

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 2.2

Contributors:

R. Tracy Wright^a, Kellie Fay^a, Amuel Kennedy^a, Kelly Mayo-Bean^b, Kendra Moran-Bruce^b,
William Meylan^c, Peter Ranslow^c, Michele Lock^c, J. Vince Nabholz^b, Justine Von Runnen^c,
Lauren M. Cassidy^c, Jay Tunkel^d

^aOffice of Pollution Prevention and Toxics
U.S. Environmental Protection Agency
1200 Pennsylvania Ave. N.W.
Washington, DC 20460

^bFormerly at U.S. EPA

^cSRC, Inc.
7502 Round Pond Rd
North Syracuse, New York 13212

^dFormerly at SRC, Inc.

^eConsortium for Environmental Risk Management, LLC
Evansville, IN 47708

February 2022

DISCLAIMER

This document has been reviewed and approved for publication by the Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency (U.S. 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 ECOlogical Structure-Activity Relationship (ECOSAR) model and underlying methodology presented in this document have been developed over a period of more than 25 years by U.S. EPA/OPPT, EPA contractors, and/or others in the scientific and technical community to screen chemicals in the absence of data. U.S. 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 assessment efforts. Models provide estimations with an inherent degree of uncertainty and therefore, valid measured data are always preferred over estimated data. If no measured or analogue data are available, 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 OF CONTENTS

1. Introduction.....	1
2. Computer-Software Requirements.....	2
3. Installing the ECOSAR Class Program	3
4. Starting the ECOSAR Class Program.....	3
5. User Inputs For Generation Of ECOSAR Predictions.....	6
5.1. Structure Entry	7
5.1.1. Direct Entry.....	8
5.1.2. CAS Number Entry.....	9
5.1.3. Chemical Name Entry.....	9
5.1.4. Entry Using The Drawing Tool	10
5.1.5. Importing Structures	11
6. Results Window	12
6.1. Output Tabs.....	16
6.2. Chemical Attributes	17
6.2.1. Entering Measured Data for Variables	17
6.2.2. Log K_{ow}	18
6.2.3. Water Solubility.....	18
6.2.4. Melting Point	19
6.3. Structure Window	19
6.4. Experimental Data Tab	19
6.5. Organic Module Result Tab.....	20
6.6. Comprehensive Approach for Determining the Most Representative QSAR Class.....	21
6.6.1. General Classes vs. Sub-Classes.....	23
6.6.2. Correlation of K_{ow} with Toxicity in QSAR Training Data Sets	24
6.6.3. Robustness and Distribution of Supporting Datasets for QSAR Classes	26
6.6.4. Selecting the Most Representative Class	27
7. Batch Runs	28
7.1. Batch Input Files	29
7.1.1. SMILES, CAS, and Chemical Name Strings.....	29
7.1.2. MDL SD Files.....	30
7.2. Conducting Batch Runs	31
7.2.1. Obtaining Individual Report Outputs.....	32
7.2.2. Obtaining Table Data Outputs	33
8. Special Class Calculations	33
9. Bibliography	35

Appendix A. Glossary of Terms and Abbreviations Associated with ECOSAR..... 36

Appendix B. Summary of Function Keys 38

1. Introduction

The ECOlogical Structure-Activity Relationship Model (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 (EPA) since 1981 to predict the aquatic toxicity of new industrial chemicals in the absence of test data. ECOSAR 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 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 over 111 organic chemical classes, along with additional QSARs for surfactants, polymers, 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 a QSARs folder within the ‘Helpful files’ bundled with the executable. For more information on QSAR development, see the ECOSAR Methodology Document within the Helpful files downloaded with the executable.

The ECOSAR program is designed for the skilled 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 the use of ECOSAR to predict aquatic toxicity is more suitable than an analogue approach. 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 Contact

R. Tracy Wright

U.S. Environmental Protection Agency

Office of Chemical Safety and Pollution Prevention (6334W)

1200 Pennsylvania Avenue, NW

Washington, DC 20460-0001

Phone: 202-564-9923

E-mail: Wright.Tracy@epa.gov

Model Development Contact

Bill Meylan

Environmental Health Analysis

SRC, Inc.

7502 Round Pond Road

North Syracuse, NY 13212-2558

Phone: 315-452-8421

E-mail: meylan@srcinc.com

2. Computer-Software Requirements




The ECOSAR Class Program is designed for use on PC and Mac devices running Microsoft Windows (Windows 7, and higher), iOS, or OS X. The program may also work for various mobile and Unix platforms; however, it was not specifically optimized for those environments. ECOSAR is 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 can be run in the background until they are completed.

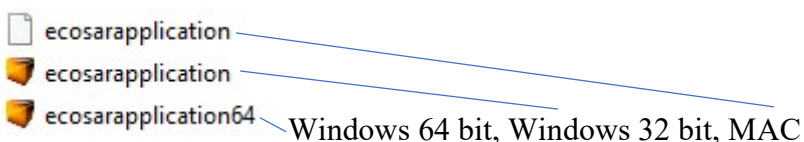
Although a mouse/touch-pad or other pointing device¹ is not required, it is highly recommended. If a mouse/touch-pad is not available, knowledge of the following keys is necessary: **Enter:**

¹ Program developers have not optimized this interface for touch-screens to act as another type of pointing device and, thus, do not know whether the ECOSAR program will function properly on tablet-type devices.




pressing the Enter (Return) key will cause the program to run if sufficient data are entered into the Chemical Input field; and **Tab** or **Shift-Tab**: changes entry fields (see Appendix B, Summary of Function Keys). ECOSAR v2.2 requires approximately 176 MB of hard disk space for the unzipped package; the executable takes up 1,458, 474, and 3 KB for the 64-bit, 32-bit, and Mac compatible executables, respectively. Included in the package is the Estimation Programs Interface (EPI) EPI_Unified Database (a database of over 110,000 Simplified Molecular Input Line Entry System [SMILES] notations indexed by Chemical Abstracts Service [CAS] number for program retrieval), which requires 84,655 KB of disk space.

3. Installing the ECOSAR Class Program

Users can download the ECOSAR Class Program for free from the EPA's website at: <https://www.epa.gov/tsca-screening-tools/ecological-structure-activity-relationships-ecosar-predictive-model>. ECOSAR is a self-extracting file. Once it is unzipped to the C: drive, desktop, or wherever the user wants to install the program, the user can access the executable one of a few ways. The most direct route is to open the ECOSAR_v2.2 file, where it was unzipped to, and click the "ecosarapplication" folder. From there, select the "ecosar.bat" file , the ECOSAR_v2.2_32 shortcut , or the ECOSAR_v2.2_64 shortcut . If the user would prefer to assign a shortcut of their own, they can go into the "bin" folder and select the "ecosarapplication" file compatible with their system (see below). A 'Helpful' file containing supporting documentation is also bundled with the zip file.



4. Starting the ECOSAR Class Program

The ECOSAR Class Program is started by clicking the "ecosar.bat" file , the ECOSAR_v2.2_32 shortcut , the ECOSAR_v2.2_64 shortcut , or any other shortcut that they have established. The user can also run this program from the "bin" folder by selecting the "ecosarapplication" file compatible with their system, as noted above.

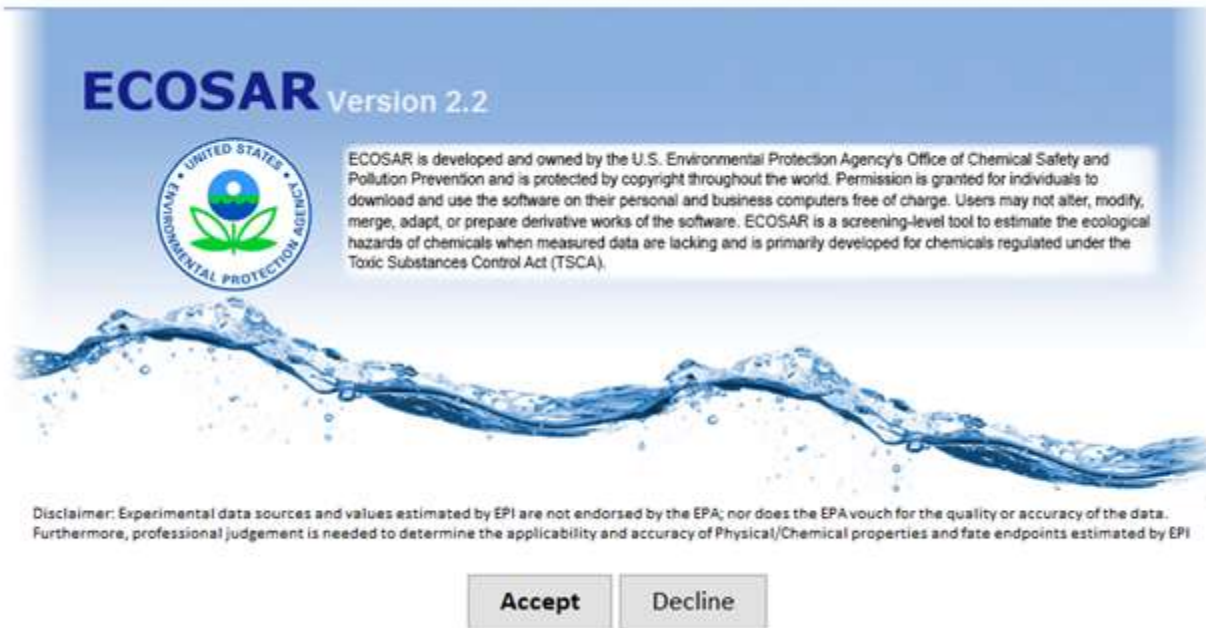
- ecosarapplication
- ecosarapplication
- ecosarapplication64

ECOSAR can be accessed through the U.S. EPA EPI Suite; however, the version associated with EPI Suite may not be the most recent version of ECOSAR. For additional information on starting Windows programs, consult your Windows documentation.

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



This will be replaced by the introduction screen where the user is asked to accept or decline the terms of service and the disclaimer.



Users should note the Disclaimer on the introduction screen regarding the need for professional judgement in determining the applicability and accuracy of the ecotoxicity endpoint values estimated by this program. Furthermore, inorganic and organometallic chemical classes (among others as noted in the Help file) are currently outside the domain of ECOSAR v2.2. Clicking the 'Decline' button will cause the program to close.

After clicking 'Accept' on the introduction screen, the data entry screen for the Organic Module is displayed (see Figure 1). The user may enter a CAS number, SMILES, or chemical name into the Chemical Input field to generate an ECOSAR prediction based on organic QSARs (described in Section 6). The user may also choose to draw a structure by clicking the 'Draw' button (described in Section 5.1.4, Entry Using the Drawing Tool) or enter multiple chemicals via the batch mode by clicking the 'Batch' button (described in Section 7, Batch Runs).

Certain user inputs, described in Section 6, can be entered, if needed, in the output tab for the generated prediction. QSAR predictions for surfactants, dyes, and polymers are generated from a separate data entry screen, which is described in Section 8. Supporting documentation for ECOSAR can be obtained from the Helpful files index and includes QSAR class definition and

equation sheets, the ECOSAR v2.2 User's Guide, the ECOSAR Methodology Document, and tutorials for using the draw tool and formulating SMILES structures.

Figure 1. Example Data Entry Screen

The screenshot shows the ECOSAR v2.2 Organic Module interface. At the top, the window title is "Ecosar Application 2.2". Below the title bar, there are two tabs: "ECOSAR" and "Special Cases". The main area has a sidebar on the left with "Organic Module" and "Organic" options. The "Organic Module" is selected, and the "Chemical Input" section is active. This section contains a large text input field with the placeholder text "Please enter CAS Number or SMILES". To the right of this field are two buttons: "Draw" and "Submit". Below the input field is a "User Entry Fields" section with a "Batch" button and four input fields: "CAS Number" (with example text "50-00-0, 000050-00-0, 50000"), "SMILES" (with example text "O=C"), "Log Kow", "Water Solubility (mg/L)", and "Melting Point (°C)".

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 user is required to enter one of the three subsequently described data types into the Chemical Input field on the main data entry screen or draw a structure for a single chemical estimation. Data types accepted:

- (1) **Enter SMILES:** Field for SMILES notation of the structure to be estimated. A maximum of 360 characters is allowed.

- (2) **Enter Chemical Name:** Field for the name and/or description of the structure. A maximum of 120 characters is allowed.
- (3) **Enter CAS Number:** The CAS number.
- (4) **User Entered Log Kow:** The octanol/water partition coefficient entered by the user.
- (5) **User Entered Water Solubility:** The water solubility in mg/L entered by the user.
- (6) **User Entered Melting Point:** The melting point in °C entered by the user.

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 Helpful files (SMILES_Help folder bundled in the zip file), <http://www.daylight.com> (Daylight Chemical Information Systems), or <https://www.epa.gov/sites/default/files/2015-05/documents/appendf.pdf>. An online SMILES translator is available from the National Cancer Institute (NCI) at <https://cactus.nci.nih.gov/translate/>.

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

- (1) Direct entry of the SMILES notation by the user from the keyboard (see Section 5.1.1 for direct entry of special cases such as salts).
- (2) Use of the structure drawing tool to convert a structure to SMILES notation and transfer for entry (see Section 5.1.4).
- (3) Import of structures in MDL MOL file formats through the drawing tool to convert to SMILES notation and transfer for entry (see Section 5.1.5).

The program can usually generate estimates for only one chemical at a time from this main screen and separate data entry is required for each chemical. If a chemical entry (name, SMILES) matches

more than one entry, the user will be prompted to select from a menu of available chemical matches. It is possible to select more than one chemical from the disambiguation list, which will result in a multiple chemical analysis (see Figure 3 in Section 5.1.3).

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 "Submit" button or hitting the "Enter" key.

5.1.1. Direct Entry

Direct entry requires knowledge of chemical structure and/or SMILES notation, as described above. For direct entry, a SMILES notation should be typed in the 'Chemical Input' 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 surrounded with brackets (e.g., Na should be entered with brackets to be [Na]). ECOSAR will not run and will output an alert if a SMILES is entered incorrectly. If more than one salt form exists in the database, the disambiguation list will prompt the user to specify.
- (2) **Organic ammonium salts:** Like the organic anionic salts described above, the conjugate salt (e.g., chloride ion) of organic ammonium salts can be directly entered. They can also be entered as amines, because 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 (i.e., four carbons are bound to the nitrogen) are an exception since reduction (e.g., loss of hydrogens) at neutral pH is unlikely; however, these compounds are generally assessed as cationic 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. All metals should be bracketed in order to be considered correctly entered (e.g., [Zn], [Fe], [Ca]). Non-carbon elements considered in organic QSARs include oxygen, phosphate, sulfur, nitrogen, silicon, fluorine, bromine, chlorine, iodine, and hydrogen. Direct hydrogen entry in a SMILES notation is unnecessary for ECOSAR v2.2.
- (4) **Charged chemicals:** Charged species (e.g., [+] and [-] signs) can be entered directly into a SMILES notation for charged chemicals, but must also be surrounded by brackets. Azido compounds (commonly written as: $N^+=N^-$) must be written as '[N+]=[N-]' or can be depicted as 'N#N' for the purposes of SMILES notation. Nitro compounds (commonly written as $N^+(O^-)=O$) are written as '[N+][O-]=O', as 'N(=O)=O', or 'T' for the purposes of SMILES notation.

5.1.2. CAS Number Entry

In previous versions, CAS numbers were retrieved through the “CAS Number Database” by selecting the ‘CAS Input’ button beneath the CAS number field on the data entry screen. This is not necessary in ECOSAR v2.2. SMILES, CAS number, or Chemical name can be entered into the ‘Chemical Input’ entry field. If the CAS number is incorrect or invalid, an alert will appear below the “Chemical Input” field (see Figure 2). The program accepts CAS numbers with or without hyphens. The user should check that the structure output matches the entered CAS number.

Figure 2. Entry Screen Error

The screenshot shows the 'Chemical Input' interface. At the top, there is a large text input field containing the placeholder text 'Please enter CAS Number or SMILES'. To the right of this field are two buttons: 'Draw' and 'Submit'. Below the input field, a red error message reads: 'Input contains an invalid character or it wasn't found in the database.' Underneath the error message, the text 'User Entry Fields:' is displayed. Below this, there are five input fields: 'CAS Number' (containing '50-00-0, 000050-00-0, 50000'), 'SMILES' (containing 'O=C'), 'Log Kow', 'Water Solubility (mg/L)', and 'Melting Point (°C)'. To the right of these fields is a 'Batch' button.

5.1.3. Chemical Name Entry

In previous versions, SMILES notations were retrieved from the Name Look-Up Database. This is not necessary in ECOSAR v2.2. SMILES, CAS number, or Chemical name are entered into the ‘Chemical Input’ entry field. If the Chemical name is incorrect or invalid, an alert will appear below the “Chemical Input” field (see Figure 2). The program accepts chemical names regardless of capitalization. If more than one entry matches the chemical name, the disambiguation list will appear (see Figure 3). The user should check that the structure matches the chemical name entered.

Figure 3. Disambiguation List

Multiple Entries Match the Entered SMILES Notation

c1ccccc1

Please select a chemical by clicking its checkbox. Multiple selections will provide multiple estimates.

<input type="checkbox"/>	71432 Benzene	<chem>c(cccc1)c1</chem>
<input type="checkbox"/>	591515 Phenyl lithium	<chem>c(ccc1[Li])cc1</chem>
<input type="checkbox"/>	1076433 Benzene-d6-	<chem>c(cccc1)c1</chem>
<input type="checkbox"/>	8007452 Tar, coal	<chem>c1ccccc1</chem>

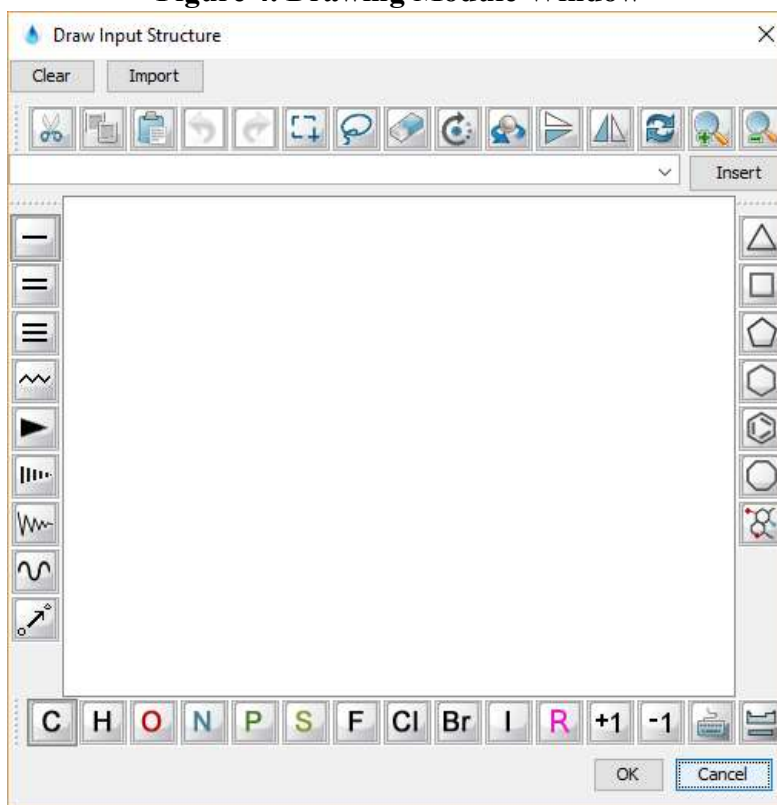
Submit

Cancel

5.1.4. Entry Using the Drawing Tool

As in version 2.0, ECOSAR v2.2 leverages the third-party package, jchempaint, for its 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. The Blank Drawing Module Window is shown in Figure 4.

Figure 4. Drawing Module Window

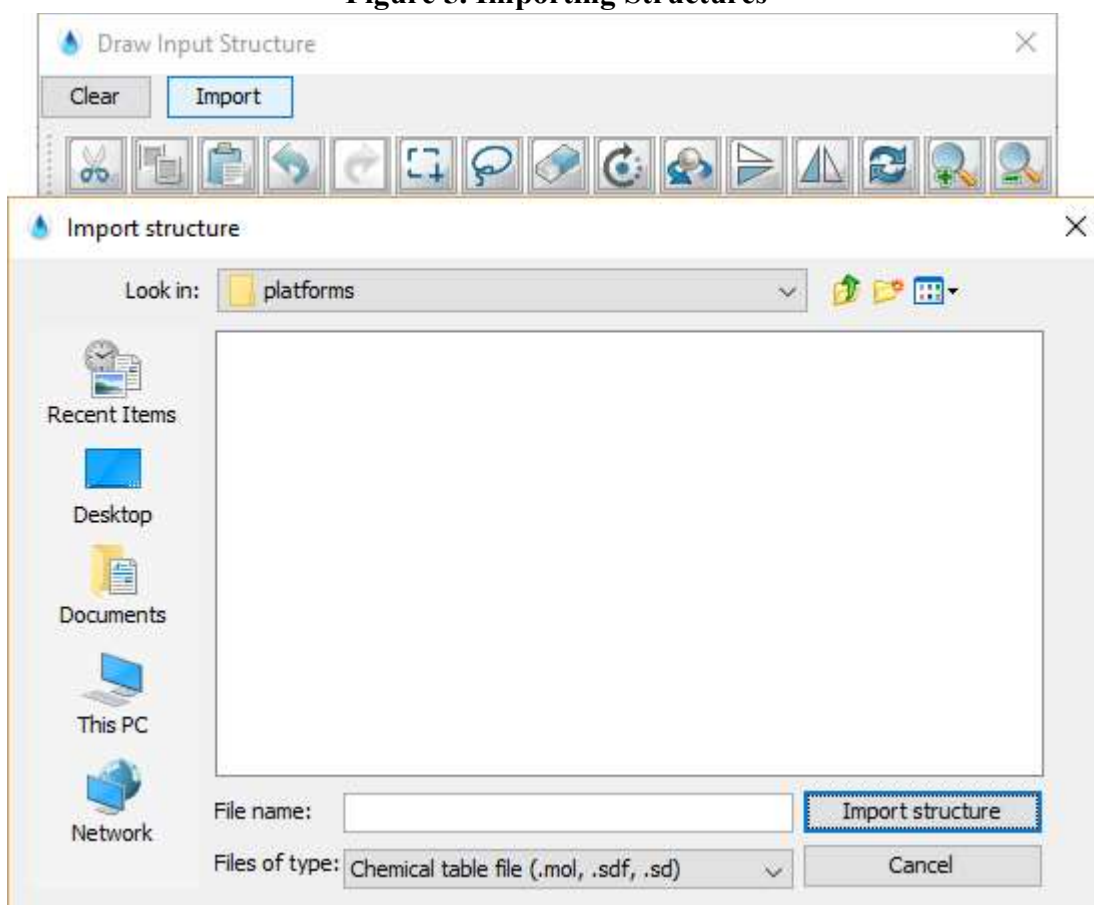


A tutorial is provided in the Helpful files folder for drawing structures with the Drawing Module (select Draw Structure Help). Once a structure is entered, the 'OK' button should be selected. The corresponding SMILES structure of the depicted chemical will be transferred to the 'Chemical Input' field on 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. Importing Structures

ECOSAR v2.2 has an "import" feature that allows MDL MOL file formats to be imported directly into ECOSAR. The "import" feature is accessed from the 'Drawing Module Window' via the 'Import' button shown in the upper left of Figures 4 and 5.

Figure 5. Importing Structures



Imported structures are converted to SMILES notations and placed in the 'Chemical Input' entry field. ECOSAR filters the conversion to make the SMILES 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.

6. Results Window

Generally, the Results Window (see Figures 6 and 7) provides the chemical structure (clicking on the small picture of the chemical structure opens a separate window with an Expanded Structure View), chemical attributes, measured training set data used in QSAR development for the query chemical, results of ECOSAR Class Program's estimations, and information specific to the interpretation of the QSAR results. Each of these is described in subsequent sections.

Figure 6. Results Window Panels

permethrin x

Chemical Name
permethrin

CAS
52645531

Log Kow
7.4267

Water Solubility (mg/L)
0.006

Melting Point (°C)
15.5

Chemical Details

SMILES
CC1(C)C(C=C(Cl)Cl)C1C(=O)OCc3ccccc(Oc

MOL WT
391.3

Log Kow
7.4267 (estimated)
6.5 (measured)

Water Solubility (mg/L)
0.042485 (estimated)
0.006 (measured)

Organic Module Result | Experimental Data | Physical Properties | **K_{ow} Estimate** | Report

Esters

Organism	Duration	End Point	Concentration (...)	Max Log Kow	Flags
Fish	96h	LC50	0.035	5.0	⚠
Daphnid	48h	LC50	0.041	5.0	⚠
Green Algae	96h	EC50	0.0074	6.4	⚠
Fish		ChV	0.00085	8.0	
Daphnid		ChV	0.0059	8.0	
Green Algae		ChV	0.011	8.0	⚠
Fish (SW)	96h	LC50	0.039	5.0	⚠
Mysid	96h	LC50	0.0026	5.0	
Fish (SW)		ChV	0.015	8.0	⚠

Vinyl/Allyl/Propargyl Halides

Organism	Duration	End Point	Concentration (...)	Max Log Kow	Flags
Fish	96h	LC50	0.00088	6.0	
Daphnid	48h	LC50	0.0010	6.0	
Green Algae	96h	EC50	0.0070	6.4	⚠
Fish		ChV	0.0000096	8.0	
Daphnid		ChV	0.0063	8.0	⚠
Green Algae		ChV	0.018	8.0	⚠
Fish (SW)	96h	LC50	0.00022	5.0	
Mysid (SW)	96h	LC50	0.00014	6.0	
Earthworm	14d	LC50	208	6.0	⚠

Pyrethroids

Organism	Duration	End Point	Concentration (...)	Max Log Kow	Flags
Fish	96h	LC50	0.00035	8.2	ⓘ
Daphnid	48h	LC50	0.00022	7.5	ⓘ
Fish		ChV	0.000017	8.0	
Daphnid		ChV	0.000045	8.0	


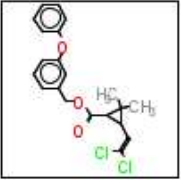



Organic Module Result | Experimental Data | Physical Properties | **K_{ow} Estimate** | Report

Organism	Duration	End Point	Concentration (mg/L)	Reference
Daphnid	48h	LC50	4.0E-5	OPP Pesticides Ecotoxicit...
Daphnid		ChV	6.0E-5	OPP Pesticides Ecotoxicit...
Mysid	96h	LC50	7.0E-5	OPP Pesticides Ecotoxicit...
Daphnid	48h	LC50	6.0E-4	OPP Pesticides Ecotoxicit...
Fish	96h	LC50	0.00252	OPP Pesticides Ecotoxicit...
Fish	96h	LC50	0.0061	OPP Pesticides Ecotoxicit...
Daphnid	48h	LC50	0.0072	OPP Pesticides Ecotoxicit...
Fish	96h	LC50	0.0098	OPP Pesticides Ecotoxicit...
Fish	96h	LC50	0.01	OPP Pesticides Ecotoxicit...
Green Algae (SW)	72/96h	EC50	0.06	Walsh et al., 1980; EC50...
Green Algae (SW)	72/96h	EC50	0.09	OPP Pesticides Ecotoxicit...

Organic Module Result	Experimental Data	Physical Properties	K _{ow} Estimate	Report
Water Solubility				
Water Solubility (mg/L)	6.00E-03			
Water Solubility Temperature (C)	20			
Water Solubility Type	EXP			
Water Solubility References	USDA PESTICIDE PROPERTIES DATABASE			
Log Kow				
Log Kow	6.50			
Log Kow Temperature (C)				
Log Kow Type	EXP			
Log Kow References	HANSCH,C ET AL. (1995)			
Vapor Pressure				
Vapor Pressure (mmHg)	2.18E-08			
Vapor Pressure Temperature (C)	25			
Vapor Pressure Type	EXP			
Vapor Pressure References	USDA PESTICIDE PROPERTIES DATABASE			
pKa				
pKa				
pKa Temperature (C)				
pKa Type				
pKa References				
Henry's Law Constant				
Henry Law Constant ((atm*m ³)/mol)	1.87E-06			
Henry Law Constant Temperature (C)	25			
Henry Law Constant Type	EST			
Henry Law Constant References	VP/WSOL			

Organic Module Result	Experimental Data	Physical Properties	K _{ow} Estimate	Report
Value	Value Type	Number	Coefficient	Total Coefficient
Fragment	-CH3 [aliphatic car...	2	0.547	1.095
Fragment	-CH2- [aliphatic car...	1	0.491	0.491
Fragment	-CH [aliphatic carb...	2	0.361	0.723
Fragment	=CH- or =C< [olefinc ...	2	0.384	0.767
Fragment	-CL [chlorine, olefi...	2	0.492	0.985
Fragment	Aromatic Carbon ...	12	0.294	3.528
Fragment	-O- [aliphatic O, tw...	1	0.292	0.292
Fragment	-C(=O)O [ester, aliph...	1	-0.951	-0.951
Fragment	-tert Carbon [3 or m...	1	0.268	0.268
Constant	Equation Constant ...	0	0	0.229
Notes				Log Kow
<input type="text"/>				7.427

Figure 7. User Entry Panel

Chemical Name	
<input type="text" value="permethrin"/>	
CAS	
<input type="text" value="52645531"/>	
Log Kow	<input type="text" value="7.4267"/>
	
Water Solubility (mg/L)	<input type="text" value="0.006"/>
	
Melting Point (°C)	<input type="text" value="45.5"/>
	
Chemical Details	
SMILES	
<input type="text" value="CC1(C)C(C=C(Cl)Cl)C1C(=O)OCc3cccc(Oc"/>	
MOL WT	
<input type="text" value="391.3"/>	
Log Kow	
<input type="text" value="7.4267"/>	(estimated)
<input type="text" value="6.5"/>	(measured)
Water Solubility (mg/L)	
<input type="text" value="0.042485"/>	(estimated)
<input type="text" value="0.006"/>	(measured)

The Results Panel can simultaneously hold the output of multiple different evaluations in separate output tabs, which can be closed or re-ordered within the panel. Each tab does not need to be removed or closed before running another chemical in the program; the Results Panel will be updated automatically. Additionally, the left side of each output tab is occupied by the User Entry Panel (see Figure 7).

The following behaviors are available in the Results tab:





Copy: All tables can be copied and pasted from any of the subtabs by first selecting all of the items in the table that the user wishes to copy, and then using the keyboard shortcut to copy


the data (Ctrl + c in Windows, Command + c MAC). This command copies the results as shown (minus the rectangle enclosing the estimate) to the Windows clipboard. The table can then be pasted wherever the user would like to paste it either using the 'Paste' function in a spreadsheet, word processing, or other program, or using the keyboard shortcut (Ctrl + v Windows, Command-V MAC).

Close tab: All results tabs can be closed by clicking the small 'x' in the upper right corner of the tab, just to the right of the tab name.

6.1. Output Tabs

The Results Window Panels hold the following subtabs for the respective chemical (see Figure 6).

- (1) Organic Module Result: This subtab holds the results of ECOSAR Class Program's estimations. This includes the class-specific Max Log K_{ow} cutoff as well as any flags or alerts associated with the values produced. Flags or alerts include: No Effect at Saturation , Acute to chronic ratio estimation used , and inexact Log K_{ow} cutoff . Additionally, Class definitions are available from the information buttons to the right of the class name .

Yellow information buttons indicate classes of special toxicological significance . Mouse over the button for more information regarding these classes.

Chemicals containing the large molecular fragment for Pyrethroids are likely designed to act as insecticides. If the query compound contains this structural fragment, it is likely to be most closely related to the training set chemicals in the Pyrethroid class irrespective of the other classes reported

- (2) Experimental Data: This subtab holds the measured ecotoxicity training set data used in QSAR development for the query chemical, when available. 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 1980s, some data incorporated into ECOSAR came from data sheets that did not clearly identify the reference; however, extensive efforts have identified most references.

- (3) **Physical Properties:** This subtab holds the experimental physical-chemical properties for the evaluated molecule when available from the Physical Properties (PhysProp) Database.
- (4) **Log K_{ow} Estimate:** This subtab shows the values used to arrive at the K_{ow} Estimate used in the calculation of the ECOSAR Class Program's toxicity endpoint estimations. This value is preferentially used in the estimated effect levels and can be changed by changes to the User Entry Panel (see Figure 1).

6.2. Chemical Attributes

The Results Tab provides the attributes of the query chemical (see Figures 1 and 7). The SMILES structure used to predict the effect levels is depicted next to the corresponding CAS number.

The field with the molecular weight ('MOL WT') is determined from the SMILES notation used for ECOSAR predictions. Chemical Name and CAS number ('CAS') appear first with a picture of the structure. The Chemical Name field is editable.

Log K_{ow}, Water Solubility, and Melting Point are all editable. Chemical Details holds the SMILES, Molecular Weight ('MOL WT'), and estimated and measured values available for Log K_{ow} and Water Solubility. Log K_{ow} preferentially relies on the estimated value, whereas Water Solubility incorporates the measured value from PhysProp when available. 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 'Melting Point' field; otherwise, the field will import a value from PhysProp, when available.

6.2.1. Entering Measured Data for Variables

In the previous version of ECOSAR (v2.0), User Entered Variables (i.e., Log K_{ow}, Water Solubility, and Melting Point) were entered in the left User Entry Panel. They are now handled in the Main Input Panel, under the SMILES entry field (Figure 1), pre-estimation in ECOSAR v2.2. Entering values allows the estimations to be calculated directly and enables users to provide data

for certain parameters that will be used in place of ECOSAR-predicted variables and/or values retrieved from PhysProp.

6.2.2. Log K_{ow}

The log K_{ow} value is used in the calculation of the predicted effect level. If a user enters a log K_{ow} value in the main data entry screen, that value will be used directly in the calculation. In the absence of a user-entered log K_{ow} value, ECOSAR will automatically use the log K_{ow} value calculated from the KOWWIN program (available from EPI Suite and embedded in ECOSAR). The variables used in calculating the estimated log K_{ow} can be found in the K_{ow} Estimate subtab in the Result tab. To minimize the potential effects of variability of measured log K_{ow} values due to factors such as variable study conditions, QSARs were developed from predicted log K_{ow} values, and it is often recommended that the predicted K_{ow} values (calculated and used by default in ECOSAR) be used in the model when there is uncertainty in the reliability of the available measured K_{ow} values (see the ECOSAR Methodology Document for further information in the Helpful files).

6.2.3. 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 a user entry field, 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 PhysProp. 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 EPI Suite. 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 EPI Suite and in the following document prepared for EPA (OPPT): Upgrade of PCGEMS Water Solubility Estimation Method (May 1994).

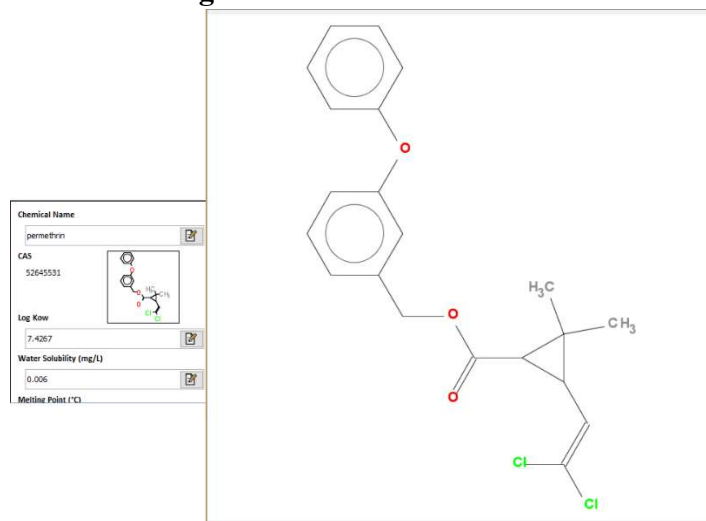
6.2.4. Melting Point

A user-entered melting point can be entered in a user entry field (see Figure 1). 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.

6.3. Structure Window

The Structure Window (see Figure 8) shows a 2-dimensional picture of the chemical structure. This can be found in the User Entry Panel (see Figure 7). Clicking on the small picture of the chemical structure opens a separate window with an Expanded Structure View. The window shows the entire structure (it does not "clip" sections of the molecule). Some particularly large molecules (e.g., Vancomycin) have a hard time rendering in the Expanded Structure View. 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 tab are printed, the accompanying structure is that of the smaller picture seen in the User Entry Panel (see Figure 7).

Figure 8. Structure Window





6.4. Experimental Data Tab

Training set data used to develop regression equations for QSAR class endpoints are presented in the training set data section of the Results Panel (see Figure 6).


Data are not presented in any particular order. If measured data used in the training set are considered to be TSCA CBI, these data are not shown in this section. Since data collection for the ECOSAR program began in the early 1980s, 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. See the Technical Reference Manual (also referred to as the ECOSAR Methodology Document in the ‘Helpful files’) for further information on data collection and selection methods.


6.5. Organic Module Result Tab

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). In this version, the neutral organics (“baseline toxicity”) class QSAR results appears at the top of the output tables. The baseline toxicity equations in the prior version ECOSAR v1.1 and in the current version (ECOSAR v2.2) are based on data collected through 2016. (Note that baseline toxicity equations in the previous ECOSAR v1.0 were based on data collected through 1999 and ECOSAR v0.99 was calculated from the 1981 Konemann Equation).

Common notations reported for predicted effect levels are used to designate predicted effect levels that exceed the water solubility limit () and to designate effects levels predicted using an acute-to-chronic (ACR) ratio ()

In the tables displaying the reported predicted effect levels, there is a column that indicates the maximum log K_{ow} cut-offs for each class (‘Max Log Kow’). The K_{ow} cut-off values signify the

point that 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 K_{ow} values that exceed the limits, results are typically reported with the flag indicating that “No Effects at Saturations” are expected, .

In general, log K_{ow} cut-offs are 5.0 for fish and daphnid acute endpoints, 6.4 for green algae 72-/96-hour EC_{50} endpoints, and 8.0 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 (>). This cannot be rendered in the table format, so any place where the cutoff would be greater than the indicated number, the flag  is used. This indicates that 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 in the ‘Helpful files’) 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 9 presents an example of a compound that fits into multiple chemical classes. The predictions section of the output is depicted in Figure 10 and described in Section 6.4, Selecting the Most Representative Class.

Figure 9. Example Chemical Classified Into Multiple Classes

Organism	Duration	End Point	Concentration (...)	Max Log Kow	Flags
Fish	96h	LC50	0.035	5.0	
Daphnid	48h	LC50	0.041	5.0	
Green Algae	96h	EC50	0.0074	6.4	
Fish		ChV	0.00085	8.0	
Daphnid		ChV	0.0059	8.0	
Green Algae		ChV	0.011	8.0	
Fish (SW)	96h	LC50	0.039	5.0	
Mysid	96h	LC50	0.0026	5.0	
Fish (SW)		ChV	0.015	8.0	

Organism	Duration	End Point	Concentration (...)	Max Log Kow	Flags
Fish	96h	LC50	0.00088	6.0	
Daphnid	48h	LC50	0.0010	6.0	
Green Algae	96h	EC50	0.0070	6.4	
Fish		ChV	0.0000096	8.0	
Daphnid		ChV	0.0063	8.0	
Green Algae		ChV	0.018	8.0	
Fish (SW)	96h	LC50	0.00022	5.0	
Mysid (SW)	96h	LC50	0.00014	6.0	
Earthworm	14d	LC50	208	6.0	

Organism	Duration	End Point	Concentration (...)	Max Log Kow	Flags
Fish	96h	LC50	0.00035	8.2	
Daphnid	48h	LC50	0.00022	7.5	
Fish		ChV	0.000017	8.0	
Daphnid		ChV	0.000045	8.0	
Fish (SW)	96h	LC50	0.00021	7.2	
Mysid	96h	LC50	0.0000074	8.2	
Fish (SW)		ChV	0.000033	6.9	
Mysid		ChV	3.0E-7	8.0	

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 Helpful files contain class supporting information that includes Definition files and QSAR equation files. These documents enable the user to evaluate adequacy of predictions.

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., haloimides). Depending on ECOSAR programming, predictions for the general classes as well as the more specific sub-classes may be displayed in the ECOSAR output.

Sub-classifications are created in ECOSAR when compounds with larger, more complex structural moieties (pyrethroids) are identified that exhibit toxicity levels that are unlike estimates for the more general classes (esters, vinyl/allyl/propargyl halides), even though those complex compounds may still contain those simple molecular features. In the example depicted in Figure 8, the general classes identified for permethrin would be esters and vinyl/allyl/propargyl halides (relating to smaller functional groups contained within permethrin). The more specific sub-class is pyrethroids, which define a much larger part of the permethrin 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, permethrin. The QSAR class supporting information in the Helpful files enables the user to assess structural features of the query compound. The user needs to determine how many molecular features of permethrin fit each class definition and whether that class is the most specific one available in ECOSAR for the query chemical. The following is EPA's interpretation of the ECOSAR output for the example depicted in Figure 8.

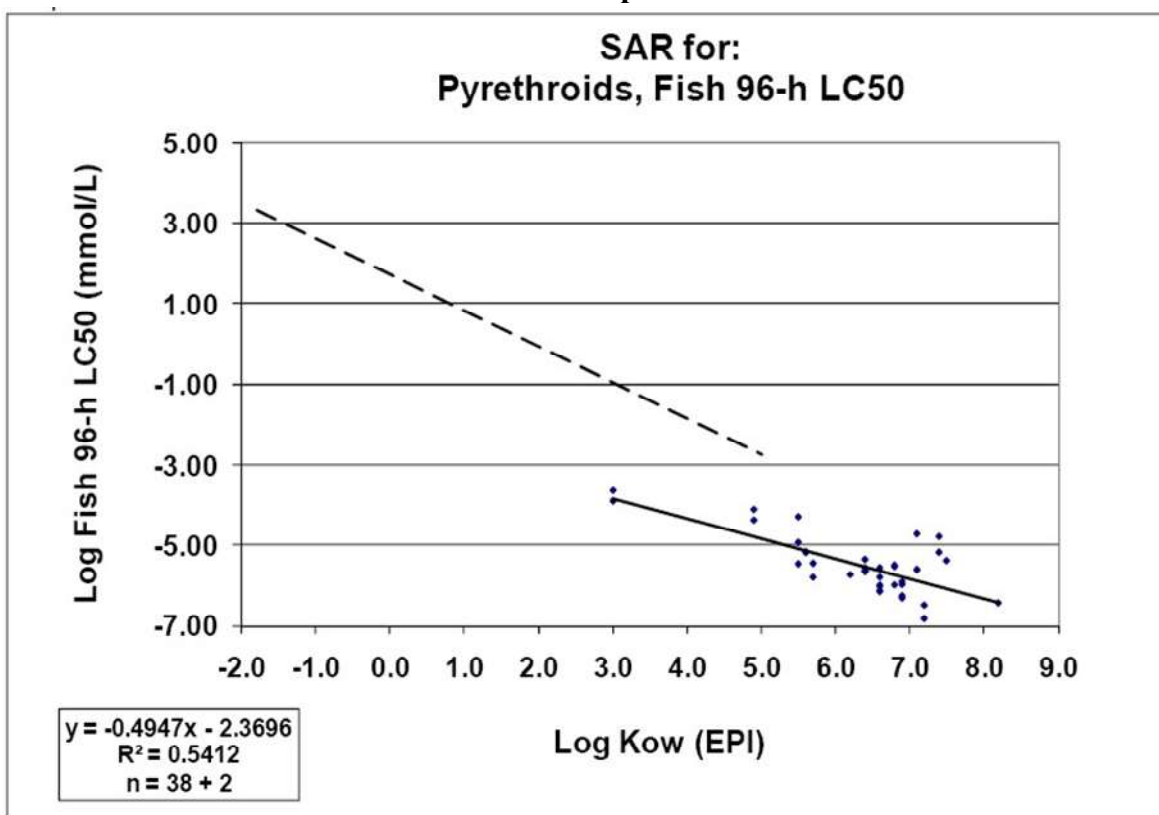
- (1) Esters: Permethrin unequivocally fits the esters definition, but this is a general class that addresses only one of the structural features of this compound.
- (2) Vinyl/Allyl/Propargyl Halides: The vinyl/allyl/propargyl halide in permethrin is part of a terminal vinyl/allyl moiety. There have 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 permethrin should or should not be excluded from this class.
- (3) Pyrethroids: Permethrin 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 permethrin 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, 6.6.3, and the Helpful files).

6.6.2. Correlation of Log K_{ow} with Toxicity in QSAR Training Data Sets

For each developed QSAR, a graph (see Figure 10) is displayed in the ECOSAR Equation Document (see the Helpful files for access to QSAR Supporting Documents including ECOSAR Equation Documents) along with a table of supporting data.

Figure 10. Graph from the Pyrethroids QSAR Equation Document for the Fish 96-Hour LC₅₀ Endpoint



A coefficient of determination (R^2) is reported in both the depicted graph (a scatterplot) and the text within 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.² However, a weak relationship does not necessarily indicate little or no correlation; if adequate data were scarce for a certain

² 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.

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-hour LC_{50} endpoint only.

- (1) Esters: Pearson's Correlation Coefficient (r) is -0.88. Using 5% uncertainty ($p = 0.05$), the correlation between the fish 96-hour LC_{50} value (mmol/L) and the log K_{ow} value is statistically significant.
- (2) Vinyl/Allyl/Propargyl Halide: Pearson's Correlation Coefficient (r) is -0.30. Using 5% uncertainty ($p = 0.05$), the correlation between the fish 96-hour LC_{50} value (mmol/L) and the log K_{ow} value is not statistically significant.
- (3) Pyrethroids: Pearson's Correlation Coefficient (r) is -0.74. Using 5% uncertainty ($p = 0.05$), the correlation between the fish 96-hour 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 K_{ow} values of the available data for a given training set, the log K_{ow} value of the queried compound may be notably less than or greater than the minimum and maximum log K_{ow} values, respectively, 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 10), 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 the above permethrin example, the following can be interpreted from the ECOSAR equation sheets.

- (1) Esters: This QSAR 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/Propargyl Halides: This QSAR may be used to estimate toxicity for vinyl/allyl/propargyl halides. The training set for this class contains 21 different chemicals in the training set.
- (3) Pyrethroids: The log K_{ow} values for data points that are within the solubility limit range from 3.0 to 8.2. Thus, if the log K_{ow} 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 the above permethrin example output (see Figure 8), available information from Class Equation and Definition Documents could support a user's decision to exclude the esters and vinyl/allyl/propargyl halides predictions. The remaining class, pyrethroids, appears to be the most representative for the query compound and also results in the most conservative effect levels (see Figure 8).

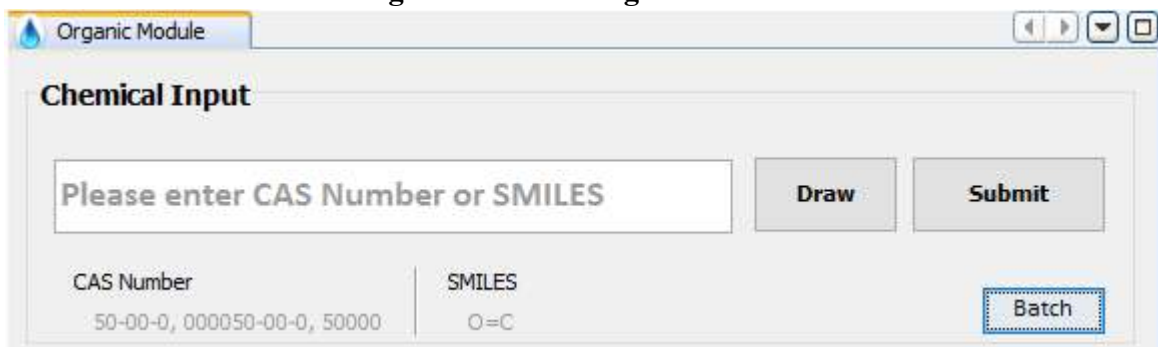
There is no one standard method for selecting the most representative predictions and, often, the best approach would be to select the most conservative effect level until measured data become available. The model developers emphasize that each predicted profile should be accompanied with a discussion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

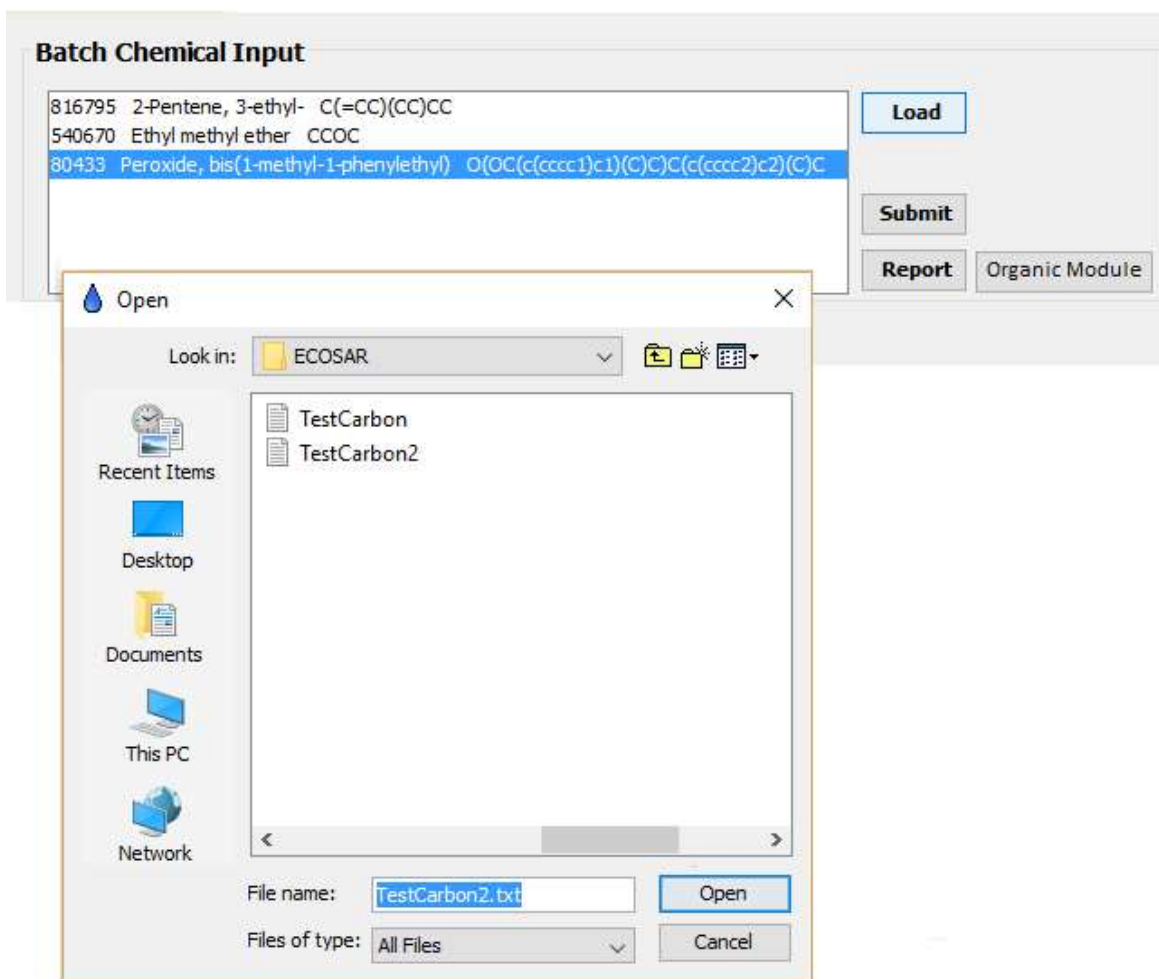
7. Batch Runs

Batch runs are used to make multiple estimates from a single input file that contains multiple chemical identifiers. The ECOSAR Class Program can accept "batch inputs" from five different data types in input files (SMILES Strings, CAS numbers, chemical name, Excel, and MDL SD files) and can output the data in two different formats (standard text output as a single run report for specifically selected chemicals, and as an Excel file). Each input file must be in a specific format; otherwise, the batch run will fail. Program access to "batch-runs" (depicted in Figure 11) is available from the main Input Panel by pressing the 'Batch' button.

Figure 11. Accessing Batch Mode



The screenshot shows a software window titled "Organic Module" with a "Chemical Input" section. It features a text input field with the placeholder text "Please enter CAS Number or SMILES", a "Draw" button, and a "Submit" button. Below the input field, there are two columns: "CAS Number" with the example values "50-00-0, 000050-00-0, 50000" and "SMILES" with the example value "O=C". A "Batch" button is located at the bottom right of the input area.



7.1. Batch Input Files

Batch run inputs are initialized by pressing the ‘Load’ button, which allows the user to upload external files. These files can include SMILES strings, CAS numbers, chemical names, Excel, and MDL SD files. The configuration of these files is described below.

7.1.1. SMILES, CAS, and Chemical Name Strings

Batch runs can be carried out with properly formatted input files. This is defined as a string format list in a plain text file (usually with a ".txt" file extension) containing a list of SMILES notations, CAS numbers, or chemical names. An example String Format follows:

```
CCCCO Butanol c1ccccc1 Benzene
```

Fc1ccccc1 Fluorobenzene

CC(=O)C Acetone

816795

540670 CCOC

000050-00-0

71-43-2

000050-02-2

7.1.2. 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.

7.1.3. Excel Files

Microsoft Excel files can be run in batch mode if the user needs to have specifically entered values (Log K_{ow}, Water Solubility, Melting Point) associated with each input SMILES. An example Excel format follows:

Smiles	Log Kow	WS	MP
CCCCCOC(CO)C	1.1	2	-30
NCC(N)CCN	1.2	3	4
C(F)CC(F)CF	1.3	4	50

Where: WS is Water Solubility in mg/L, MP is Melting Point in °C

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, chemical name, or Excel 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) Set of Individual Reports (for batch runs with less than 20 chemicals only) and (2) Excel Table Data.

- (1) Individual Reports (for batch runs with less than 20 chemicals) are the same as the individual run reports and capture results for each compound the same as they would appear in the “Result Panel” (i.e., if each compound was estimated individually); these output files can be saved in a number of file outputs.

- (2) Excel Table Data Output captures ECOSAR-predicted effect levels. Output varies slightly depending on whether data were input as SMILES strings or MDL SD files. Fields output in the Excel table include (1) a ‘CAS’ field that depicts an ID number for measured data, (2) chemical name, when available, (3) a ‘SMILES’ field that depicts the entered SMILES string, (4) an ‘ECOSAR Class’ field that lists the identified ECOSAR classes for input SMILES strings, (5) an ‘Organism’ field that identifies the organism for the predicted and measured effects, (6) a ‘Duration’ field that identifies the exposure duration for predicted and measured effects, (7) an ‘End Point’ field that identifies the endpoint (e.g., LC₅₀, EC₅₀, or ChV) for the effect level, (8) a ‘Concentration (mg/L)’ field that reports predicted effect levels, (9) Max Log K_{ow} that gives the class and organism-specific cutoff value, (10) a ‘Flags’ field that identifies flags present for predicted effect levels, and (11) an ‘Alert’ field to indicate any further alerts.

7.2.1. Obtaining Individual Report Outputs

In order to conduct batch runs yielding individual report outputs, select up to 20 of the required chemicals from the uploaded input file. Only chemicals that have successfully loaded or were formatted correctly from the input file will show up in this field. Hit the ‘Submit’ button and all of the chemicals selected will run a full estimation. Following that, go to the Report subtab in the Results Tab and select the “Generate Report” button (see Figure 12).

Figure 12. Individual Report Outputs

The screenshot shows a web interface for generating individual report outputs. At the top, there is a horizontal navigation bar with five tabs: "Organic Module Result", "Experimental Data", "Physical Properties", "K_{ow} Estimate", and "Report". The "Organic Module Result" tab is currently selected. Below the navigation bar, the main content area is titled "Report Options". It contains a list of four items, each with a checkbox: "Organic Module Result" (checked), "Experimental Data", "Physical Properties", and "K_{ow} Estimate". Below this list is a button labeled "Generate Report".

7.2.2. Obtaining Table Data Outputs

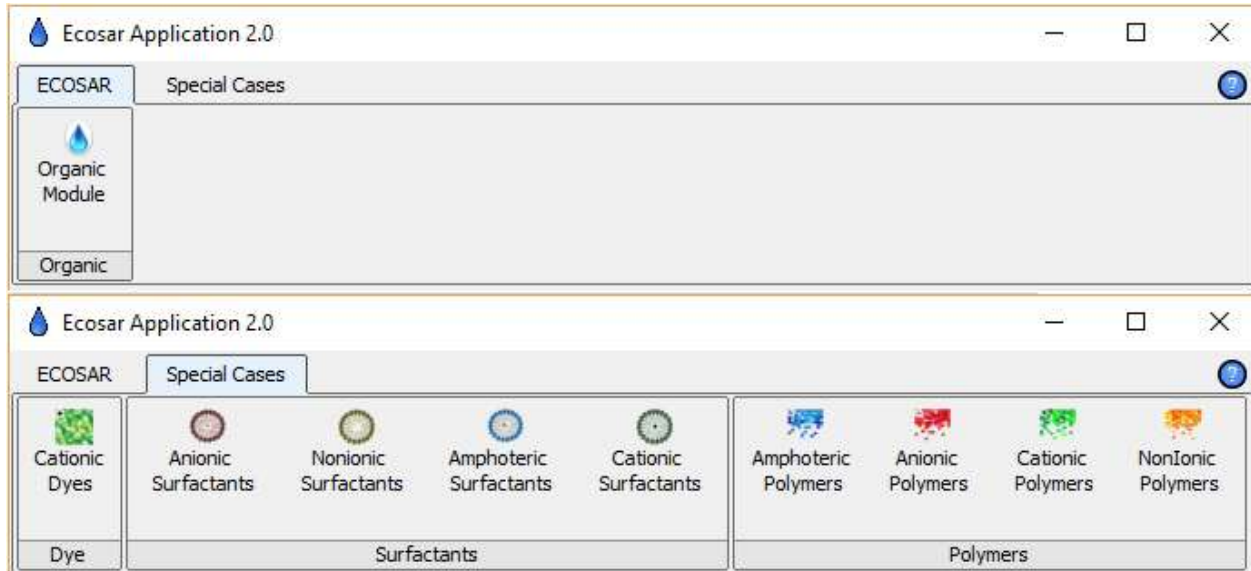
As an alternative to the Individual Report Outputs that is generated as described in Section 7.2.1, the “Report” button in the “BatchMode” entry area will generate results in a Microsoft Excel file, which can be converted into various other table formats.

8. Special Class Calculations

The ECOSAR Program has been developed primarily for the following scenario: (1) enter a SMILES, CAS number, or chemical name, (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 always use the conventional log K_{ow} value or cannot be adequately classified from the SMILES notation. These "Special Classes" include dyes, polymers, and surfactants. QSARs are available for various anionic, cationic, nonionic, and amphoteric surfactants and polymers. Instead of only using the log K_{ow} value, these surfactant QSARs may utilize the number of ethoxylate units or the average length of the carbon chain. The polymer QSARs may utilize the percent amine nitrogen, the cation to anion ratio, and the polymer type. Dye QSARs are also available for triphenylmethane dyes and ethoxylated triphenylmethane dyes, which utilize ethoxylate units for

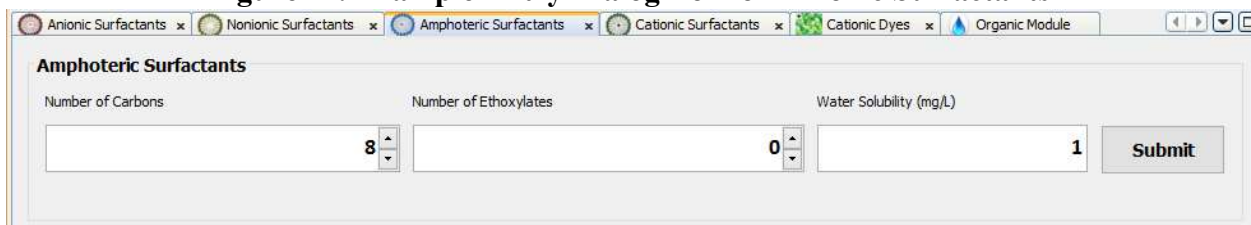
QSAR predications. These "Special Classes" are accessed from the Main Menu bar (see Figure 13).

Figure 13. Special Classes in ECOSAR



The Special Classes have their own data entry panels (see Figure 14).

Figure 14. Example Entry Dialog Box for Anionic Surfactants



The calculated results are placed in the same Results tab as results using SMILES notations (an example is illustrated in Figure 15).

Figure 15. Example Results Window for Anionic Surfactants

Organism	Duration	End Point	Concent...	Flags
Fish	96h	LC50	10.5	⚠
Daphnid	48h	LC50	10.5	⚠
Green Algae	96h	EC50	10.5	⚠

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

Meylan, W.M. and P.H. Howard. 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 (the geometric mean of the NOEC and LOEC)
EPA	U.S. Environmental Protection Agency
F96	Fish 96-Hour Endpoint
FChV	Fish Chronic Value
D48	Daphnid 48-Hour Endpoint
DChV	Daphnid Chronic Value
EC ₅₀	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- or 96-Hour Endpoint
GACHV	Green Algae Chronic Value
GM	Geometric Mean
hr	Hour
K _{ow}	Octanol-Water Partition Coefficient
LC ₅₀	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 Relationship
r	Correlation Coefficient
R ²	Coefficient of Determination
SMILES	Simplified Molecular Input Line Entry System
SW	Saltwater
TSCA	Toxic Substances 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.

Enter: Pressing the Enter (Return) key sends the cursor to the next data entry field. If an entry has been made into the Chemical Input field (CAS, SMILES, name), pressing Enter will cause the estimate to run.

Tab or Shift-Tab: Changes entry fields.

Ctrl + Tab: Return to the Chemical Input field or other first entry field in a panel.