

**OPERATION MANUAL**  
for the  
**ECOLOGICAL Structure-Activity Relationship Model**  
**(ECOSAR)**  
**Class Program**

**ESTIMATING TOXICITY OF INDUSTRIAL CHEMICALS  
TO AQUATIC ORGANISMS USING THE  
ECOSAR (ECOLOGICAL STRUCTURE ACTIVITY RELATIONSHIP) CLASS  
PROGRAM.**

**MS-Windows Version 1.11**

**Contributors:**

Kelly Mayo-Bean<sup>a</sup>, Kendra Moran-Bruce<sup>a</sup>  
J. Vince Nabholz<sup>a\*</sup>, William M. Meylan<sup>b</sup>, Philip H. Howard<sup>b</sup>,

<sup>a</sup>Risk Assessment Division (7403)  
U.S. Environmental Protection Agency  
1200 Pennsylvania Ave., N.W.  
Washington, DC 20460  
\* Deceased

<sup>b</sup>Chemical Biological Environmental Center  
SRC, Inc.  
7502 Round Pond Road  
North Syracuse, NY 13212-2558

**March 2012**

## DISCLAIMER

**This document has been reviewed and approved for publication by the Risk Assessment Division of the Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency (EPA/OPPT). Approval does not signify that the contents necessarily reflect the views and policies of all Offices/Divisions in the Environmental Protection Agency, nor does the mention of trade names or commercial products constitute endorsement or recommendation for use.**

**The ECOSAR model and underlying methodology presented in this document have been developed over a period of more than 25 years by EPA/OPPT, EPA contractors, and/or others in the scientific and technical community to screen chemicals in the absence of data. EPA/OPPT has made this screening level model, along with many other tools, available to industry and other stakeholders in the hopes that use of the models in the early stages of research and development or prior to submission of notifications to the Agency, will result in safer chemicals entering commerce.**

**Other chemical screening methodologies have been developed and are in use by other Agencies, chemical companies and other stakeholders. The U.S. EPA recognizes that other models are available and that these models can also be of value in chemical screening efforts. Screening models provide estimations with an inherent degree of uncertainty and therefore, valid measured data are always preferred over estimated data. If no measured or analog data are available, screening level models such as the ECOSAR Class Program may be used to predict toxicity values that can be used to indicate which chemicals may need further testing or characterization.**

## Table Contents

1.	Introduction.....	4
2.	Computer-Software Requirements.....	7
3.	Installing the ECOSAR Class Program.....	7
4.	Starting the ECOSAR Class Program.....	8
5.	User Inputs For Generation Of ECOSAR Predictions .....	9
5.1.	Structure Entry.....	11
5.1.1.	Direct Entry.....	12
5.1.2.	CAS Number Database .....	13
5.1.3.	Name Look-Up Database .....	13
5.1.4.	Entry Using The Drawing Tool.....	15
5.1.5.	Entry From A User File.....	16
5.1.6.	Importing Structures.....	18
5.2.	Descriptor Fields .....	19
5.3.	Entering Measured Data For Variables .....	19
5.3.1.	Log Kow .....	19
5.3.2.	Water Solubility .....	20
5.3.3.	Melting Point.....	20
5.3.4.	Molecular Weight Adjustment.....	20
6.	Results Window .....	222
6.1.	Structure Window.....	25
6.2.	Chemical Attributes.....	25
6.3.	Measured Effect Levels.....	27
6.3.1.	Adjusting Effect Level for Presence of Acid (x10).....	30
6.4.	Presentation of Predicted Toxicity.....	28
6.5.	QSAR-Specific Information .....	30
6.6.	Comprehensive approach for determining the most representative QSAR class.....	31
6.6.1.	General Classes vs. Sub-Classes .....	34
6.6.2.	Statistical Determination of Correlation of QSAR Training Data Sets .....	35
6.6.3.	Robustness and Distribution of Supporting Datasets for QSAR Classes .....	37
6.6.4.	Selecting the Most Representative Class.....	38
7.	Batch Runs.....	40
7.1.	Batch Input Files.....	41
7.1.1.	SMILES Strings .....	41
7.1.2.	CAS Number List.....	42
7.1.3.	MDL SD Files .....	42
7.2.	Conducting Batch Runs .....	43
7.2.1.	Obtaining Full Outputs and Summary Outputs .....	45
7.2.2.	Obtaining Table Data Outputs.....	46
8.	Special Class Calculations .....	49
9.	Bibliography .....	52
	Appendix A-Glossary of Terms and Abbreviations Associated with ECOSAR .....	53
	Appendix B-Summary of Function Keys .....	55
	Appendix C- Files Associated With ECOSAR.....	58

## 1. Introduction

The ECOSAR Class Program is a computerized version of the ecotoxicity analysis procedures as currently practiced by the Office of Pollution Prevention and Toxics (OPPT) when data are lacking for risk assessment development. Quantitative structure-activity relationships (QSARs) have been used by the U.S. Environmental Protection Agency since 1981 to predict the aquatic toxicity of new industrial chemicals in the absence of test data. It has been developed within the regulatory constraints of the Toxic Substances Control Act (TSCA) and is a pragmatic approach to QSAR as opposed to a theoretical approach.

The QSARs presented in this program were developed by classification/sub-classification of chemicals based on similarity of structure and similarity in measured effect levels from aquatic toxicity data. When available, modes of action (MOA) have been integrated into the classification scheme to substantiate trends seen in available data. Although acute and chronic toxicity of chemicals to fish (both fresh and saltwater), aquatic invertebrates (freshwater daphnids and saltwater mysids), and algae (freshwater green algae and saltwater algae) has been the focus of QSAR development, additional QSARs for certain chemical classes were developed for other organisms (e.g., earthworms, sea urchins, and *Lemna gibba*). QSAR development for chemical classes is based on measured test data submitted by industry under TSCA or collected from publicly available sources. To date, 709 QSARs have been developed for 111 organic chemical classes, along with additional QSARs for surfactants and dyes that can be obtained through the “special classes” menu. The supporting data sets (training sets) used to derive QSARs within a chemical class range from the very large, e.g., neutral organics, to the very small, e.g., aromatic diazoniums and have been made public, when permitted, in the ECOSAR Equation documents located in the ‘Help’ menu under ‘Online ECOSAR Help’. For more information on QSAR development, see the ECOSAR Methodology Document within the ECOSAR Model ‘Help’ menu.

The ECOSAR program is designed for the expert user. Specifically, users are expected to have some knowledge of environmental toxicology and organic chemistry, and are expected to use this knowledge to determine appropriate classifications when multiple classes are identified for a given chemical. Users are also expected to determine whether use of ECOSAR to predict aquatic toxicity or use of an analogue approach is more suitable. ECOSAR is menu-driven and contains various help functions to assist the user. Users cannot change any of the equations or data stored within the program or accidentally erase any important information. The following pages show how to install, access, and use the ECOSAR Class Program. If users have any questions or comments on the ECOSAR program, or find any errors, please contact:

### **U.S. EPA Technical Contacts**

Kelly E. Mayo-Bean, U.S. EPA  
Risk Assessment Division (7403M)  
1200 Pennsylvania Ave., N.W.  
Washington, DC 20460-0001  
Phone: 202-564-7662  
Fax: 202-564-9063  
email: [mayo.kelly@epa.gov](mailto:mayo.kelly@epa.gov)

Kendra Moran, U.S. EPA  
Risk Assessment Division (7403)  
U.S. Environmental Protection Agency  
1200 Pennsylvania Avenue, NW  
Washington, DC 20460-0001  
Phone: 202-564-4322  
Fax: 202-564-9063  
E-mail: [moran.kendra@epa.gov](mailto:moran.kendra@epa.gov)

**Model Development Contact**

Bill Meylan

Chemical Biological Environmental Center

SRC, Inc.

7502 Round Pond Road

North Syracuse, NY 13212-2558

phone: 315-452-8421

fax: 315-452-8440

e-mail: [meylan@srcinc.com](mailto:meylan@srcinc.com)

## 2. Computer-Software Requirements

The ECOSAR Class Program is designed for use on the IBM and IBM-compatible series of personal computers running Microsoft Windows 95 (and higher including Windows 98, 2000, ME, NT, Vista, and 7). ECOSAR is not currently designed to run as a multi-tasking program (e.g. running ECOSAR batch-mode runs in the background while running another program in the foreground) and, thus, batch-modes should be run in the foreground until they are completed. Although a mouse/touch-pad or other pointing device<sup>1</sup> is not required, it is highly recommended. If a mouse/touch-pad is not available, knowledge of the following keys is necessary: **Esc**: During data entry, pressing the Esc key exits the program or results window; **Enter**: Pressing the Enter (Return) key sends the cursor to the next data entry field; and **Tab** or **Shift-Tab**: changes entry fields (Appendix B-Summary of Function Keys). ECOSAR v1.1 requires approximately 60 MB of hard disk space; this includes the SMILECAS Database (a database of over 108,000 SMILES notations indexed by CAS number for program retrieval) which requires 11 MB of disk space.

## 3. Installing the ECOSAR Class Program

Users can download for free the ECOSAR Class Program from the U.S. Environmental Protection Agency's website at: <http://www.epa.gov/oppt/newchems/tools/21ecosar.htm>. ECOSAR is a self-extracting file. Once it is copied to a diskette or hard drive, execute (double-click) the file to install the program. A comprehensive list of files installed with the ECOSAR program is provided in Appendix C- Files Associated With ECOSAR.

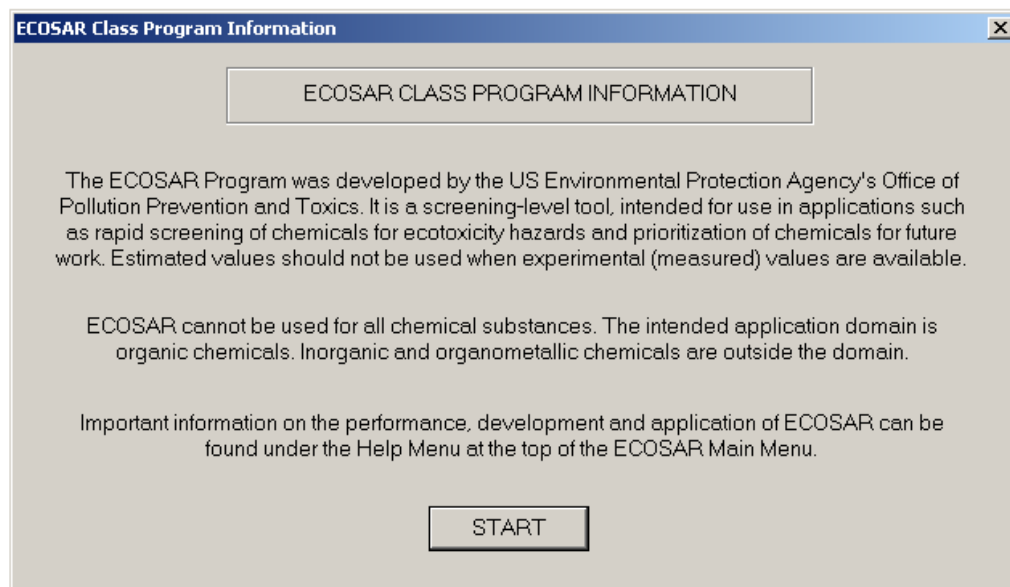
---

<sup>1</sup> Program developers do not yet know whether touch-screen devices will act as another type of pointing device and, thus, do not know whether the ECOSAR program will function properly on a Windows based tablet.

#### 4. Starting the ECOSAR Class Program

The ECOSAR Class Program is started like any other Microsoft Windows program. The easiest way to start the ECOSAR Class Program is to initiate the program from the Windows program list or by double-clicking the ECOSAR icon/short-cut. ECOSAR can be accessed through EpiSuite; however, the version associated with EpiSuite may not be the most recent version of ECOSAR. Version information can be obtained from the 'Help' menu by selecting 'About ECOSAR'. For additional information on starting Windows programs, consult your Windows documentation.

Once the ECOSAR program has been initiated, the following introductory screen is displayed:



Users should note the language on the opening screen regarding inorganic and organometallic chemicals in ECOSAR. These classes are outside the domain of ECOSAR V1.1.

After clicking 'Start' on the above pictured introductory screen, the data entry screen displays. Certain user inputs, depicted in the data entry screen and described in Section 5, are needed in



order to generate ECOSAR predictions based on organic QSARs (described in Section 6). QSAR predictions for surfactants are generated from a separate data entry screen, which is described in Section 8. Supporting documentation for ECOSAR can be obtained from the 'Help' menu and includes QSAR class definition and equation sheets (see 'Online ECOSAR Help'), the ECOSAR V1.1 Users guide, the ECOSAR Methodology Document, and tutorials for using the draw tool and for formulating SMILES structures.

**Figure 1. Example Data Entry Screen**

The screenshot shows the ECOSAR v1.10 data entry interface. The window title is "Ecosar v1.10" and the menu bar includes "File", "Edit", "Functions", "BatchMode", "ShowStructure", "Special\_Classes", and "Help". The interface features several input fields and buttons. At the top, there are buttons for "Clear Screen", "Previous", "Get User", "Save User", and "Calculate". Below these, there is a large text input field for "Enter SMILES:" with a "DRAW" button to its left. Underneath is another large text input field for "Enter NAME:" with a "NameLookup" button to its left. To the left of the "User Entered Values" section, there are input fields for "CAS Number:" (with a "CAS Input" button below it) and "Chemical ID 1:". The "User Entered Values" section on the right contains four input fields: "Water Solubility (mg/L):", "Melting Point (deg C):", "Log Kow:", and "Mol Wt [special use only:]".

Note: The appearance of the screen may vary somewhat due to screen resolution (e.g. 640 X 480 vs. 800 X 600), user selection of MS-Windows attributes (e.g. colors, font size, etc.), etc. In addition, Figure 1 illustrates how the entry screen appears when using Windows XP.

## 5. User Inputs For Generation Of ECOSAR Predictions

The information in section 5 applies to the main data entry screen shown in Figure 1. The required and optional data entry fields/inputs on the main data entry screen are subsequently described (pressing the F1 key where the edit cursor is located gives a brief description of that field):

(1) **Enter SMILES:** Field for SMILES notation of the structure to be estimated. A maximum of 360 characters is allowed. This field is required, but can be obtained by methods other than direct entry (see Section 5.1).

(2) **Enter Name:** Field for the name and/or description of the structure. This field is optional; not required. A maximum of 120 characters is allowed.

(3) **CAS Number:** The CAS (Chemical Abstract Service Registry) Number. This field is optional; not required.

(4) **Chemical ID 1:** Optional description / identity field; not required.

(5) **User-entered Log Kow:** The user-entered log octanol-water partition coefficient. This field is optional; not required.

(6) User-entered **Water Solubility:** The user-entered water solubility value in mg/L. This field is optional; not required.

(7) **User-entered Melting Point:** The user-entered melting point in degrees Celsius (°C). This field is optional; not required.

## 5.1. Structure Entry

Calculations in ECOSAR from the main data entry screen require the chemical structure to be interpreted using SMILES notation. Briefly, a SMILES notation depicts a molecular structure as a two-dimensional picture. SMILES notations are comprised of atoms (designated by atomic symbols), bonds, parentheses (used to show branching), and numbers (used to designate ring opening and closing positions). Users unfamiliar with SMILES notations can consult a descriptive journal article (Weininger, 1988), the ECOSAR Class Program help file (accessed by selecting "Help" from the top menu), <http://www.daylight.com> (Daylight Information Services), or <http://www.epa.gov/ncct/dsstox/MoreonSMILES.html>. An online SMILES translator is available from the National Cancer Institute at <http://cactus.nci.nih.gov/services/translate/>.

Seven different methods can be used to directly enter or retrieve the SMILES notation into the ECOSAR data entry screen:

- (1) Direct entry by the user from the keyboard (see Section 5.1.1 for direct entry of special cases such as salts)
- (2) Entry through a supplementary CAS "look-up" database (see Section 5.1.2).
- (3) Entry through a supplementary Name "look-up" database (see Section 5.1.3).
- (4) Entry through the structure drawing tool (see Section 5.1.4)
- (5) Entry from a previously created user file (see Section 5.1.5)
- (6) Importing of structures in MDL MOL file formats (see Section 5.1.6).
- (7) Recall of a previously entered smiles structure by pressing the F2 key or selecting the 'Previous' button on the main data entry screen.

The program can generate estimates for only one chemical at a time from this main screen and separate data entry is required for each chemical. Batch mode runs are possible and are described in detail in Section 7. Once a structure is entered, estimation for the entered structure is started by pressing the "Calculate" button or selecting the 'Page Down' key at any time during data entry.

### 5.1.1. Direct Entry

Direct entry requires knowledge of chemical structure and SMILES notation, as described in the introduction to Section 5.1. For direct entry, SMILES notation should be typed in the 'Enter SMILES' field on the data entry screen (see Figure 1). A SMILES notation is considered terminated at the first blank space. Characters following the first blank space are ignored! Methodology for entering certain compounds is highlighted below.

- (1) Organic anionic salts (e.g, Haloacids conjugated with Na, K, Li) – When entering an organic anionic salt, the conjugate cation (e.g., sodium, potassium, and lithium) should be replaced with a hydrogen ('H'). ECOSAR will automatically convert the conjugate cation to hydrogen if a SMILES structure is entered incorrectly.
- (2) Organic ammonium salts – Like the organic anionic salts described above, the conjugate salt (e.g., chloride ion) of organic ammonium salts should be replaced with a hydrogen; however, at a neutral pH ammonium moieties are reduced (e.g., loss of two hydrogens) to amines. Since QSARs are developed from test data using neutralized test solutions to replicate environmental conditions, ammonium salts should be entered into ECOSAR in the reduced amine form. Quaternary ammonium compounds (four carbons are bound to the nitrogen) are an exception since reduction (e.g., loss of hydrogens) at neutral pH is unlikely; however, these are classified as surfactants.
- (3) Inorganics - If metals and some elements are included in entered SMILES structures, the user will receive a warning that the chemical should not be profiled. Non-carbon elements considered in organic QSARs include oxygen, phosphate, sulfur, nitrogen,

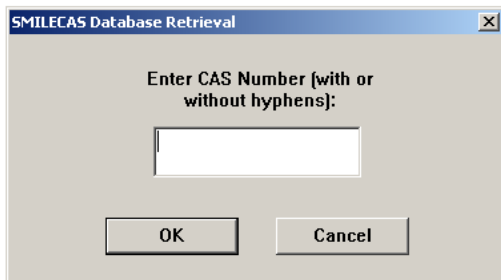
silicon, fluorine, bromine, chlorine, iodine, and hydrogen. Generally, direct hydrogen entry in a SMILES notation is unnecessary for ECOSAR v1.1.

(4) Charged Chemicals - Charged species (e.g., [+] and [-] signs) cannot be entered directly into a SMILES notation for charged chemicals. Azido compounds (commonly written as:  $N^+=N^-$ ) are depicted as 'N#N' for the purposes of SMILES notation. Nitro compounds (commonly written as  $N^+(O^-)=O$ ) are depicted as 'N(=O)=O' or 'T' for the purposes of SMILES notation.

### 5.1.2. CAS Number Database

To retrieve SMILES structures from the CAS Number Database, select the 'CAS Input' button beneath the CAS number field on the data entry screen or press the F8 key. A pop-up window will appear requesting entry of the CAS number. A valid CAS number should then be entered into the SMILES Look-Up Entry Screen (see Figure 2), followed by selection of the 'OK' button. The corresponding name and SMILES structure of the entered CAS number will be transferred to the Data Entry screen. The program accepts CAS numbers with or without hyphens. If an invalid CAS number is entered, the message 'NO CAS Match Found' will display. The user should check that the structure output by the CAS Number Database is correct.

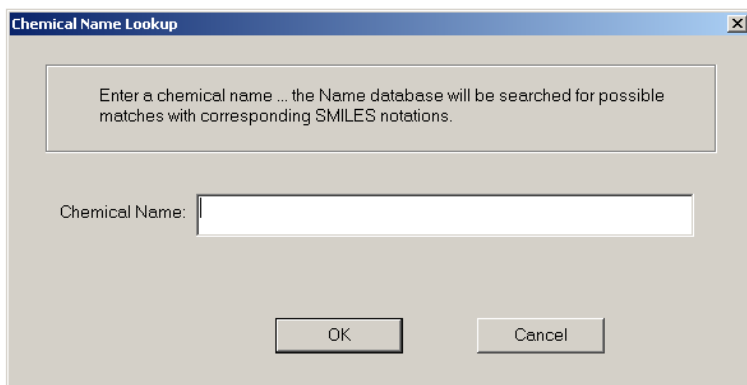
**Figure 2: SMILES Look-Up Entry Screen**



### 5.1.3. Name Look-Up Database

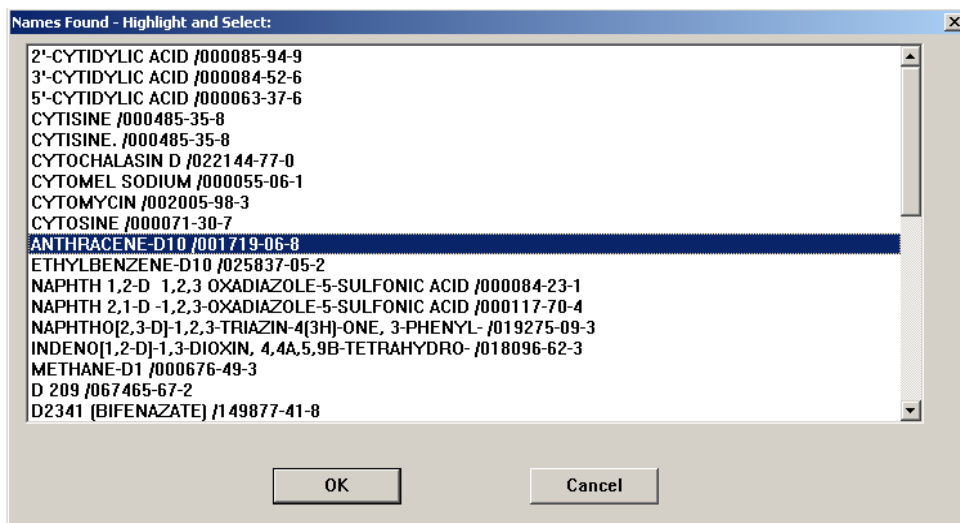
To retrieve SMILES structures from the Name Look-Up Database, select the 'NameLookup' button underneath the 'Enter Name' field on the data entry screen or Ctrl-N on the keyboard. Enter the name of the chemical in the 'Chemical Name' field of the entry screen (see Figure 3).

**Figure 3: Chemical Name Lookup Entry Screen**



After clicking the OK button, names that correspond to the entered name will display in a new 'Names Found' box (Figure 4). If no name is entered or if only a partial name is entered in the 'Chemical Name' field of the entry screen, then the first corresponding 40 chemical names in the Name Look-Up Database will display.

**Figure 4: Names Found Box**

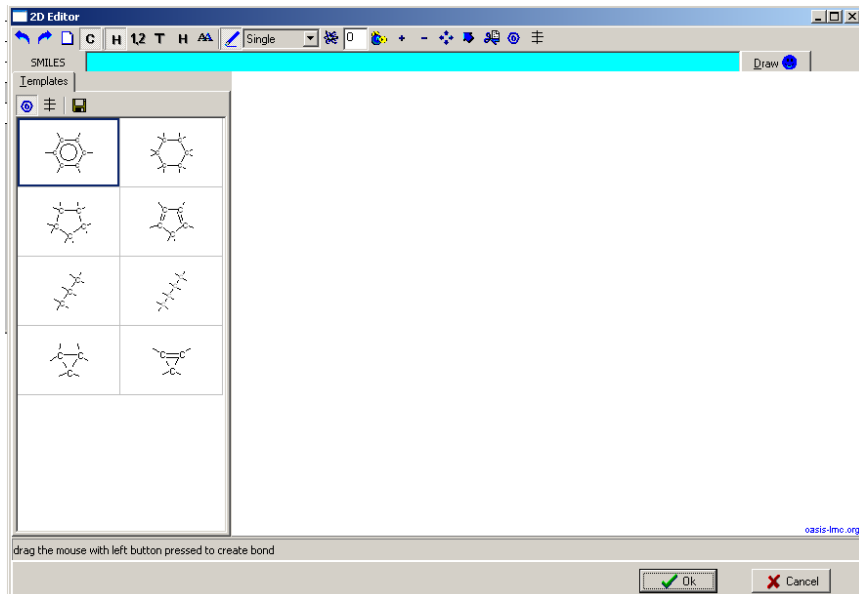


Highlight a chemical name from the 'Names Found' box and select 'OK'. The corresponding name, SMILES structure, and CAS number of the selected chemical will be transferred to the Data Entry screen. The user should check that the structure output by the Name Look-Up Database is correct.

#### 5.1.4. Entry Using The Drawing Tool

ECOSAR v1.10 now includes a structure drawing module that allows users to draw chemical structures and modify structures to generate SMILES notations for direct entry into the ECOSAR program. To initiate the module, select the 'DRAW' button on the main data entry screen or press the F11 key. The Blank Drawing Module Window is shown above in Figure 5.

**Figure 5 - Drawing Module Window**



A tutorial is provided in the 'Help' menu for drawing structures with the Drawing Module (select Draw Structure Help). Once a structure is entered, the 'OK' button should be selected. The corresponding SMILES structure of the depicted chemical will be transferred to the Data Entry screen and the drawing window will be shut. Pressing the 'Cancel' Button exits the Drawing Module with no transfer of SMILES and closes the Drawing Module.

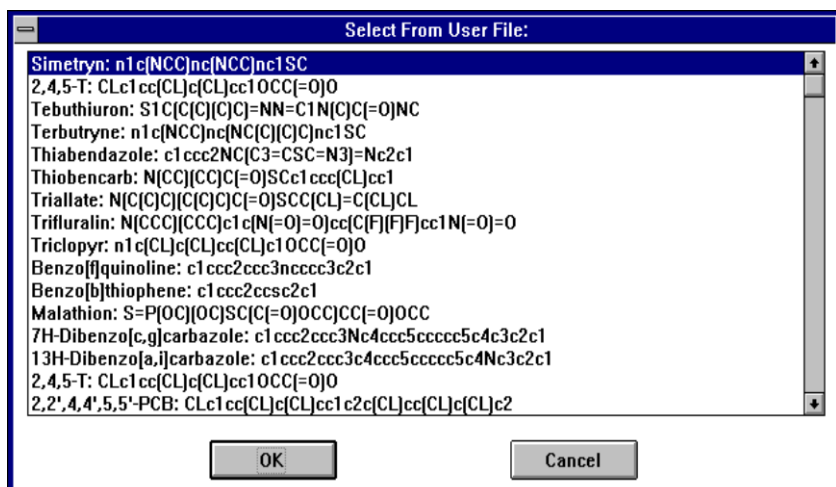
#### **5.1.5. Entry From A User File**

The User Input File is a file that can be created by a user and can contain up to 1500 SMILES notation and chemical names. This option enables users to store entered structures and can also be used for entering data into other estimation programs available from EPA such as EPISuite [<http://epa.gov/opptintr/exposure/pubs/episuite.htm>]. Once created, previously saved SMILES notations and chemical names can be accessed from a dialog box (see Figure 6) by pressing the F4 key or clicking the "Get User" button. Previously saved SMILES notations and chemical names in the User Input File can also be accessed for running batch modes (see Section 7). Until a User Input File is created, the dialogue box will appear empty. The user-highlighted name and



SMILES structure of a selected chemical will be transferred to the Data Entry screen when the 'OK' button is clicked.

**Figure 6.** Example User Input File



To create a User Input File, press the F6 key or click the "Save User" button displayed on the data entry screen. The SMILES notation and chemical name currently showing on the data entry screen will be saved to a file with the default name of SMILES.INP; this is for compatibility with similar estimation programs and cannot be changed by the user. After a file is selected (or entered by the user) and the 'Open' button displayed on the file window is pressed, ECOSAR appends the SMILES notation and chemical name currently showing on the data entry screen to the file. To add additional chemicals to the User Input File, just press the F6 key or click the "Save User" button again. If the file does not already exist, ECOSAR will create it and append the current SMILES and name as the first entry.

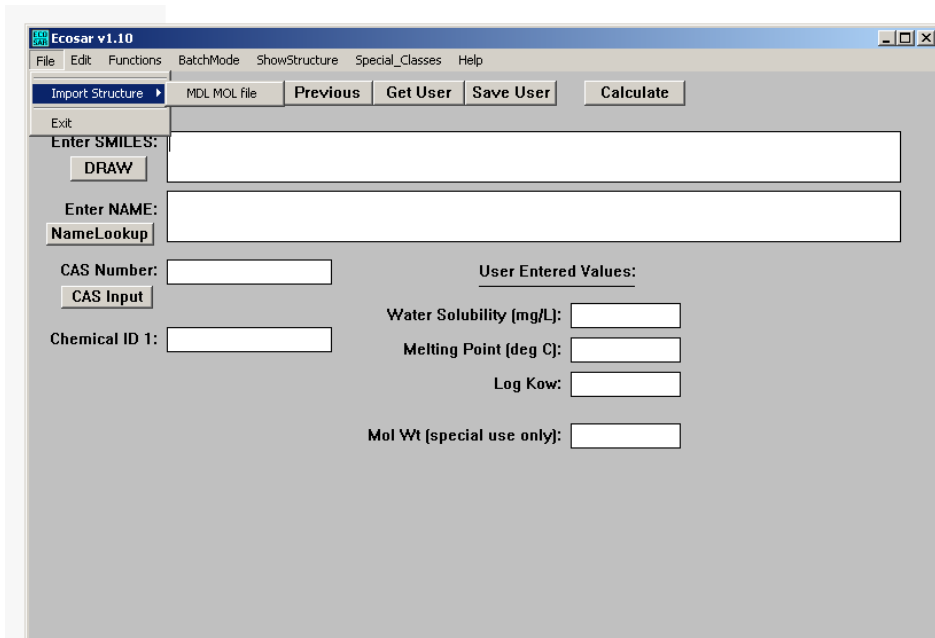
Currently, there is no way to edit or delete entries to SMILES.INP within the ECOSAR Class Program. However, SMILES.INP is a plain text file and it can be edited with any text editor or word processing program (as long as it is imported and saved as a DOS text file). Any text editor or word processing program can be used to create and add entries to SMILES.INP as long as the format is correct. The correct format is the following: the chemical name (up to 60 characters)

followed by a colon (:), then one space (and only one space) followed by the SMILES notation and a carriage return.

### 5.1.6. Importing Structures

ECOSAR v. 1.1 has an "import" features that allows MDL MOL file formats to be imported directly into ECOSAR. The "import" feature is accessed from the Menu Bar via: File/Import Structure as shown in Figure 7.

**Figure 7: Importing Structures**



Imported structures are converted to SMILES notations and placed in the SMILES data entry field of ECOSAR. ECOSAR filters the conversion to make the ECOSAR notation as compatible as possible with ECOSAR. However, some converted SMILES notations (especially SMILES with charged ions) will require some user modification before ECOSAR can estimate the structure.

## **5.2. Descriptor Fields**

The 'Enter Name', 'CAS Number', and 'Chemical ID1' fields are optional. Entering information into these fields does not alter the results. However, knowledge of these fields is necessary for use of certain SMILES look-up features such as the CAS number look-up database and the Name look-up database.

## **5.3. Entering Measured Data for Variables**

ECOSAR v1.1 enables users to provide data for certain parameters that will be used in place of ECOSAR predicted variables and/or values retrieved from Physprop. These variables are the Log Kow value, the water solubility value, and the melting point.

### **5.3.1. Log Kow**

The Log Kow value is used in the calculation of the predicted effect level. If a user enters a Log Kow value on the data entry screen, that value will be used in subsequent calculations. In the absence of a user-entered Log Kow value, ECOSAR will automatically use the Log Kow value calculated from the KOWWIN program (available from EPISuite). To minimize the potential effects of variability of measured Log Kow values due to factors such as variable study conditions, QSARS were developed from predicted Log Kow values, and it is often recommended that the predicted Kow values (calculated and used by default in ECOSAR) be used in the model when there is uncertainty in reliability of the available measured Kow values (see the ECOSAR Methodology Document for further information).

### 5.3.2. Water Solubility

Water solubility values (mg/L) are compared with predicted effect levels in order to identify effect levels that exceed the limit of water solubility. If a user enters a water solubility value in the data entry screen, that value will be used to determine if the predicted effect level exceeds the water solubility. In the absence of a user-entered water solubility value, ECOSAR will use the measured water solubility value retrieved from the Physprop Experimental Database (<http://www.syrres.com/what-we-do/product.aspx?id=133> ). If no measured water solubility values are available, ECOSAR will calculate the water solubility value for all query compounds using the WSKOWWIN program available from EPISuite. If available, a user-entered melting point (°C) will be used to calculate water solubility; otherwise, water solubility will be calculated without use of a melting point. Further details on the calculation of water solubility with WSKOWWIN are available from the help menu of Episuite and in the following document prepared for the U.S. Environmental Protection Agency (OPPT): Upgrade of PCGEMS Water Solubility Estimation Method (May 1994).

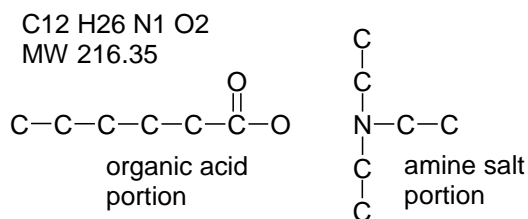
### 5.3.3. Melting Point

A user-entered melting point is used only in the calculation of water solubility. If a melting point is not entered, then the alternative method for water solubility calculation without a melting point value is used automatically by WSKOWWIN.

### 5.3.4. Molecular Weight Adjustment

ECOSAR calculates endpoints initially in units of mmol / L (micromole per litre), in certain cases (usually with complex salts) the ECOSAR input structure may have a different calculated MW than the molecule as a whole. Due to the mmol/L units, this may result in less accurate predicitions. Using the **Molecular Weight (special use only)** function can aid in refining the toxicity estimate.

**Example 1:** You are evaluating a chemical that is an aliphatic amine salt of an organic acid (shown to the right). This substance may disassociate in aqueous media. You are concerned that the amine portion could have some toxicity. If you evaluate the complete molecule in ECOSAR the results indicate the chemical is a neutral organic with low toxicity. You will also get a structural alert telling you this chemical should be evaluated as a surfactant.



ECOSAR v1.10

File Edit Functions BatchMode ShowStructure Special\_Classes Help

Previous Get User Save User CAS Input Calculate

Enter SMILES:  DRAW

Enter NAME:  NameLookup

CAS Number:

Chemical ID 1:

User Entered Values:

Water Solubility (mg/L):

Melting Point (deg C):

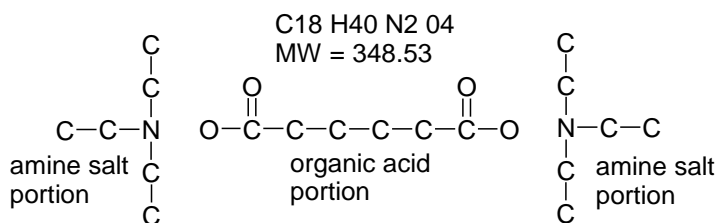
Log Kow:

Mol Wt [special use only]:

Enter the SMILES, draw, or otherwise input the amine structural unit as if it were a discrete chemical. Next, enter the molecular weight of the *entire* chemical (216.35) in the Mol. wt. (special use only) box. This is the molecular weight of the overall structure that contains one amine unit (as input). ECOSAR automatically adjusts for the input MW and will give you the predicted toxicity for the entire chemical based on the separate amine portion.

The results indicate the chemical may have a moderate concern for acute and chronic exposures. You should remember to clear the Mol wt box before you evaluate another chemical because ECOSAR doesn't automatically clear the box when enter a new structure.

**Example 2:** In the above example, you evaluated the toxicity of the portion of the molecule that contains *one* amine. In some cases the molecule you are evaluating may



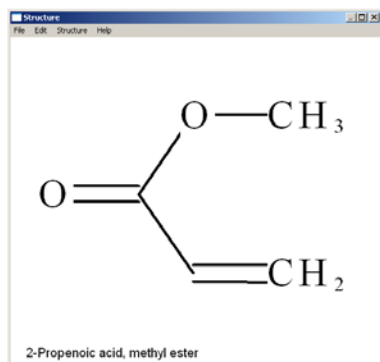
have multiple amines. For example, when you run the neutral form of the diacid shown to the right in ECOSAR you get results indicating it is low toxicity and you get the structural alert

telling you this chemical should be evaluated as a surfactant. In order to evaluate the toxicity of the amine portions, just like the example above, you run the amine portion as if it were a discrete chemical then enter *one half* the molecular weight of the entire chemical ( $348.53 / 2 = 174.265$ ) in the Mol. wt. (special use only) box. This is the molecular weight of the subunit that contains only one amine. ECOSAR results show a moderate concern for the diacid. If the molecule were to contain three amines, enter one third the molecular weight of the entire molecule in the Mol. wt. (special use only) box, etc

## 6. Results Window

Generally, the Results Window (see Figures 8 and 9) provides the chemical structure (separate window), chemical attributes, measured training set data used in QSAR development for the query chemical, the results of ECOSAR Class Program's estimations, baseline toxicity values, and information specific to the interpretation of the QSAR results. Each of these is described in subsequent sections.

**Figure 8:** Structure Window



**Figure 9: Results Window**

**Ecosar Results**  
 Print Save Results TableSave Copy Remove Window Help

SMILES : O=C(OC)C=C  
 CHEM : 2-Propenoic acid, methyl ester  
 CAS Num: 000096-33-3  
 ChemID1:  
 MOL FOR: C4 H6 O2  
 MOL WT : 86.09  
 Log Kow: 0.728 (EPISuite Kowwin Estimate)  
 Log Kow: (User Entered)  
 Log Kow: 0.80 (PhysProp DB exp value - for comparison only)  
 Heit Pt: (User Entered for Wat Sol estimate)  
 Wat Sol: 2.636E+004 (ng/L, EPISuite MSKowwin Estimate)  
 Wat Sol: (User Entered)  
 Wat Sol: 4.94E+004 (ng/L, PhysProp DB exp value)

-----  
 Values used to Generate ECOSAR Profile  
 -----  
 Log Kow: 0.728 (EPISuite Kowwin Estimate)  
 Wat Sol: 4.94E+004 (ng/L, PhysProp DB exp value)

-----  
 Available Measured Data From ECOSAR Training Set  
 -----

CAS No	Organism	Duration	End Pt	Measured ng/L (ppm)	Ecosar Class	Reference
000096-33-3	Fish	96-hr	LC50	3.4	Acrylates	Datasheet
000096-33-3	Fish (SW)	96-hr	LC50	1.1	Acrylates	Drottar, 1995
000096-33-3	Daphnid	48-hr	LC50	2.6	Acrylates	Drottar, 1995
000096-33-3	Mysid	96-hr	LC50	1.6	Acrylates	Drottar, 1996
000096-33-3	Green Algae	72/96-hr	ChU	1.9	Acrylates	Thompson, 1995
000096-33-3	Green Algae	72/96-hr	EC50	4.8	Acrylates	Thompson, 1995

-----  
 ECOSAR v1.1 Class-specific Estimations  
 -----  
 Acrylates

ECOSAR Class	Organism	Duration	End Pt	Predicted mg/L (ppm)
Acrylates	: Fish	96-hr	LC50	2.519
Acrylates	: Daphnid	48-hr	LC50	8.465
Acrylates	: Green Algae	96-hr	EC50	1.086
Acrylates	: Fish		ChU	0.098
Acrylates	: Daphnid		ChU	0.634 †
Acrylates	: Green Algae		ChU	0.348
Acrylates	: Fish (SW)	96-hr	LC50	1.102
Acrylates	: Mysid (SW)	96-hr	LC50	1.605
Acrylates	: Fish (SW)		ChU	0.110 †
Acrylates	: Mysid (SW)		ChU	0.161 †

-----  
 Neutral Organic SAR  
 (Baseline Toxicity)

: Fish	96-hr	LC50	982.023
: Daphnid	48-hr	LC50	408.124
: Green Algae	96-hr	EC50	375.761
: Fish		ChU	83.569
: Daphnid		ChU	34.868
: Green Algae		ChU	45.811

NOTE: \* = asterisk designates: Chemical may not be soluble enough to measure this predicted effect.  
 NOTE: † = exclamation designates: The toxicity value was estimated through application of acute-to-chronic ratios per methods outlined in the ECOSAR Technical Reference Manual posted on the ECOSAR webpage.

-----  
 Class Specific LogKow Cut-Offs  
 -----  
 Acrylates:  
 -----  
 Maximum LogKow: 5.0 (LC50)  
 Maximum LogKow: 6.4 (EC50)  
 Maximum LogKow: 8.0 (ChU)

Baseline Toxicity SAR Limitations:  
 -----  
 Maximum LogKow: 5.0 (Fish 96-hr LC50; Daphnid LC50)  
 Maximum LogKow: 6.4 (Green Algae EC50)  
 Maximum LogKow: 8.0 (ChU)

The Results Windows can be moved, sized and placed anywhere on the Microsoft Windows desktop. It does not need to be removed or closed before running another chemical in the

program; the Results Window will be updated automatically. The following menu choices are available when the Results Window is active:

**Print:** Prints the results with or without the structure.

**Save Results:** This command saves the summary output to a file. The output files are named ECOW\*.DAT where "\*" is a number from 1 to 100. Numbering begins at 1 and automatically proceeds to number 100. Currently, all results are appended to the same file number until the program is exited. The next time the program is started, the next available number is used; therefore, different files are used from session to session! If all numbers have been used in existing files, then number 1 will be used and the existing file ECOW001.DAT will be overwritten.

**Copy:** This command copies the results as shown (minus the rectangle enclosing the estimate) to the Windows clipboard. The results can then be copied into other Windows programs such as word processors. When copied to a word processor (such as Word Perfect or Microsoft Word), a non-proportional font (such as courier) must be used for correct formatting and the page width must be wide enough.

**Remove Window:** This command deletes the Results Windows; a new Results Window will appear with the next estimation. It may be more convenient to move and size the Results Window for personal preference (after the first estimation) rather than to remove it after each estimation. If the Results Window is left on the screen, the next estimation results will simply replace the existing results.



## 6.1. Structure Window

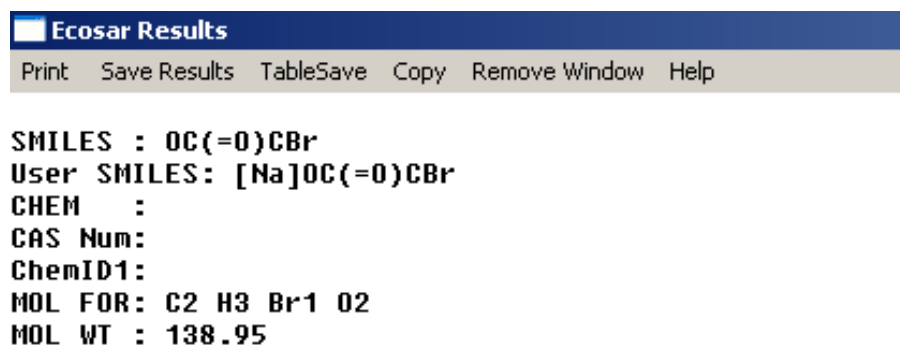
The Structure Window (see Figure 8) shows a 2-dimensional plot of the chemical structure. The window shows the entire structure (it does not "clip" sections of the molecule). In order to fit the entire structure in the window, the aspect ratio of the MS-Windows metafile depiction has been rendered proportional (that is, by changing the height or width of the window, the structure scaling changes). At times, the height or width of the window may need to be changed to give a better structure depiction. When results from the Results Window are printed with the "Print results with structure" option, the aspect ratio of the structure will be printed (if possible) with the same aspect ratio as shown in the Structure Window.

The Structure Window Menu Bar gives access to printing the structure, saving the structure as an MDL MOL file, copying the structure to the MS-Windows clipboard, or changing selected window parameters. Changeable windows parameters include background colors of the structure or bottom text areas. Double clicking the text at the bottom of the window allows the text to be changed. Copying the structure (from the menu bar Edit) to the Windows clipboard can be carried out in two different ways: (1) "Copy (as a placeable metafile)" copies both structure and text to the clipboard; and (2) "Copy Structure (as metafile)" copies only the structure to the clipboard. Most commercial word processors and drawing programs will recognize at least one of these format.

## 6.2. Chemical Attributes

The ECOSAR results page provides the attributes of the query chemical first (see Figure 10 and 11). The SMILES structure used to predict the effect levels is depicted first. If a structure different from the user-entered structure (e.g., a sodium molecule is included in the SMILES notation) is used to determine predictions, then an additional field titled 'User SMILES' will appear that represents the structure entered by the user (see Figure 10).

**Figure 10:** Example of ‘User SMILES’ Field



The screenshot shows a window titled "Ecosar Results" with a menu bar containing "Print", "Save Results", "TableSave", "Copy", "Remove Window", and "Help". Below the menu bar, the following text is displayed:

```
SMILES : OC(=O)CBr
User SMILES: [Na]OC(=O)CBr
CHEM   :
CAS Num:
ChemID1:
MOL FOR: C2 H3 Br1 O2
MOL WT : 138.95
```

The field with the molecular formula (‘MOL FOR’) and the molecular weight (‘MOL WT’) are determined from the SMILES notation used for ECOSAR predictions. Chemical name (‘CHEM’), CAS number (‘CAS Num’), and ‘ChemID1’ are unnecessary descriptors of the query chemical, and blank fields in the output signify that this information was not entered and/or not available.

Log Kow and water solubility shown in Figure 11 are presented from EpiSuite (Kowwin and WSKowwin estimations, respectively), from the PhysProp database if available, and for values entered by the user. For water solubility, units of measurement are always milligrams per liter (mg/L). If the user enters a melting point, the value will display next to the ‘Melt Pt’ field; otherwise, the field will be blank.

Below the chemical properties section, as shown in Figure 11, the ECOSAR results page presents “Values used to Generate ECOSAR Profile”. These values represent the variables used for ECOSAR estimations. Section 5.3 describes methods for selecting the Log Kow and water solubility values used for ECOSAR estimations and how these values are used for ECOSAR predictions.

**Figure 11: Chemical Attribute Information**

```

Ecosar Results
Print Save Results TableSave Copy Remove Window Help

SMILES : O=C(O)C=C
CHEM   : 2-Propenoic acid, methyl ester
CAS Num: 000096-33-3
ChemID1:
MOL FOR: C4 H6 O2
MOL WT : 86.09
Log Kow: 0.728      (EPISuite Kowwin Estimate)
Log Kow:           (User Entered)
Log Kow: 0.80      (PhysProp DB exp value - for comparison only)
Melt Pt:           (User Entered for Wat Sol estimate)
Wat Sol: 2.636E+004 (mg/L, EPISuite WSKowwin Estimate)
Wat Sol:           (User Entered)
Wat Sol: 4.94E+004 (mg/L, PhysProp DB exp value)

-----
Values used to Generate ECOSAR Profile
-----
Log Kow: 0.728      (EPISuite Kowwin Estimate)
Wat Sol: 4.94E+004 (mg/L, PhysProp DB exp value)

```

### 6.3. Measured Effect Levels

Training set data used to develop regression equations for QSAR class endpoints are presented in the training set data section of the results page (see Figure 12).

**Figure 12: Training Set Data for Query Chemical**

```

-----
Available Measured Data from ECOSAR Training Set
-----

```

CAS No	Organism	Duration	End Pt	Measured ng/L (ppm)	Ecosar Class	Reference
000096-33-3	Fish	96-hr	LC50	3.4	Acrylates	Datasheet
000096-33-3	Fish (SW)	96-hr	LC50	1.1	Acrylates	Drottar, 1995
000096-33-3	Daphnid	48-hr	LC50	2.6	Acrylates	Drottar, 1995
000096-33-3	Mysid	96-hr	LC50	1.6	Acrylates	Drottar, 1996
000096-33-3	Green Algae	72/96-hr	ChU	1.9	Acrylates	Thompson, 1995
000096-33-3	Green Algae	72/96-hr	EC50	4.8	Acrylates	Thompson, 1995

Data are not presented in any particular order. If measured data used in the training set are considered to be TSCA confidential business information (CBI), these data are not shown in this section. Since data collection for the ECOSAR program began in the early 1980's, some data incorporated into ECOSAR came from data sheets that did not clearly identify the reference; however, extensive efforts have identified most references. Data considered 'supplemental' or 'adequate with restrictions' may have been included in some training sets in the absence of better

data. The class identified in the 'ECOSAR Class' column represents the class(es) of training set data that were used in to develop the QSAR equation. See the Technical Reference Manual (also referred to as the ECOSAR Methodology Document) for further information on data collection and selection methods.

#### **6.4. Presentation of Predicted Toxicity**

The mode of toxic action for most neutral organic chemicals is narcosis, and many types of chemical classes present toxicity to organisms via narcosis (i.e ethers, alcohols, ketones). However, some organic chemical classes have been identified as having a more specific mode of toxicity. These are typically organics that are reactive and/or ionizable and exhibit excess toxicity in addition to narcosis (i.e. acrylates, epoxides, anilines). At the bottom of the ECOSAR v1.1 Class-Specific Estimation Section (see Figure 13) the user will be provided the neutral organics class QSAR results for the 6 primary endpoints (F96 LC<sub>50</sub>, D48 LC<sub>50</sub>, GA96 EC<sub>50</sub>, FChV, DChV, and GACHV), often referred to as "baseline toxicity", even when the compound falls into a separate class with excess toxicity. The baseline toxicity equations in ECOSAR v. 1.1 are based on data collected through 2010. (Please note that baseline toxicity equations in the previous ECOSAR v 1.0 were based on data collected through 1999 and ECOSAR v0.99 was calculated from the 1981 Konemann Equation). The purpose for presenting the baseline toxicity values is so the user can quantify the amount of excess toxicity above baseline narcosis for the chemical class, if interested.

**Figure 13:** Predicted ECOSAR Toxicity Profile for a Chemical with Excess Toxicity

```

-----
ECOSAR v1.1 Class-specific Estimations
-----
Acrylates

ECOSAR Class          Organism          Duration  End Pt  Predicted
-----
Acrylates             : Fish            96-hr    LC50    2.519
Acrylates             : Daphnid         48-hr    LC50    8.465
Acrylates             : Green Algae     96-hr    EC50    1.086
Acrylates             : Fish            ChU      0.098
Acrylates             : Daphnid         ChU      0.634 *
Acrylates             : Green Algae     ChU      0.348
Acrylates             : Fish (SW)       96-hr    LC50    1.102
Acrylates             : Mysid (SW)      96-hr    LC50    1.605
Acrylates             : Fish (SW)       ChU      0.110 *
Acrylates             : Mysid (SW)      ChU      0.161 *

-----
Neutral Organic SAR   : Fish            96-hr    LC50    982.023
(Baseline Toxicity)  : Daphnid         48-hr    LC50    408.124
                    : Green Algae     96-hr    EC50    375.761
                    : Fish            ChU      83.569
                    : Daphnid         ChU      34.868
                    : Green Algae     ChU      45.811

NOTE: * = asterisk designates: Chemical may not be soluble
       enough to measure this predicted effect.

NOTE: ! = exclamation designates: The toxicity value was estimated through
       application of acute-to-chronic ratios per methods outlined in the
       ECOSAR Technical Reference Manual posted on the ECOSAR webpage.

```

Output of predicted ECOSAR values is slightly different for query compounds which are only identified as neutral organics since the user will only be provided with neutral organics QSAR predictions (see Figure 14).

**Figure 14:** Predicted ECOSAR Toxicity Profile for a Neutral Organics Compound

```

-----
ECOSAR v1.1 Class-specific Estimations
-----
Neutral Organics

ECOSAR Class          Organism          Duration  End Pt  Predicted
-----
Neutral Organics     : Fish            96-hr    LC50    65.106
Neutral Organics     : Fish            14-day   LC50    67.838
Neutral Organics     : Daphnid         48-hr    LC50    36.910
Neutral Organics     : Green Algae     96-hr    EC50    31.902
Neutral Organics     : Fish            30-day   ChU      6.359
Neutral Organics     : Daphnid         ChU      3.596
Neutral Organics     : Green Algae     ChU      7.177
Neutral Organics     : Fish (SW)       96-hr    LC50    81.960
Neutral Organics     : Mysid           96-hr    LC50    61.201
Neutral Organics     : Fish (SW)       ChU      8.858
Neutral Organics     : Mysid (SW)     ChU      5.301
Neutral Organics     : Earthworm       14-day   LC50    136.037

NOTE: * = asterisk designates: Chemical may not be soluble
       enough to measure this predicted effect.

```

Two common notations reported for predicted effect levels are an asterisk (\*) used to designate predicted effect levels that exceed the water solubility limit and an exclamation (!) used to designate effects levels predicted using an acute-to-chronic (ACR) ratio.

#### 6.4.1 Adjusting Effect Level for Presence of an Acid (x10)

When you enter an acid into ECOSAR it is represented as the neutral form (SMILES interpreter does not recognize charges) and resulting estimates for Kow and toxicity are calculated accordingly. However, acids are typically fully ionized at environmental pH's which mean the acid would be represented as "COO-" and not "COOH". The ionized form of the acid would have a lower Kow value (more soluble in water) than the neutral form. Based on the observed correlation between Kow and toxicity, the ionized form (with a lower Kow value) would be less toxic to aquatic organisms than the neutral form. Therefore, the estimates for acids are adjusted (multiplied by 10) to account for ionization. The adjustment factor is not chemical or class-specific, but represents a reasonable adjustment across the whole spectrum of chemicals. When the factor is initiated, the results from ECOSAR will indicate the class name with "-acid" added and then present the flag "Acid moiety found: Predicted values multiplied by 10" as shown below:

```
-----
ECOSAR v1.1 Class-specific Estimations
-----
Esters-acid

ECOSAR Class          Organism          Duration  End Pt  Predicted
=====
--> Acid moiety found: Predicted values multiplied by 10
```

#### 6.5. QSAR-Specific Information

Following the reported predicted effect levels is a section that states maximum Log K<sub>ow</sub> cut-offs for each class (see Figure 15). The K<sub>ow</sub> cut-off values signify the point a chemical is no longer particularly soluble and not likely to result in toxicity to the organism for the given duration. For chemicals with log Kow values that exceed the limits, results are typically reported as "No Effects at Saturation".

**Figure 15: QSAR Specific Information**

-----  
Class Specific LogKow Cut-Offs  
-----

Acrylates:

-----  
Maximum LogKow: 5.0 (LC50)  
Maximum LogKow: 6.4 (EC50)  
Maximum LogKow: 8.0 (ChU)

Baseline Toxicity SAR Limitations:

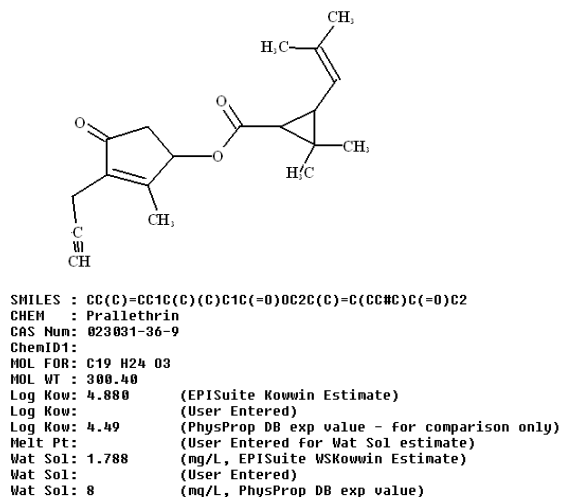
-----  
Maximum LogKow: 5.0 (Fish 96-hr LC50; Daphnid LC50)  
Maximum LogKow: 6.4 (Green Algae EC50)  
Maximum LogKow: 8.0 (ChU)

In general, Log  $K_{ow}$  cut-offs are 5 for fish and daphnid acute endpoints, 6.4 for green algae 72/96-h EC50 endpoints, and 8 for all chronic endpoints.. However, when available training set data are more robust, attempts have been made to tailor Log  $K_{ow}$  cut-offs to the specific QSAR. In some cases, Log  $K_{ow}$  cut-offs have been depicted with a greater than sign (>) since available data exhibited toxicity for class members having Log  $K_{ow}$  values above the  $K_{ow}$  limit, but within the limit of water solubility. See the Technical Reference Manual (also referred to as the ECOSAR Methodology Document) for further discussion of Log  $K_{ow}$  cut-offs.

**6.6. Comprehensive approach for determining the most representative QSAR class.**

The ECOSAR program may provide results for multiple classes if the entered structure contains the defined base-structure from each of those classes identified in the ECOSAR class definition sheets. Figure 16 presents an example of a compound that fits multiple classes. The predictions section of the output is depicted in Figure 18 and described in the ‘Selecting the Most Representative Class’ section (see Section 6.6.4).

**Figure 16:** Example Chemical Classified Into Multiple Classes



-----  
Values used to Generate ECOSAR Profile

Log Kow: 4.880 (EPISuite Kowwin Estimate)  
Wat Sol: 8 (mg/L, PhysProp DB exp value)

-----  
Available Measured Data from ECOSAR Training Set

CAS No	Organism	Duration	End Pt	Measured mg/L (ppm)	Ecosar Class	Reference
023031-36-9	Daphnid		ChU	0.00092	Pyrethroids	OPP Pesticides Ecotoxicity DB
023031-36-9	Daphnid	48-hr	LC50	0.0062	Pyrethroids	OPP Pesticides Ecotoxicity DB
023031-36-9	Fish	96-hr	LC50	0.022	Pyrethroids	OPP Pesticides Ecotoxicity DB
023031-36-9	Fish	96-hr	LC50	0.012	Pyrethroids	OPP Pesticides Ecotoxicity DB
023031-36-9	Fish (SW)	96-hr	LC50	0.026	Pyrethroids	OPP Pesticides Ecotoxicity DB
023031-36-9	Mysid	96-hr	LC50	0.003900	Pyrethroids	OPP Pesticides Ecotoxicity DB

-----  
ECOSAR v1.1 Class-specific Estimations

Esters  
Vinyl/Allyl Ketones  
Vinyl/Allyl Esters  
Pyrethroids

When the program identifies multiple classes, the user must determine the most suitable class for estimating toxicity using knowledge of environmental toxicology, organic chemistry, and statistics. If available, measured data should be used over predicted data as long as the measured data are considered adequate (determination of study adequacy is the responsibility of the user). Additionally, the Online ECOSAR Help menu (Figure 17) contains ‘Class Supporting Information’ that includes definitions and equation documents for all QSARs (Figure 18). These documents enable the user to evaluate adequacy of predictions.

Figure 17: Selection Bar for Online ECOSAR Help



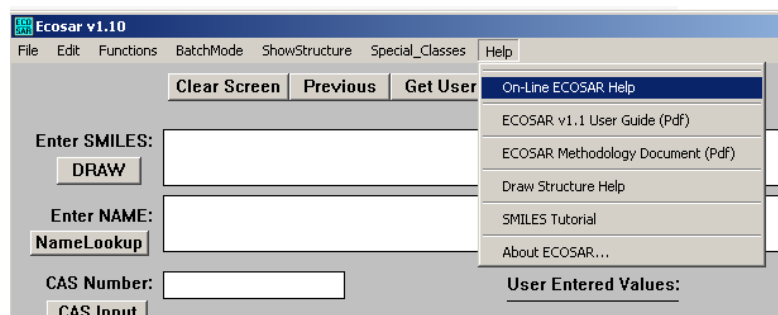
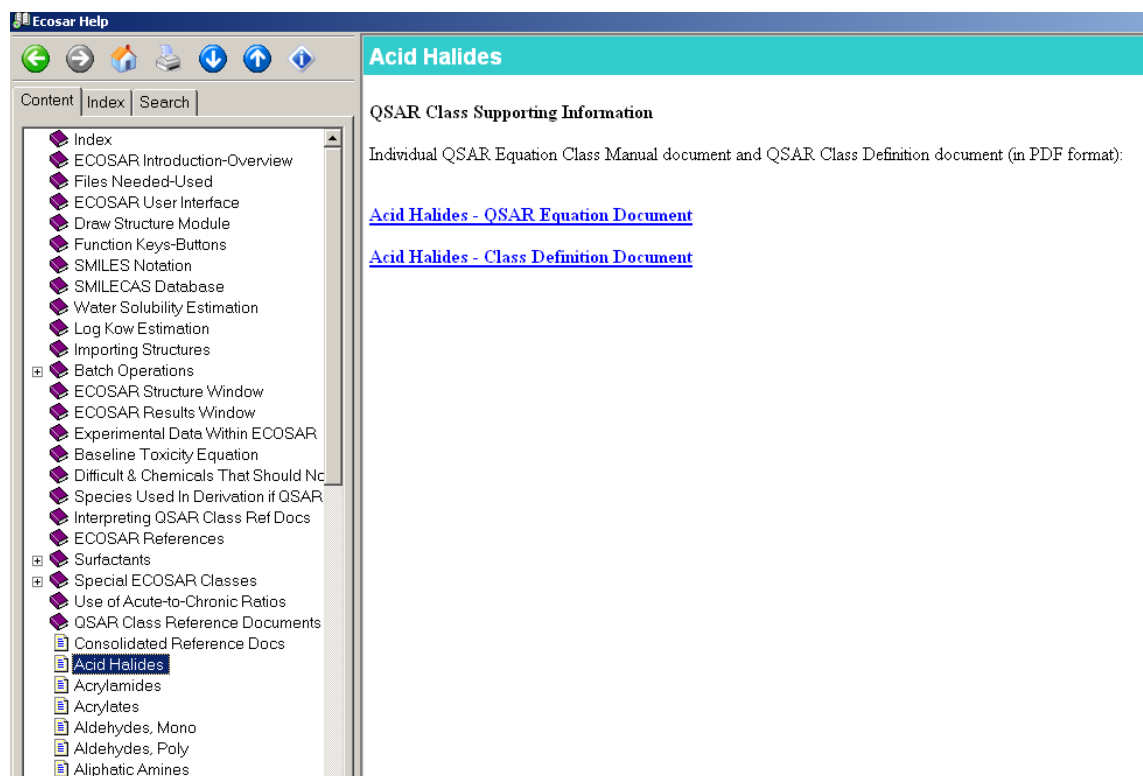


Figure 18: Online ECOSAR Help Screen With QSAR Class Supporting Information Shown



In the absence of adequate measured data, the traditional approach has been to use the most conservative effect level from any of the multiple classes listed. However, this is not always the best approach. The methods described in sections 6.6.1 through 6.6.4 provide useful guidance for eliminating classes that are not truly representative of the query compound or classes with insufficient data to fully support a regression equation. Considerations for class determination are described below and are correlated with the above example.

### 6.6.1. General Classes vs. Sub-Classes

Some classes within ECOSAR are considered general classes and represent a simple molecular moiety (e.g., Esters, Aliphatic Amines, Phenols, Amides). Other sub-classes define more specific and complex molecular configurations (e.g. Nicotinoids, Pyrethroids) or define explicit molecular attachments to otherwise general classes (e.g., Haloamides). Depending on ECOSAR programming, predictions for the general classes as well as the more specific sub-classes may be displayed in the ECOSAR output. In the example depicted in Figure 16, Prallethrin is identified as an Ester, Vinyl/Allyl Ketone, Vinyl/Allyl Ester, and a Pyrethroid.

Sub-classifications are created in ECOSAR when compounds with larger, more complex structural moieties (pyrethroids) are identified that exhibit toxicity levels which are unlike estimates for the more general classes (esters, vinyl/allyl ketones, vinyl allyl esters), even though those complex compounds may still contain those simple molecular features. In the example depicted in Figure 16, the general classes identified for prallethrin would be esters, vinyl/allyl ketones, and vinyl/allyl esters (relating to smaller functional groups contained within prallethrin). The more specific sub-class is pyrethroids which define a much larger part of the prallethrin molecule. The first step to identifying the optimum prediction would be to compare the chemical class definition with the structural features of the query compound, Prallethrin. Figures 17 and 18 show how to access the QSAR class supporting information that enables the user to assess structural features of the query compound. The user needs to determine how many molecular features of Prallethrin fit each class definition and whether that class is the most specific available in ECOSAR for the query chemical. The following is the author's interpretation of the ECOSAR output for the example depicted in Figure 16.

- (1) Esters-Prallethrin unequivocally fits the esters definition, but this is a general class that addresses only one of the structural features of this compound.
- (2) Vinyl/Allyl Ketones-The vinyl/allyl ketone in prallethrin is within a 5-carbon ring. Although there have been scientific discussions on whether to restrict vinyl/allyl classes to only terminal vinyl/allyl moieties, ECOSAR definitions for

these classes have not yet been restricted due to uncertainty. Thus, the user must decide whether Prallethrin should be excluded from this class.

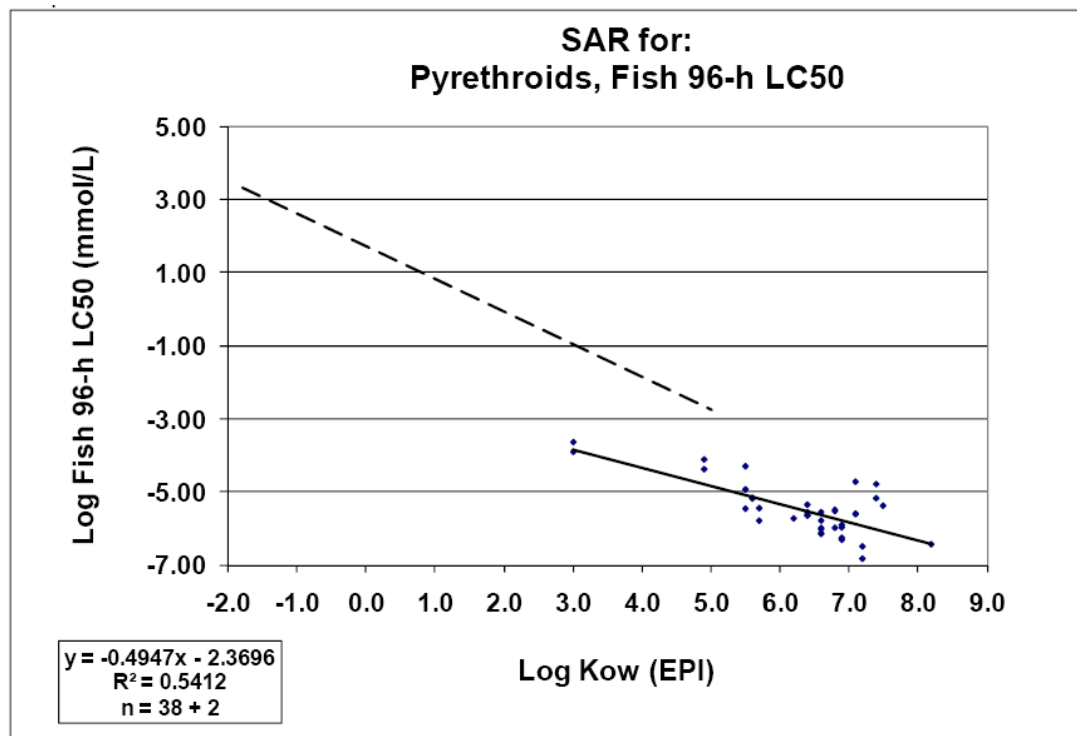
- (3) Vinyl/Allyl Esters-The allyl moiety of Prallethrin is within a ring structure. As discussed for Vinyl/Allyl Ketones, some scientists indicate that the vinyl/allyl moiety must be terminal and/or cannot be within a ring. Thus, the user must decide whether prallethrin should be excluded from this class.
- (4) Pyrethroids- Prallethrin unequivocally fits the pyrethroids definition and literature resources consistently identify the compound as a pyrethroid pesticide for which the class is modeled. The structure features of the pyrethroid class are also more specific to the structural features of prallethrin than to the features of the esters QSAR class.

The next step involves looking at the equation documents to determine the quality of the QSARs (see Sections 6.6.2 and 6.6.3).

#### **6.6.2. Correlation of Kow with Toxicity in QSAR Training Data Sets**

For each developed QSAR, a graph (see Figure 19) is displayed in the ECOSAR Equation document (see Figures 17 and 18 for access QSAR Supporting Documents including ECOSAR Equation Documents) along with a table of supporting data.

**Figure 19:** Graph from the Pyrethroids QSAR Equation Document for the F96 Endpoint.



A coefficient of determination ( $r^2$ ) is reported in both the depicted graph (a scatterplot) and in the text of the ECOSAR Equation Documents. The coefficient of determination is a numeric representation of how much variation in one variable is directly related to the variation in another variable (e.g., endpoint effect level (mmol/L) vs. Log  $K_{ow}$ ). A correlation coefficient can be determined by taking the square root of the presented coefficient of determination. Users should consult the ECOSAR Equation documents of identified classes to quantitatively determine correlation of variables within training data sets based on the coefficient of determination. Depending on the user's knowledge and understanding of statistics, a level of significance can be determined for relationships observed in each QSAR class for each endpoint using a correlation coefficient derived from the presented coefficient of determination.<sup>2</sup> However, a weak relationship does not necessarily indicate little or no correlation; if adequate data were scarce for

<sup>2</sup> This discussion is beyond the scope of this document. Since these methods are a simple correlation of two variables, there is an abundance of material for determining significance using Pearson's correlation coefficient including a publically available educational document from the Smithsonian National Zoo (<http://nationalzoo.si.edu/Education/ClassroomPartnerships/BioDivMonPro/TrainingCourseandManuals/trainingmanual/SA%203.pdf>).

a certain endpoint, low correlation may be a product of insufficient supporting data and/or may indicate that further sub-classification or reclassification is needed.

For simplicity, significance of the relationship of effect levels (mmol/L) vs. Log  $K_{ow}$  values for each identified class will be determined for the fish 96-hr  $LC_{50}$  endpoint only.

- (1) Esters- Pearson's Correlation Coefficient ( $r$ ) is -0.88. Using 5% uncertainty ( $p = 0.05$ ), the correlation between the F96-hr  $LC_{50}$  value (mmol/L) and the Log  $K_{ow}$  value is statistically significant.
- (2) Vinyl/Allyl Ketones- Pearson's Correlation Coefficient ( $r$ ) is -0.84. Using 5% uncertainty ( $p = 0.05$ ), the correlation between the F96-hr  $LC_{50}$  value (mmol/L) and the Log  $K_{ow}$  value is statistically significant.
- (3) Vinyl/Allyl Esters- Pearson's Correlation Coefficient ( $r$ ) is -0.44. Using 5% uncertainty ( $p = 0.05$ ), the correlation between the F96-hr  $LC_{50}$  value (mmol/L) and the Log  $K_{ow}$  value is not statistically significant.
- (4) Pyrethroids- Pearson's Correlation Coefficient ( $r$ ) is -0.74. Using 5% uncertainty ( $p = 0.05$ ), the correlation between the F96-hr  $LC_{50}$  value (mmol/L) and the Log  $K_{ow}$  value is statistically significant.

### 6.6.3. Robustness and Distribution of Supporting Datasets for QSAR Classes

The supporting data sets (training sets) used to derive QSARs within a chemical class range from the very large, e.g., neutral organics, to the very small, e.g., aromatic diazoniums. If a classification or sub-classification is supported by a large dataset that is well correlated, then strength of the association is increased and adequacy of the resulting regression equation is better substantiated. Additionally, depending on the range of Log  $Kow$  values of the available data for a given training set, the Log  $Kow$  value of the queried compound may be notably less than or greater than the minimum and maximum Log  $Kow$  values of the training set. Sometimes data are distributed so that the regression line overlaps or crosses over the depicted neutral organic line (dashed line, see Figure 19), which may be an artifact of the training set data and/or may indicate that excess toxicity for that particular endpoint was not observed. These issues are not

always apparent from the ECOSAR results output and may result in predictions that seem anomalous. Users should consult the ECOSAR Equation documents of identified classes to visually determine correlation from the depicted graphs of each endpoint. For our prallethrin example, the following can be interpreted from the ECOSAR equation sheets.

- (1) Esters- This SAR may be used to estimate toxicity for a variety of esters that include acetates (non-acids), benzoates, dicarboxylic aliphatics, and phthalates derived from aliphatic alcohols and phenol.
- (2) Vinyl/Allyl Ketones- The class-specific SAR equation is yielding lower toxicity values (i.e., less toxic) than estimated from baseline toxicity (neutral organics), as indicated in the text of the ECOSAR Equation document. This may be an artifact of the training data set, but may also indicate that the vinyl/allyl ketone moiety does not exhibit excess toxicity.
- (3) Vinyl/Allyl Esters- The training data set consists of 4 chemicals. All data are considered TSCA CBI and, thus, their identities are unknown. One of these data points appears to be an outlier, which may have contributed to low correlation reported in Section 6.6.2.
- (4) Pyrethroids- The Log K<sub>ow</sub> values for data points that are within the solubility limit range from 3 to 8.2. Thus, if the Log K<sub>ow</sub> value of the query compound is much less than 3, there may be some uncertainty with the prediction. However, the pyrethroid QSAR class, which by definition contains an ester moiety, appears to exhibit much greater toxicity than the esters QSAR class.

#### **6.6.4. Selecting the Most Representative Class**

Traditionally, in the absence of adequate measured data, the most conservative effect level is used when predictions are identified from multiple classes. The methods described in Sections 6.6.1 through 6.6.3 are useful for identifying classes that are not representative of the query compound or classes with insufficient data to fully support the regression equation. In our prallethrin example output (see Figures 16 and 20), available information from Class Equation and Definition documents could support a user's decision to exclude the esters, Vinyl/Allyl

Ketones and Vinyl/Allyl Esters predictions. Of the remaining classes, the pyrethroid class appears to be the most representative for the query compound and also results in the most conservative effect levels (see Figure 20).

**Figure 20:** Predicted Effect Levels for Classes Identified for Example Chemical.

ECOSAR Class	Organism	Duration	End Pt	Predicted mg/L (ppm)
Esters	: Fish	96-hr	LC50	0.835
Esters	: Fish	14-day	LC50	1.773
Esters	: Daphnid	48-hr	LC50	1.275
Esters	: Green Algae	96-hr	EC50	0.344
Esters	: Fish	32/33-d	ChU	0.034
Esters	: Daphnid	21-day	ChU	0.381
Esters	: Green Algae		ChU	0.225
Esters	: Fish (SW)	96-hr	LC50	1.065
Esters	: Mysid	96-hr	LC50	0.253
Esters	: Fish (SW)		ChU	0.256
Esters	: Mysid (SW)		ChU	0.053
Esters	: Earthworm	14-day	LC50	461.663 *
Vinyl/Allyl Ketones	: Fish	96-hr	LC50	1.941
Vinyl/Allyl Ketones	: Daphnid	48-hr	LC50	0.527
Vinyl/Allyl Ketones	: Green Algae	96-hr	EC50	0.551
Vinyl/Allyl Ketones	: Fish		ChU	0.171
Vinyl/Allyl Ketones	: Daphnid		ChU	0.148 †
Vinyl/Allyl Ketones	: Green Algae		ChU	0.599 †
Vinyl/Allyl Ketones	: Fish (SW)	96-hr	LC50	0.907
Vinyl/Allyl Ketones	: Mysid (SW)	96-hr	LC50	0.085
Vinyl/Allyl Ketones	: Fish (SW)		ChU	2.651
Vinyl/Allyl Ketones	: Mysid (SW)		ChU	0.003 †
Vinyl/Allyl Esters	: Fish	96-hr	LC50	0.722
Vinyl/Allyl Esters	: Daphnid	48-hr	LC50	2.087
Vinyl/Allyl Esters	: Green Algae	96-hr	EC50	0.438
Vinyl/Allyl Esters	: Fish		ChU	0.004 †
Vinyl/Allyl Esters	: Daphnid		ChU	0.026 †
Vinyl/Allyl Esters	: Green Algae		ChU	0.116
Pyrethroids	: Fish	96-hr	LC50	0.005
Pyrethroids	: Daphnid	48-hr	LC50	0.004
Pyrethroids	: Fish		ChU	0.000212
Pyrethroids	: Daphnid		ChU	9.49e-005
Pyrethroids	: Fish (SW)	96-hr	LC50	0.017
Pyrethroids	: Mysid	96-hr	LC50	3.7e-005
Pyrethroids	: Fish (SW)		ChU	0.002
Pyrethroids	: Mysid		ChU	3.73e-006
Neutral Organic SAR (Baseline Toxicity)	: Fish	96-hr	LC50	0.639
	: Daphnid	48-hr	LC50	0.584
	: Green Algae	96-hr	EC50	0.551
	: Fish		ChU	0.085
	: Daphnid		ChU	0.097
	: Green Algae		ChU	0.501

NOTE: \* = asterisk designates: Chemical may not be soluble enough to measure this predicted effect.

NOTE: † = exclamation designates: The toxicity value was estimated through application of acute-to-chronic ratios per methods outlined in the ECOSAR Technical Reference Manual posted on the ECOSAR webpage.

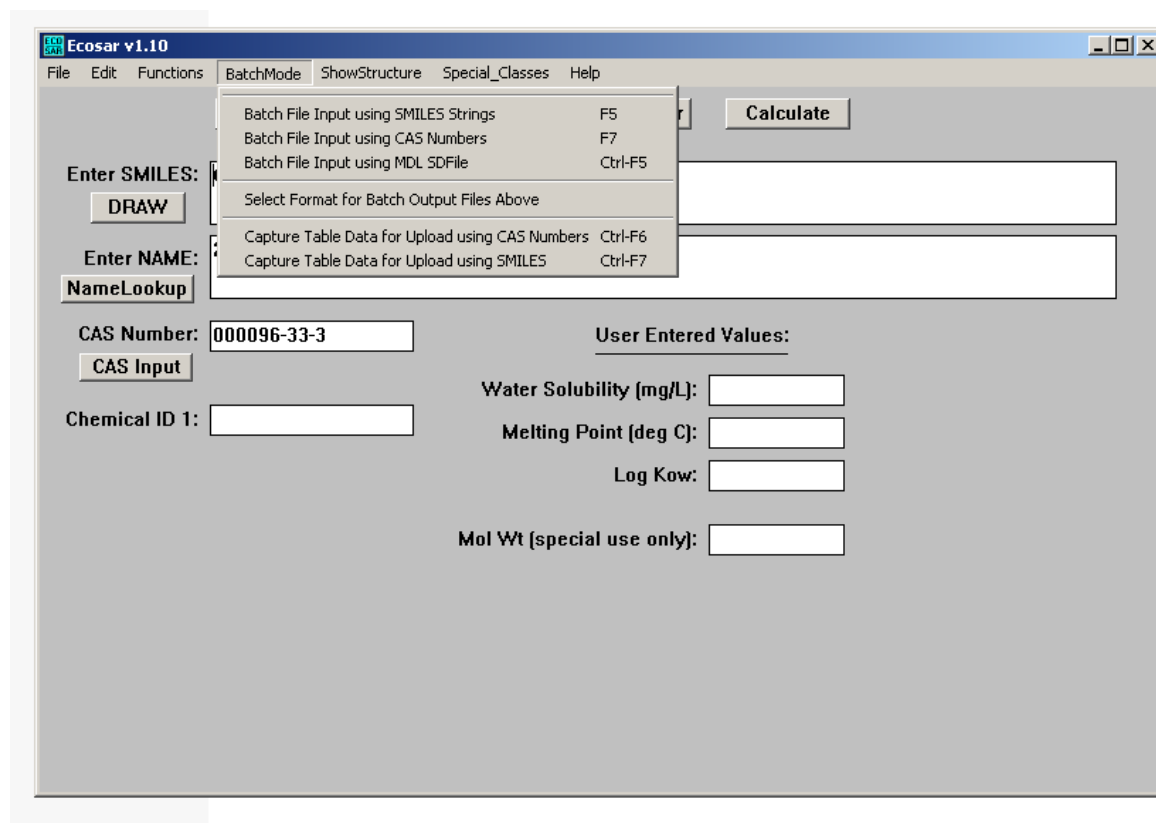
There is no one standard method for selecting the most representative predictions and, often, the best approach would be to select the most conservative effect level until measured data become available. The model developers emphasize that each predicted profile should be accompanied

with a discussion on the appropriateness of the estimates and a description of the identified uncertainties. The complexity of the discussion will vary depending on the expertise of the user and the significance of the chemical management decision being made.

## 7. Batch Runs

Batch runs are used to make multiple estimates from a single input file that contains multiple chemical identifiers. The ECOSAR Class Program can make "batch runs" from three different types of input files (SMILES Strings, CAS Numbers, and MDL SDFfiles) and can output the data in two different formats (standard text output offered in previous versions of ECOSAR and a text delimited table format). Each input file must be in a specific format, otherwise, the batch run will fail. Program access to "batch-runs" (depicted in Figure 21) is available from (a) the top menu option "BatchMode", (b) various options under the top menu option "Functions", and (c) the F5, F7, Ctrl-F5, Ctrl-F6, and Ctrl-F7 function keys.

**Figure 21:** Accessing Batch Mode





## 7.1. Batch Input Files

Batch run inputs recognized by the 'Batch File Input' and/or the 'Capture Table Data' methods for conducting batch runs include SMILES strings, CAS numbers, and MDL SD files. The configuration of these files is described below.

### 7.1.1. SMILES Strings

Batch runs can be carried out with two types of SMILES input files: (1) a SMILES string format list and (2) EcoWin format described in Section 5.1.5 as a method for single structure entry.

- (1) A SMILES string format list is a plain text file (usually with a ".txt" file extension) containing a list of SMILES notations. A "String Format" list must have the SMILES string at the beginning of each line in the file; it can then be followed by a space(s) and then the name or other ID. The SMILES string is considered terminated at the first space. An example String Format is as follows:

```
CCCCO Butanol  
c1ccccc1 Benzene  
Fc1ccccc1 Fluorobenzene  
CC(=O)C Acetone
```

- (2) An EcoWin format list is a plain text file (usually with an ".inp" file extension) containing a list of SMILES notations. EcoWin format is the same format used when creating the User Input files, which is designated as "SMILES.INP" (described in section 5.1.5). Therefore, the "SMILES.INP" file can be used directly to run batch file outputs. In this format, the name comes first (maximum of 60 characters) followed by a colon and one space, and then the SMILES notation. An example format is as follows:

Butanol: CCCCCO

Benzene: c1ccccc1

Fluorobenzene: c1ccccc1F

Acetone: CC(=O)C

### 7.1.2. CAS Number List

This is a plain text file (usually with a ".txt" file extension) containing a list of CAS (Chemical Abstract Service) Registry numbers. The format of the ascii text file is: no spaces in front of the CAS number, hyphens and leading zeros are optional, and a trailing carriage return. There is no limit to the number of CAS numbers in the file. An example format is as follows:

000050-00-0

71-43-2

108883

000050-02-2

### 7.1.3. MDL SD Files

SD Files (Structure Definition files) are text files containing chemical structures (stored as MOL files) and other data that can be uploaded and generated by various commercial chemistry programs such as ISIS/Base, ChemFinder and Accord for Excel. The following example is a section from an SD File:

-ISIS- 04010908242D

4 3 0 0 0 0 0 0 0 0999 V2000

2.4667 -0.0833 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0

2.4667 -0.9125 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0

1.7500 -1.3292 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0

```

3.1833 -1.3292 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 1 2 0 0 0 0
3 2 1 0 0 0 0
4 2 1 0 0 0 0
M END
> <CAS> (000050-00-0)
000050-00-0

> <NAME> (000050-00-0)
FORMALDEHYDE

> <Kow> (000050-00-0)
3.5000000000000000e-001

$$$$

```

The various fields are delimited with "<" and ">" brackets. This formaldehyde example includes <CAS>, <NAME>, and <Kow> fields. To extract ID (Name) for an ECOSAR batch run, in the SD File Option Box, enter the field name exactly as it appears between the brackets.

## 7.2. Conducting Batch Runs

To conduct a batch run, a user must determine the input format of the query compounds (e.g., SMILES strings, CAS Number List, or MDL SD Files) and decide how the results will be captured in the output. The format of the input query compounds is described in Section 7.1. Batch runs can capture results as (1) Full Output, (2) Summary Output, or (3) Table Data.

- (1) Full Outputs capture results for each compound the same as they would appear in the "Result Window" (if each compound was estimated individually); these output files can get very large for large numbers of compounds. Structures of query compounds are not depicted in the output.

- (2) Summary Output captures the ECOSAR class, CAS number, and query chemical name and places these results on a single line for each compound. Predicted effect levels are not included. The format of the output can be selected by the user before running a batch by selecting the ‘Select Format for Batch Output Files Above’ in the ‘BatchMode’ drop down menu. The output default is “space filled” with required identifiers to identify various results. Output can also be “Comma de-limited” or “Tab de-limited”. These output selections separate results on each line with either commas or tabs, respectively. This is useful for importing batch output file directly into other programs (such as Microsoft Excel™ or Lotus123™ spreadsheets).
- (3) Table Data Output captures both ECOSAR predicted effect levels and experimental data used in the ECOSAR class training data sets. Output varies slightly depending on whether data were input as SMILES string or CAS number. MDL SD files cannot be used as input files if the user wants a table output. Fields output when SMILES string are used as the input include (1) an ‘ID’ field that depicts a CAS number for measured data, (2) a ‘SMILES’ field that depicts the entered SMILES string or the name of available measured data, (3) an ‘ECOSAR Class’ field that lists the identified ECOSAR classes for input SMILES strings or states ‘Experimental’ if the data are measured, (4) an ‘Organism’ field that identifies the organism for the predicted and measured effects, (5) a ‘Duration’ field that identifies the exposure duration for predicted and measured effects, (6) an ‘End Pt’ field that identifies the endpoint (e.g., LC<sub>50</sub>, EC<sub>50</sub>, or ChV) for the effect level, (7) a ‘Predicted mg/L (PPM)’ field that reports predicted effect levels, (8) a ‘Flag’ field that identifies flag notation (e.g., an asterisk or exclamation mark as described at the end of Section 6.4) for predicted effect levels, and (9) a ‘Measured mg/L (PPM)’ field that reports experimental effect levels. Field outputs when CAS numbers are input include (1) a ‘CAS number’ field that identifies the CAS number, (2) a ‘Name’ field that identifies the name of the query compounds, (3) an ‘ECOSAR class’ field that identifies the ECOSAR class used for the prediction or states ‘experimental’ for measured data, (4) an ‘Organism’ field that identifies the organism for the predicted and measured effects,

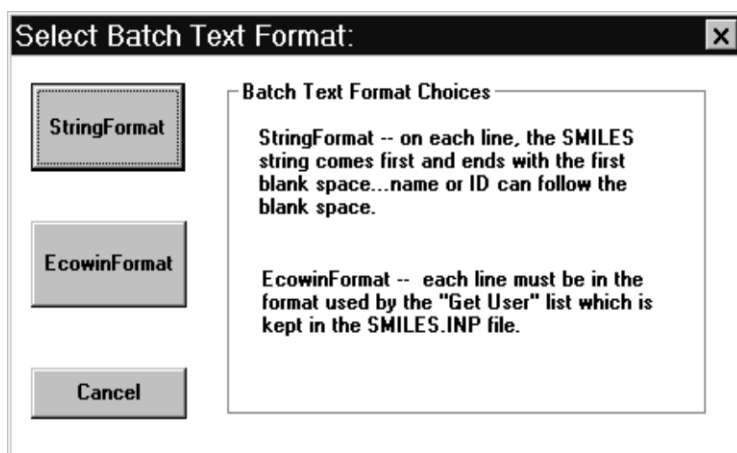
(5) a 'Duration' field that identifies the exposure duration for predicted and measured effects, (6) an 'End Pt' field that identifies the endpoint (e.g., LC<sub>50</sub>, EC<sub>50</sub>, or ChV) for the effect level, (7) a 'Predicted mg/L (PPM)' field that reports predicted effect levels, (8) a 'Flag' field that identifies flag notation (e.g., an asterisk or exclamation mark as described at the end of Section 6.4) for predicted effect levels, and (9) a 'Measured mg/L (PPM)' field that reports experimental effect levels.

### 7.2.1. Obtaining Full Outputs and Summary Outputs

Similar methods are used to conduct batch runs in order to obtain full outputs and summary outputs. If an error screen is displayed following a batch run, there may be a problem with the data input file.

For batch entry using SMILES string format, select 'Batch File Input Using SMILES Strings' from the 'BatchMode' drop down menu or press the F5 key. The user is given the option of using string format or Ecwin format (see Figure 22) to input SMILES structures, as described in Section 7.1.1. After selecting the input type, a file selection screen displays from which the user selects a file by highlighting the file name, then presses the 'OK' button, and then confirms the selection in the subsequent window by selecting 'YES'.

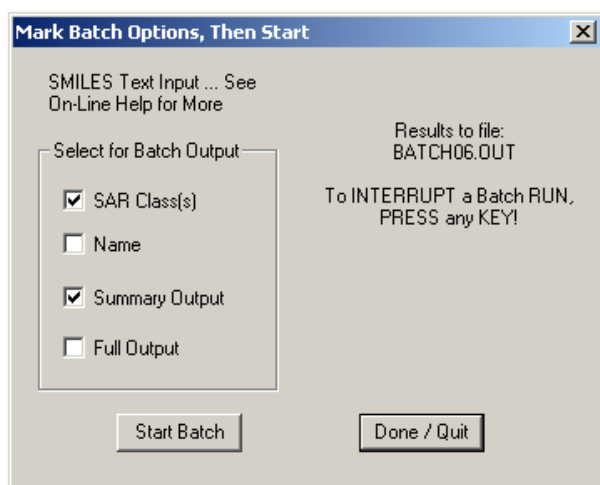
**Figure 22: Batch File input Screen for SMILES strings**



For batch entry using CAS number input files or MDL SD input files, select ‘Batch File Input Using CAS Number’ or ‘Batch File Input Using MDL SDfiles’ from the ‘BatchMode’ drop down menu or press the F7 key or Ctrl-F5 key, respectively. A file selection screen displays from which the user selects a file by highlighting the file name, then presses the ‘OK’ button, and then confirms the selection in the subsequent window by selecting ‘YES’.

Once an input file is selected, the user can then select whether the output will be a ‘Summary Output’ or a ‘Full Output’ from the screen depicted in Figure 23. The user also can select whether the ‘SAR Class(s)’ and the ‘Name’ will be included in the output file.

**Figure 23:** Batch Output Option Screen



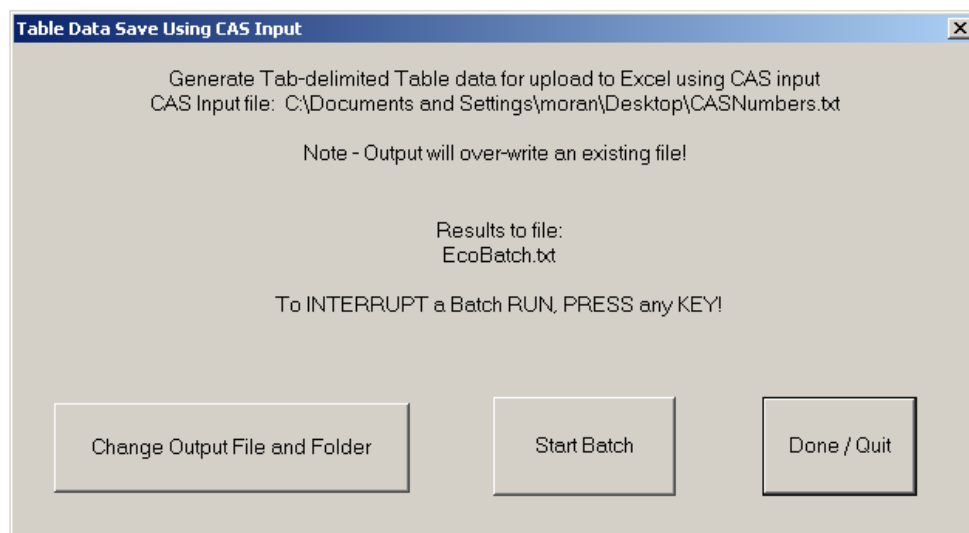
After selecting ‘Start Batch’, a file is saved into the same location as the executable ECOSAR files, and given a name that is identified under ‘Results to file’ (see Figure 21). The screen depicted in Figure 23 will not close after the batch run is completed; to close select ‘Done/Quit’.

### 7.2.2. Obtaining Table Data Outputs

As an alternative to the Full or Summary Output that is generated as described in Section 7.2.1, the ‘Capture Table Data’ selection from the ‘BatchMode’ drop down menu will generate results in a text-delimited table format that can be imported into programs such as Microsoft EXCEL.

To capture table data outputs from a batch run using CAS numbers, select 'Capture Table Data for Upload using CAS Numbers' in the 'BatchMode' drop down menu' or press the Ctrl-F6 key. The user will then be prompted to select an input file from a file window. Once an input file containing CAS numbers is selected, a screen (see Figure 24) will give the user an opportunity to change the output file and folder from the default (same location as the ECOSAR executable file) by selecting 'Change Output File and Folder'. 'Start Batch' can be selected when the user is ready to start the batch run. The screen depicted in Figure 24 will not close after the batch run is completed; to close select 'Done/Quit'.

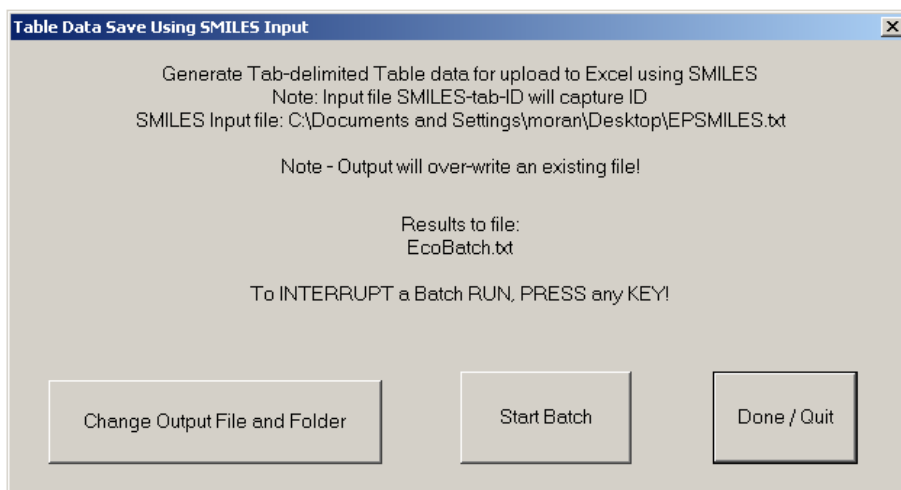
**Figure 24:** User Screen for Batch Runs using CAS Input with Table Output



To capture table data outputs from a batch run using SMILES strings, select 'Capture Table Data for Upload using SMILES' in the 'BatchMode' drop down menu or press the Ctrl-F7 key. The user will then be prompted to select an input file from a file window. Once an input file containing CAS numbers is selected, a screen (see Figure 25) will give the user an opportunity to change the output file and folder from the default (same location as the ECOSAR executable file) by selecting 'Change Output File and Folder'. 'Start Batch' (shown in Figure 24) can be selected when the user is ready to start the batch run. The screen depicted in Figure 25 will not close after the batch run is completed; to close select 'Done/Quit'.

**Figure 25:** User Screen for Batch Runs using SMILES Input with Table Output

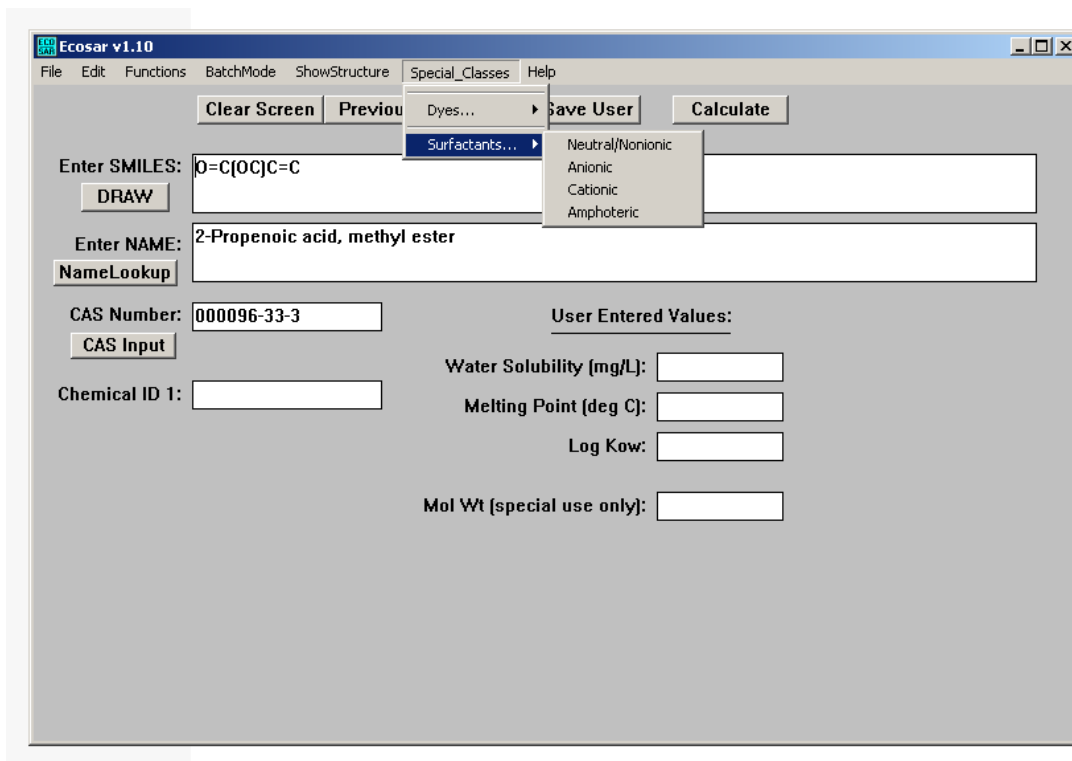




## 8. Special Class Calculations

The ECOSAR Class Program has been developed primarily for the following scenario: (1) enter a SMILES notation, (2) the ECOSAR program determines the appropriate ECOSAR class(es) from the SMILES notation, and (3) ECOSAR calculates the ecotoxicity QSARs using a log  $K_{ow}$  value. Several "Special Classes" of ECOSAR QSARs or classifications do not use the log  $K_{ow}$  value or can not be adequately classified from the SMILES notation. These "Special Classes" include Dyes and Surfactants. QSARs are available for various Anionic, Cationic, Nonionic, and Amphoteric Surfactants. However, instead of the log  $K_{ow}$  value, these surfactant QSARs may utilize the number of ethoxylate units or the average length of the carbon chain. Dye QSARs are also available for triphenylmethane dyes and ethoxylated triphenylmethane dyes and utilize ethoxylate units for QSAR predictions. These "Special Classes" are accessed from the Main Menu bar (see Figure 26).

**Figure 26: Special Classes in ECOSAR**



The Special Classes have their own data entry dialog box (see Figure 27).

**Figure 27. Example Entry Dialog Box for Surfactants**

The calculated results are placed in the same Results Windows as results using SMILES notations (an example is illustrated in Figure 28). **Note:** the Water Solubility or Water Dispersibility fields in the data entry dialogs are not used in QSAR calculations.

**Figure 28. Example Results Window for Surfactants**

**Ecosar Results**  
 Print Save Results TableSave Copy Remove Window Help

CHEM : S100  
 CAS Num:  
 ChemID1:  
 ChemID2:  
 ChemID3:  
 WatDisp: 0 mg/L

Avg Length Carbon Chain: 12.00

ECOSAR Class: Alkyl Benzene Sulfonates

Organism	Duration	End Pt	Predicted mg/L (ppm)
Fish	96-hr	LC50	2.595
Daphnid	48-hr	LC50	2.595
Green Algae	96-hr	EC50	0.007
Fish	28-day	NEC	0.399
Daphnid	21-day	NEC	0.399
Green Algae	21-day	NEC	0.005

Note: \* = asterisk designates: Chemical may not be soluble enough to measure this predicted effect.  
 Fish and daphnid acute toxicity log Kow cutoff: none  
 Green algal EC50 toxicity log Kow cutoff: none  
 Chronic toxicity log Kow cutoff: none  
 MW cutoff: none

## 9. Bibliography

Koneman, H. 1981. Fish toxicity tests with mixtures of more than two chemicals: a proposal for a quantitative approach and experimental results. *Toxicology* 19: 229-238.

Meylan, W.M. and P.H. Howard. 1994a. Upgrade of PCGEMS Water Solubility Estimation Method (May 1994 Draft). prepared for Robert S. Boethling, U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Washington, DC; prepared by Syracuse Research Corporation, Environmental Science Center, Syracuse, NY 13210.

Meylan, W.M. and P.H. Howard. 1994b. Validation of Water Solubility Estimation Methods Using Log  $K_{ow}$  for Application in PCGEMS & EPI (Sept 1994, Final Report). Prepared for Robert S. Boethling, U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Washington, DC; prepared by Syracuse Research Corporation, Environmental Science Center, Syracuse, NY 13210.

Meylan, W.M. and Howard, P.H. 1995. Atom/Fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci.* 84: 83-92.

Meylan, W.M. and Howard, P.H. 1996. Improved method for estimating water solubility from octanol/water partition coefficient. *Environ. Toxicol. Chem.* 15: 100-106.

Weininger, D. 1988. SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **28**: 31-36.

## Appendix A-Glossary of Terms and Abbreviations Associated with ECOSAR

ACR	Acute-To-Chronic Ratio
CAS	Chemical Abstracts Service
CBI	Confidential Business Information
ChV	Chronic Value
EPA	U.S. Environmental Protection Agency
F96	Fish 96-hr Endpoint
FChV	Fish Chronic Value
D48	Daphnid 48-hr Endpoint
DChV	Daphnid Chronic Value
EC <sub>50</sub>	Median Effect Concentration. A statistically derived concentration of a substance that can be expected to cause a specific effect (e.g., growth inhibition) in 50% of test animals. It is usually expressed as milligrams (mg) of substance per liter (L) water.
ECOSAR	ECOLOGical Structure Activity Relationship
FW	Freshwater
GA72/96	Green Algae 72-hr or 96-hr Endpoint
GACHV	Green Algae Chronic Value
GM	Geometric Mean
hr	Hour
K <sub>ow</sub>	Octanol-Water Partition Coefficient
LC <sub>50</sub>	Median Lethal Concentration. A statistically derived concentration of a substance that can be expected to cause death in 50% of test animals. It is usually expressed as milligrams (mg) of substance per liter (L) water.
Log	Logarithm; a base of 10 is used in ECOSAR calculations.
LOEC	Lowest Observed Effect Concentration. Lowest concentration of a substance that produced statistically significant effects.
MATC	Maximum Acceptable Toxicant Concentration. Value is equivalent to the ChV.
mg/L	Milligrams Per Liter

mmol/L	Millimoles Per Liter
n	The number of measured data points used to develop the QSAR equation sometimes followed by a '+' and then a number that represents the number of predicted data points and the number of measured data points above the limit of solubility.
NA	Not Applicable
N/A	Not Applicable
No	Number
NO	Neutral Organic
NOEC	No Observed Effect Concentration. Highest tested concentration of a substance that produced no statistically significant effects.
PPM	Parts Per Million
QSAR	Quantitative Structure Activity Relationships
r	Correlation Coefficient
r <sup>2</sup>	Coefficient of Determination
SMILES	Simplified Molecular Input Line Entry System
SW	Saltwater
TSCA	Toxic Substance Control Act

## Appendix B-Summary of Function Keys

Provided below is a summary of function keys that correspond to buttons and/or functions accessible from the ECOSAR data entry screen. These function keys can be used as alternatives to the buttons, which may require a mouse/touch pad to work.

**F1:** Pressing the F1 key accesses a help message for the individual field where the blinking cursor is located. General Help is available from "Help" on the Menu Bar at the top of the screen. It is a standard Windows help system; to access a specific help topic, simply click on the topic (or keyword) that is highlighted in green where the mouse pointer changes to a hand.

**F2:** Pressing the F2 key or clicking the "Previous" button recalls the most recent SMILES and chemical name that was calculated or attempted to be calculated by the program. It can save a lot of time when making small changes to large SMILES and names. It is especially useful after a SMILES notation error occurs, then the incorrect SMILES can be recalled and edited.

**F3:** Pressing F3 clears the currently displayed SMILES Notation, Chemical Name and other data. All entry fields are filled with blank spaces.

**F4:** Pressing the F4 key or clicking the "Get User" button displays a file selection dialog box that allows the user to open a file of previously saved SMILES notations and chemical names. The default name of the file is SMILES.INP; this is for compatibility with similar programs. The file selection box looks for files with the extension ".INP", so it is best to name files with this extension when creating them with the F6 key ("Save User"). A "Get User" file can contain up to 1500 SMILES and names and the user can select any single SMILES and name for input. The SMILES.INP file can be created one chemical at a time by using the F6 key as described below.

**F5:** Pressing the F5 key (or clicking the "BatchMode" option on the main menu and selecting "Batch File Input Using SMILES Strings") brings up the selection box in Figure 22. This

method for batch entry uses SMILES strings from ascii text files to create output screens (non-tabular) similar to the output screens produced from individual ECOSAR runs.

**F6:** Pressing the F6 key or clicking the "Save User" button displays a file selection dialog box that allows the user to save the SMILES notation and chemical name currently showing on the data entry screen to the file. The default name of the file is SMILES.INP; this is for compatibility with similar estimation programs. After a file is selected (or entered by the user), ECOSAR appends the SMILES notation and chemical name currently showing on the data entry screen to the file. If the file does not already exist, ECOSAR will create it and append the current SMILES and name as the first entry. The SMILES and names in a "Saver User" file can be accessed from the data input screen by pressing the F4 key.

**F7:** The F7 key is used to enter CAS numbers from an ascii text file. The number of CAS numbers in the file is not limited. The user must enter the file name and an election menu is not currently available. The F7 key is used primarily for batch-mode runs and the non-tabular output is written to files named "CASLOG#.OUT" where "#" is a number determined by the program.

The format of the ascii text file is: no spaces in front of the CAS number, hyphens and leading zeros are optional, and a trailing carriage return.

Example:

000050-00-0

71-43-2

108883

000050-02-2

**F8:** Pressing the F8 key or clicking the "CAS Input" button requires the presence of a supplemental database file (SMILECAS.DB) and index file in the current subdirectory. A small data entry window is created on the data entry screen which asks for the CAS number of the chemical. An error message will appear in the window if the program can not find the database



or index file. The database file contains over 103,000 entries, but not all chemicals with CAS numbers are included in the file. If the chemical is not in the database, an appropriate message is displayed. The program can identify impossible CAS numbers by examining the check digit (the final number of the CAS).

**F11:** Pressing the F11 or clicking the 'DRAW' button executes the draw tool (see Section 5.1.4).

**PgDn:** Pressing The PgDn key or clicking the "Calculate" button calculates the SMILES currently showing on the data entry screen. If an acceptable SMILES has been entered, the Results Window will either appear or be updated. If an incorrect SMILES has been entered, an error message box will appear. After removing the error message box, the incorrect SMILES can be recalled and then edited by pressing the F2 key or clicking the "Previous" button.

**Esc:** During data entry, pressing the Esc key exits the program. When the Results Window is active, pressing the Esc key removes the Results Window.

**Enter:** Pressing the Enter (Return) key sends the cursor to the next data entry field.

**Tab** or **Shift-Tab:** changes entry fields.

**Ctrl-F5** – MDL SD File Batch Input

**Ctrl-F6** - batch with CAS

**Ctrl-F7** - batch with SMILES

**Ctrl-F10** or **Alt-** Selects the menu bar at the top. Type in 'F' for file, 'E' for Edit, 'U' for Function, 'B' for Batchmode, 'S' for Showstructure or Special\_Class, and 'H' for Help.

## **Appendix C- Files Associated With ECOSAR**

### **DrawStructure\_Help.exe**

**ECOSAR\_Help\_Files:** Folder that Contains files associated with the 'Help' menu including QSAR Equation Sheets and Definition Sheets.

**ECOS\_EXP.DB:** A database of experimental training set data used in ECOSAR.

**ECOS\_EXP.IDX:** The structure index file for ECOS\_EXP.DB.

**ECOWINNT.EXE:** The necessary ECOSAR executable file

**EXPKOW.DB:** A database file of more than 13200 experimental log Kow values; WSKOWWIN searches this database for each structure entered to see if there are experimental values. If there are experimental values (and the user has not entered a log Kow value), WSKOWWIN estimates the water solubility using the experimental log Kow value.

**EXPKOW.IDX:** The structure index file for EXPKOW.DB

**EXPWSOL.DB:** A database file of more than 6230 experimental water solubility values; WSKOWWIN searches this database for each structure entered to see if there are experimental values. If there are experimental values (and the user has not entered a measured water solubility value), ECOSAR uses the experimental water solubility.

**EXPWSOL.IDX:** The structure index files for EXPWSOL.DB

**ECOWHELP\_Help.EXE:** An on-line help file located in the ECOSAR\_Help\_Files folder containing information about the ECOSAR program and the QSARs used by ECOSAR.

**KOWHELP\_Help.EXE:** A help file located in the ECOSAR\_Help\_Files folder containing information about the KOWWIN Program its log Kow estimations which are used by ECOSAR.

**NAMESEPI.DB:** The NAMESEPI data base contains a list of compounds organized alphabetically by name with corresponding SMILES Notations. Although the number of entries is large, various chemicals that may be of interest may not be included in the data base.

**NAMESEPI.IDX:** the structure index file for NAMESEPI.DB

### **OnLineHelp\_Word\_RTF\_Manual\_Version**

**SMILHELP\_Help.EXE:** A SMILES Notation help file located in the ECOSAR\_Help\_Files folder containing an on-line version of the document "A Brief Description of SMILES Notation".

**SMICONV.DLL:** A dynamic link library that converts scientific structure formats (MDL MOL files) to SMILES notation and vice versa, and is also used to display chemical structures in the Structure Window. ECOSAR will not run without it.

**SMILECAS.DB:** The SMILECAS data base contains more than 104,000 CAS numbers with corresponding SMILES Notations. Although the number of entries is large, various chemicals that may be of interest may not be included in the data base.

**SMILECAS.IDX:** The structure index file for SMILECAS.DB

**TwoDEdit.DLL:** The dynamic-link-library (DLL) module of the Drawing Program (two-dimensional structure editor).

**WCT32D.DLL:** A commercial dynamic link library used to display the chemical structure. ECOSAR will not run without it.

**WSKOWDLL.DLL:** A dynamic link library of the WSKOWWIN Program that estimates log Kow values, water solubilities and retrieves available experimental water solubilities. ECOSAR will not run without it. The log Kow estimation methodology comes from the KOWWIN Program (described in a journal article (Meylan and Howard, 1995). The WSKOWWIN estimation methodology for water solubility is described in a journal article Meylan et al., 1996.

**WSKOWWIN\_Help.EXE:** A help file containing information about the WSKOWWIN Program