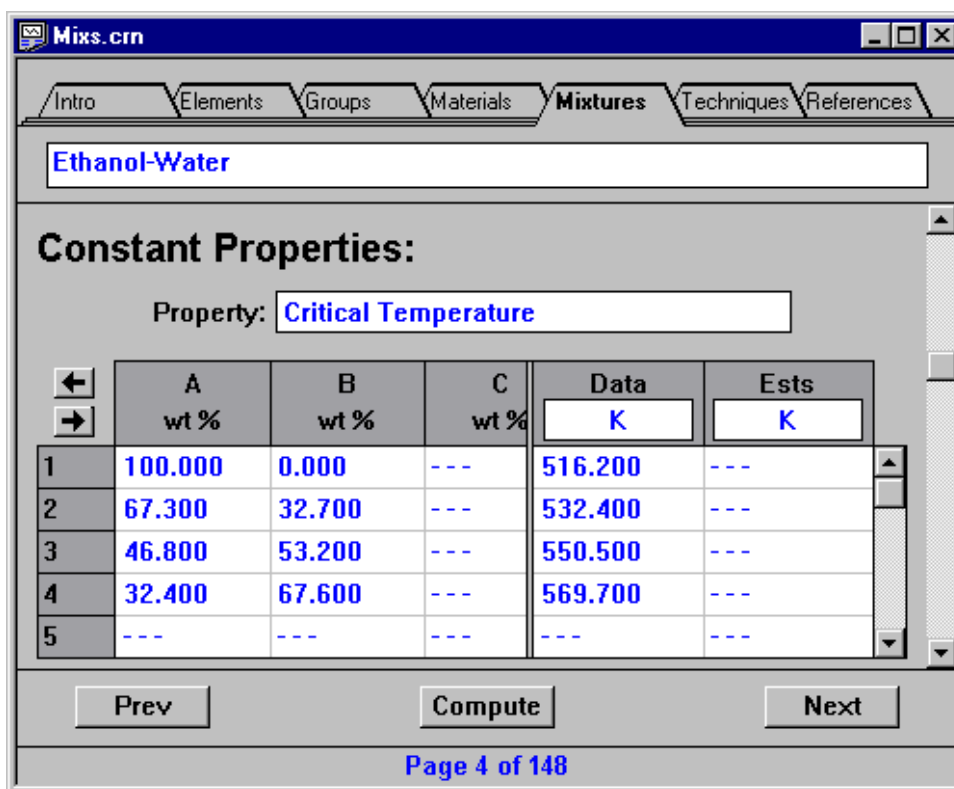


Figure 10: Cranium's Knowledge Base Interface

Contact Information

For discussions about this paper you may contact Dr. Kevin G. Joback at Molecular Knowledge Systems, Inc., PO Box 10755, Bedford, NH 03110-0755, USA, Phone: 1-603-472-5315, FAX: 1-603-472-5359, eMail: kevin@MolKnow.com. Additional information on Cranium can also be found at Molecular Knowledge Systems' web site: <http://www.MolKnow.com>.