

Utah State University

DigitalCommons@USU

---

Reports

Utah Water Research Laboratory

---

1-1-1990

## PEP Property Estimation Program and Chemical Property Database

William J. Doucette

Mark S. Holt

Follow this and additional works at: [https://digitalcommons.usu.edu/water\\_rep](https://digitalcommons.usu.edu/water_rep)



Part of the [Civil and Environmental Engineering Commons](#), and the [Water Resource Management Commons](#)

---

### Recommended Citation

Doucette, William J. and Holt, Mark S., "PEP Property Estimation Program and Chemical Property Database" (1990). *Reports*. Paper 510.

[https://digitalcommons.usu.edu/water\\_rep/510](https://digitalcommons.usu.edu/water_rep/510)

This Report is brought to you for free and open access by the Utah Water Research Laboratory at DigitalCommons@USU. It has been accepted for inclusion in Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



**PEP**  
**Property Estimation Program**  
**and**  
**Chemical Property Database**

**Utah Water Research Laboratory**  
**Utah State University**  
**Logan, Utah 84321**

---

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . .	vii
DISCLAIMER . . . . .	viii
INTRODUCTION . . . . .	1
Background . . . . .	1
PEP Overview . . . . .	2
PEP Features . . . . .	3
What Do I Need to Use PEP? . . . . .	4
Installation of PEP . . . . .	4
General Programming Description . . . . .	5
Starting the PEP Software System . . . . .	5
Menus, Buttons and Icons . . . . .	7
Tutorial . . . . .	8
REFERENCE SECTION . . . . .	10
PEP Processor . . . . .	10
MCI Module . . . . .	11
Overview . . . . .	11
Entering Chemical Structure . . . . .	12
Calculating MCIs . . . . .	14
How PEP Calculates MCIs . . . . .	15
Choosing the Properties . . . . .	16
Choosing the Regression Models and Chemical Classes . . . . .	16
Statistics Cards . . . . .	18
Estimating the Properties and Viewing the Results . . . . .	19
Adding or Deleting MCI-Property Regression Models . . . . .	19
Limitations of MCI-Property Regression Models . . . . .	20
TSA Module . . . . .	22
Overview . . . . .	22
Entering the Structural Information . . . . .	23
Calculating Total Surface Area (TSA) . . . . .	25
Choosing the Most Appropriate TSA-Property Regression Model . . . . .	27
Estimating the Properties . . . . .	28
Development of TSA-Property Relationships . . . . .	28
Limitations of TSA-Property Relationships . . . . .	29

---

## TABLE OF CONTENTS (Cont'd)

UNIFAC Module . . . . .	30
Overview . . . . .	30
Entering Structural Information . . . . .	31
Calculate Activity Coefficients . . . . .	34
Editing Parameters . . . . .	34
Estimating Properties . . . . .	34
Limitations of UNIFAC Approach to Estimating S and $K_{ow}$ . . . . .	34
Property/Property Module . . . . .	35
Overview . . . . .	35
Selecting the Properties to be Estimated . . . . .	36
Choosing the Property-Property Regression Model . . . . .	36
Viewing the Calculated Values . . . . .	36
Limitations of the Property/Property Module . . . . .	37
PEP Batch . . . . .	38
Overview . . . . .	38
Input Structure . . . . .	39
Output Options . . . . .	40
Start Batch Driver . . . . .	40
CHEMICAL PROPERTY DATABASE . . . . .	41
Overview . . . . .	41
Searching for Chemical Compounds in PEP's Database . . . . .	43
Sorting the Database . . . . .	43
Adding or Deleting Data . . . . .	43
Printing Information from the Database . . . . .	44
Exporting Information from the Database . . . . .	44
Moving to Other PEP Modules . . . . .	44
Changing the Units of Measurement . . . . .	44
PEP Models . . . . .	45
Overview . . . . .	45
Input Property Values . . . . .	47
Input Environmental Values . . . . .	48
Calculate Distribution . . . . .	48
PEP Help . . . . .	50
REFERENCES . . . . .	51

---

## LIST OF TABLES

	<u>Page</u>
1. MCIs Calculated by PEP . . . . .	15
2. UNIFAC Groups . . . . .	33

---

## LIST OF FIGURES

	<u>Page</u>
1. PEP stack icons . . . . .	1
2. PEP opening screen . . . . .	6
3. The steps for using a menu for command selection . . . . .	7
4. Buttons and icons used in PEP . . . . .	8
5. Opening screen of PEP tutorial . . . . .	9
6. Example illustrating the use of the PEP Processor's flow chart interface . . . . .	10
7. Screen display of PEP MCI module . . . . .	11
8. Standard Macintosh file selection dialog box . . . . .	12
9. Example connection table . . . . .	13
10. Statistics card associated with the MCI module . . . . .	18
11. Example dialog box for the input of new MCI-property relationships . . . . .	20
12. TSA module card from PEP . . . . .	22
13. Example Alchemy file . . . . .	24
14. Example Cartesian coordinate file . . . . .	24
15. Example card for the entry of atomic coordinates . . . . .	25
16. Dialog box for editing van der Waal radii . . . . .	26
17. Example TSA card . . . . .	27
18. Example card from the PEP UNIFAC module . . . . .	31
19. UNIFAC module card used to select UNIFAC groups . . . . .	32
20. Example card from PEP property-property module . . . . .	35
21. Example card for PEP Batch, MCI module . . . . .	38

---

## LIST OF FIGURES (Cont'd)

	<u>Page</u>
22. PEP Batch, explain SMILES file card . . . . .	39
23. Example card from the Chemical Property data base . . . . .	42
24. PEP database degradation properties . . . . .	42
25. Representation of Fugacity Level 1 compartments . . . . .	45
26. PEP Models card, input for Fugacity Level 1 . . . . .	46
27. PEP Models card, input for Fugacity Level 2 . . . . .	47
28. Example dialog box resulting rom the "Look for Values in Prop.DB" option . . . . .	48
29. PEP Models results card . . . . .	49
30. Example card from PEP Help stack for MCI module . . . . .	50

---

## ***ACKNOWLEDGMENTS***

PEP was initially developed with funding provided by the Air Force Office of Scientific Research, Bolling Air Force Base, DC 20332-6448, Grant No. AFOSR-89-0509

Project Leader: Lt. Col. T. Jan Cerveny, Life Science Directorate, Department of the Air Force.

Authors: William J. Doucette and Mark S. Holt

Project Investigators: W. J. Doucette (Principal Investigator), D. K. Stevens, R. R. Dupont, R. C. Sims, and J. E. McLean

Graduate Students: Mark Holt, Doug Denne, Rick Miles and Joe Frazier

Programmers: Mark Holt, Joe Frazier and Mike Jablonski (NR Systems, Inc.)

The authors would also like to acknowledge the following individuals for their assistance with the SMILES interpreter: Eric Anderson, Gil Veith, Chris Russom (U.S. EPA, ERL-Duluth)



---

## ***DISCLAIMER***

Utah State University and the authors of PEP make no warranties, either expressed or implied with respect to the operation or subsequent use of property values obtained through the use of PEP. In no event will Utah State University or the authors of PEP be liable for direct, indirect, or consequential damages resulting from the use of this software.

---

## ***INTRODUCTION***

### **Background**

Mathematical models are often used by environmental scientists and engineers to estimate the fate and impact of organic chemicals in the environment. Use of these models requires a variety of parameters describing site and chemical characteristics. Aqueous solubility (S), the octanol/water partition coefficient (K<sub>ow</sub>), the organic carbon normalized soil/water sorption coefficient (K<sub>oc</sub>), vapor pressure (P<sub>v</sub>), Henry's Law constant (H), and bioconcentration factor (BCF) are considered key properties used to assess the mobility and distribution of a organic chemical in environmental systems.

One major limitation to the use of environmental fate models has been the lack of suitable values for many of these properties. The scarcity of data, due mainly to the difficulty and cost involved in experimental determination of such properties for an increasing number of synthetic chemicals, has resulted in an increased reliance on the use of estimated values.

Quantitative Property-Property Relationships (QPPRs), based on the relationship between two properties as determined by regression analysis, are used to predict the property of interest from another more easily obtained property. Quantitative Structure-Property Relationships (QSPRs) often take the form of a correlation between a structurally derived parameter(s), such as molecular connectivity indices (MCIs) or total molecular surface area (TSA) and the property of interest.

Selection and application of the most appropriate QPPRs or QSPRs for a given compound is based on several factors including: the availability of required input, the methodology for calculating the necessary structural or topological information, the appropriateness of correlation to chemical of interest and an understanding of the mechanisms controlling the property being estimated.

Incorporation of QPPRs and QSPRs into a computer format is a logical and necessary step to gain full advantage of the methodologies for simplifying fate assessment.

## PEP Overview

A Property Estimation Program (PEP), utilizing MCI-property, TSA-property and property-property correlations and UNIFAC-derived activity coefficients, has been developed for the Apple Macintosh microcomputer to provide the user with several approaches to estimate S, Kow, Pv, H, Koc and BCF depending on the information available.

Structural information required for the MCI and UNIFAC calculation routines can be entered using either Simplified Molecular Identification and Line Entry System (SMILES) notation or connection tables generated with commercially available two-dimensional drawing programs. The TSA module accepts 3-D atomic coordinates entered manually or directly reads coordinate files generated by molecular modeling software. The program's built-in intelligence helps the user choose the most appropriate QSPR or QPPR based on the structure of the chemical of interest. In addition, the statistical information associated with each QSPR or QPPR in PEP can be displayed to help the user determine the model's validity. For the regression-based property estimation models, assessments of accuracy based on the 95% confidence interval and estimated precision of the experimental values are also provided along with the estimated property value.

PEP also provides a batch mode that provides users with a method for the convenient, unattended calculation of MCIs, TSA and UNIFAC activity coefficients and the subsequent estimation of physical properties for large numbers of compounds.

A chemical property database, containing experimental values of S, Kow, H, Pv, Koc, and BCF compiled from a variety of literature sources and computerized databases was used for developing the MCI-property, TSA-property and property-property relationships used in PEP. This database, which currently contains over 800 chemicals, is linked directly to PEP.

The property estimation modules in PEP are also linked directly to the Level 1 and 2 Fugacity Models. The combination of the various property estimation methods, chemical property database, and simple environmental fate models provides users with a methodology for predicting the environmental distribution of an organic chemical in a multi-phase system requiring only the structure of the chemical of interest as input.

PEP was designed to be intuitive and user friendly. The easiest way to become familiar with the PEP is to try clicking on the buttons and pull down menus found on each card. Any comments or suggestions regarding improving the operation of PEP would be greatly appreciated by the authors.

- **PEP Features**

- Developed using Hypercard™ for the Apple Macintosh series of personal computers
- Comprised of a chemical property database and four property estimation modules
- Uses standard Macintosh operations (buttons, menus, windows)
- Simple user interface based on flow chart design
- Four property estimation methods are available:
  - Molecular Connectivity Indices (MCIs)-property correlations
  - Total Surface Area Regressions (TSA)-property correlations
  - Property-Property Correlations
  - UNIFAC derived activity coefficients
- PEP can be used to estimate six chemical/physical properties
  - Solubility (S)
  - Octanol-water partition coefficients (Kow)
  - Henry's Law Constant (H)
  - Vapor Pressure (Pv)
  - Organic carbon normalized soil-water distribution coefficients (Koc)
  - Bioconcentration factors (BCF)
- Universal and class specific regression models are available
- PEP uses decision support for determination of chemical class.
- Estimates include 95% prediction interval for each regression based estimated value
- Statistical information readily available for each regression
- New regression models can be easily added
- Database contains over 800 chemicals having at least one property values
- Each chemical has at least one property and a two-dimensional SMILES string
- Chemicals in database can be search for by chemical name, CAS number, synonym, or selected from an alphabetized list
- Property estimation modules and property database are linked directly to the Fugacity Level 1 and 2 environmental fate models
- Published and on-line documentation
- Includes PEP tutorial
- Includes PEP Batch for estimating properties or calculating MCIs, TSA, or UNIFAC activity coefficients for large numbers of chemicals without continuous user input

• **What Do I Need To Use PEP (i.e. Hardware requirements)?**

1. Macintosh II computer or better with
2. 4,000 Kbytes (4 Megabytes) of usable hard disk space,
3. 3 Meg of RAM installed,
4. running system software 6.0.5 or higher, and
5. HyperCard 2.0 software or higher installed with the size allocated to 1500 MB.

• **Installation of PEP**

PEP is typically shipped on one 3.5 inch 1.44 Megabyte floppy disk. To install PEP:

1. Insert the PEP disk into the disk drive.
2. Drag the file PEP.sea to the hard drive that PEP is to be installed on. When the PEP.sea file has been copied to the hard drive you can eject the disk by dragging the "PEP" disk icon to the trash.
3. Double click on the icon of the "PEP.sea" file. This will start the installation process by first creating a new folder called "PEP system" on the hard drive and then uncompacting five HyperCard stacks: (1) "Chemical Property Data Base", (2) "PEP Processor", (3) "PEP Help", (4) "PEP Models", and (5) "PEP Batch" (not necessarily in that order).
4. Drag the "PEP.sea" icon to the trash and remove it from your hard drive using the "empty the trash" command which can be found under the Special pull down menu located at the top of the screen.

The installation process is now complete. Test the installation by double clicking on the "PEP system" folder, then double clicking on the file "Chemical Property Data Base". If installation was successful the opening card of PEP will appear. If you have a problem with the installation please contact Mark Holt at (801)750-3916 or Bill Doucette at (801)750-3178. *Note: PEP can also be sent on two 3.5 inch 800k floppy disks if requested.*

## General Programming Description

The PEP software system is a HyperCard™ based program that runs on Apple Macintosh computers. HyperCard, which is bundled with most Macintosh computers sold, offers graphics, information storage, the means to display information in a variety of formats, the ability to establish links between related information, a high level language (HyperTalk), the ability to extend HyperTalk by writing new commands in a compiled language (i.e. C or Fortran) and a mechanism to transfer control to other Macintosh applications. The PEP system uses all these features.

HyperCard treats each screen full of information as a card and each set of related cards as a stack. Cards can contain fields for data and buttons for action procedures to operate on the data in the fields. This allows the standard Macintosh interface to be used without the direct use of the Macintosh toolbox routines, greatly simplifying programming. In order to create a user interface, the programmer simply draws, or creates the buttons or fields that are to be used. The link between buttons, fields and cards is done through HyperTalk. HyperTalk is an high-level, interpreted language used to establish links between related information and perform simple calculations within HyperCard. However, large repetitive tasks and complicated computations can be very slow if HyperTalk is used. HyperCard also allows the programmer to create extensions of HyperTalk in a lower level language. These extensions, called external functions (XFCN) and external commands (XCMD), greatly increase the speed of repetitive and calculation intensive algorithms over using HyperTalk itself and can also be used to implement custom Macintosh features such as popup menus and custom dialog boxes.

## Starting the PEP Software System

If the installation of PEP (as described on the previous page) was successful, a new folder called "PEP system", containing five HyperCard stacks ("Chemical Property Data Base", "PEP Processor", "PEP Help", "PEP Models", and "PEP Batch"), should appear on your hard drive.

PEP is started from the Macintosh operating system by clicking twice (double clicking) on any one of the five stack icons, except the PEP batch stack, shown in Figure 1. The five stacks must be in the same folder on a hard disk. This will open the PEP Processor stack and display the opening card shown in Figure 2. PEP can also be used by opening any HyperCard stack and then choosing "Open" from the "File" menu and selecting the "PEP Processor" stack.

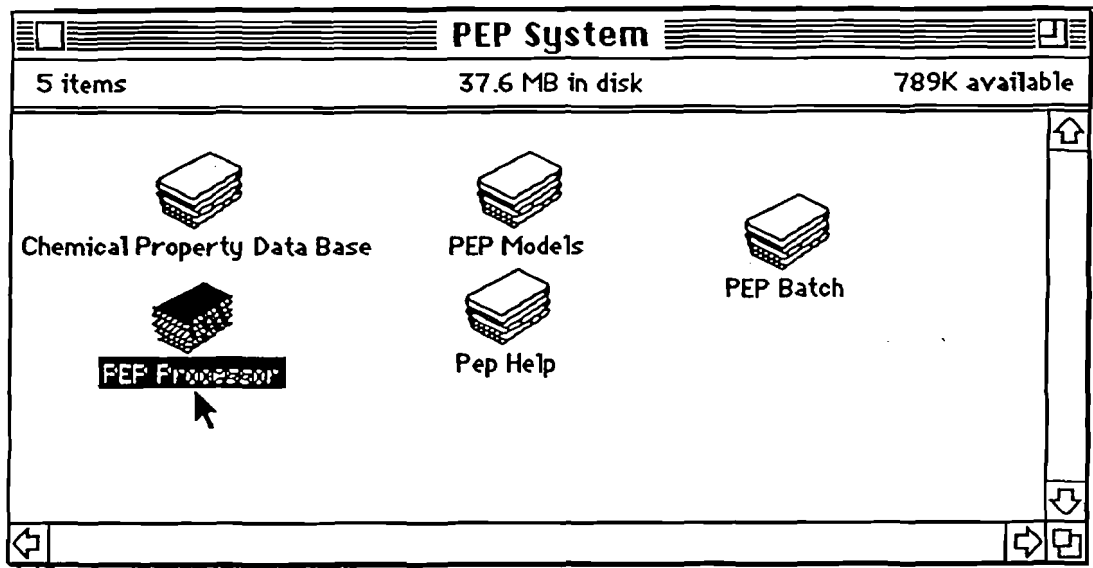


Figure 1. PEP stack icons.

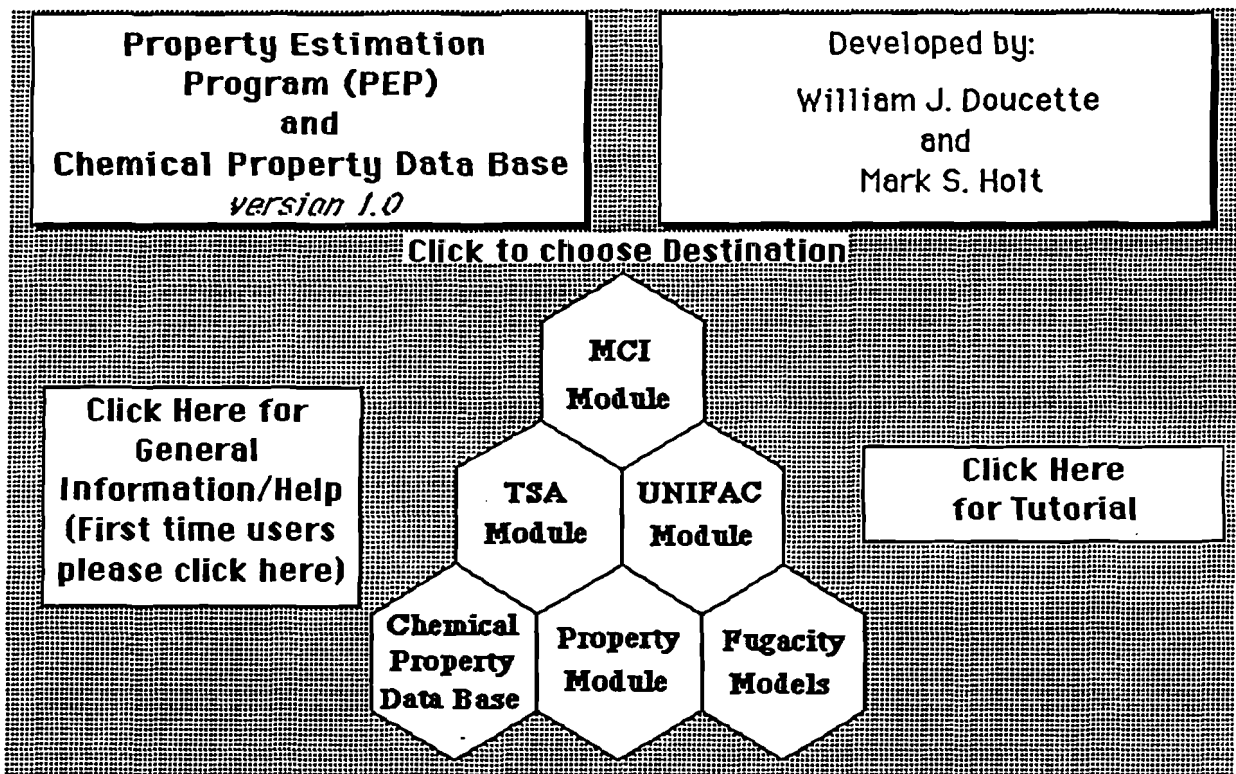
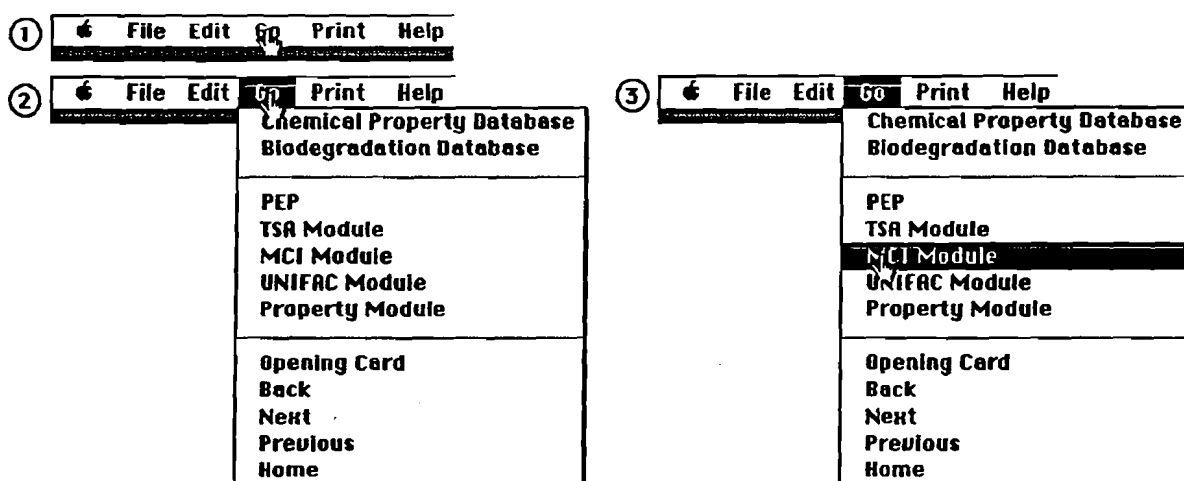


Figure 2. PEP opening screen.

## Menus, Buttons and Icons

In Macintosh applications most cursor movements are accomplished by use of the mouse or trackball. Action buttons are operated by positioning the "hand" cursor over the button area and then depressing and releasing the mouse button. This is referred to as clicking. Popup and pull down type menus are operated by positioning the hand cursor over the button and then depressing and holding the mouse button. While holding down the mouse button move the cursor over the desired menu item and release (Note: on slower machines such as the Mac Plus, the menu selections may be slow to appear, but be patient, they will eventually show up.). Figure 3 shows the steps involved in using a menu for the selection of a command. The action of moving the mouse while holding down the mouse button is called dragging.



1. Menu Bar before the mouse is pressed
2. Menu Bar after the mouse is pressed
3. Menu Bar after the mouse is moved over the desired command but before the mouse button is released.

Figure 3. The steps for using a menu for command selection

The PEP software system is comprised of five HyperCard stacks that are linked together by various menus and buttons. The pull down menus, located at the top of each card and buttons positioned at various card locations allow the user to navigate through PEP.

The File and Edit menus at the top of each PEP card duplicate the general File and Edit menus in Hypercard (Please refer to the Hypercard manual for complete instructions.) The Go menu allows the user to move to the either of the four property estimation modules, the chemical property database, the fugacity model or the batch mode.



In addition to the menu items found at the top of each PEP card, buttons, such as those displayed below, can be found on each of the various PEP cards. Figure 4 shows each button icon, its title and its action.










<b>Icon</b>	<b>Title</b>	<b>Action</b>
	PEP Icon	Shows Opening Screen of PEP
	Return Arrow	Takes you back to the card you were at prior to this one
	Help	Shows Help for the current card
	Information	Shows general information
	Eye	Shows the equations or statistics
	Book	Shows the reference
	First Card	Takes you to the opening card of HyperCard
	Pop Up Button	Lets the user choose from a popup menu list
	Action Button	Initiates some action or calculation

Figure 4. Buttons and icons used in PEP.

## Tutorial

The "Click Here for Tutorial" button, accessible on the opening screen of PEP, takes the user to the opening screen of the tutorial. As shown in Figure 5, this screen takes the form of a flow chart depicting the overall design of PEP. Individual tutorials, available for the chemical property database and each of the four property estimation modules, can be activated by clicking on the appropriate button in the flow chart. Each tutorial automatically runs the chosen component of the PEP system while illustrating its operation in a step-by-step manner.

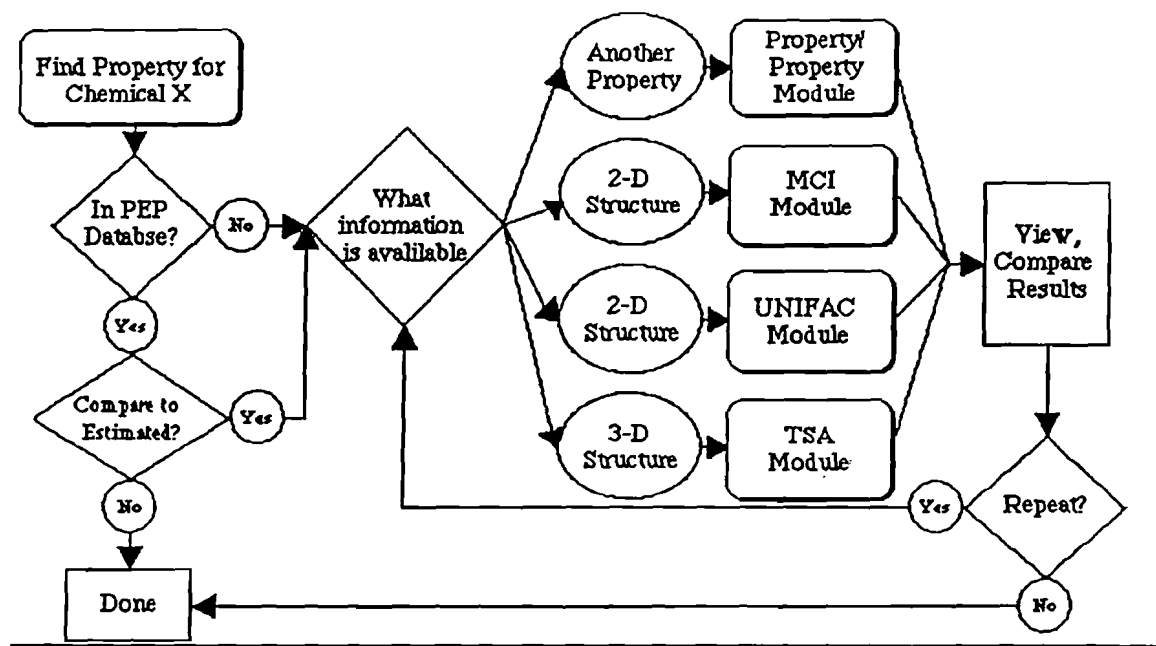


Figure 5. Opening screen of PEP tutorial (Flow chart illustrating overall design of PEP system).

---

## REFERENCE SECTION

### PEP Processor

The PEP Processor stack contains the algorithms for data input, calculations and output of the physical-chemical properties estimated using MCI-Property, TSA-Property, and Property-Property correlations and UNIFAC derived activity coefficients. The PEP Processor is divided into four modules, one module for each of these estimation methods.

The user interface for each module is designed as a flow chart consisting of a series of numbered steps or actions required to estimate the physical-chemical property(s) of interest. The steps are numbered sequentially from left to right and top down. Initially, all but the first step is dimmed. As the user completes the first step the next step will darken, indicating the appropriate progression. However, during any step a darkened, previously accomplished step can be redone or the selection changed. Figure 6 shows this progression of steps.

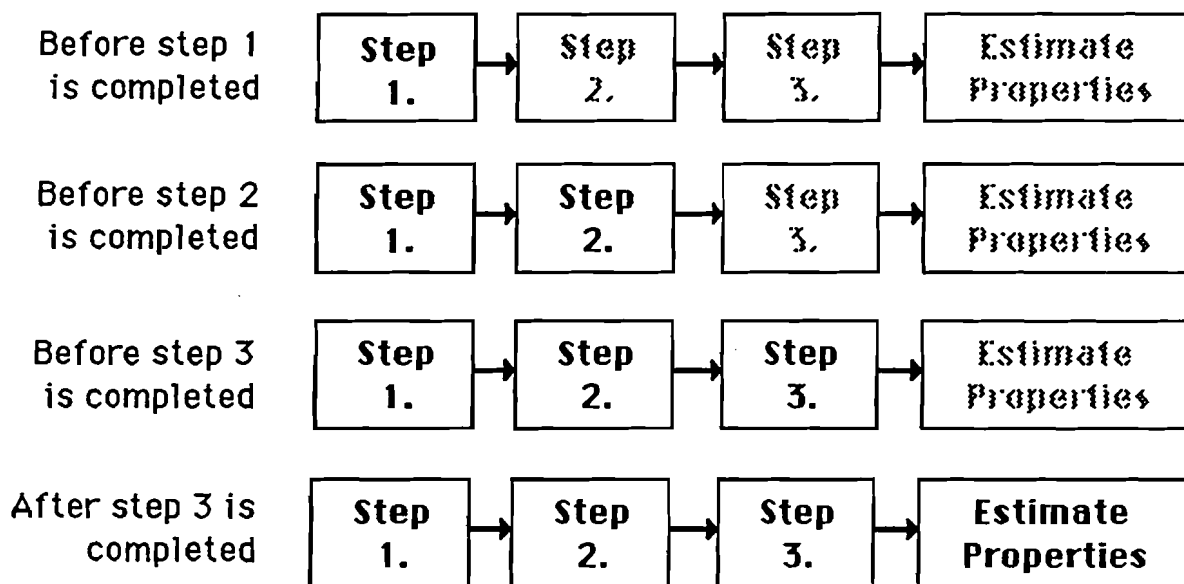


Figure 6. Example illustrating the use of the PEP Processor's flow chart interface.

# MCI Module

## Overview

The user interface of the MCI module, like the other property estimation modules in PEP, is designed in the form of a flow chart depicting the steps that must be completed in order to use the module. As illustrated in Figure 7 by the fully darkened MCI module card, the four steps that must be completed before the selected physical-chemical properties can be estimated using are: (1) input the necessary structural information for the chemical of interest using either SMILES notation or a connection file, (2) calculate and display the MCIs, (3) select the properties to estimate and (4) choose the most appropriate MCI-property regression model.

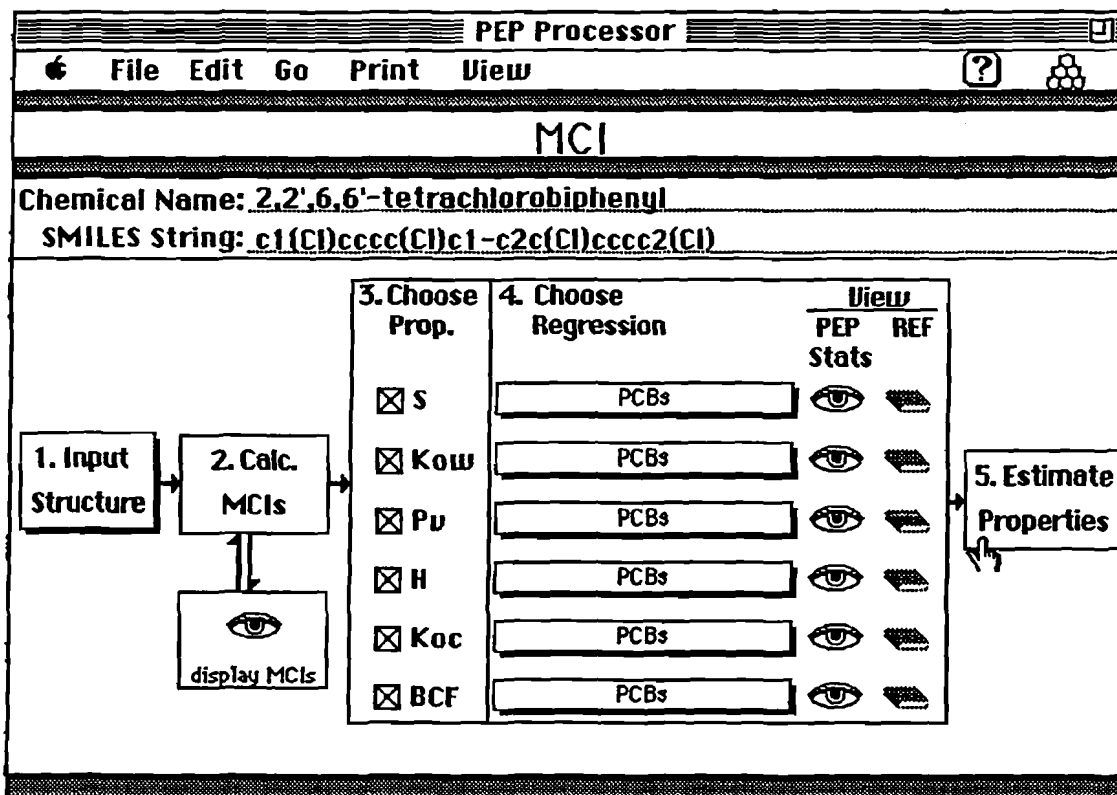


Figure 7. Screen display of PEP MCI module.

## Entering Chemical Structure

Upon entering the MCI module, only the first step in the flow chart is active as indicated by its darkened status. The two dimensional molecular structure of the chemical of interest is needed to calculate the MCIs. You must first input the necessary structural information using either SMILES [33,34] notation or connection files before you can continue to the second step. Select either option from the "Input Structure" popup button.

If you select SMILES, two blank lines appear, one for the chemical name (optional) and one for the SMILES string. Once you have entered the SMILES string, click the "OK" button or the carriage return. The SMILES string must conform to the standard set by Anderson, Veith, and Weininger (1987) and Weininger (1988) with the following exceptions: a single bond connecting aromatic rings must be explicitly denoted by an "-", the SMILES string must be Hydrogen suppressed, and the SMILES string cannot contain any "{" or "]" qualifiers. SMILES is a chemical notation language specifically designed for computer use. It is a method of "unfolding" a 2D chemical structure into a single line of characters containing the structural information. For users unfamiliar with SMILES notation, a detailed description describing its use can be found in scrollable window directly below the SMILES string input line.

If a connection table is chosen as the input method, the standard Macintosh file selection dialog box, as shown in Figure 8, will be used to select the file. This requires that a connection table has already been created for the chemical of interest.

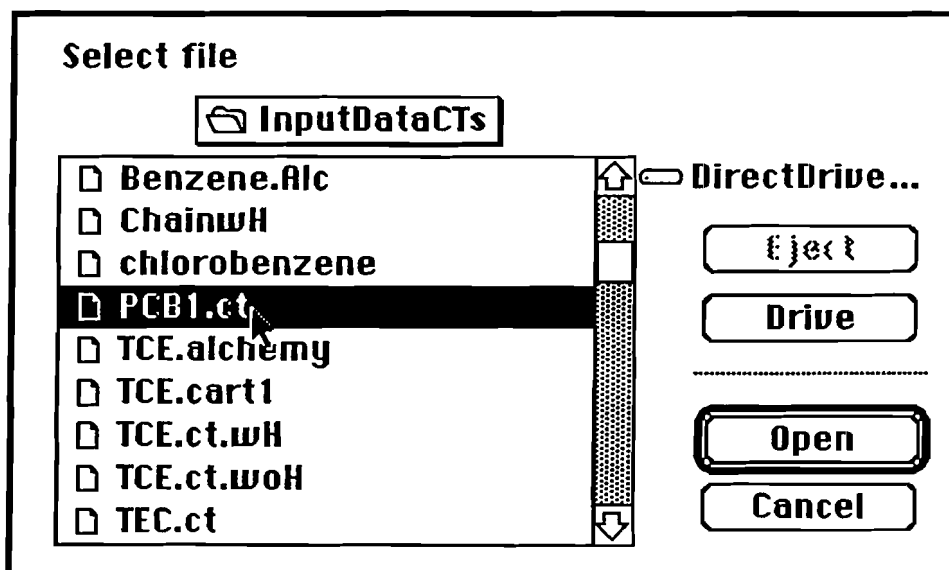


Figure 8. Standard Macintosh file selection dialog box.

Connection tables can be generated from commercially available, two-dimensional (2D) chemical drawing programs, such as ChemDraw™, Chemintosh™, or ISIS/Draw™, that have the ability to save the structure as a connection table file. The connection table file must be formatted the same as a connection table from ChemDraw (1989). An example of a ChemDraw compatible connection table is shown in Figure 9.

Title line	Example connection table			
Number of Atoms	10,10			
Number of Connections	7.62500	2.29167	0.00000	C
	7.62500	3.12500	0.00000	C
X,Y,Z Coordinates (Not used)	8.33333	3.54167	0.00000	C
	9.04167	3.12500	0.00000	C
	9.04167	2.29167	0.00000	C
Atom Symbol	8.33333	1.87500	0.00000	C
	8.33333	1.04167	0.00000	C
	9.16667	1.04167	0.00000	C
	10.00000	1.04167	0.00000	C
	10.83333	1.04167	0.00000	C
	2	1	4	1
Atom Numbers	3	2	4	1
	4	3	4	1
	5	4	4	1
Bond Type	6	5	4	1
	1	6	4	1
Not Used	7	6	1	1
	8	7	1	1
	9	8	2	1
	10	9	1	1

Figure 9. Example connection table.

The first data line contains the title of the chemical or any other identifier. The second line consists of two numbers separated by a comma. The first number is the number of atoms in the connection table and the second is the number of connections described in the connection table. The remaining lines describe the type of atoms in the molecule and their location (atom information) and the atoms to which they are connected (connection information). The total number of lines depends on the number of atoms in the molecule and the number of connections.

Atom information is contained in four columns separated by one or more spaces. Columns one through three are the X,Y,Z coordinates of the atom (not used for calculation of MCIs) and column four contains the atom symbol (e.g. C, Cl, Br, N etc.). Connection information is contained in four columns of whole numbers. Columns one and two contain the atom numbers of the atoms that are connected (atoms are numbered consecutively), column three contains the bond type (1, 2, 3, or 4), and column four usually contains a 1 (not used). The bond type in column three can be either "1" for a single bond, "2" for a double bond, "3" for a triple bond, or "4" for an aromatic bond. (NOTE: *The Macintosh compatible molecular modeling program, Alchemy II™, generates files containing connection table information along with 3D atomic coordinates. To use alchemy files for input into the MCI module simply treat the alchemy file as a "ChemDraw Connection Table".*)

As described in the next section, MCIs are calculated from the hydrogen suppressed structure of a chemical. Consequently, for the most efficient calculation of MCI, the connection tables created for input into the MCI module should be created from hydrogen suppressed structures. If the connection tables are not hydrogen suppressed PEP will automatically remove them. This can result in a significant increase in the time required to calculate the MCIs.

### Calculating MCIs

As soon as the structure is entered, the second step becomes active. Clicking on the "Calc. MCIs" button starts the calculation of the MCIs. If the structure was entered as a SMILES string the "mciSmile" XFCN converts the SMILES string into the proper format for the "mciChi" XFCN which calculates the MCIs. Similarly, if a connection table was used for input the "mciConvert" XFCN converts the connection table into the proper format. The MCIs that are calculated are listed along with the variable names in Table 1. Once the MCIs have been calculated, they can be viewed, exported, or printed by using the "Display MCIs" button under the "Calc. MCIs" button.

Table 1. MCIs calculated by PEP.

Variable Name	MCI Title	Orders calculated
np	normal path	0 through 6
ncl	normal cluster	3 through 6
nch	normal chain	3 through 6
npc	normal path/cluster	4 through 6
bp	bond path	0 through 6
bcl	bond cluster	3 through 6
bch	bond chain	3 through 6
bpc	bond path/cluster	4 through 6
vp	valence path	0 through 6
vcl	valence cluster	3 through 6
vch	valence chain	3 through 6
vpc	valence path/cluster	4 through 6
$\Delta vp$	delta valence path	0 through 6

### How PEP calculates MCIs

To calculate the MCIs for a given compound, a delta (d) value are first assigned to each non-hydrogen atom in the structure. Three d values were computed in this study: normal, bond, and valence. Normal deltas are computed by summing the number of bonds (single, double, etc. are counted as one bond) connected to the atom whose delta is being calculated. The bond deltas are calculated the same as the normal deltas except the bonds were taken at their face value (single is one, double is two, etc.) instead of each bond being equal to one. Valence deltas for each atom are computed according to equations (1) and (2) (Kier and Hall, 1986):

$$d_v = Z_v - h \quad (1)$$

$$d_v = (Z_v - h)/(z - Z_v) \quad (2)$$

where  $d_v$  is the valence delta,  $Z_v$  is the number of valence electrons in the atom,  $h$  is the number of hydrogen atoms bound to the atom, and  $Z$  is the atomic number of the atom. Equation (1) is used for those atoms in the first row of the periodic chart, and equation (2) is used for all other atoms.

Once the delta values have been calculated for each atom in the molecule, simple, bond and valence indices of different orders and types can be calculated. The order refers to the number of bonds in the skeletal substructure of fragment used in computing the index: zero order defines individual atoms, first order used individual bond lengths, second order uses two adjacent bond combinations, and so on. The type refers to the structural fragment (path, cluster, path/cluster or



chain) used in computing the index. A more detailed explanation of the calculation of MCIs can be found in Kier and Hall (1986).

The MCI calculation routine in PEP calculates simple, bond and valence indices of several types (path, cluster, chain, and path/cluster) and orders (0 through 6), if possible, for each molecule, resulting in a maximum of 54 index values for each molecule.

To account for non-dispersive force effects on aqueous solubility and solubility related properties zero through six order  $\Delta$  valence path indices ( $\Delta\chi$ ), as described by Bahnick and Doucette (1988), are calculated by PEP, in addition to the 54 indices described above. To calculate  $\Delta\chi$  indices, a nonpolar equivalent is made by substituting C for O or N atoms. MCIs are calculated for the nonpolar equivalent and values for  $\Delta\chi$  can be computed for each type of index by:

$$\Delta\chi = (\chi)_{np} - \chi \quad (3)$$

where  $\Delta\chi$  is the delta index,  $(\chi)_{np}$  is the index for the non-polar molecule and  $\chi$  is the index for the original molecule.

### Choosing the Properties

After the MCIs have been calculated, the third step becomes active. Select the property or properties you would like to estimate by clicking on the check box button next to it. You can simultaneously select all the properties by holding down the shift key and clicking any one of the property buttons.

### Choosing the Regression Models and Chemical Classes

After the properties are selected, step four becomes active and the regression models available for each property are displayed in a popup menu. Two categories of MCI-property relationships are displayed for each property. The first category of MCIs property relationships, preceded with the word PEP, were developed in this project using the experimental values reported in the PEP property database. "Universal" MCI-property relationships were developed using all available experimental data for a given property regardless of chemical class. "Class-specific" MCI-property relationships were developed if property values were available for a sufficient number (10 or greater) of compounds within a particular chemical class (i.e. PCBs, PAHs, ureas). In addition, several multi-class MCI-property correlations were developed for more broad classes of compounds such as: halogenated aliphatics and halogenated aromatics.

The second category of MCI-property relationships displayed for each property were obtained directly from the literature and are located below the PEP relationships, separated by a gray line, in the popup menu. By clicking the “book” found at the left of each literature MCI-property regression model, the coefficients,  $r^2$  value and the appropriate citation can be displayed

To illustrate the potential hierarchy of MCI-property relationships available to the user, an example for the predicting the vapor pressure ( $P_v$ ) of a polychlorinated biphenyl (PCB) is provided below. There are three appropriate PEP-derived MCI-property relationships available to the user, one developed using only PCBs, one using halogenated aromatics including PCBs and one using all compound types:

$$\log P_v = 5.814 (\text{nc5}) - 2.428 (\text{np3}) + 9.479 \quad (\text{PCBs})$$

$$\log P_v = -1.559 (\text{bp1}) + 6.622 \quad (\text{Halogenated aromatics})$$

$$\log P_v = -1.275 (\text{np3}) + 5.261 \quad (\text{Universal})$$

Generally, the use of a “class-specific” relationship, if available, should provide the best estimate (i.e. the estimate associated with the least amount of uncertainty). To automatically aid you in choosing the most appropriate MCI-property relationships, PEP looks in the SMILES string or connection file for groups of atoms and bonds that distinguish various chemical classes. The number of MCI-property relationships or chemical classes that are chosen by the program is determined by the number of different distinguishing subgroups that are found. For the example shown above, the most appropriate regression model, PCBs, would be made the default equation. The two other appropriate models, halogenated aromatics and Universal, would be denoted with a ♦ in the popup menu. If the compound entered into PEP does not fit one of the class specific models the “Universal” equation is selected as the default. You may also choose to ignore the regression model chosen by PEP and select your own.

MCI-property regression models are available for the following “classes” of chemicals: Universal, Universal Nonionizable, Universal Ionizable, Alcohols, Anilines, Carbamates, Halogenated, Aliphatics, Nonhalogenated Aliphatics, Halogenated Aromatics, PCBs, PAHs, Phenols, Triazines, and Ureas. Examples of representative chemicals for each of the classes can be found in the View menu. To see the structures of the chemicals click on the class name. Not all of the classes listed above are implemented for each property. The models not available for the properties to be estimated are dimmed in the popup menu.

## Statistics Cards

A summary of the regression statistics and list of compounds used to develop and evaluate each MCI-property relationship can be displayed by clicking the “eye” or “view statistics option” found at the left of each regression model. Information displayed on the statistics card includes: the MCI-property regression equation, the list of chemicals used in developing the regression model, the standard errors of the coefficients in the regression equation, the Analysis of Variance (ANOVA) table, the  $r^2$  value, a graph of the the predicted vs. estimated values, a graph of the residuals vs. the predicted values, a graph of the residuals vs. the number of standard deviations, a normal probability plot of the residuals, the  $X'X$  inverse matrix and appropriate reference. An example of the statistical information provided for each MCI-property relationship is shown in Figure 10.

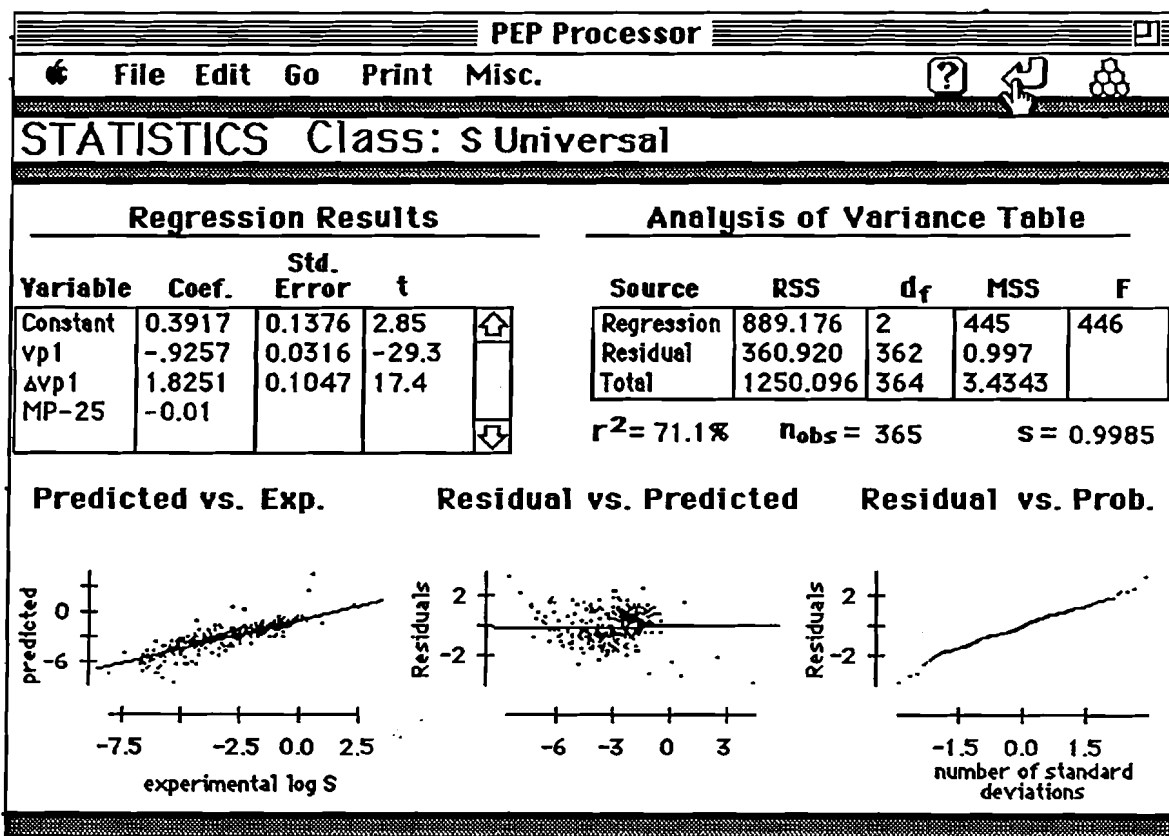


Figure 10. Statistics card associated with the MCI module.

The ANOVA table contains the degrees of freedom, the residual sum of squares, the residual mean square, and the variance ratio (F) for regression, residual and the total source of errors.

The  $X'X$  inverse matrix is used in PEP to calculate the prediction interval of an estimate. The matrix is derived by pre-multiplying the  $X$  matrix by its transpose and then inverting the result. The  $X$  matrix has a column for each variable in the regression equation and a row for each observation used to calculate the regression equation. Each row contains the value used for each of the variables in the regression equation. For example, if the regression equation is

$$Y_j = b_0 + b_1X_{j1} + b_2X_{j2}$$

where  $j$  is 1 to the number of observations,  $Y_j$  is the estimated values for each observation, the  $b_i$  are the regression coefficients, and the  $X_{ji}$  are the values of the variables used then the  $X$  matrix is

1	X11	X12
1	X21	x22
:	:	:
:	:	:
1	Xj1	Xj2.

The resulting  $X'X$  inverse is a square matrix with the number of rows and columns equal to the number of variables in the regression equations.

If the QSPR or QPPR was taken from the literature only the input variables and the statistical information provided in the original reference is included.

### Estimating the Properties and Viewing the Results

Click the "Est. Property" button to calculate the selected properties. The results card displays the estimated properties and their respective 95% prediction interval *Note: 95% prediction intervals are not available for MCI-property relationships taking from the literature.* You can return to the previous card by clicking on the "return" button at the upper right corner of the results card. The "Go" menu can be used to move to another module. You can compare the estimate property values with those contained in the "Chemical Property Database", if available, by clicking on the "Look in DB" button. Clicking this button activates a database "search by name" routine. The name on the results card must match exactly the name in the database for the search routine to find the compound. If a property value is not found an NA will be displayed.

### Adding or Deleting MCI-Property Regression Models

Additional MCI-property regression models can be added to the PEP MCI module through the statistics cards. Choosing the "New Stat. Card" option from the "Misc." menu of a statistics cards

while the type of statistics card that is to be added is the current card The statistics for a new regression equation can be added to PEP b. First the new card must be titled. This title will be used in the popup menus and on the results card. The first 28 characters of the title must be unique. The second step is to enter the regression equation. The user will be prompted for the number of terms in the equation, then prompted for each coefficient and the associated variable with the dialog box shown in Figure 11. The variables which are available for that type of statistics card will be in a popup menu for easy, consistent selection.

The dialog box has a title bar that reads "Choose MCI and Input Coefficient". Below the title, there are two input fields. The first is labeled "Choose MCI :" and contains the text "up0". The second is labeled "Coefficient:" and contains the text "-2.3". At the bottom of the dialog box, there are two buttons: "Cancel" on the left and "OK" on the right.

Figure 11. Example dialog box for the input of new MCI-property relationships

A relationship can easily be deleted from PEP by first making that statistic card the current card, and then choosing "Del. Stat. Card" from the "Misc." menu. The user will be prompted to confirm the deletion and then the regression list will be rebuilt.

### Limitations of MCI-Property Regression Models

Selection of the most appropriate MCI-property relationship depends on the structure of a particular chemical of interest, knowledge of the mechanism of the process, and the extent of the database used to develop the MCI-property relationship. For example, some MCI-property relationships are broader than others in the range of chemicals that are covered, and some have been established with a better understanding of the mechanisms or properties involved.

One problem that has limited the widespread acceptance of MCI-property correlations is that the actual physical meaning associated with the individual indices is not well understood. Frazier

(1990) and Doucette and Holt (1991), however, have shown a strong correlation between calculated molecular surface area and several MCIs for a variety of organic chemicals. MCI-property correlations tend to be class specific and thus are highly dependent on the type and range of compounds that were used to derive a particular correlation. Indiscriminate use of such models without an examination of number and type of compounds used to develop the model can result in considerable error.

## TSA Module

### Overview

As shown in Figure 12, the TSA module is similar in design to the MCI module. However, unlike molecular connectivity, the calculation of TSA requires information describing the geometry of the molecule in terms of its 3-D atomic coordinates.

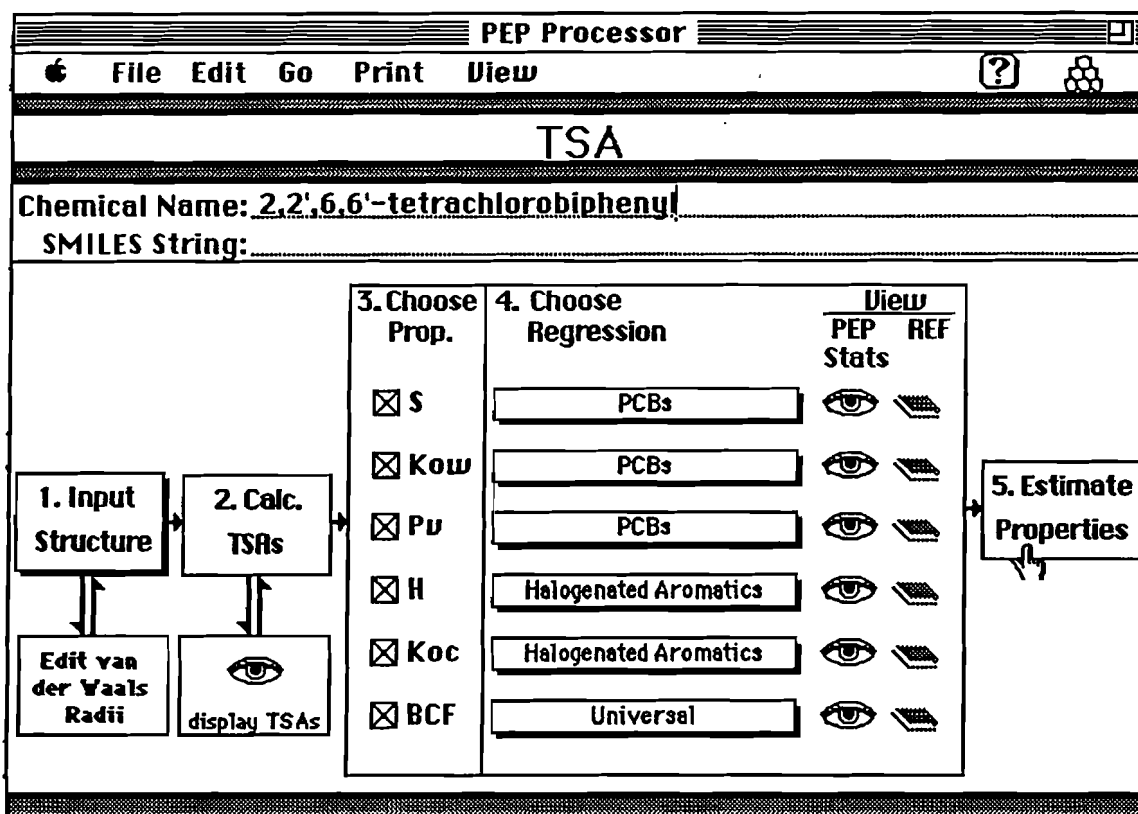


Figure 12. TSA module card from PEP

Three-dimensional coordinates can be obtained from X-ray crystallography data or from molecular modeling. Alchemy from Tripos Associates (1989) and Chem3D+ from Cambridge Scientific (1989) are examples of Macintosh compatible molecular modeling software that allows the user to draw a chemical structure, energy minimize the structure, and produce a file containing the three-dimensional coordinates. The TSA module is also designed to accept files generated by other hardware/software combinations including UNIX or VAX versions of CONCORD (Tripos

Associates, Inc., 1990), a hybrid expert system and molecular modeling software designed for the rapid generation of high quality approximate 3-D molecular structures.

### **Entering the Structural Information**

To estimate properties using the TSA module, you must first input the necessary structural information by using one of the three options available under the popup menu titled "Input Structure": (1) Alchemy file, (2) cartesian coordinates file, or (3) manually entered cartesian coordinates. The preferred method of structural input is via Alchemy files because they contain both the three-dimensional structure and the connection information. This information allows PEP to calculate the chemical's TSA and determine the most appropriate TSA-property relationship based on chemical class. Therefore, if an Alchemy file is chosen for the input then only the file selection dialog box will be presented for selection of the file. However, if either of the cartesian coordinate options are selected for input, both the standard file selection dialog box will be presented and an opportunity to select the connection table file or enter a SMILES string will be presented.

The format of an Alchemy file, shown in Figure 13, is similar to that of a ChemDraw connection table discussed previously. The first line of an Alchemy file contains the number of atoms followed by "ATOMS," the number of bonds followed by "BONDS," the number of charges followed by "CHARGES," and then the title of the file. The next set of lines contains six columns of information for each atom in the molecule including the hydrogen atoms. Column one contains the atom numbers, column two contains the atomic symbols, and columns three through five contain the X,Y,Z coordinates of the atom. The set of lines describing the atoms is followed by a series of lines containing four columns describing the bonds. The first column contains the bond number, the second and third columns contain the atom numbers of the two atoms connected, and the fourth column contains the type of bond either "SINGLE," "DOUBLE," "TRIPLE," or "AROMATIC."



	Number of Atoms	Number of Bonds	not used		title
	6 ATOMS, 5 BONDS, 0 CHARGES, TCE				
Atom numbers	1	C2	0.0008	0.0149	0.0048 0.0000
Atom symbols	2	Cl	0.7042	1.5686	0.4047 0.0000
X,Y,Z coordinates	3	C2	0.6631	1.1153	0.2713 0.0000
	4	Cl	1.5811	0.0700	0.7463 0.0000
not used	5	Cl	2.2442	1.0786	1.0242 0.0000
	6	H	0.2306	2.0846	0.0245 0.0000
Bond numbers	1	1	2	SINGLE	
Atom numbers	2	1	3	DOUBLE	
Bond type	3	1	4	SINGLE	
	4	3	5	SINGLE	
	5	3	6	SINGLE	

Figure 13. Example Alchemy file.

PEP accepts cartesian coordinates files having the following format. The file has one header line indicating the number of atoms in the molecule. The rest of the lines in the file describe each atom in the molecule. Only the first five columns of each line are used. Column one contains the atom symbols, column two contains the atom numbers, and columns two through five contain the X,Y,Z coordinates of the atom. An example cartesian coordinates file is shown in Figure 14.

Number of Atoms	Atom Symbols	Atom Numbers	X Y Z coordinates			Not used	
6	Cl	1	1.658981	1.412048	0.220383	12	2
	C	2	0.678864	0.010635	0.004303	2	1
	C	3	0.661346	0.015198	0.109131	2	2
	Cl	4	1.498978	1.476791	0.333939	12	3
	Cl	5	1.607437	1.456451	0.031769	12	3
	H	6	1.277542	0.912277	0.055084	5	2

Figure 14. Example Cartesian coordinate file.

If the option to manually enter the coordinates is chosen then a card is presented, as shown in Figure 14, that allows the coordinates and atom symbol of each atom to be entered individually from the keyboard. The X,Y,Z, coordinate values and atom symbols for each atom in the molecule are entered in the appropriate labeled boxes one at a time. After the information for each line is correctly entered the "Line OK" button is clicked. This enters the information into the scrollable window below. This process is repeated for each atom in the molecule. When all the atoms have been entered clicking the "Done" button will send the structural information to the TSA module.

File Edit Go Print Misc. ?

**Enter XYZ coordinates for the chemical**

Chemical Name: Benzene

Fill in one atom at a time at the bottom, type Return or Tab to put it into the table. To Edit an entry click on the line in the table. Click "All Done" when the table is complete.

Atom	Symbol	X coord	Y coord	Z coord
2	C			

Line OK

1 C 1 1 1

Clear

Cancel

All Done

Figure 15. Example card for the entry of atomic coordinates.

### Calculating Total Surface Area (TSA)

In addition to the 3-D molecular structure, the user must also input van der Waals radii for each of the atoms before the TSA of the molecule can be calculated. PEP automatically enters a van der Waals radius for each atom using values from Pauling (1960). However, when the "Calc. TSA" button is clicked the user has the opportunity to edit the van der Waals radii using the dialog box shown in Figure 16.

Atom Symbol	Radius Å	Default values taken from: Pauling, 1960. "The Nature of the Chemical Bond." Cornell University Press. Ithaca, New York.
C	1.7	<input type="button" value="Reset Defaults"/>  <input type="button" value="Cancel"/>  <input type="button" value="OK"/>
CL	1.8	
BR	1.95	
O	1.4	
N	1.5	
H	1.2	
S	1.85	
P	1.9	
F	1.35	
Solvent Radius	<input type="text" value="0.0"/>	

Figure 16. Dialog box for editing van der Waal radii.

This dialog box also contains a place to enter a solvent radius. If it is left at "0.0" then the total surface area will be calculated. Some relationships from the literature require the solvent accessible surface area to be calculated. If this is the case, the desired solvent radius can be entered.

Once the molecular geometry and the van der Waal radii are input, the "Calc. TSA" button becomes active and TSA can be calculated using a XFCN that was adapted from the SALVO2 algorithm developed by Pearlman (1980). This algorithm represents each atom of a molecule by a sphere centered at the equilibrium position of the nucleus. The radius of the sphere is equal to that of the van der Waals radius. Planes of intersection between spheres are used to estimate the contribution to surface area from the individual atoms or groups. The program computes the surface area of individual atoms or group by numerical integration, and the overlap due to intersecting spheres is excluded from the calculation. TSA is calculated by the summation of individual group contributions. These areas are then imported to the "TSAs" card shown in Figure 17.

TSA's					
Chemical Name: Benzene.flc					
TSA's Calculated with 0.0 Solvent radius					
atom #	sym.	isolated area	accessible area	isolated volume	accessible volume
1	ArC	36.317	15.391	20.580	12.027
2	ArC	36.317	15.397	20.580	12.028
3	ArC	36.317	15.382	20.580	12.023
4	ArC	36.317	15.394	20.580	12.027
5	ArC	36.317	15.390	20.580	12.027
6	ArC	36.317	15.394	20.580	12.026

Using a 0.0 Solvent Radius	
total area (Å <sup>2</sup> )	92.348
total volume (Å <sup>3</sup> )	72.158
Not Using a 0.0 Solvent Radius	
total area (Å <sup>2</sup> )	Not Calc.
total volume (Å <sup>3</sup> )	Not Calc.

N-TSA	O-TSA	P-TSA	S-TSA	ArN-TSA
0	0	0	0	0

Figure 17. Example TSA card.

To quantify the surface area attributed to the polar portions of the molecule the surface areas of nitrogen, oxygen, phosphorous, sulfur, and aromatic nitrogen atoms are individually separated from the TSA and placed on the "TSA's" card. A more detailed description of the TSA calculation method is provided by Pearlman (1980).

### Choosing the Most Appropriate TSA-Property Regression Model

After the TSA has been calculated, you can display the values and/or choose the properties of interest and a corresponding regression model using the same approach described in the MCI module. As discussed previously, "class-specific" regression models generally yield estimates associated with the least amount of uncertainty. If an Alchemy file is used to enter the 3D structure information or if a SMILES string or connection file is entered along with the cartesian coordinates, the decision support system in PEP will choose the most appropriate TSA-property regression model(s) as described in the MCI module. The most appropriate regression model will be made the default and a ♦ will be placed next to the other appropriate class name in the popup menu containing a list of the regressions. If no class specific regression models are available, the

“Universal” equation will be made the default. The operation of the TSA module from this point on is identical to that of the MCI module. Note: for solid solutes the melting point is needed to estimate the solubility. When the “S” button is clicked, the user is prompted with a dialog box to enter the melting point. Because the solubility will be estimated at 25° C, only if the melting point is above 25° C (solid solute) will the value be required. If the melting point is below 25° C it is set equal to 25° C. At this time only the “≤ 25° C” and “known” options are useable in the dialog box.

TSA-property regression models available within PEP are: Universal, Universal Nonionizable, Universal Ionizable, Alcohols, Anilines, Carbamates halogenated Aliphatics, Nonhalogenated Aliphatic, Halogenate Aromatics, Nonhalogenated Aliphatic, Halogenate Aromatics, PCBs, PAHs, Phenols, Triazine, and Ureas. Not all models are available for each property. The models not available for the properties to be estimated are dimmed in the popup menu. You can also view the statistical information associated with each model by clicking on the “eye” next to the regression equation.

### **Estimating the Properties**

After the appropriate TSA-property relationships have been selected you can now click on the “Estimate Property” button to calculate the estimated properties. The estimated properties and their respective 95% prediction interval will be displayed on the results card. Return from the results card by clicking on the “Return arrow” button at the upper right hand corner of the results card. Use the “Go” menu to move to another module. View values in the Chemical Property Database by clicking on the “Look in DB” button. If a property value is not available in the database, an NA will be displayed in the property field.

### **Development of TSA-Property Relationships**

The PEP TSA-Property relationships that were were developed using stepwise regression techniques and the data in the Chemical Physical Data Base . The stepwise regression was stopped when the coefficient of determination did not improve by at least 0.05 when the next variable was added.

For compounds containing polar functional groups, the addition partial TSA terms (i.e. nitrogen (N-TSA), oxygen (O-TSA), aromatic nitrogen (ArN-TSA), sulfur (S-TSA), or phosphorous (P-TSA)) significantly improved the TSA-property regression models. To view the regression information by clicking on the “eye” next to the regression title on the results card or on the TSA card.

## **Limitations of TSA-Property Relationships**

A major factor in the solubilitzation process is the energy required to create a cavity in the solvent into which the solute is placed. The energy needed for the hole formation is considered to be proportional to the surface area of the solute. TSA has been found to be linearly related to the logarithm of solubility for many classes of non-ionizable organic chemicals.

As with the MCI module, selection of the most appropriate TSA-property relationship depends on the structure of a particular chemical of interest, knowledge of the mechanism of the process, and the extent of the database used to develop the TSA-property relationship. Some TSA-property relationships are broader than others in the range of chemicals that are covered and some have been established with a better understanding of the mechanisms or properties involved.

## UNIFAC Module

### Overview

The UNIFAC (UNIQUE Functional Group Activity Coefficient) group contribution method for calculating activity coefficients, as described by Grain (1990), is implemented using both HyperTalk and an XFCN. The functional group interaction parameters, presented by Gmehling et al. (1982) and derived from vapor-liquid equilibria (VLE), are used in the calculation routine but can be changed by the user. After the activity coefficients are calculated they can be displayed along with relevant intermediate values and used to estimate S and  $K_{OW}$  by the following expressions (Arbuckle, 1986):

$$K_{OW} = 0.115 \gamma^{\infty W} / \gamma^{\infty O} \quad (2)$$

$$S \text{ (mol/L)} = 55.6 / \gamma^{\infty W} \quad (3)$$

where  $\gamma^{\infty W}$  is the activity coefficient of the chemical infinitely dilute in water and  $\gamma^{\infty O}$  is the activity coefficient of the chemical infinitely dilute in octanol.

The operation of the UNIFAC module, illustrated in Figure 18, is considerably different than the correlation modules previously described. To use the UNIFAC module you must input structural information, calculate the activity coefficients, choose then choose the properties to be estimated. Currently the only properties that can be directly estimated using the UNIFAC module are S and  $K_{OW}$ .

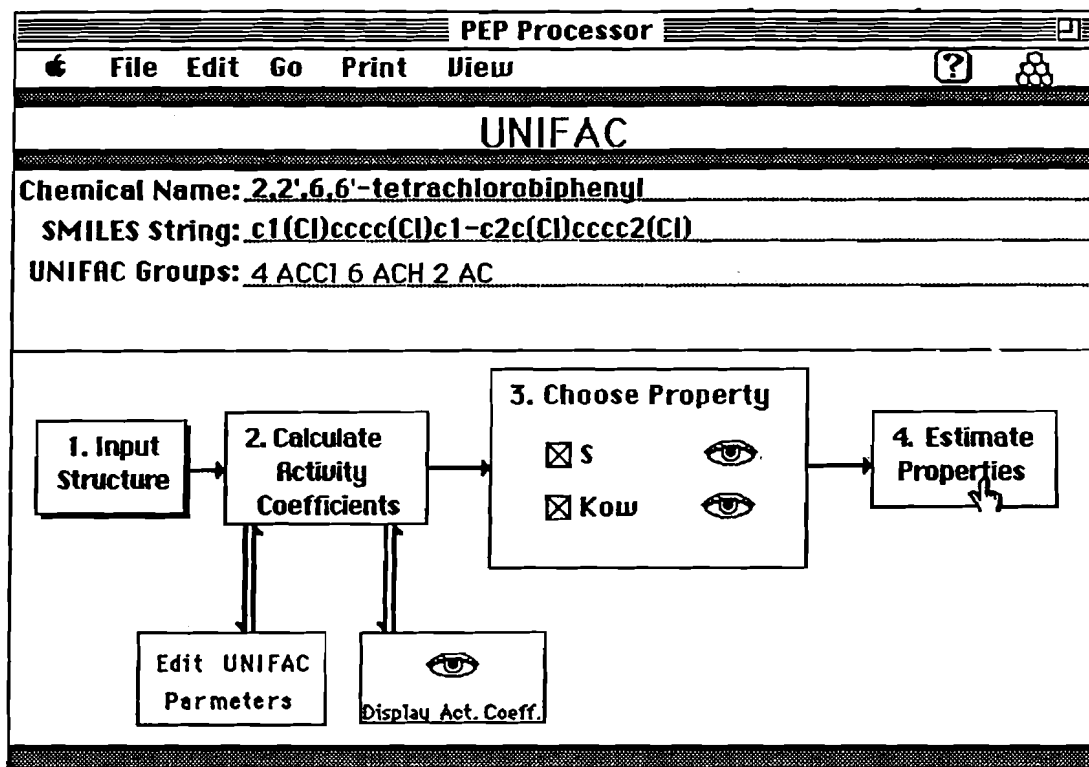


Figure 18. Example card from the PEP UNIFAC module.

### Entering Structural Information

Calculation of activity coefficients via the UNIFAC approach requires that the user input the valid UNIFAC groups that make up the chemical of interest and the number of each group present. PEP provides the user with three options for entering the appropriate UNIFAC groups using the popup menu under the button titled "Input Structure": (1) hand selecting the groups from a list in HyperCard, (2) using a connection table, and (3) using SMILES.

If the connection table option is chosen, a the standard file selection dialog box is presented where the user can select the desired connection table file. The "Struct" XFCN then uses the connection information to dissect the molecule into its UNIFAC groups.

If SMILES is chosen as the input method, the user is then prompted to input the chemical name and the SMILES string. The Struct XFCN uses the connection information contained in the SMILES string to build the UNIFAC input string.



When the user chooses to hand select the UNIFAC groups, a card is displayed showing the first 37 groups as shown in Figure 19.

### Main Groups

(1) $\text{-CH}_2\text{-}$	(4) $\text{-ACCH}_2\text{-}$	(10) $\text{O CH0}$	(15) $\text{CNH}$	
<input type="radio"/> CH3	<input type="radio"/> ACCH3	(11) $\text{-CCOO-}$	<input type="radio"/> CH3NH	
<input type="radio"/> CH2	<input type="radio"/> ACCH2	<input type="radio"/> CH3COO	<input type="radio"/> CHNH	
<input type="radio"/> CH	<input type="radio"/> ACCH	<input type="radio"/> CH2COO	<input type="radio"/> CH2NH	
<input type="radio"/> C				
(2) $\text{-C=C-}$	(5) $\text{O OH}$	(12) $\text{O HCOO}$		
<input type="radio"/> CH2=CH	(6) $\text{O CH3OH}$	(13) $\text{-CH}_2\text{O-}$	(16) $\text{-C}_3\text{N-}$	<b>More</b>
<input type="radio"/> CH=CH	(7) $\text{O H2O}$	<input type="radio"/> CH3O	<input type="radio"/> CH2N	
<input type="radio"/> CH2=C	(8) $\text{O ACOH}$	<input type="radio"/> CH2O	<input type="radio"/> CH3N	<b>DONE</b>
<input type="radio"/> CH=C		<input type="radio"/> CH-O		
<input type="radio"/> C=C	(9) $\text{-CH}_2\text{CO-}$	<input type="radio"/> FCH2O	(17) $\text{O ACNH2}$	<b>?</b>
(3) $\text{-ACH-}$	<input type="radio"/> CH3CO	(14) $\text{-CNH}_2\text{-}$		
<input checked="" type="radio"/> 6 ACH	<input type="radio"/> CH2CO	<input type="radio"/> CH3NH2		
<input type="radio"/> AC		<input type="radio"/> CH2NH2		
		<input type="radio"/> CHNH2		

Figure 19. UNIFAC module card used to select UNIFAC groups.

This card is connected to two other similar cards showing the rest of the available groups. To select a group, the user clicks on the group symbol and a dialog box will then appear to enable the user to input the number of this group that is in the molecule. The user then continues to select groups until all the groups that are in the molecule have been selected. When the user clicks the "Done" button the UNIFAC input string is built and returned to the UNIFAC card. In addition, for users familiar with the UNIFAC approach, the appropriate subgroups can also be entered directly by simply typing the number of a group followed by a space then the symbol of the group for each group in the chemical.

The final form of the input is "# group # group ...". The # represents the number of the functional group in the molecule, the group is the group symbol of the functional group. For example the UNIFAC input string for Toluene is "5 ACH 1 ACCH3", meaning five aromatic carbons with one hydrogen atom (5 ACH) and one aromatic carbon connected to a methyl group (1

ACCH3). The UNIFAC groups available to PEP are shown in Table 2. Remember UNIFAC subgroups may not be available for every compound. In these cases the activity coefficient can not be calculated using the UNIFAC.

Table 2. UNIFAC groups

Group Name	Group Symbol	Group Name	Group Symbol
CH2	CH3 CH2 CH	CCN	CH3CN CH2CN
C=C	CH2=CH CH=CH CH2=C CH=C	COOH	COOH HCOOH
ACH	ACH AC	CCl	CH2Cl CHCl CCl
ACCH2	ACCH3 ACCH2 ACCH	CCl2	CH2Cl2 CHCl2
OH	OH	CCl3	CHCl3 CCl3 CCl2
CH3OH	CH3OH	CCl4	CCl4
H2O	H2O	ACCl	ACCl
ACOH	ACOH	CNO2	CH3NO2 CH2NO2 CHNO2
CH2CO	CH3CO CH2CO	ACNO2	ACNO2
CHO	CHO	CS2	CS2
CCOO	CH3CO CH2COO	CH3SH	CH3SH CH2SH
HCOO	HCOO	Furfural	Furfural
CH2O	CH3O CH2O CH-O FCH2O	DOH	(CH2OH)2
CNH2	CH3NH2 CH2NH2 CHNH2	I	I
CNH	CH3NH CH2NH CHNH	Br	Br
(C)3N	CH3N CH2N	C=C	CH=C C=C
ACHN2	ACHN2	DMSO	DMSO
Pyridine	C5H5N C5H4N C5H3N	ACRY	ACRY
		ClCC	Cl-(C=C)
		ACF	ACF
		DMF	DMF-1 DMF-2
		CF2	CF3 CF2 CF

## **Calculate Activity Coefficients**

After the functional groups are chosen, the activity coefficients can be calculated using the procedure described by Grain (1990) by clicking on the "Calc. Activity Coefficients" button. Once the activity coefficients have been calculated they can be displayed, along with values from several intermediate steps, by clicking on the "Display Act. Coeff." button. The equations used to calculate the activity coefficients can also be displayed by clicking the "balloon" on the "UNIFAC Calculations" card.

## **Editing Parameters**

The UNIFAC group values can be edited by clicking on the "Edit Parameters" button. This will take you to an Index card containing a button for each group. To edit the value for that group, click on the corresponding group button. To edit the Q and R values, click on the left arrow at the bottom of the index card.

## **Estimating Properties**

To estimate S and/or  $K_{OW}$  click on the "Estimate Properties" button. Clicking the "eye" next to the property will display the equations used to estimate S or  $K_{OW}$ . The results will be reported on the results card.

## **Limitations of UNIFAC Approach to Estimating S and $K_{OW}$**

The use of UNIFAC is limited to chemicals that have subgroups contained in a UNIFAC data base. If a chemical has subgroups for which no UNIFAC values are available, the activity coefficients cannot be calculated and, therefore, the properties cannot be estimated. In addition, it has been observed that the errors associated with estimating S and  $K_{OW}$  via the UNIFAC approach tend to increase as the compound becomes less solubility (larger  $K_{OW}$ ). Correction factors have been presented in the literature to correct for this tendency.

# Property/Property Module

## Overview

Quantitative Property-Property Relationships (QPPRs), based on the relationship between two properties as determined by regression analysis, are used to predict the property of interest from another more easily obtained property without a specific concern for molecular structure. Frequently, the regression expressions are expressed in terms of the log of the two properties.

The operation of the Property-Property correlation module, illustrated in Figure 20, is considerably different than the other modules because the choice of the regression models used for the property estimation determines the required inputs.

The screenshot shows a software window titled "PEP Processor" with a menu bar (File, Edit, Go, Print, View) and a title bar. The main window is titled "Property/Property" and contains the following elements:

- Chemical Name:** 2,2',6,6'-tetrachlorobiphenyl
- SMILES String:** \_\_\_\_\_
- 1. Choose Property:** A list of checkboxes for properties: S (checked), Kow (unchecked), Koc (checked), BCF (checked), and H (unchecked).
- 2. Regression Model Selection:** Two radio buttons: "USER choose regression" (selected) and "PEP choose regression". Below are four regression models, each with a "View" column containing "PEP" and "REF" icons:
  - Universal from Kow
  - Universal from S & K<sub>ow</sub>
  - Universal from Kow
  - Universal from Kow
- 3. Input Properties:** A section titled "Look in Prop. DB" with input fields for: log S, log Kow (value: 5.936), log P<sub>v</sub>, log H, log K<sub>oc</sub>, log BCF, M.P., and M.W.
- 4. Estimate Properties:** A button with a mouse cursor icon.
- Note:** All values are at 25°C

Figure 20. Example card from PEP property-property module.

## **Selecting the Properties to be Estimated**

Since the input for the Property/Property module depends on the properties to be estimated and the regression models used, the user must first select the properties to be estimated and the regression equations to be used before any input values are requested. The program keeps track of the properties required and provides the appropriate input fields. If available, the required properties can be imported directly from the associated chemical property database. Currently property-property relationships for the estimation of S, Kow, Koc, and BCF are available and will darken as the properties are selected.

## **Choosing the Property-Property Regression Model**

The most appropriate model to use in estimating the property of interest depends on the class of the chemical and what property-property relationships are available. The property-property relationships that are available in PEP are divided into two general categories; regressions developed using data from the PEP Chemical Physical Data Base and regressions obtained from the literature. The two categories are separated by a grey line in the popup menu. To choose the property-property model, use the popup menu directly under the current regression by holding the mouse button down while over the title of the current regression. The "eye" icon to the right of the regression title will display the regression equation and statistics associated with the current relationship. The "book" icon will show the reference for the current regression if the regression is from the literature.

## **Viewing the Calculated Values**

Once the properties and the relationships are chosen, the properties required for input need to be entered. The properties that are required are shown on the right side of the card with a line next to the symbol. The values can be retrieved from the PEP Chemical Property Database by making sure the property name is correct and clicking on the "Look in Prop. DB" button. The current units of the inputs will be shown if appropriate, and will be changed as needed by PEP for the different regression equations.

When the "Estimate Properties" button is clicked the properties that were input will be used along with the chosen regression equations to estimate the selected properties. If the values for the input properties have not been entered on the Property-Property card the user will be prompted to supply the required values. The property estimates along with the calculated 95% prediction interval, if the statistical information is available, will be placed on the Results card.

## **Limitations of the Property/Property Module**

One major limitation to the use of property/property relationships is the lack of suitable property values required for input into the regression models. For many new chemicals, no physical/chemical property information is available. In addition, the quality of the property-property regression model depends on the extent of the database used to develop the relationship. For example, some property-property models are broader than others in the type and number of compounds used in its development.

# PEP BATCH

## Overview

PEP Batch provides users with a method for the convenient, unattended calculation of MCIs, TSA and UNIFAC activity coefficients and subsequent estimation of physical properties for large numbers of compounds via the PEP processor described earlier. Like the PEP Processor, PEP Batch is divided into MCI, TSA and UNIFAC modules. Each module, as illustrated in Figure 21 for the MCI module, requires the user to select the appropriate input file (i.e. SMILES string, connection table or 3-D atomic coordinates), choose to information to be sent to the output file (i.e. chemical name, SMILES string, properties) and start the batch driver.

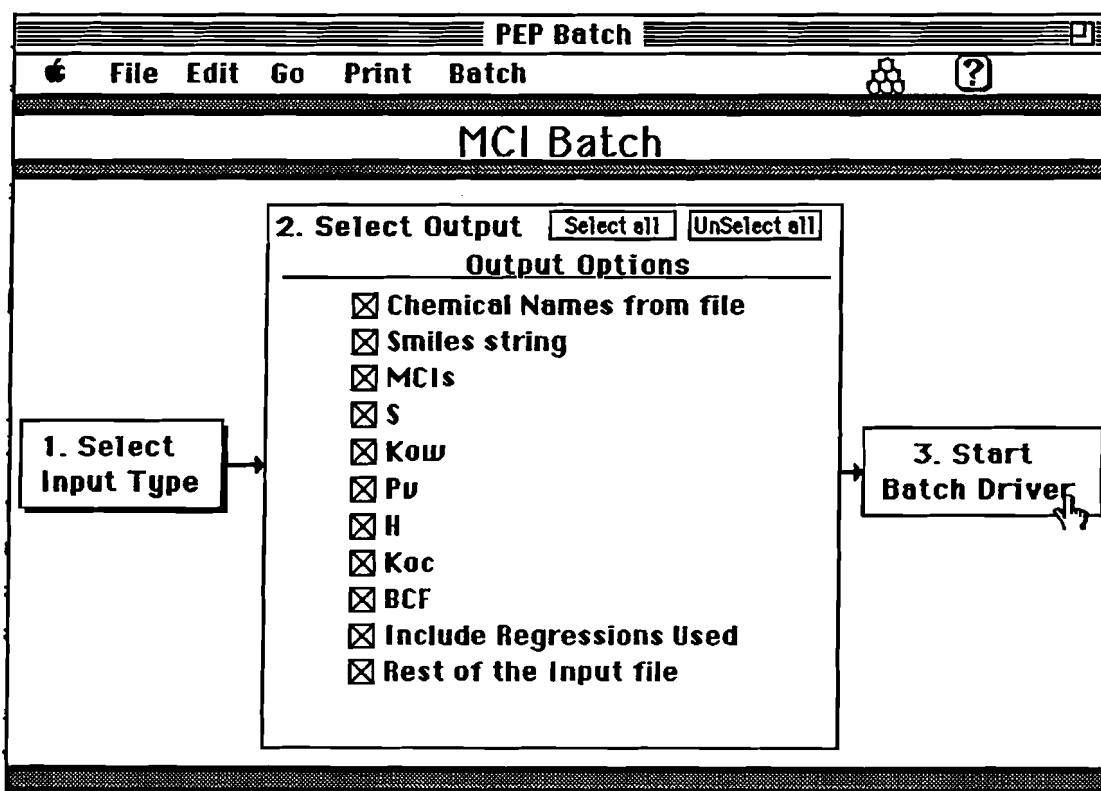


Figure 21. Example card for PEP Batch, MCI module.

## Input Structure

The MCI and UNIFAC modules require the two-dimensional molecular structure of the chemical to be entered using either SMILES strings or connection tables.

To enter SMILES strings into the MCI or UNIFAC batch modules, you must create a text file containing name of the chemical and its corresponding SMILES strings in columns separated by tabs and hard returns at the end of each row. The text file can contain additional tab-delimited information, but the SMILES strings and chemical names must be in the first or second column. When you select the SMILES input option, you must indicate the column order on the "Explain SMILES file card shown in Figure 22. The text file can be created with word processing or spreadsheet programs or you can also edit or create a file containing the SMILES strings and chemical names within PEP by selecting SMILES from the "input structure type" pulldown menu.

PEP Batch

File Edit Go Print Batch

### Explain SMILES file

**View File Contents**

```
benzene c1ccccc1
toluene c1cc(C)ccc1
o-xylene c1c(C)c(C)ccc1
ethylbenzene c1c(CC)ccc1
```

Column 1	Column 2
<input checked="" type="radio"/> Chemical Name	<input type="radio"/> Chemical Name
<input type="radio"/> SMILES	<input checked="" type="radio"/> SMILES

Ignore First Line

Figure 22. PEP Batch, explain SMILES file card.



To input connection table files into the MCI or UNIFAC batch modules, you must first place them in a single folder. Select the "Connection tables" option from the "Select Input Type" popup button and choose the folder that contains the connection table files using the standard "open file" dialog box that appears. Highlighting any one of the files in the folder selects all of the files in that folder and allows you to view or delete specific files. The files can be any valid type for the MCI module as they will be converted if possible (see MCI module helps). After selecting the input files, click on the advance arrow located at the upper right of the card to advance to the next step.

The TSA batch module operates in the same manner as the MCI and UNIFAC modules except that the calculation of TSA requires the three-dimensional cartesian coordinates for each atom in the chemical of interest. The TSA batch module accepts Cartesian Coordinates or Alchemy files. Alchemy files contain both the two-dimensional chemical structure and the coordinates. This allows PEP to calculate the chemical's TSA and determine the most appropriate TSA-property relationship based on chemical class. From within the standard dialog box, you can click on any file in a folder to select all of the files in that folder. The files will then be displayed. You can also view or delete files.

### **Output Options**

Once the input files have been selected, the "output option" step becomes active. This allows the user to select the properties to be calculated and any additional information that is available (i.e. chemical name, SMILES string, MCIs, TSA, or UNIFAC activity coefficients) to be exported by to a tab-delimited text file.

### **Start Batch Driver**

After the output information is selected, the "Start Batch Driver" button becomes active. Clicking this button brings up the standard Macintosh "save file" dialog box that allows the user to specify the name of the data file to be exported by PEP batch and the location that the file will be saved.

---

## **CHEMICAL PROPERTY DATABASE**

### **Overview**

Experimentally determined physical property data for about 800 compounds, having at least one value of aqueous solubility (S), octanol/water partition coefficient (Kow), vapor pressure (Pv), organic carbon normalized soil sorption coefficient (Koc), bioconcentration factor (BCF), or Henry's law constant (H), was compiled from a variety of literature sources and computerized databases. Using this information, a chemical property database was constructed using HyperCard™ and subsequently used for developing MCI-property, TSA-property and property-property relationships. In addition to the properties listed above, the database includes the following information: compound name and synonyms, a diagram of the 2-D chemical structure, SMILES notation, uses, CAS number, chemical formula, molecular weight (MW), boiling point (BP), melting point (MP), and appropriate references for each value. A built-in unit conversion utility enables users to quickly view property values in a variety of commonly used units. The database is directly connected to the PEP Processor stack.

The Chemical Property Database also provides the means for the user to search for chemical compounds by full or partial name or synonym, to sort the compounds by name, boiling point, melting point, or molecular weight, and the ability to transfer to any of the property estimation modules. In addition, the user can easily edit exiting values, add new values or export information to a text file or another database.

In addition to the physical properties, information describing the environmental persistence and toxicity of specific chemicals can also be entered into the database. Placeholders for biodegradation rates, hydrolysis rates, photolysis rates and LC50s, along with the appropriate references and comments have been incorporated into the database. This feature was added to the database after requests from test users, however currently no degradation or toxicity data has been entered into the database. The chemical property database is illustrated in Figures 23 and 24.

Chemical Property Data Base

File Edit Go Find Sort Values Print Export

2,2',6,6'-tetrachlorobiphenyl

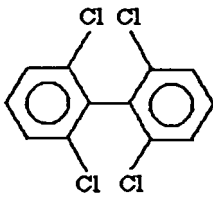
View: Structure

CAS #: 15968-05-5

Formula: \_\_\_\_\_ MW: 292

MP: 200 °C BP: \_\_\_\_\_ °C

Synonyms:



View: Chemical\Physical Properties

log Antilog	Physical/Chemical Properties	Values	Temp °C	Units	Ref
	log Aqueous Solubility (S)				
	log Octanol/Water Partition Coefficients (Kow)	5.936	25		
	log Vapor Pressure (Pv)				
	log Henry's Law Constant (H)	-1.648	25	dimensionless	
	log Soil/Water Sorption Coefficient (oc-based) (Koc)				
	log Bioconcentration Factor (BCF)				
	log Acid Dissociation (Ka)				

Figure 23 Example card from the Chemical Property Data Base

chemical property data base

File Edit Go Find Sort Values Print Export

2,2',6,6'-tetrachlorobiphenyl

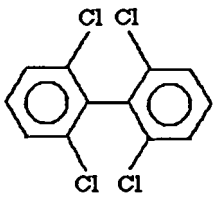
View: Structure

CAS #: 15968-05-5

Formula: \_\_\_\_\_ MW: 292

MP: 200 °C BP: \_\_\_\_\_ °C

Synonyms:



View: Chemical\Physical Properties

Degradation Properties

Degradation	t 1/2	Matrix	Temp (°C)	Comments
Hydrolysis				
Photolysis				
Biodegradation				

Figure 24. PEP database degradation properties.

## Searching for Chemical Compounds in PEP's Database

Chemicals in the database can be found by full chemical name, partial chemical name, synonyms, or CAS number searches, or by selecting the chemical from a current list of compounds. The user selects the desired option from the "Find" menu.

Searching or Finding by either chemical name, partial chemical name, synonyms or CAS number is done using the finding tools built into HyperCard. When the finding option is chosen a standard dialog box is used to input the string that is to be found. Finding by full chemical name or CAS number requires that the entire contents of the appropriate field match the string of characters that was entered. Finding by partial chemical name or synonym requires that all of the string that was entered be present in the appropriate field.

The select from list option allows the user to select a compound from the current list of chemical names. The chemical names appear alphabetically sorted in a dialog box. Typing the first letter or number of the chemical will scroll the list to the first chemical starting with that character. The chemical card is located and displayed by clicking twice on the same chemical name or selecting the chemical name and then clicking the "Go There" button.

## Sorting the Database

The "Sort" menu at the top of the card allows you to sort the database by name, melting point (MP), boiling point (BP), or molecular weight (MW). NOTE: the alphabetic sort function built into HyperCard places all chemicals with numbers preceding the letters in the chemical name before those chemicals having only letters in the chemical name. For example, after sorting by name, 1-octanol would come before acetone and 2-octanol, but after 1-decanol.

## Adding or Deleting Data

New chemicals may be added to the data base by choosing "Add new card" option from the Values menu. The user will then be prompted for the chemical name, CAS number, molecular formula, melting point, boiling point, molecular weight, and the SMILES string.

Additional values for melting point, boiling point, S,  $K_{OW}$ ,  $P_v$ , H,  $K_{OC}$ , BCF, and  $K_a$  can be added to a chemical already in the database using the "Add value" option accessed from the Value menu. As each new value is entered the user is prompted for the reference. The reference can be an existing reference or the user can add a new reference. If a new reference is to be added, the user will be prompted for the author, year, title, publication, volume, issue, and the page numbers. The user will also be given a chance to edit the new reference.

An entire card or specific values contained on a card can be deleted by choosing the appropriate item from the Value menu. A reference can also be deleted if it was entered incorrectly. The data will not be deleted until the user verifies the request through a standard dialog box.

### **Printing Information From the Database**

The user can print the current card, or a series of cards using the print features built into HyperCard. To print a series of cards, first the user must select that option from the "Print" menu. Then while each card that is to be printed is the currently displayed card, the select "Print this card" option from the print menu must be chosen. When the desired cards have been printed, the user then selects "Stop printing" from the "Print" menu. The user can also print a list of all the references contained in the PEP data base.

### **Exporting Information From the Database**

Information in the database can either be exported as text in a line or as a spreadsheet format using the "Export" menu. If the line format option is used, the data will appear with each data field on a line and each chemical separated by a solid line. This option is most appropriate if a word processor will be used to print or manipulate the exported data. If the spreadsheet format is used the data will be put in columns delimited by a tab.

### **Moving to Other PEP Modules**

The "Go" menu at the top of the card allows you to move to any PEP module. To use the Go option, move the pointer to Go menu at the top of the card. Hold the pointer down on the button until the selections appear. Then move the pointer until the selection of choice is highlighted and release the mouse button.

### **Changing the Units of Measurement**

The popup menu under each property's units allows you to change the units for solubility, Henry's law, and vapor pressure. All values of the property will be changed to reflect the new units.

## PEP Models

### Overview

To illustrate the practical application of PEP, an additional stack called PEP Models was developed. This stack, which contains the algorithms for the Level 1 and 2 Fugacity Models (Mackay, 1980), is linked directly to the the PEP Processor, but can also be used independently. The Level 1 Fugacity Model considers a unit world consisting of six compartments: air, water, soil, suspended solids, sediment, and biota (as shown in Figure 25).

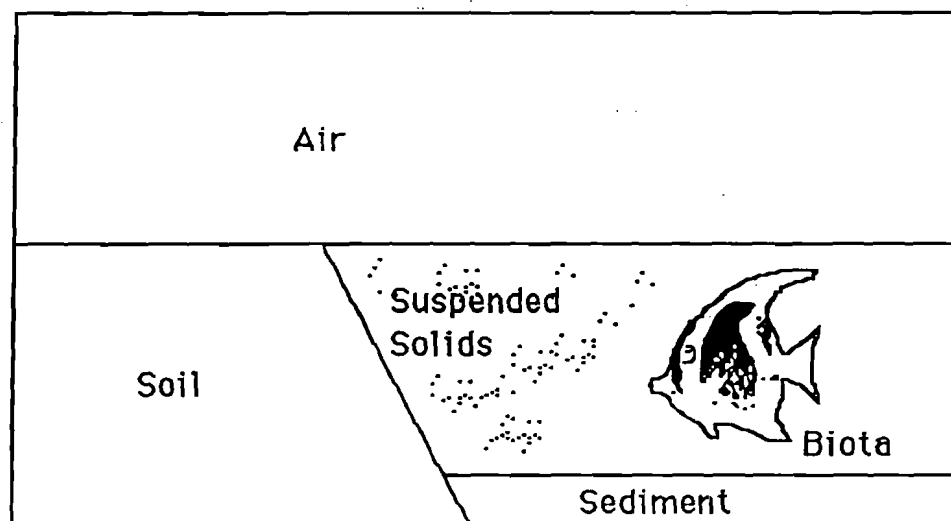


Figure 25. Representation of Fugacity Level 1 compartments.

The model predicts the equilibrium concentrations of the chemical of interest in each compartment using the fugacity approach described by Mackay (1980). The model requires the input of  $K_{OC}$ ,  $H$  and  $BCF$  which can be read directly from the PEP processor or the PEP chemical property database, if available. In addition to the chemical specific properties, the density and volume of each compartment must be specified along with the organic carbon content of the soil, sediment and suspended sediment. An editable set of default values for compartment density, volume and organic carbon content, as shown in Figure 26.

**PEP Models**

File Edit Go Print Compartments

## Fugacity Model

**Chemical Name: 2,2',6,6'-tetrachlorobiphenyl**

**1. Input Property Values**

Look for Values in Prop.DB

log Koc   
3.92

log H   
(dimensionless)  
-1.648

log BCF   
4.69

Indicates value from DB

**2. Input Environmental Compartment Values**

Compartment	Density kg/m <sup>3</sup>	Volume m <sup>3</sup>	% Organic Carbon
<input checked="" type="checkbox"/> Air	1.19	1e10	
<input checked="" type="checkbox"/> Water	1000	7e6	
<input checked="" type="checkbox"/> Soil	1500	9e3	2
<input checked="" type="checkbox"/> Susp. Solids	1500	35	4
<input checked="" type="checkbox"/> Sediment	1500	2.1e4	4
<input checked="" type="checkbox"/> Biota	1000	3.5	

**3. Calculate Distribution**

OR

**Input Fugacity Level 2 Data**

Figure 26. PEP Models card, input for Fugacity Level 1

The Level 2 Fugacity Model allows for the chemical of interest to degrade in each compartment, move by advection through the water and air phases, and be emitted into the unit world. The rate values for each of these processes must be entered by the user. The degradation rates for each compartment can be entered either by  $t_{1/2}$  values in hours or by first order reaction rate constants in 1/hours, as shown in Figure 27. The advection rate data can be entered either by residence time or flow rate and the concentration or by directly entering the mass flow rate in moles per hour. The emission rate is entered in the units of moles per hour.

**PEP Models**

**File Edit Go Print Rates**

## Fugacity Level 2

**Chemical Name: 2,2',6,6'-tetrachlorobiphenyl**

**3. Input Reaction Rate Data**  
(input reaction 1/2 life or reaction rate constant)

	Reaction 1/2 life hours	OR	Reaction rate const. hours <sup>-1</sup>
<input checked="" type="checkbox"/> Air	_____		_____
<input checked="" type="checkbox"/> Water	_____		_____
<input checked="" type="checkbox"/> Soil	_____		_____
<input checked="" type="checkbox"/> Susp. Solids	_____		_____
<input checked="" type="checkbox"/> Sediment	_____		_____
<input checked="" type="checkbox"/> Biota	_____		_____

**4. Input Advective Data**  
(input mass flow rate or conc. and residence time or flow rate)

	Residence time hours	OR	Flow Rate m <sup>3</sup> /hr	Conc. mol/m <sup>3</sup>	Mass flow rate mol/hr
	_____		_____	_____	0.0E+00
	_____		_____	_____	0.0E+00

**5. Input Emission Rate**  
Emission rate mol/hr \_\_\_\_\_

OR

Figure 27. PEP Models card, input for Fugacity Level 2

### Input Property Values

To calculate the environmental distribution of a chemical the user must enter the several physical-chemical properties along with the size and densities of each compartment. The physical-chemical properties that are need to be input for the six compartment unit world are Koc, H, and BCF. However, in step 2 the user may choose to eliminate one or more compartments to simplify the model or to more closely examine the distribution of a chemical in a sub-environment such as the hypolimnion of a lake (i.e., water, air, biota). This would eliminate the need for a value for Koc. PEP determines which physical-chemical property values are needed depending on the unit world which the user has defined. The chemical-physical properties that are not needed will be dimmed.

When available, experimental values are preferred over estimated values in environmental fate modeling. Experimental values for the physical-chemical properties can be searched for in the Chemical Property Data Base by using the "Look for Values in Prop.DB" option. When clicked, a



full chemical name search is initiated, and if the chemical is found, the values of the three properties required by the Fugacity level 1 model will be displayed in a dialog box. The values can be copied to the Fugacity card using the buttons on the dialog box as illustrated in Figure 28. The check boxes next to the property names of the Fugacity card indicate if the property came from the data base.

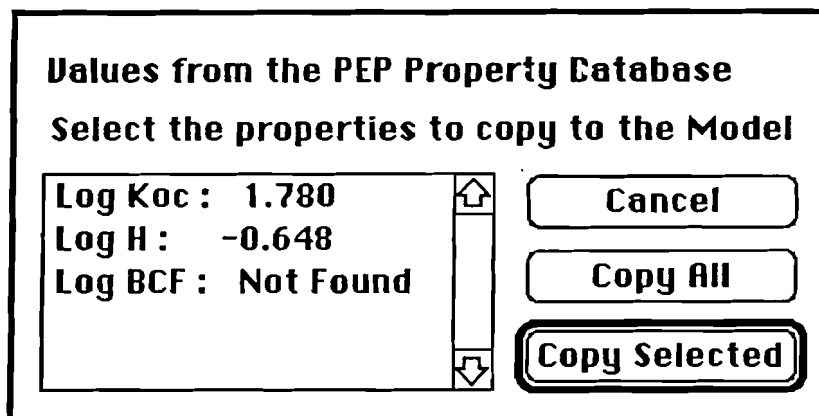


Figure 28. Example dialog box resulting from the "Look for Values in Prop.DB" option

### Input Environmental Values

The values used for the unit world can be changed in step 2. The available compartments are air, water, soil, suspended solids, sediment, and biota. Water must always be part of the unit world because of the partition coefficients are defined as the ratio of the chemical in the non-aqueous phase to the aqueous phase. The rest of the compartments can be eliminated by clicking the check box next the compartment name. The densities, volumes, and the percent organic carbon (if appropriate) of each compartment can be changed.

Different user defined environments or unit worlds can also be saved using the "Save Current Values" button and later recalled using the "Values:" popup menu. The "Delete Saved Values" button can be used to remove a saved set of values from the list that will popup when the "Values" button is used. The "Set Default Values" button can be used to set the unit world that is used when the PEP Models stack is opened.

### Calculate Distribution

After the user inputs all the required information and hits the "calculate distribution" button, the model calculations are performed in HyperTalk and the results are presented in both in tabular and

graphical form as illustrated in Figure 29. The graphical display can be changed from bar (concentration of the chemical in each phase) to pie (percent of the chemical in each phase) chart forms using the "Graph:" popup menu. The values of the distribution coefficients that were used in the calculations are also shown on the results card.

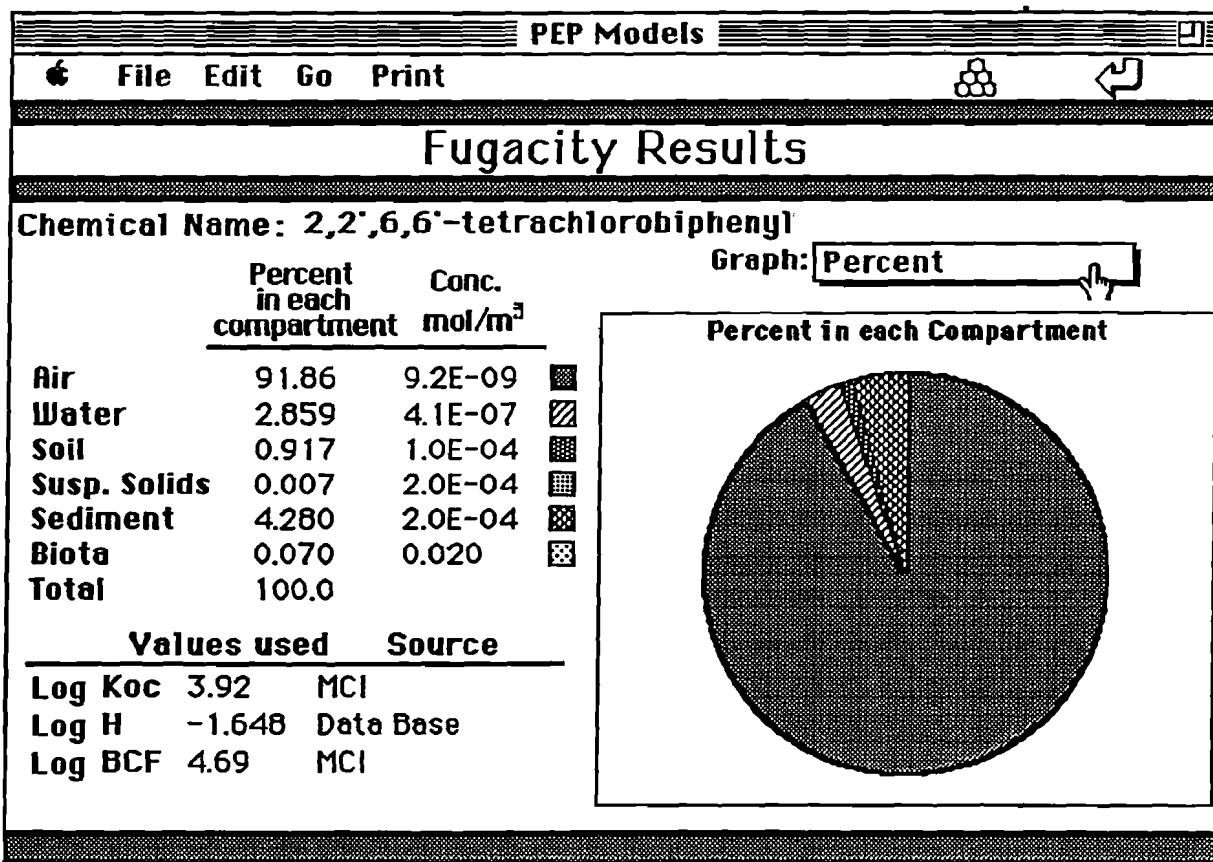


Figure 29. PEP Models results card

## PEP Help

Information regarding the operation of the chemical property database and the property estimation, models and batch modules is available in the PEP Help stack. This stack easily accessed at any time within the PEP system. The organization and layout of each help card is similar to that illustrated in Figure 30 for the MCI module. The user can select the topic of interest by clicking on the appropriate radio button and the information on that subject will be displayed in the scrolling field.

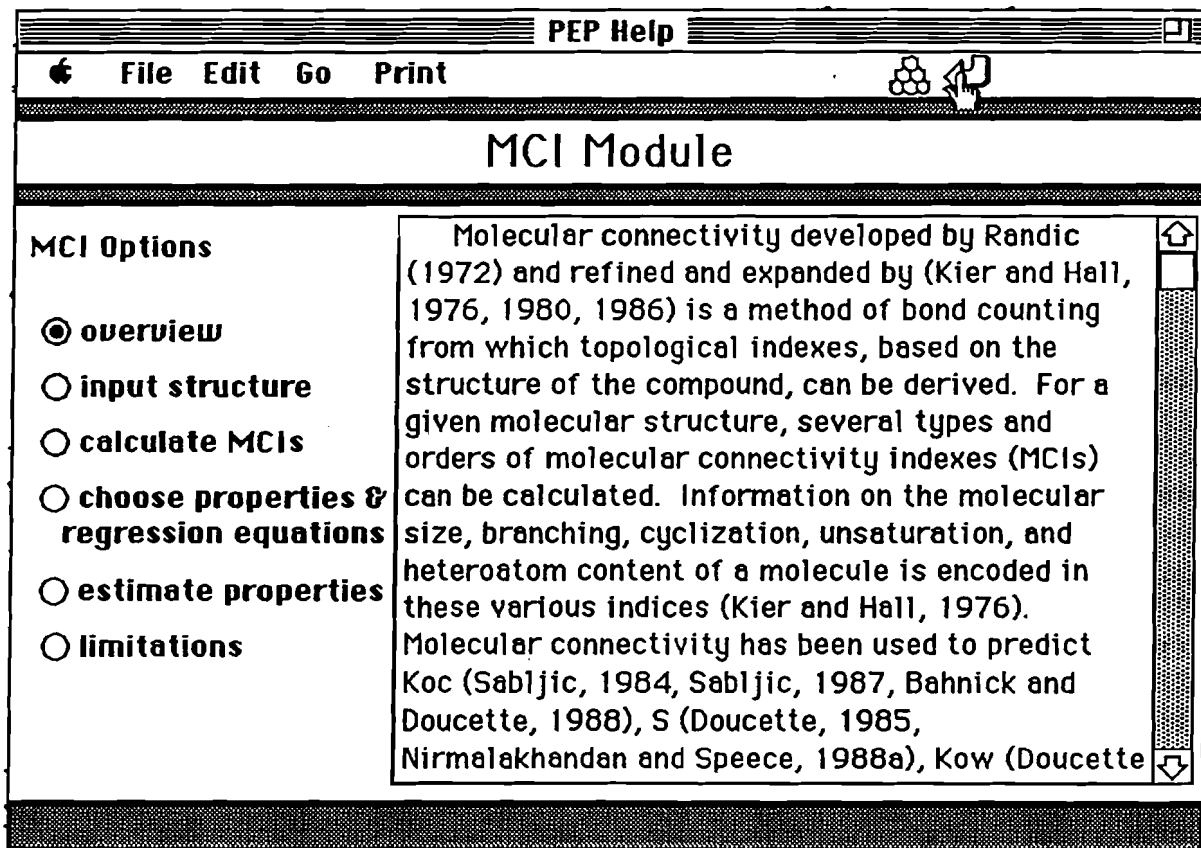


Figure 30. Example card from PEP Help stack for MCI module.

---

## REFERENCES

- Anderson, E., G.D. Veith, and D. Weininger. 1987. SMILES: A Line Notation and Computerized Interpreter for Chemical Structures. EPA/600/M-87/021.
- Arrouckle, W.B. 1986. Using UNIFAC to Calculate Aqueous Solubilities. *Environ. Sci. Technol.*, 20: 1060-1064.
- Bahnick, D.A. and W.J. Doucette. 1988. Use of Molecular Connectivity Indices to Estimate Soil Sorption Coefficients for Organic Chemicals. *Chemosphere*. 17(9): 1703-1715
- Banerjee, S. 1985. Calculation of Water Solubility of Organic Compounds with UNIFAC-Derived Parameters. *Environ. Sci. Technol.* 19(4): 369-370.
- Banerjee, S. and P.H. Howard. 1988. Improved Estimation of solubility and Partitioning Through Correction of UNIFAC-Derived Activity Coefficients. *Environ. Sci. Technol.* 22(7): 839-841.
- Cambridge Scientific. 1989. ChemDraw and Chem3D+. Cambridge, Massachusetts [Computer Programs]
- Gmehling, J., P. Rasmussen, and A. Fredenslund. 1982. Vapor-liquid equilibria by group contribution. Revision and Extension. 2. *Ind. Eng. Chem. Process Des. Dev.* . 21: 118-127.
- Grain, C. F. 1990. Activity Coefficient, Ch. 11. In Lyman, W.J. W.F. Reehl and D.H. Rosenblatt (Eds.). *Handbook of chemical and property estimation methods*. American Chemical Society. Washington, D. C .
- Kier, L.B., Hall, and L. H., *Molecular Connectivity in Structure-Activity Analysis*. 1986, Letchworth, Hertfordshire, England.: Research Studies Press, LTD., England, and John Wiley & Sons, Inc., New York. 262.
- Mackay, D. 1979. Finding fugacity feasible. *Environ. Sci. Technol.* 13(12): 1218-1223.
- Mackay, D. and S. Paterson. 1981. Calculating Fugacity. *Environ. Sci. Technol.* 15(9): 1006-1014.

Pauling, L. 1960. The nature of the chemical bond. Cornell University Press, Ithaca, New York  
262 p.

Pearlman, R.S. 1980. Molecular surface areas and volumes and their use in structure/activity relationships, p. 331-346. In S.H. Yalkowsky, A.A. Sikula and S.C. Valvani (Eds.). Physical chemical properties of drugs. Marcel Dekker, Inc., New York.

Tripos Associates Inc. 1989. Alchemy and CONCORD. St. Louis, Missouri [Computer Programs]

Weininger, David. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28:31-36.