

OECD QSAR Toolbox v.4.1

Step by step example how to predict acute aquatic toxicity to Daphnia for the 3-ethyl-5-methyl-3-methoxyphenol by the trend analysis approach

Outlook

- **Background**
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- Workflow of the exercise

Background

- This is a step-by-step presentation designed to take the user of the Toolbox through the workflow of a data gap filling exercise by trend analysis approach.

Outlook

- Background
- **Objectives**
- Specific Aims
- Trend analysis
- The exercise
- Workflow of the exercise

Objectives

- **This presentation reviews a number of functionalities of the Toolbox:**
 - Identify analogues for a target chemical
 - Retrieve experimental results available for those analogues
 - Fill data gaps by trend-analysis

Outlook

- Background
- Objectives
- **Specific Aims**
- Trend analysis
- The exercise
- Workflow of the exercise

Specific Aims

- To review the workflow of the Toolbox.
- To review the six modules of the Toolbox.
- To reacquaint the user with the basic functionalities within each module.
- To explain the rationale behind each step of the exercise.

Outlook

- Background
- Objectives
- Specific Aims
- **Trend analysis**
- The exercise
- Workflow of the exercise

Trend Analysis Overview

- For a given (eco)toxicological endpoint, the members of a category are often related by a trend (e.g. increasing, decreasing or constant). The trend could be related to molecular mass, carbon chain length, or to some other physicochemical property.
- A demonstration of consistent trends in the behaviour of a group of chemicals is one of the desirable attributes of a chemical category and one of the indicators that a common mechanism for all chemicals is involved. When some chemicals in a category have measured values and a consistent trend is observed, missing values can be estimated by simple scaling from the measured values to unmeasured values as a means of filling data gaps.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- **The exercise**
- Workflow of the exercise

Exercise

- In this exercise we will predict the acute toxicity to daphnids for an untested compound, (3-ethyl-5-methyl-4-methoxyphenol), which is the “target” chemical.
- This prediction will be accomplished by collecting a set of test data for chemicals considered to be in the same category as the target molecule.
- The category will be defined using the following categorization schemes:
 - *Acute aquatic toxicity classification by ECOSAR* – for structural grouping.
 - *Acute aquatic toxicity MOA by OASIS* – for mechanistic grouping.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**

Workflow

- **Toolbox has six modules which are used in a sequential workflow:**
 - Chemical Input
 - Profiling
 - Data
 - Category Definition
 - Filling Data Gaps
 - Report

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
 - **Chemical Input**

Chemical Input Overview

- This module provides several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.

Chemical Input

Ways of Entering a Chemical

User Alternatives for Chemical ID:

A. Single target chemical

- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases
- Substructure by using SMART

B. Group of chemicals

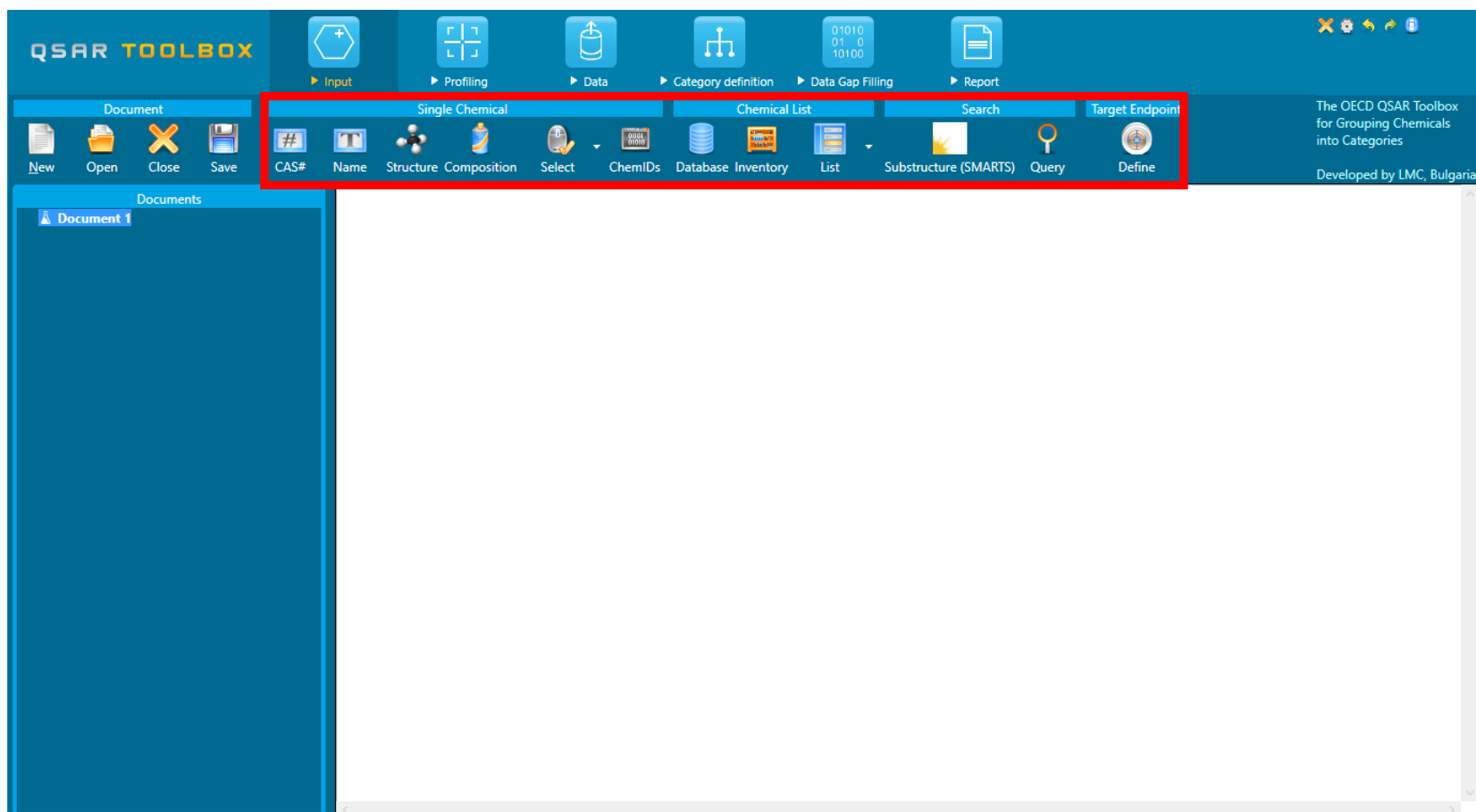
- User List/Inventory
- Specialized Databases

Getting Started

- Open Toolbox.
- The six modules in the workflow are seen listed next to “QSAR TOOLBOX”.
- Click “Input” (see next slide).

Chemical Input Screen

Input screen



Chemical Input by Drawing

- Inputting the target chemical by drawing varies in difficulty with the structural complexity of the molecule.
- It is accomplished by a series of point-click-move-click operations within the 2D-editor which drops down when you click on "structure" (see next screen shot).
- The subsequent series of slides will take you through the process for the target chemical.

Chemical Input Screen

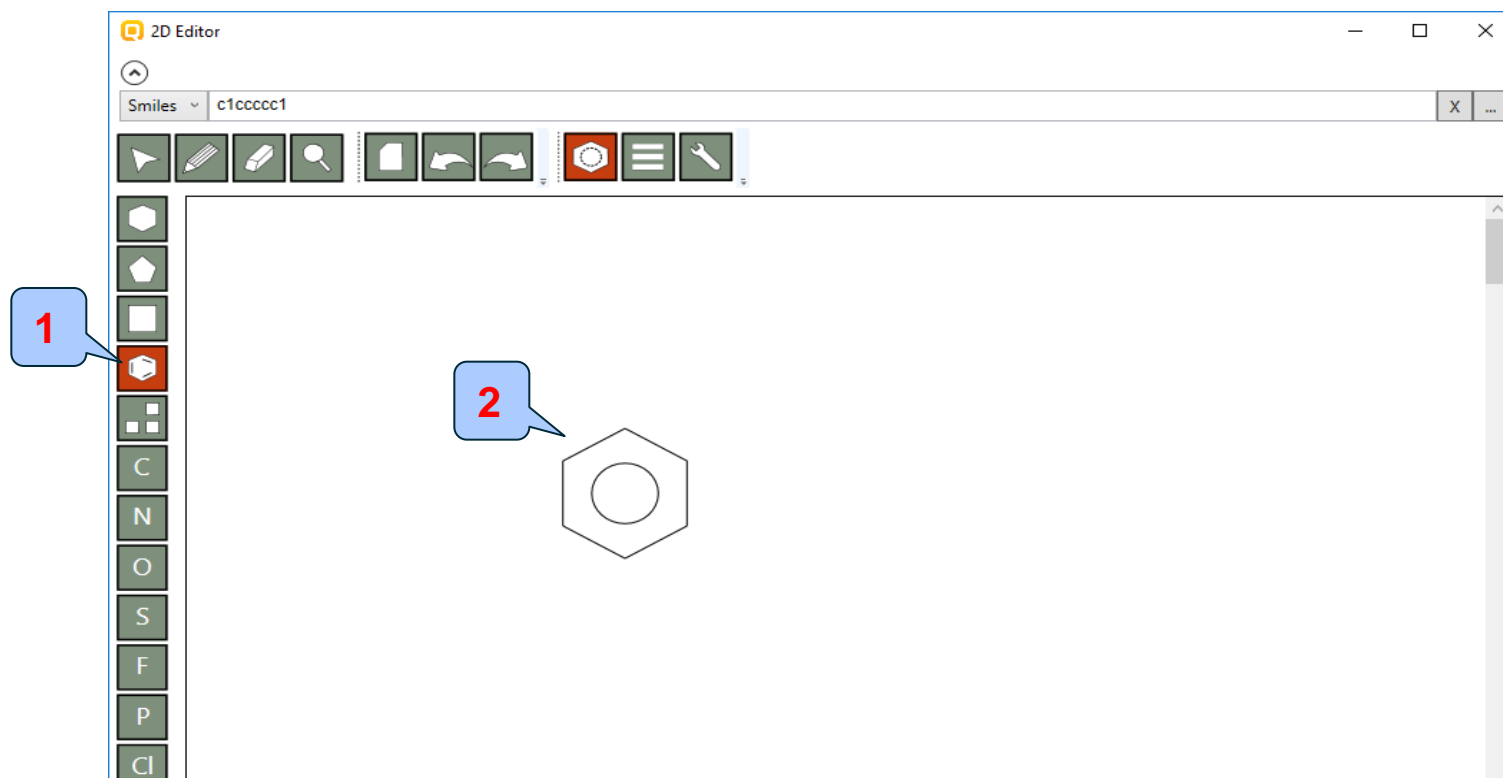
Input target chemical by drawing

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options for Document, Single Chemical, Chemical List, Search, and Target Endpoint. The 'Single Chemical' sub-menu is expanded, showing options like Name, Structure, Composition, Select, ChemIDs, Database, Inventory, List, Substructure (SMARTS), Query, and Define. The 'Structure' button, which features a ball-and-stick molecular model icon, is highlighted with a red rectangular box. A blue callout bubble with the number '1' points to this button. The main workspace is currently empty.

1. Click **Structure**.

Chemical Input

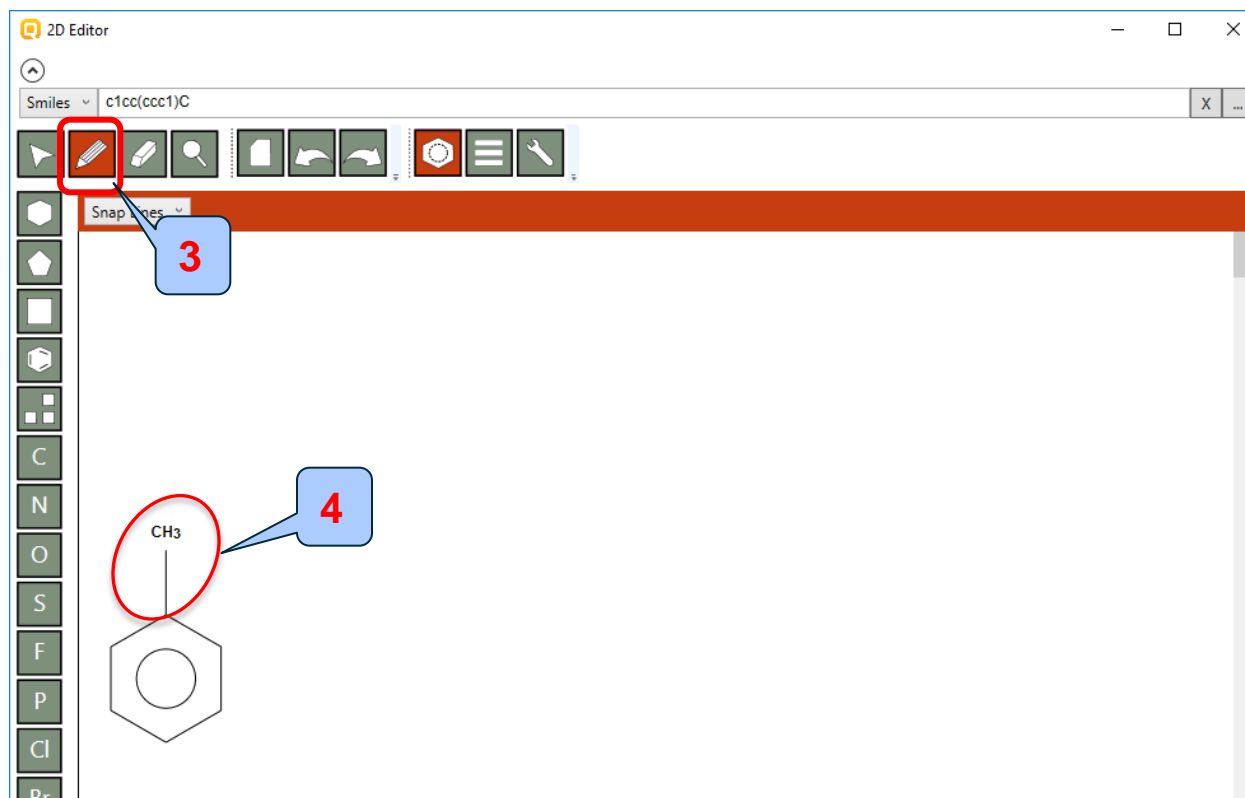
Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor




1. **Left Click** on the appropriate template form.
2. Move the cursor to the large clear area and **left click** again, this puts the selected template on the plot.


Chemical Input

Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor



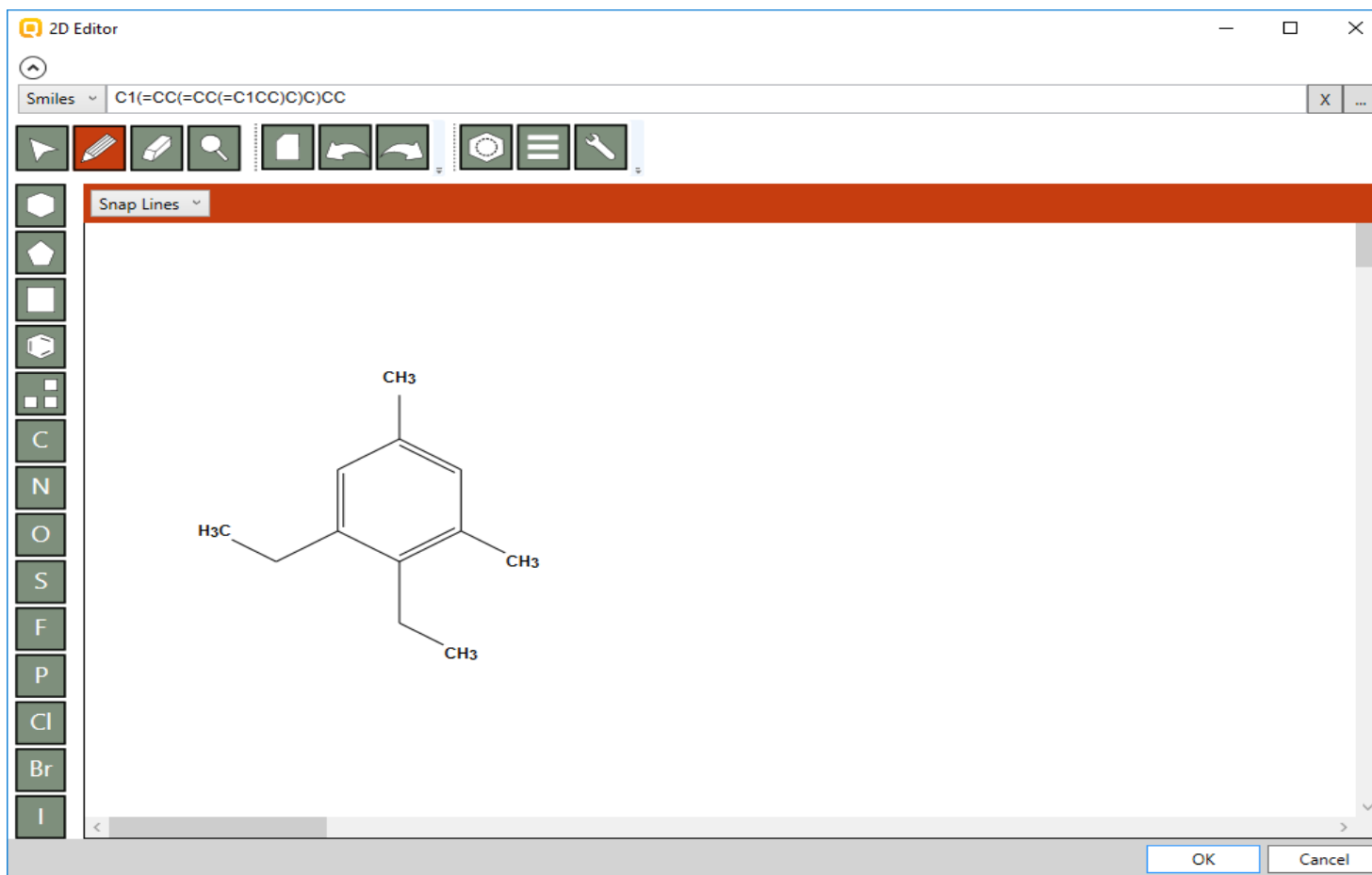
3. Click  pencil button to draw a single bond.
4. Drag the mouse (pointing finger) to the appropriate atom and **left click** to create a single bond.

Chemical Input by Drawing

- Note the default is addition of a CH₃-group.
- By moving the 'finger' to other C-atoms and left clicking the mouse adds other hydrocarbon fragments.
- If you make an incorrect entry you can click the 'undo' icon () in the upper corner of the screen to remove the addition.
- This process allows you to build the hydrocarbon skeleton of the target molecule (see next screen shot).

Chemical Input

Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor



Chemical Input

Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor

The screenshot shows the 2D Editor window with the SMILES string CC(=CC(=CC(=C1CC)C)C)CC in the top bar. The main drawing area displays the chemical structure of 3-ethyl-5-methyl-4-methoxyphenol. A vertical palette on the left contains elements C, N, O, S, F, P, Cl, Br, and I. The 'O' element is highlighted with a red box and labeled '1'. A callout box labeled '2' points to a methyl group on the structure. Another callout box labeled '2' points to the oxygen atom being inserted into the structure. A callout box labeled '3' points to the 'OK' button at the bottom right of the drawing area.

1. Click **O** to add an oxygen atom; 2. Left click with mouse over the methyl group to insert an oxygen atom; 3. Click **OK**.

Chemical Input

Target chemical identity

- The already drawn target structure automatically appears on the data matrix
- Note that no CAS number or name is displayed for this chemical. This means the target chemical is not listed in the chemical inventories/databases implemented in the Toolbox (see next slide).

Chemical Input

Target chemical identity

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Input' module is active, and the 'Profiling' button is highlighted with a red dashed box and a red circle containing the number '1'. The main window displays a chemical structure of 2-methoxy-3,4-dimethylphenol. Below the structure, the 'Structure info' section shows 'Invalid CAS number: 0-00-0' circled in red.

The workflow on the first module is now complete, and the user can proceed to the next module. Click **Profiling** (1).

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - **Profiling**

Profiling Overview

- *Profiling* module refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- Available information includes likely mechanism(s) of action, as well as observed or simulated metabolites.

Profiling

Summary information of the different profilers are provided in the "About".

1. Highlight the profiler, then perform right click; 2. Select **About**;

3. After acquiring the information you desire, **close** the window;

Name	Value
Scheme type	Linear
Scheme nature	Predefined
Number of categories	66
Number of help files	66
Adopted in version	QSAR Toolbox 2.0 beta, April 2010
Last modified	12/15/2016
Hash	5477187
Counter category	Not categorized

Profiling

Summary information of the different profilers are provided in the "About".

The screenshot shows the QSAR Toolbox software interface. On the left, a sidebar contains a tree view of profilers. Callout 1 points to the 'US-EPA New Chemical Categories' profiler. Callout 2 points to the 'About' button in the 'Options' menu. The main window displays the 'US-EPA New Chemical Categories' profiler details, including a list of categories and a disclaimer. Callout 3 points to the close button (X) of the 'About' dialog box. The 'About' dialog box contains the following text:

Name: US-EPA New Chemical Categories

Short Description:

The rules coded in the US-EPA New Chemical Program profiler reproduces the original categories cited in the document "TSCA New Chemicals Program (NCP)/ Chemical Categories" - an official document of U.S. EPA Office of Pollution Prevention and Toxics. However, not all of the categories have been coded because they are very broad and limits are mostly based on physical considerations and do not include structure based rules. These categories are listed below:

- Category: Acid Dyes and Amphoteric Dyes
- Category: Cationic Dyes
- Category: Polyaniionic Polymers (& Monomers)
- Category: Polycationic Polymers
- Category: Respirable, Poorly soluble Particulates

Disclaimer:

The U.S. EPA/OPPT New Chemical Categories include classes of chemicals for which sufficient regulatory history has been accumulated so that hazard and risk concerns and testing recommendations vary little from chemical to chemical within the new chemicals program. It is important to note that the U.S. EPA/OPPT New Chemical Categories do not cover the universe of potential industrial molecules. Also, the categories do not necessarily represent substances of greatest concern to the Agency. The categories are also NOT intended to be a comprehensive list of all substances that may be subject to further action by the U.S. EPA/OPPT New Chemicals Program.

The structural boundaries used to define the chemical classes (e.g. "Alcohol" - chemical class from "Organic functional group" profiler) or alerting groups responsible for the binding with biological macromolecules (e.g. "Aldehydes" - structural alert for protein binding), represent structural functionalities in the molecule which could be used for building chemical categories for subsequent data gap filling. They are not recommended to be used directly for prediction purposes (as SARs).

Donator(s): United States Environmental Protection Agency (EPA), USA

Author(s):

Website: <http://www.epa.gov/oppt/newchems/pubs/chemcat.htm>

Details:

Name	Value
Scheme type	Linear
Scheme nature	Predefined
Version	2.0
Number of categories	66
Number of help files	66
Adopted in version	QSAR Toolbox 2.0 beta, April 2010
Last modified	15-Dec-16
Counter category	Not categorized

1. Click the profiler of interest; 2. Click **About** or right-click the profiler and select *About*; 3. **Close** the window.

Profiling

- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, *DNA binding by OECD*) and clicking **View** (see next slide).

Profiling

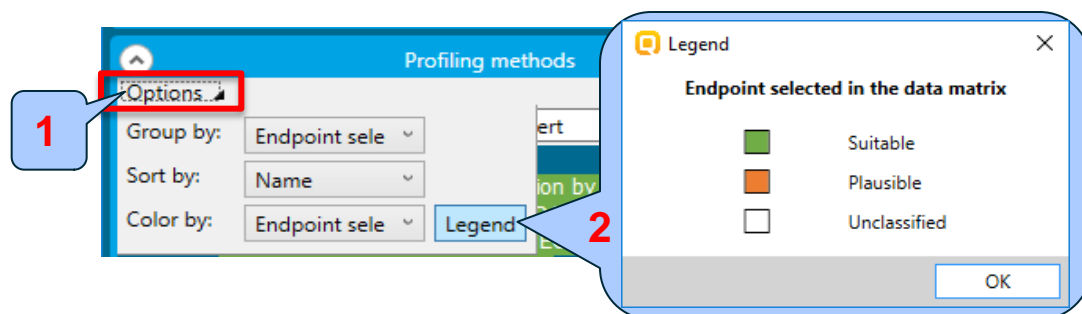
The screenshot illustrates the steps for accessing structural alert literature in the QSAR Toolbox. Step 1: Select the 'DNA binding by OECD' method in the 'Profiling methods' list. Step 2: Click the 'View' button to open the 'Profiling Scheme Browser'. Step 3: Select the 'Isocyanates' structural alert from the 'Categories' list. Step 4: Switch to the 'Literature' tab to view the background information, including the chemical structure $R-N=C=O$ and the reaction mechanism showing the nucleophilic attack of a biological nucleophile (Nu) on the carbon atom of the isocyanate group.

1. Click the profiler; 2. Click **View**; 3. Click one of the structural alerts (for example **Isocyanates**); 4. Go to the **Literature** tab to see the background information.

Profiling

Profiling the target chemical

- Once the endpoint is selected, the relevant profiles and metabolic transformations are highlighted. Meaning of the colors could be seen within the **Options** (1) by click **Legend** (2).



- **Suitable** - developed using data/knowledge for the target endpoint;
 - **Plausible** – not endpoint specific; structure-based; form broader group of analogues;
 - **Unclassified** – all profilers, which are not classified in any of the categories above.
- Select the Profiling methods related to the target endpoint by ticking the checkbox next to the profilers name.
- This selects (a green check mark appears) or deselects (the green check disappears) profilers.

Profiling

Profiling the target chemical

The screenshot shows the QSAR Toolbox interface during the Profiling step. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Profiling' sub-menu is active, showing 'Apply', 'View', 'New', and 'Delete' options. The 'Documents' panel on the left lists various 'Profiling methods' with checkboxes. The 'Filter endpoint tree...' panel on the right shows a tree structure with 'Aquatic Toxicity' selected and circled in red. The 'Structure' panel on the right displays the chemical structure of 2,4,6-trimethylphenol. A blue callout box at the bottom contains the following instructions:

1. Select the row corresponding to the *Aquatic Toxicity*;
2. Tick the checkboxes of the suitable and plausible profilers;
3. Click **Apply**.

Profiling

Profiling the target chemical

- The actual profiling will take several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appear as a dropdown box under the target chemical.
- Please note the specific profiling results of *Classification by ECOSAR* and *MOA by OASIS* (see next slide).
- These results will be used to search for suitable analogues in the next steps of the exercise.

Profiling

Profiles of the target "3-ethyl-5-methyl-4-methoxyphenol"

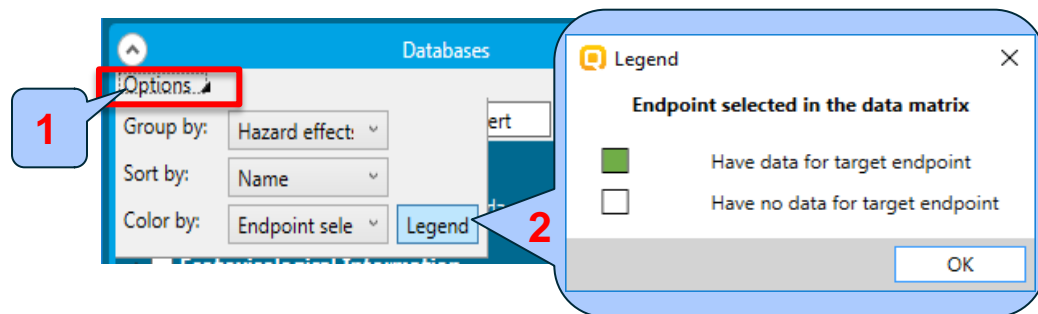
The screenshot displays the QSAR Toolbox interface for profiling the target compound, 3-ethyl-5-methyl-4-methoxyphenol. The top navigation bar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows 'Profiling methods' with a list of options, including 'Suitable' and 'Plausible' categories. The central 'Filter endpoint tree...' panel shows a hierarchical list of endpoints, with 'Acute aquatic toxicity MOA by OASIS' and 'Aquatic toxicity classification by ECOSAR' circled in red. The right panel shows the chemical structure of the target and a list of classification results, including 'Class 3 (unspecific reactivity)', 'Phenols and Anilines', and 'Phenols'. The 'Phenols and Anilines' result is circled in red.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - **Data**

Data

- *Data* module refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox databases.
- Data gathering can be executed in a global fashion (i.e. collecting all data of all endpoints) or on a more narrowly defined basis (i.e. collecting data for a single or limited number of endpoints).
- Once the endpoint is selected, the relevant databases are highlighted. Meaning of the colors could be seen within the **Options** (1) by click **Legend** (2).



- In this example, we limit our data gathering to the common aquatic toxicity endpoints from databases containing aquatic toxicity data (Aquatic ECETOC, Aquatic Japan MoE, ECOTOX, and Aquatic OASIS).

Data

Gather data using database` relevancy

The screenshot shows the QSAR Toolbox interface. At the top, there is a navigation bar with icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this is a menu bar with 'Data', 'Import', and 'Export'. The 'Data' menu is open, showing 'Gather', 'Import', 'IUCLID6', and 'IUCLID6'. A red circle and callout '3' highlight the 'Gather' button. On the left, a 'Databases' list is shown with several options checked: 'Aquatic ECETOC', 'Aquatic Japan MoE', 'Aquatic OASIS', and 'ECOTOX'. A red circle and callout '2' highlight these four checked items. In the center, the 'Filter endpoint tree...' is open, showing a tree structure with 'Aquatic Toxicity' selected. A red circle and callout '1' highlight this selection. On the right, a table titled '1 [target]' shows a chemical structure of 2,4-dimethylphenol and a row for 'Aquatic Toxicity' with the value 'AW SW'. A red circle and callout '1' highlight this row.

1. Select the row corresponding to the *Aquatic toxicity*; 2. Select all highlighted databases except ECHA CHEM; 3. Click **Gather**.

Data

Process of collecting data

Toxicity information on the target chemical is electronically collected from the selected datasets.

A window with "Read data?" appears. Now the user could choose to collect "all" or "endpoint specific" data.

The screenshot displays the QSAR Toolbox interface. The top menu bar includes options like Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, there are sections for Documents, Databases, and Options. The 'Options' section is expanded, showing a list of databases and endpoints. The 'Ecotoxicological Information' section is checked, and 'Aquatic ECETOC' is selected. A 'Read data?' dialog box is open in the foreground, with the 'All endpoints' radio button selected. The 'from Tautomers' checkbox is also visible. The 'OK' button is highlighted with a blue callout box containing the number '1'. In the background, a 'Filter endpoint tree...' window shows a tree structure with 'Environmental Fate and Transport' and 'Ecotoxicological Information' expanded. A chemical structure of 4-(2-methoxyphenyl)phenol is shown in the top right corner.

1. Click **OK** to read all available data

Data

Process of collecting data

In this example, an insert window appears stating that no experimental data is available for the chemical of interest.

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Data', 'Import', and 'Export'. Below the menu bar, there are icons for 'Gather', 'Import', 'IUCLID6', and 'IUCLID6'. The main window is divided into several panes. On the left, there is a 'Documents' pane and a 'Databases' pane with a list of options including 'Hydrolysis rate constant OASIS', 'kM database Environment Canada', 'Phys-chem EPISUITE', 'Ecotoxicological Information', and 'Human Health Hazards'. The 'Ecotoxicological Information' section is expanded, showing 'Aquatic ECETOC', 'Aquatic Japan MoE', 'Aquatic OASIS', 'ECHA CHEM', and 'ECOTOX' as checked options. The 'Human Health Hazards' section is also expanded, showing 'Acute Oral toxicity', 'Bacterial mutagenicity ISSSTY', 'Biocides and plant protection ISSBIOC', 'Carcinogenic Potency Database (CPDB)', 'Carcinogenicity&mutagenicity ISSCAN', 'Cell Transformation Assay ISSCTA', and 'Genotoxic cells CPDTA'. The 'Filter endpoint tree...' pane is visible, showing a tree structure with 'Aquatic Toxicity', 'Sediment toxicity', 'Terrestrial Toxicity', and 'Human Health Hazards'. The 'Structure' pane shows a chemical structure of 2-(4-hydroxyphenyl)propane-1-ol. The '1 [target]' pane shows the chemical structure. A dialog box is open in the center of the screen with the message: 'There is no experimental data available for the chemicals of interest.' The dialog box has an 'OK' button. A callout box with the number '1' points to the 'OK' button.

1. Click **OK** to close the window

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Data
 - **Category definition**

Recap

- You have entered the target chemical being sure of the correct structure.
- You have profiled the target chemical and found no experimental data is currently available for this structure.
- In other words, you have identified a data gap, which you would like to fill.
- Now you are ready to continue with next step of the workflow - *Category Definition*.

Category Definition Overview

- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- This is the critical step in the workflow.
- Several options are available in the Toolbox to assist the user in defining the category definition.

Category Definition

Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity” so that within a category data gaps can be filled by trend-analysis.
- For this example, starting from the target chemical a specific ECOSAR classification is identified, subsequently analogues are found within the same specific classification for which experimental results are available.

Category Definition

ECOSAR categories

- ECOSAR has 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.
- The *Aquatic toxicity classification by ECOSAR* profiling scheme in the Toolbox is used for grouping of chemicals by structural similarity which may or may not have mechanistic meaning. Experience has shown ECOSAR to be a robust profiler which makes it a logical choice in an initial profiling scheme.

Category Definition

Defining ECOSAR category

The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Category definition' step is active. In the 'Categorize' section, the 'Define' button is circled in red and labeled with a '2'. The main list of categories is titled 'Aquatic toxicity classification by ECOSAR'. The 'Aquatic toxicity classification by ECOSAR' entry is highlighted in blue and labeled with a '1'. The 'Grouping options' dialog box is open, showing 'Phenols' in the 'Target' field, which is circled in red. The 'Combine profiles' section has 'AND' selected. The 'OK' button is labeled with a '3'. A chemical structure of a phenol derivative is shown on the right.

1. Select the highlighted **Aquatic toxicity classification by ECOSAR**; 2. Click **Define**; 3. Confirm the category "Phenols" by clicking **OK**.

Category Definition

Defining ECOSAR category

The screenshot shows the QSAR Toolbox interface with the 'Category definition' workflow selected. The 'Filter endpoint tree...' window is open, showing a list of endpoints. A 'Grouping results' dialog box is displayed, indicating that 768 chemicals were found. A blue callout box with the number '1' points to the 'OK' button in the dialog box. A large blue box at the bottom of the image contains the instruction: '1. Click **OK** to confirm the result and to gather experimental data'.

Category Definition

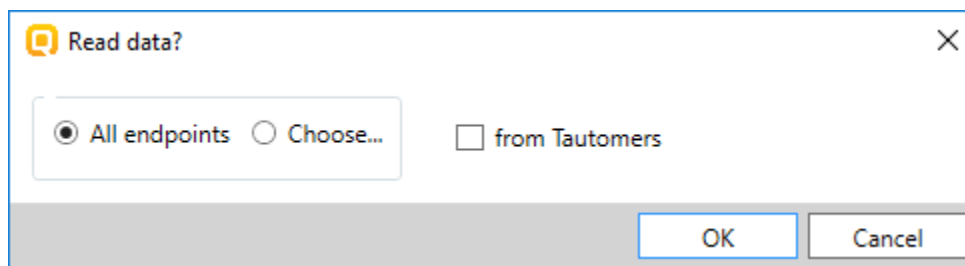
Analogues

- The Toolbox now identifies all chemicals corresponding to the ECOSAR classification of *Phenols* which are listed in the selected databases within the *Data* module.
- 768 analogues are identified. Along with the target they form a category (Phenols) which can be used for data gap filling.

Category Definition

Read data for Analogues

- The Toolbox automatically request the user to select the endpoint that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).



- In this example, since only databases that contain information for Eco-toxicological endpoints are selected, both options give the same results.
- As the Toolbox must search the database, this may take some time.

Category Definition

Summary information of Analogues

The screenshot shows the QSAR Toolbox software interface. The main window displays the 'Category definition' workflow. The 'Filter endpoint tree...' window is open, showing a table of chemical structures and their associated parameters. A 'Gather data' dialog box is overlaid on the table, showing '49927 points added across 767 chemicals.' A red '1' in a blue box points to the 'OK' button in the dialog.

Filter endpoint tree...	1 [target]	2	3	4	5	6
Structure						
Structure info						
Parameters						
Physical Chemical Properties						
Environmental Fate and Transport	(50/1039)		M: >0.699+1 log(L/k)			
Ecotoxicological Information	(756/45713)	M: 10 mg/L	M: <4 g/eu	M: 0.5 mg/L	M: 0.25 mg/kg bdwt	M: 10 mg/l
Human Health Hazards	(94/3175)		M: ≥100 mg/kg bdw		M: 0.25 mg/kg bdwt	

Gather data dialog box: 49927 points added across 767 chemicals. OK

1. Click **OK** on window that provides information for common number gathered data across the number of chemicals.

Category Definition

Summary information of Analogues

The screenshot displays the QSAR Toolbox interface with the following components:

- Navigation Bar:** Includes icons for Input, Profiling, Data, Category definition (active), Data Gap Filling, and Report.
- Documents Panel:** Shows 'Aquatic toxicity classification by ECOSAR' with various options like 'Predefined', 'General Mechanistic', and 'Hydrolysis half-life'.
- Filter endpoint tree...:** A tree view where 'Aquatic Toxicity' is expanded. A callout points to 'AW SW (688/39881)'.
- Data Matrix Table:**

	1 [target]	2	3
Structure			
Structure info			
Parameters			
Physical Chemical Properties			
Environmental Fate and Transport (50/1039)			M: >0.699+1
Ecotoxicological Information			
Aquatic Toxicity		M: 10 mg/L	M: >30 mg/L
Sediment toxicity			
Terrestrial Toxicity (233/5832)			M: <4 g/eu
Human Health Hazards (94/3175)			M: ≥100 mg/
- Callout Box:**

Chemical statistics presents the number of chemicals and the available experimental data. This is statistics for the current row on data matrix.

Category Definition

Summary information of Analogues

The screenshot shows the QSAR Toolbox software interface. The main window is titled 'Aquatic toxicity classification by ECOSAR'. The 'Filter endpoint tree...' panel is open, showing a list of endpoints. A red circle highlights the 'M: 10 mg/L' value in the 'Aquatic Toxicity' row. A blue callout box with the number '1' points to this cell. A 'Data points' dialog box is open, showing a table of data points for the selected chemical.

Datapoints	#	Value	Original value	Additional comments
Ecotoxicological Information/Aquatic Toxicity	1	M: 5 ppm (Volume concentration)	5 ppm (Volume concentration)	
Ecotoxicological Information/Aquatic Toxicity	2	M: 5 ppm (Volume concentration)	5 ppm (Volume concentration)	
Ecotoxicological Information/Aquatic Toxicity	3	M: 5 ppm (Volume concentration)	5 ppm (Volume concentration)	
Ecotoxicological Information/Aquatic Toxicity	4	M: 10 mg/L (Mass concentration)	1E+04 µg/L (Mass concentration)	AERATED//TESTED IN POLYETHYLENE BAGS//CONC. TESTED//EFFECT/LOSS OF EFFICACY OCCURRED IN 0-1 H//
Ecotoxicological Information/Aquatic Toxicity	5	M: 10 mg/L (Mass concentration)	1E+04 µg/L (Mass concentration)	AERATED//TESTED IN POLYETHYLENE BAGS//CONC. TESTED//EFFECT/DEATH OCCURRED IN 0-1 H//
Ecotoxicological Information/Aquatic Toxicity	6	M: 10 mg/L (Mass concentration)	1E+04 µg/L (Mass concentration)	AERATED//TESTED IN POLYETHYLENE BAGS//CONC. TESTED//EFFECT/DEATH OCCURRED IN 0-1 H//

1. Double-click on the cell with measured data opens a table which provides detailed information for all experimental data of the focused chemical.

Recap

- You have identified a category (*Phenols*) with the *Aquatic toxicity classification by ECOSAR* profiler for the target chemical 3-ethyl-5-methyl-4-methoxyphenol.
- The available experimental results for these 767 analogues have been collected from the selected databases (Aquatic ECETOC, Aquatic Japan MoE, ECOTOX, and Aquatic OASIS).
- But before the user can proceed with the *Filling Data Gap* module, he/she should navigate through the endpoint tree and find the specific gap that will be filled.

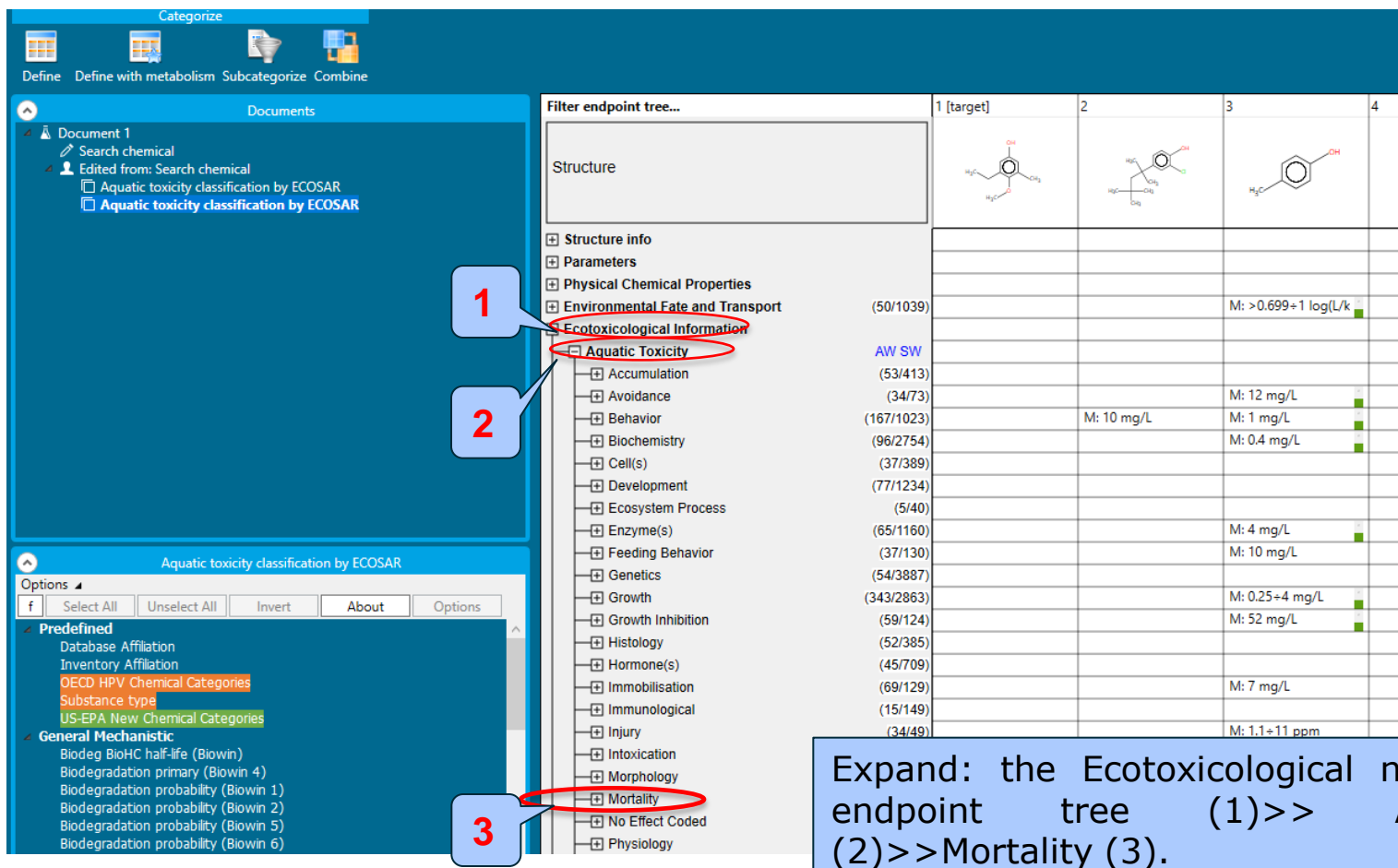
Category Definition

Navigation through the endpoint tree




- The user can navigate through the data tree by opening (or closing) the nodes of the tree.
- The data tree is extensive but logically constructed. It can be mastered with a practice.
- In this example, the “48 h LC50 Mortality for *Daphnia magna*” is the target endpoint.
- You can navigate through the endpoint tree by typing the species “*Daphnia magna*” in the “Filter endpoint tree...” box and clicking (Aquatic Toxicity, Mortality, LC50, 48 h, Animalia, etc to *Daphnia magna* - the specific endpoint (see next two slides).

Category Definition

Navigation through the endpoint tree



1 Expand: the Ecotoxicological node of endpoint tree (1)>> Aquatic (2)>>Mortality (3).

Filter endpoint tree...	1 [target]	2	3	4
Structure				
Structure info				
Parameters				
Physical Chemical Properties				
Environmental Fate and Transport (50/1039)			M: >0.699+1 log(L/k)	
Ecotoxicological Information				
Aquatic Toxicity (AW SW)				
Accumulation (53/413)				
Avoidance (34/73)			M: 12 mg/L	
Behavior (167/1023)		M: 10 mg/L	M: 1 mg/L	
Biochemistry (96/2754)			M: 0.4 mg/L	
Cell(s) (37/389)				
Development (77/1234)				
Ecosystem Process (5/40)				
Enzyme(s) (65/1160)			M: 4 mg/L	
Feeding Behavior (37/130)			M: 10 mg/L	
Genetics (54/3887)				
Growth (343/2863)			M: 0.25+4 mg/L	
Growth Inhibition (59/124)			M: 52 mg/L	
Histology (52/385)				
Hormone(s) (45/709)				
Immobilisation (69/129)			M: 7 mg/L	
Immunological (15/149)				
Injury (34/49)			M: 1.1+11 ppm	
Intoxication				
Morphology				
Mortality				
No Effect Coded				
Physiology				

Category Definition

Navigation through the endpoint tree

The image displays two screenshots of the QSAR Toolbox interface. The left screenshot shows the 'Filter endpoint tree...' dialog box with a 'Structure' field. Below it, a tree view lists various endpoints. The 'LC50' endpoint is highlighted with a red circle and a callout box labeled '4'. The right screenshot shows the expanded tree view. The '48 h' endpoint is highlighted with a red circle and a callout box labeled '5'. Below it, the 'Arthropoda (arthropods)' category is highlighted with a red circle and a callout box labeled '6'. Under 'Arthropoda', the 'Branchiopoda (branchiopods)' category is highlighted with a red circle and a callout box labeled '7'. Finally, the 'Daphnia magna' species is highlighted with a red circle and a callout box labeled '8'.

>> LC50 (4) >> 48h (5) >> Arthropoda (arthropods) (6) >> Branchiopoda (branchiopods) (7) >> **Daphnia magna** (8) - this is the species related to target endpoint.

Recap

- You have now retrieved the available experimental data on aquatic toxicity for 767 chemicals classified as “phenols” by the “Aquatic toxicity classification by ECOSAR” profiler found in the databases Aquatic ECETOC, Aquatic Japan MoE, ECOTOX, and Aquatic OASIS.
- You have identified the target endpoint of “48 h LC50 Mortality for *Daphnia magna*”.
- You are ready to fill in the data gap so click *Data Gap Filling* (see next slide).

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Data
 - Category definition
 - **Data Gap Filling**

Data Gap Filling

Overview

- *Data Gap Filling* module gives access to five different data gap filling tools:
 - Read-across
 - Trend analysis
 - (Q)SAR models
 - Standardized workflow (SW)
 - Automated workflow (AW)
- The most relevant data gap mechanism is used , taking into account the following considerations:
 - *Read-across* is the appropriate data-gap filling method for “qualitative” endpoints like skin sensitisation or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore read-across is recommended for “quantitative endpoints” (e.g., 96h-LC50 for fish) if only a low number of analogues with experimental results are identified.
 - *Trend analysis* is the appropriate data-gap filling method for “quantitative endpoints” (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.
 - *(Q)SAR models* can be used to fill a data gap if no adequate analogues are found for a target chemical.
 - *Automated and standardized workflows* follow preliminary implemented logic. The AW is not affected by the user activities (proceeding or subsequent), while the SW stops at the each step of the workflows allowing the user to make different selection.
- In this example we use trend analysis.

Data Gap Filling

Apply Trend analysis

The screenshot displays the QSAR Toolbox interface during a Data Gap Filling session. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' section is active, showing 'Trend analysis' and 'Standardized Automated' options. A 'Documents' panel on the left shows a document titled 'Aquatic toxicity classification by ECOSAR'. The main workspace is divided into a 'Filter endpoint tree...' on the left and a data table on the right. The tree lists various taxonomic endpoints, with 'Daphnia magna' highlighted in blue. The data table shows columns for target, chemical, and data values. A 'Possible data inconsistency' dialog box is open, listing units such as M, mg/L, mmol/L, mol/L, ppm, μM, and μg/L. The 'log(1/mol/L)' unit is selected. The dialog also shows 'Data 239/250; Chemicals 56/56' and 'OK'/'Cancel' buttons.

1. Click the "data endpoint box" corresponding to *Daphnia magna*/ LC50 /48h under the target chemical; 2. Click **Trend analysis**; 3. A pop-up window alerting you to possible data inconsistencies appears. Click **OK**.

Data Gap Filling

Results of Trend analysis

The screenshot displays the QSAR Toolbox interface. On the left is a 'Filter endpoint tree...' with a 'Structure' section and a list of taxonomic groups and endpoints. The main area is a table with columns for chemical structures and numerical data. A pop-up 'Information' dialog box is overlaid on the table, stating: '4 observed values for 1 chemical were excluded due to missing X descriptor value(s)'. An 'OK' button is visible in the dialog, with a blue callout box containing the number '1' pointing to it.

Filter endpoint tree...	1 [target]	3	23	40	56	72	83	95	
Structure									
Malacostraca (21/68)			M: 9.89 (8.5+11.2) ppm			M: 0.0913 mg/L	M: > 1.8 mg/L		
Maxillopoda (3/4)									
Ostracoda (ostracods) (4/4)			M: 26 ppm						
Chordata (chordates) (37/503)		M: 14 mg/L	M: > 5+10 mg/L			M: 0.352 mg/L	M: > 26 mg/L		
Cnidaria (cnidarians, coelenterates) (3/6)									
Mollusca (molluscs, mollusks) (11/28)					M: 550 (520+ 579) mg/L				
Nematoda (2/7)									
Platyhelminthes (flatworms) (4/6)									
Rotifera (rotifers, wheel animalcul) (3/13)									
Undefined Kingdom (19/206)									
50 h (2/4)									
52 h (1/1)									
54 h (1/1)									
59 h (1/1)									
66 h (1/1)									
68 h (1/1)									
72 h (28/346)							M: > 2 mg/L	M: 6.3	
84 h (2/3)									
85 h (1/3)									
96 h (52/1494)		M: 14 (11.8+16.6) mg/L	M: 11.7 mg/L	M: 1.6 (1.4+1.8) mg/L	M: 206 (179+238) mg/L	M: 0.0734 mg/L	M: 28 mg/L	M: 1.5 mg/L	M: 17
100 h (2/3)									
4.2 d (1/1)									
4.67 d (1/2)									
5 d (6/23)							M: > 28 mg/L		
5.83 d (1/1)									
6 d (6/7)							M: 11 (9.4+14) mg/L		
6.33 d (1/1)									
7 d (17/72)			M: > 22 mg/L				M: 3.73 mg/L	M: 1.1	
7.21 d (1/1)									
7.5 d (1/2)									
4+11 d (1/2)									
8 d (14/33)			M: 49 (46+53) mg/L				M: 6.3 (6.1+6.6) mg/L	M: 4.5 (3.8+5.4) mg/L	
9 d (3/4)							M: 6.8 (5.8+8) mg/L		

A pop-up message informs about the number of data points across the number of chemicals (e.g. mixtures or UVCB substances) that will not be included in the Trend analysis prediction due to missing X descriptor value(s), which by default is LogKow. Click OK (1).

Data Gap Filling

Interpreting Trend analysis

- The resulting plot outlines the log of the experimental LC50 results of all analogues (Y axis) according to a descriptor (X axis) with Log *Kow* being the default descriptor (see next slide).
- The **RED** points represents the predicted value for the target chemical.
- The **BLUE** points represent the experimental results available for the analogues used in the trend analysis.
- The square-shaped signs in the right side of the data gap filling window are the so-called “helpers”. The helpers are notifying messages that provide different type of information related to the used data points in the prediction.
- Before accepting the estimated result for the target chemical, the trend analysis should be further refined by subcategorization (see the next slides).

Data Gap Filling

Results of Trend analysis

The screenshot displays the QSAR Toolbox interface during the 'Data Gap Filling' process. The top navigation bar shows the current step. The main area is divided into a 'Filter endpoint tree...' on the left, a central data table, and a right-hand panel with warning messages. The data table lists various chemicals and their corresponding values for different endpoints. The right-hand panel contains several yellow warning boxes with exclamation marks, providing information about endpoint values and substance types. A red bracket on the right side of the interface is labeled with a blue callout box containing the number '1'. A blue callout box with the number '2' points to a 'Remove data points' button in one of the warning boxes. Below the table, a 'Trend analysis prediction for LC50' plot is shown, featuring a scatter plot of log(C/mo/L) vs log Kow with a red regression line. The plot includes a predicted value of 1.61 mg/L and a model equation. A 'Statistics' sidebar is visible on the far right, and a green checkmark with the text 'Accept prediction' is at the bottom right.

1. Click the helpers to see the information that they provide; 2. Click **Remove data points** in order to eliminate the data points that have values bigger than WS. Once the data points are removed the helpers disappear.

Data Gap Filling

Subcategorization

- Remember in the Toolbox, a category refers to a group of chemicals which have the same profiling result according to one of the profilers listed in the module *Profiling*.
- Subcategorization refers to the process of applying additional profilers to the previously defined category. The subcategorization identifies chemicals which have differing profiling results and eventually eliminating these chemicals from the final category.

Data Gap Filling

Subcategorization

In this example, subcategorization allows for the elimination of analogues which are dissimilar to the target chemical with respect to:

- Substance type (mixtures and hydrolyzing chemicals)

The categorisation based on substance type allows keeping among the analogues only those that are of the same chemical type: discrete chemical, organic, mixture, polymer(predefined), inorganic, mono constituent(predefined) or multi constituent(predefined). The current target is a discrete chemical, organic, mono constituent(predefined) hence the analogues should also be discrete chemicals.

- OASIS Mode of action (all except phenols and anilines)

The categorization based on mode of action identifies analogues having the same mode of action as the target which is in the group of phenols and anilines.

- Chemical elements

The profiler aimed to identify analogues consisting of same elements as those presented in the target chemical

Subcategorization is demonstrated in the next 4 slides.

Data Gap Filling

Subcategorization by Substance type

The screenshot displays the 'Subcategorization' window in the QSAR Toolbox. It is divided into several sections:

- Options Panel (Left):** Contains 'Predefined' and 'General Mechanistic' categories. 'Substance type' is highlighted with a red circle and callout '3'. Below it, 'Documented' and 'Simulated' categories are listed.
- Adjust options Panel (Top Right):** Shows 'Discrete chemical' and 'Mono constituent (predefined)' as selected categories.
- Data Table (Top Right):** A table with columns for chemical properties and predicted values. Callout '1' points to the table area.
- Trend Analysis Plot (Bottom Center):** A scatter plot titled 'Trend analysis prediction for LC50, based on 54 values'. The predicted value is 1.61 mg/L. The model equation is $LC50 = 3.77 (\pm 0.358) + 0.392 (\pm 0.113) * \log Kow, \log(1/\text{mol/L})$. Callout '2' points to the plot area.
- Control Panel (Bottom Right):** Contains buttons for 'Select / filter data', 'Subcategorize', 'Mark chemicals by WS', 'Mark chemicals by descriptor value', 'Mark outliers', 'Mark chemicals by test conditions', 'Mark focused chemical', 'Mark focused points', 'Remove marked data', and 'Accept prediction'.

1. Open **Select/filter data**; 2. Select **Subcategorize**; 3. Select **Substance type**

Data Gap Filling

Subcategorization by Substance type

The screenshot displays the QSAR Toolbox interface for subcategorization. On the left, the 'Subcategorization' panel shows various options, with 'Substance type' highlighted under 'Predefined' (callout 3). The main area features a scatter plot of log Kow values. A red line indicates a target trend, and a single outlier point is highlighted with a blue callout (1). A callout (2) points to the 'Data points' window for this outlier, which shows its SMILES, name, CAS number, and a table of properties.

Data points window details:

- SMILES: [Na]Oc1c(Cl)c(Cl)c(Cl)c(Cl)c1Cl
- Chem.name(s): 2,3,4,5,6-Pentachlorophenol sodium salt (1:1), PENTACHLOROPHENOL, NA SALT, Pentachlorophenol, Sodium salt
- CAS Number: 131-52-2

Name	Source	Units	Value	Endpoint	Reference	Units	Value
log Kow	Toolbox calculator		2.05	LC50	Endpoint obs. dat	log(1/mol/L)	5.86
Molecular Weight	Toolbox calculator	Da	288.3				

At the bottom of the interface, a red circle highlights the 'Remove selected' button (callout 3), and a green checkmark indicates 'Accept prediction'.

1. **Double click** on the highlighted outlier to see why this chemical is different to the target. The chemical is dissociating chemical, mono constituent and organic and should to be eliminated being different substance type compared to the target. 2. Click **Close**; 3. Click **Remove selected** to eliminate dissimilar chemicals.

Data Gap Filling

Subcategorization by Acute-aquatic toxicity MOA

The screenshot displays the 'Subcategorization' window in the QSAR Toolbox. On the left, the 'Endpoint-specific' section is expanded, and 'Acute aquatic toxicity MOA by OASIS' is circled in red, with a callout '1'. Below it, the 'Documented' and 'Simulated' sections are visible. In the center, a list of categories is shown, with '(49) Phenols and Anilines' and '(3) Reactive unspecified' selected. A blue box highlights the 'Remove selected' button, with a callout '2'. On the right, a scatter plot titled 'Trend analysis prediction for LC50, based on 53 values' shows a positive correlation between log Kow and LC50. The predicted LC50 is 1.69 mg/L. A table at the top of the window lists chemical data with columns for LC50 values and their corresponding log Kow values.

1. Select **Acute aquatic toxicity MOA by OASIS**; 2. Click **Remove selected** to eliminate the dissimilar chemicals.

Data Gap Filling

Subcategorization by Chemical elements

3

4

1

2

674	675	677	695	696	702
<chem>Oc1ccc(Cl)c(Cl)c1</chem>	<chem>Oc1cc(Cl)c(Cl)c(Cl)c1</chem>	<chem>Oc1c(Cl)c(Cl)c(Cl)c(Cl)c1</chem>	<chem>Oc1ccc(Cl)cc1</chem>	<chem>Oc1ccc(Cl)cc1</chem>	<chem>Oc1ccc(Cl)cc1</chem>
M: 2.21 mg/L	M: 0.68 mg/L	M: 0.038 mg/L	M: 3.1 (2.7+3.7) mg/	M: 3.66 mg/L	M: 0.45 mg/L
		M: 0.246 mg/L			
		M: 0.111 mg/L			
		M: 0.17 mg/L			
		M: 0.11 mg/L			
		M: 0.0056 mg/L	M: 0.56 (0.33+0.95) mg/		
		M: 0.2 mg/L	M: 0.194 mg/L		
		M: 13 (11.8+13.1) mg/L			
		M: <0.1 mg/L			
		M: 0.73 mg/L			
		M: 0.3 mg/L			
		M: >9.19 mg/L			
		M: 0.13 mg/L			

analysis prediction for LC50, based on 49 values
 predicted: 1.76 mg/L
 $ln LC50 = 3.18 (\pm 0.415) + 0.565 (\pm 0.127) * \log Kow, \log(1/mol/L)$

Select / filter data
 Subcategorize
 Mark chemicals by WS

1. Right click over the outlier; 2. Select **Differences to target**; 3. Appearing of new subcategorization window and then select **Chemical elements**; 4. Click **Remove selected** to eliminate the dissimilar

Data Gap Filling Results

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Gap Filling Workflow

Trend analysis Read across (QSAR) Standardized Automated

Documents

Document 1
Created from: Search chemical
Aquatic toxicity classification by ECOSAR
Enter GF(TA) with 56 chemicals, 235 data points
Ch: 50] Data: 55 Filter by WS - Water Solubility
Ch: 55] Data: 54 Filter by WS - Water Solubility (fragments)
Ch: 54] Data: 53 Subcategorized: Substance type
Ch: 50] Data: 49 Subcategorized: Acute aquatic toxicity MOA by OASIS
Ch: 21] Data: 20 Subcategorized: Chemical elements

Data Gap Filling Settings

Only endpoint relevant
 Only chemical relevant

At this position:

Select a cell with a rigid (bold) path
Automated workflows
Standardized workflows

Filter endpoint tree...

Structure

Structure	1 [target]	3	134	144	146	151	The prediction is acceptable according to the statistics (Interpolation and R2 ≥ 0.7 and analogues ≥ 10)
Daphnia magna (20/66)	M: 1.4 mg/L	M: 3.54 (3.17+3.95) mg/L	M: 9.93 mg/L	M: 3.64 mg/L	M: 25.9 mg/L	M: 4.23 mg/L	M: 0.0857 mg/L
Daphnia pulex (3/12)	M: 22.7 mg/L						
Daphnia pulex (4/5)	M: 22.7 mg/L						
Lynceus brachyurus (1/1)							
Sida crystallina (1/1)							
Streptocephalus torvicornis(1/1)							
Insecta (insects) (2/59)		M: 4.32 (1.15+26.3) mg/L					
Malacostraca (6/30)		M: 0.85 (0.7+1.03) mg/L					M
Mastilopoda (2/3)							
Ostracoda (ostracods) (1/1)							
Chordata (chordates) (14/160)	M: 14 mg/L	M: 3.2 mg/L		M: 4.28 mg/L		M: >4.5+6.25 mg/L	M: 0.94 mg/L
Cnidaria (cnidarians,coelenterates)(1/2)							
Mollusca (molluscs,mollusks) (4/14)		M: 24.7 (22.4+27.3) mg/L					
Platyhelminthes (flatworms) (2/4)							
Rotifera (rotifers,wheel animalcules)(2/9)							
Undefined Kingdom (5/77)							

Descriptors

Prediction

Adequacy

Cumulative frequency

Residuals

Statistics

Trend analysis prediction for LC50, based on 20 values
Predicted: 3.23 mg/L
Model equation: $LC50 = 2.64 (\pm 0.380) + 0.650 (\pm 0.114) \cdot \log Kow, \log(1/mol/L)$

Select / filter data

Subcategorize

Mark chemicals by WS

Mark chemicals by descriptor value

Mark outliers

Filter points by test conditions

Mark focused chemical

Mark focused points

Remove marked data

Accept prediction

Helper indicating that the prediction is acceptable according to the statistic appears.

Data Gap Filling Results

- The remaining chemicals in the graph now all have a consistent profile relevant for aquatic toxicity (i.e. Substance type, Classification by ECOSAR, MOA by OASIS and Chemical elements).
- By **accepting the prediction** the data gap is filled (see next screen shot).

Data Gap Filling

Accepting prediction result

Filter endpoint tree... 1 [target] 3 134 144 146 151

The prediction is acceptable according to the statistics (interpolation and $R^2 \geq 0.7$ and analogues ≥ 10)

Structure	1 [target]	3	134	144	146	151			
Daphnia magna (20/66)		M: 1.4 mg/L	M: 3.54 (3.17+3.95)	M: 9.93 mg/L	M: 3.64 mg/L	M: 25.9 mg/L	M: 4.23 mg/L	M: 0.0857 mg/L	M
Daphnia pulex (3/12)		M: 22.7 mg/L							
Daphnia pulicaria (4/5)		M: 22.7 mg/L							
Lynceus brachyurus (1/1)									
Sida crystallina (1/1)									
Streptocephalus torvicornis (1/1)									
Insecta (insects) (3/59)			M: 4.32 (1.15+26.3) mg						
Malacostraca (6/30)			M: 0.85 (0.7+1.03) mg/L						M
Maxillopoda (2/3)									
Ostracoda (ostracods) (1/1)									
Chordata (chordates) (14/160)		M: 14 mg/L	M: 3.2 mg/L		M: 4.28 mg/L		M: >4.5+6.25 mg/L	M: 0.94 mg/L	M
Cnidaria (cnidarians,coelenterates)(1/2)									
Mollusca (molluscs,mollusks) (4/14)			M: 24.7 (22.4+27.3) mg						
Platyhelminthes (flatworms) (2/4)									
Rotifera (rotifers,wheel animalcule)(2/9)									
Undefined Kingdom (5/77)									

Descriptors

Prediction

Adequacy

Cumulative frequency

Residuals

Statistics

Trend analysis prediction for LC50, based on 20 values
Predicted: 3.23 mg/L

Confirm

Are you sure you want to accept this prediction?

Yes No

1

2

Accept prediction

1. Click **Accept prediction**; 2. Click **Yes** to confirm the prediction. The prediction is accepted successfully and the system automatically returns you to the data matrix.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Data
 - Category definition
 - Data Gap Filling
 - **Export a prediction to IUCLID6**

Export prediction to the IUCLID 6

Overview

- The OECD QSAR Toolbox allows the users to export predicted data (by means of the Filling Data Gap tools) to IUCLID 6.
- The way of exporting is connect to an IUCLID 6 server (via WebServices) and assigning the predicted endpoint data to a selected substance.
- A wizard will guide the user through the different steps of exporting (see next screenshot).

Exporting the prediction to IUCLID 6 Case study

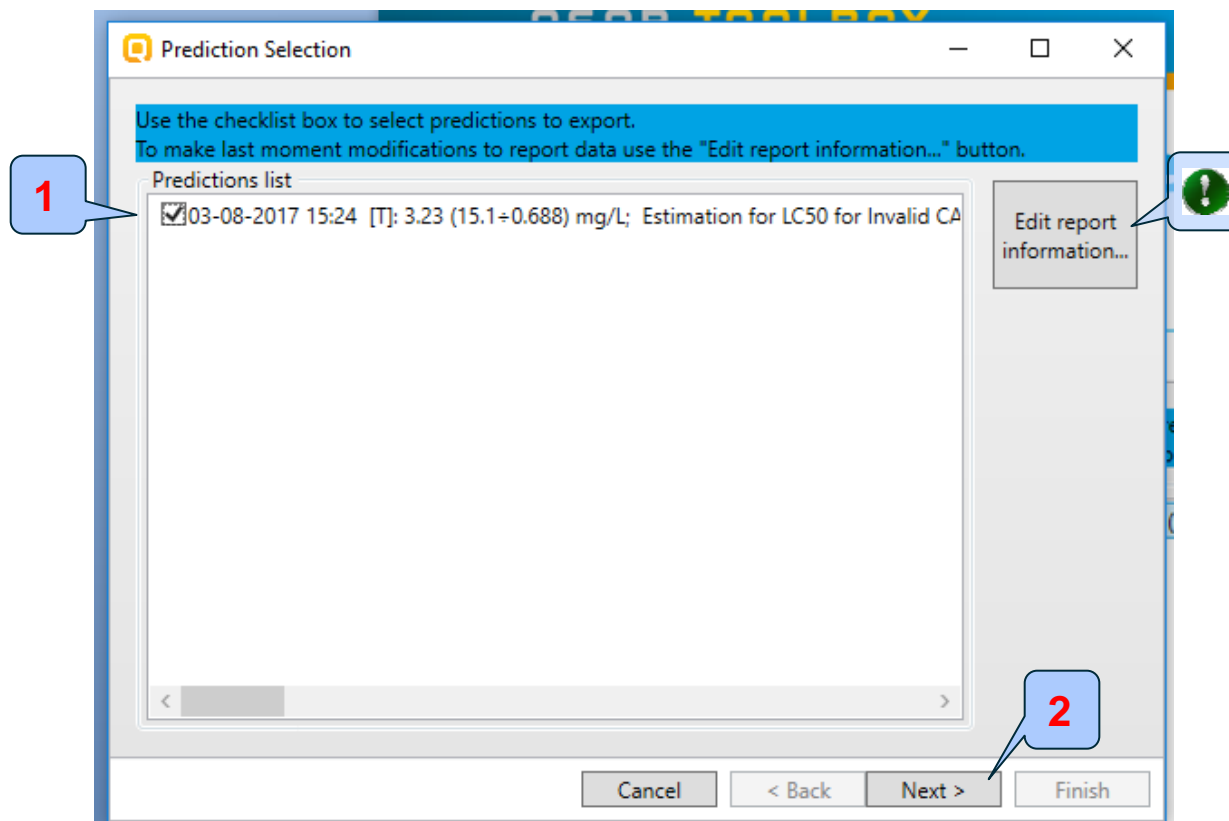
The screenshot shows the QSAR Toolbox interface. In the top menu, the 'Data' section is highlighted with a red dashed box and a callout '1'. In the 'Documents' panel on the left, the 'IUCLID6' icon is highlighted with a red box and a callout '3'. In the 'Filter endpoint tree...' panel, the 'Daphnia magna' entry is highlighted with a blue box and a callout '2'. The table on the right shows the prediction results for 'Daphnia magna'.

Structure	1 [target]	2	3	4	5
35 h (1/1)					
36 h (3/3)					
44 h (3/3)					
46 h (1/1)					
48 h					
Animalia (animals)					
Annelida (annelids) (8/34)					
Arthropoda (arthropods)					
Arachnida (arachnids) (1/15)					
Branchiopoda (branchiopods)					
Artemia franciscana (1/1)					
Artemia parthenogenetica (1/1)					
Artemia salina (8/8)					
Ceriodaphnia cornuta (1/1)					
Ceriodaphnia dubia (12/42)					
Ceriodaphnia pulchella (1/1)					
Ceriodaphnia reticulata (2/5)					
Chydorus sphaericus (1/1)					
Daphnia carinata (2/2)					
Daphnia cucullata (3/4)					
Daphnia longispina (1/1)					
Daphnia magna (57/251)	3.23 (15.1-0.688) mg/L				
Daphnia pulex (9/64)					
Daphnia pulicaria (4/5)					
Lynceus brachyurus (1/1)					
Moina macrocopa (1/1)					
Sida crystallina (1/1)					
Simocephalus vetulus (2/29)					
Streptocephalus torvicornis(2/2)					
Insecta (insects) (11/85)					
Malacostraca (27/84)					
Maxillopoda (6/7)					

1. Go to **Data** section; 2. Click the prediction; 3. Click **IUCLID6**.

Exporting the prediction to IUCLID 6

Case study



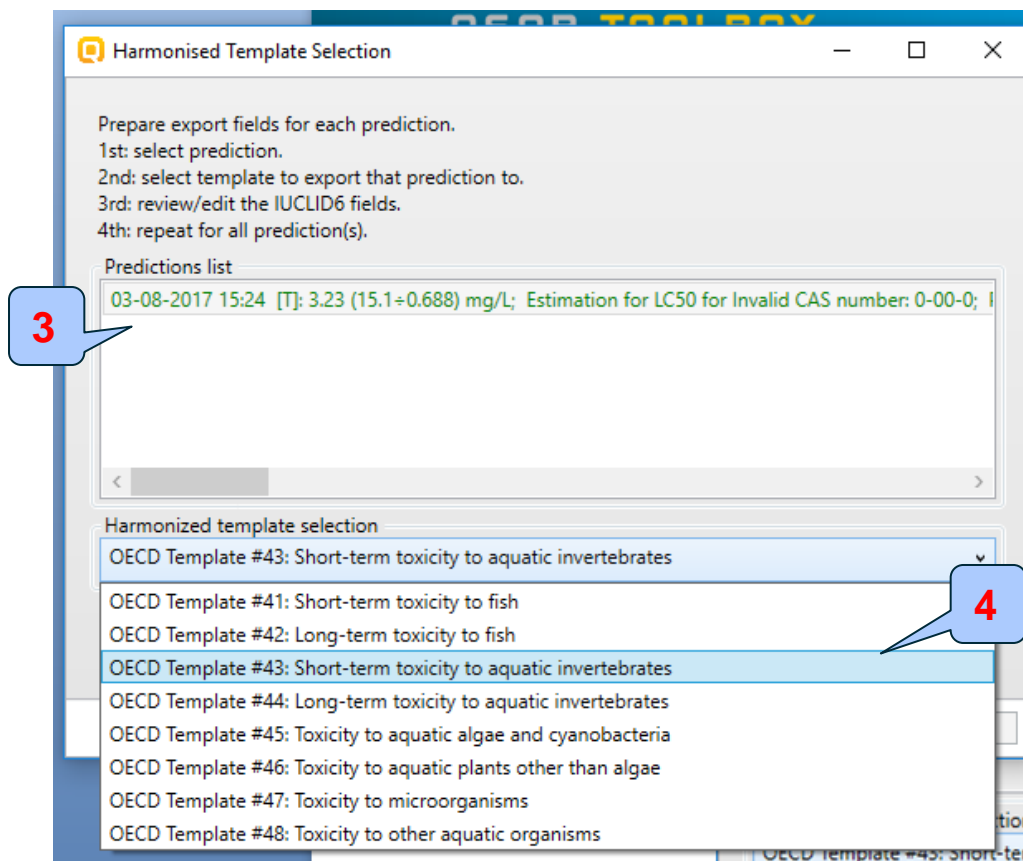
1. Tick the prediction to export 2. Click **Next** to move through the next step of the export.



The user could also edit the report information

Exporting the prediction to IUCLID 6

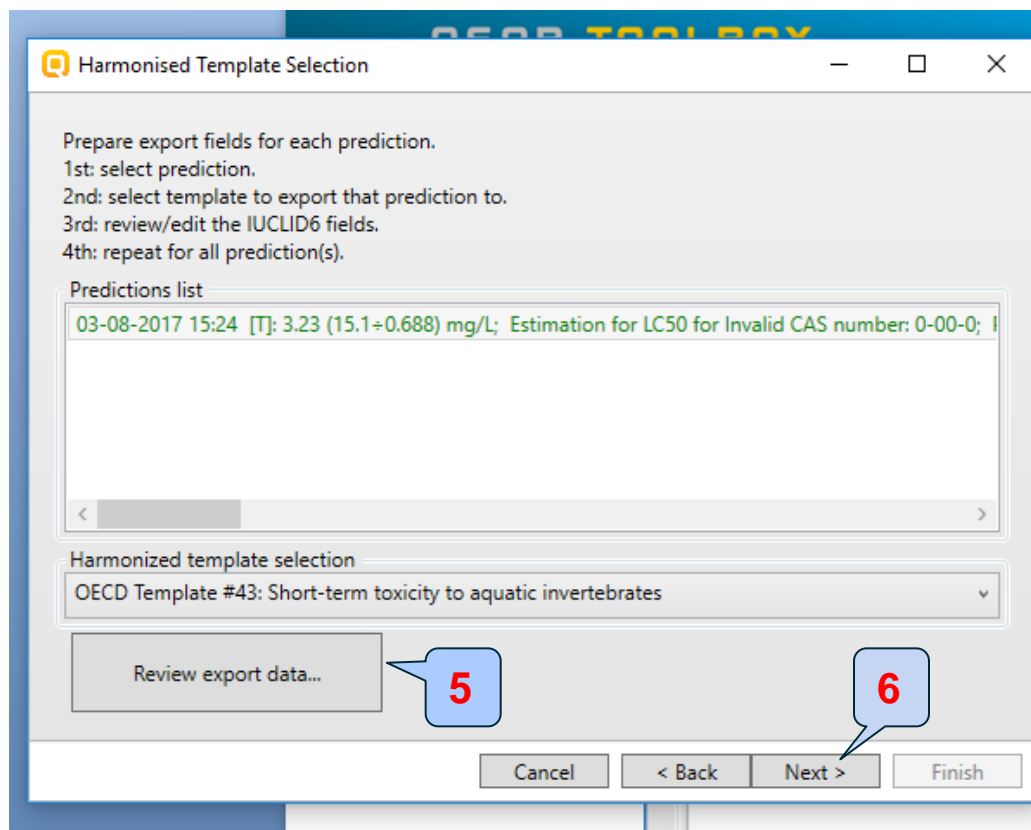
Case study



3. Select a prediction; 4. Select a template to export the prediction

Exporting the prediction to IUCLID 6

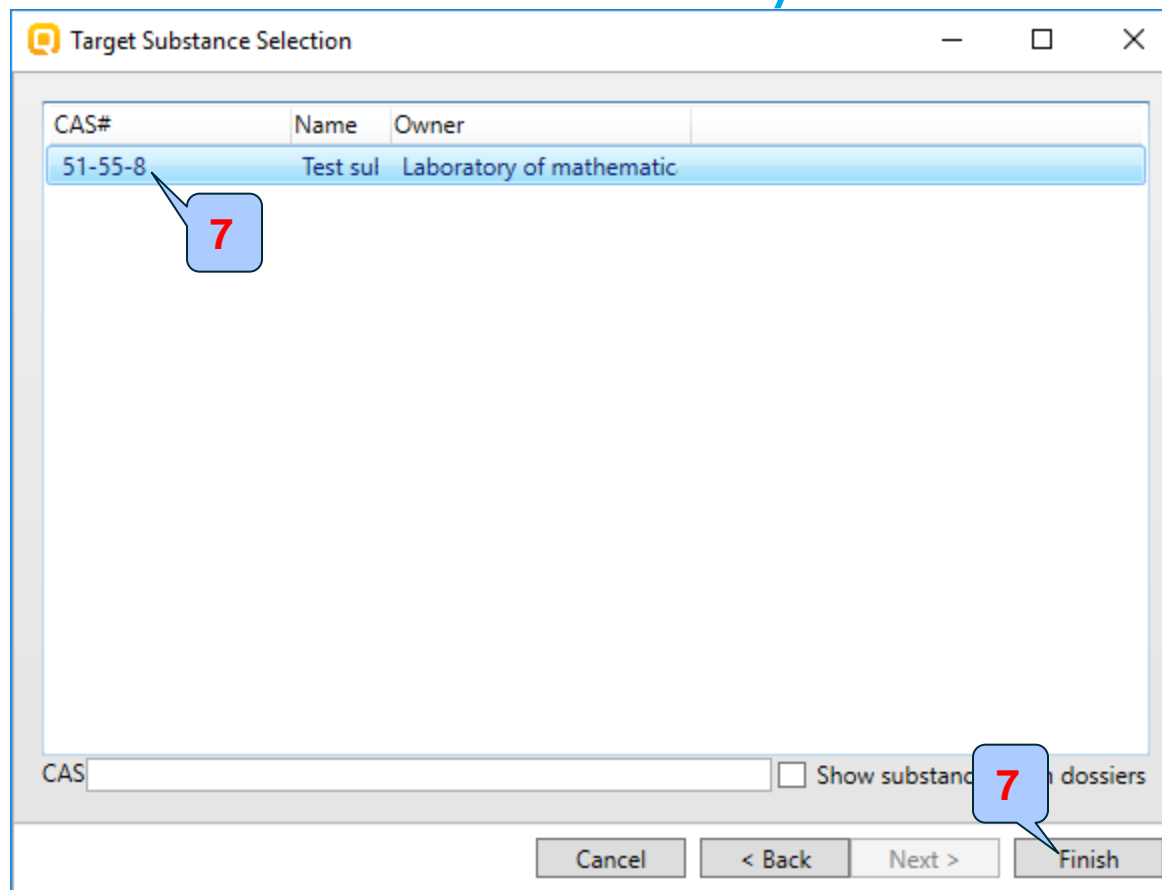
Case study



5. Review/edit the IUCLID6 fields; 6. Click **Next**

Exporting the prediction to IUCLID 6

Case study



7. Select a substance to assigning the predicted endpoint data; 8. Click **Finish**

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Data
 - Category definition
 - Data Gap Filling
 - Export a prediction to IUCLID5
 - **Report**

Report Overview

- Report module could generate report on any of predictions performed with the Toolbox.
- Report module contains predefined report templates, which could be customized.
- The report can then be printed or saved in different formats.

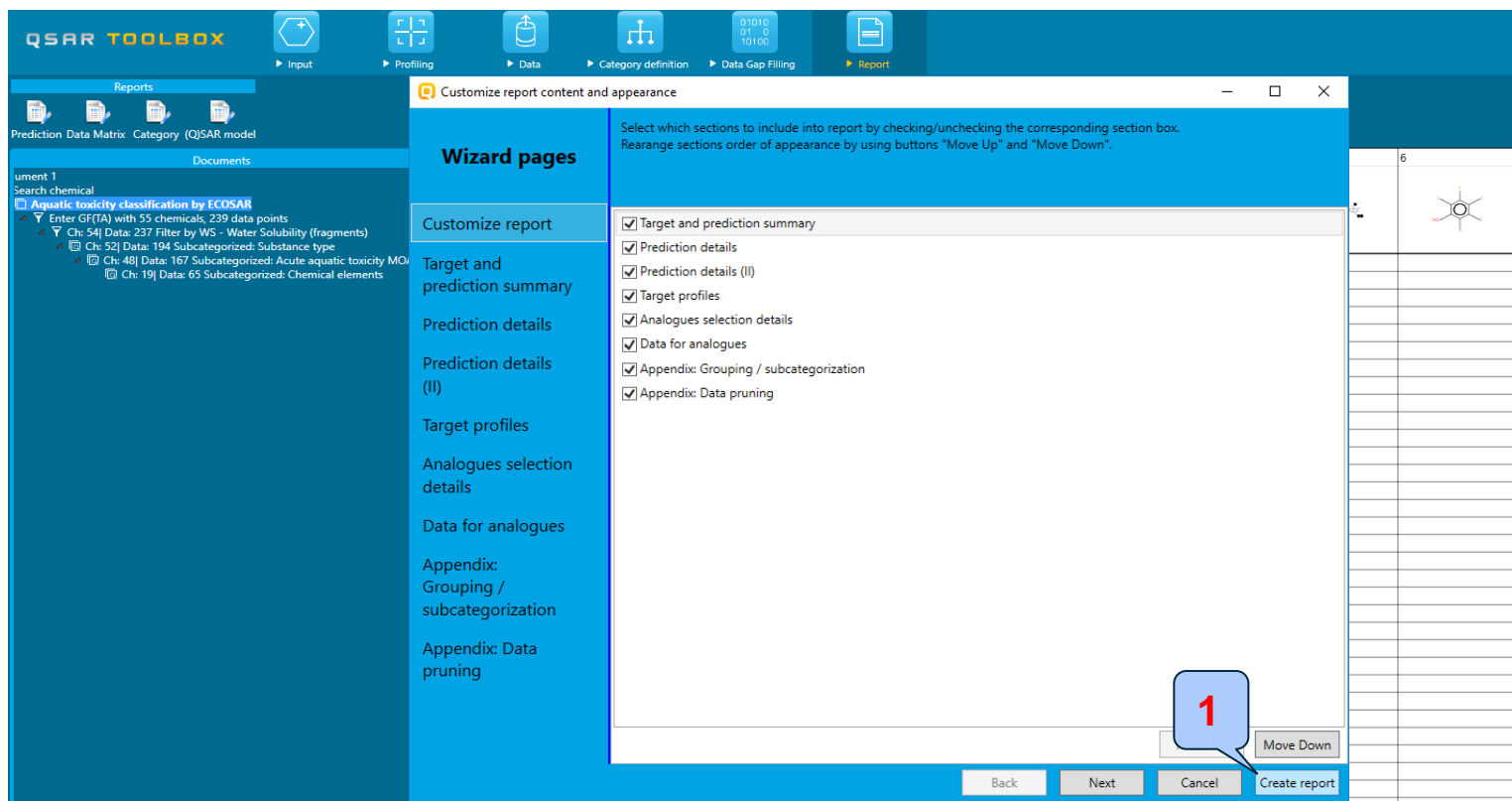
Report Generation report

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Report' module is selected, indicated by a red dashed box and a callout '1'. The left sidebar shows a 'Documents' panel with a tree view of data points, where the 'Prediction' option is highlighted by a red box and callout '3'. The main workspace is divided into three sections: a 'Filter endpoint tree...' on the left, a 'Structure' view in the center, and a table of prediction results on the right. The table has columns for target, structure, and numerical results. The prediction result for 'Daphnia magna' is highlighted by a red oval and callout '2', showing a value of 3.23 (15.1 ± 0.688) mg/L. The table also includes molecular weight (M) values for various species.

Filter endpoint tree...	Structure	1 [target]	2	3	4	5
35 h (1/1)	<chem>Oc1ccc(O)c(O)c1</chem>					
36 h (3/3)						
44 h (3/3)						
46 h (1/1)						
48 h						
Animalia (animals)						
Annelida (annelids) (8/34)						
Arthropoda (arthropods)						
Arachnida (arachnids) (1/15)						
Branchiopoda (branchiopods)						
Artemia franciscana (1/1)						
Artemia parthenogenetica (1/1)						
Artemia salina (6/8)						
Ceriodaphnia cornuta (1/1)						
Ceriodaphnia dubia (12/42)						
Ceriodaphnia pulchella (1/1)						
Ceriodaphnia reticulata (2/5)						
Chydorus sphaericus (1/1)						
Daphnia carinata (2/2)						
Daphnia cucullata (3/4)						
Daphnia longispina (1/1)						
Daphnia magna (57/25)		3.23 (15.1 ± 0.688) mg/L				
Daphnia pulex (9/64)						M: 1.4 mg/L
Daphnia pulicaria (4/5)						M: 17.1 mg/L
Lynceus brachyurus (1/1)						M: 22.7 mg/L
Moina macrocopa (1/1)						M: 22.7 mg/L
Sida crystallina (1/1)						
Simocephalus vetulus (2/29)						
Streptocephalus torvicornis(2/2)						
Insecta (insects) (11/85)						
Malacostraca (27/84)						
Maxillopoda (6/7)						

1. Go to the **Report** module; 2. Highlight the prediction result; 3. Click **Prediction** to create a report

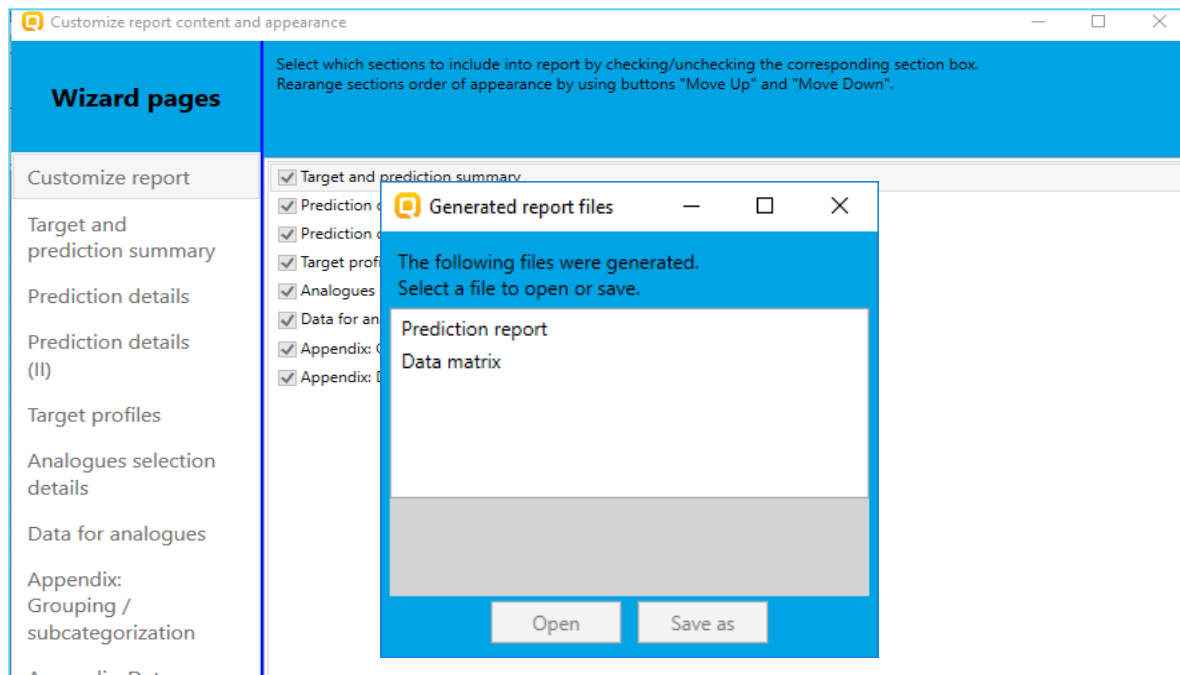
Report Generation report



Report wizard pages appear, where the user could customize the report content and appearance. Some of the fields in the report are automatically populated by the system.

1. Click **Create report** to generate the report.

Report Generation report



Additional window appears with two options: **Prediction report** or **Data matrix**. The user can open and/or save the files.

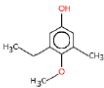
- The **Prediction report** is a PDF file containing the prediction information related to the target.
- The **Data matrix** is a MS Excel file containing chemicals used for prediction along with their data for selected parameters, profiles and endpoint tree positions.

Prediction report Overview

Prediction of LCS0 for CCc1cc(O)cc(C)c1OC 1 / 12

QSAR Toolbox prediction for single chemical

Date: 3 Aug 2017
 Author(s):
 Contact details:

Target information		
Structural information	Numerical identifiers	Chemical names
SMILES: <chem>CCc1cc(O)cc(C)c1OC</chem> Structure 	EC#: N/A CAS#: Invalid CAS number: 0-00-0 Other: N/A	

Prediction summary
Predicted endpoint: LCS0; Mortality; Daphnia magna; 48h; No guideline specified Predicted value: 3.23 (from 0.688 to 15.1) Unit/scale: mg/L Data gap filling method: Trend analysis Summary: manually editable field Not provided by the user

QSAR Toolbox 4.1 Database version: 4.1 **QSAR TOOLBOX** TPRF v4.1

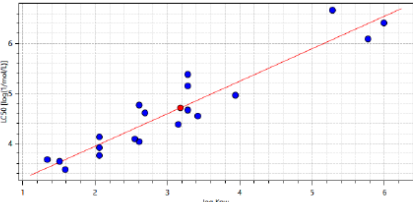
Prediction of LCS0 for CCc1cc(O)cc(C)c1OC 2 / 12

Prediction details (I)

Predicted value: 4.71 log(1/mol/L), conf.range: (4.04 ; 5.38) at 95.0%

Predicted endpoint (OECD Principle 1 - Defined endpoint): Ecotoxicological Information -> Aquatic Toxicity -> Daphnia magna -> LCS0 -> Mortality -> Animalia (animals) -> Arthropoda (arthropods) -> Branchiopoda (branchiopods) -> 48 h

Prediction plot:



Trend analysis prediction for LCS0, based on 20 values
 Predicted: 3.23 mg/L
 Model equation: $LCS0 = 2.64 (+0.380) + 0.650 (+0.114) * \log Kow, \log(1/mol/L)$

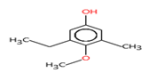
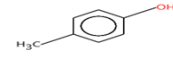
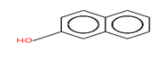
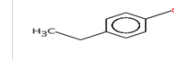
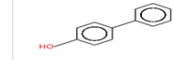
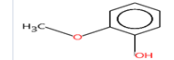
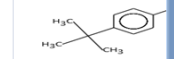
Calculation approach (OECD principle 2 - Unambiguous algorithm): Linear approximation
 Model equation: $LCS0 = 2.64 (+0.380) + 0.650 (+0.114) * \log Kow, \log(1/mol/L)$
 Active descriptor: log Kow (calculated)
 Data usage: Arithmetic mean (average) value*
 Statistics of the prediction model:
 N = 20; count of data points
 R2 = 0.889; coefficient of determination
 R2adj = 0.883; adjusted coefficient of determination
 SSR = 1.75; sum of squared residuals
 s = 0.296; sample standard deviation of residuals
 F = 145; Fisher function

*When multiple values are available for the same chemical, their arithmetic mean (average) value is taken in prediction calculations

QSAR Toolbox 4.1 Database version: 4.1 **QSAR TOOLBOX** TPRF v4.1

Data matrix Overview

Data matrix_3_8_17_16_39_11.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Substance identity																						
2	Structure																						
3	CAS number	Invalid CAS number: 0-00-0	106-44-5	135-19-3	123-07-9	92-69-3	90-05-1	98-54-4															
4	Chemical name		Cresol	2Naphthol	4-Ethylphenol	biphenyl-4-ol	guaiacol	Butylphenol															
5	Other identifier																						
6	SMILES	CC1cc(O)cc(O)c1OC	Cc1ccc(O)cc1	Oc1ccc2ccccc2c1	CCc1ccc(O)cc1	Oc1ccc(cc1)-c1ccccc1	COc1ccccc1O	CC(C)(O)c1ccc(O)cc1															
7	Parameters	unit																					
8	Profilers																						
9	Profiles used for grouping/subcategorization																						
10	Aquatic toxicity classification by ECOSAR																						
11	Substance type (subcategorization)	Phenols Discrete chemical; Mono constituent (predefined); Organic	Phenols Discrete chemical; Mono constituent (predefined); Organic	Phenols Discrete chemical; Mono constituent (predefined); Organic	Phenols Discrete chemical; Mono constituent (predefined); Organic	Phenols Discrete chemical; Mono constituent (predefined); Organic	Phenols Discrete chemical; Mono constituent (predefined); Organic	Phenols Discrete chemical; Mono constituent (predefined); Organic															
12	Acute aquatic toxicity MOA by OASIS	Phenols and Anilines Group 14 - Carbon C; Group 16 - Oxygen O	Phenols and Anilines Group 14 - Carbon C; Group 16 - Oxygen O	Phenols and Anilines Group 14 - Carbon C; Group 16 - Oxygen O	Phenols and Anilines Group 14 - Carbon C; Group 16 - Oxygen O	Phenols and Anilines Group 14 - Carbon C; Group 16 - Oxygen O	Phenols and Anilines Group 14 - Carbon C; Group 16 - Oxygen O	Phenols and Anilines Group 14 - Carbon C; Group 16 - Oxygen O															
13	Chemical elements (subcategorization)																						
14	Measured and predicted data																						
15	Data used for prediction																						
16	environment	endpoint	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen
17	Aquatic Toxicity	LC50	1.4	mg/L	Daphnia magna 48 h Bull. Environ. Contam. Toxicol.23(3): 349-356 Parkhurst,B.R., A.S. Bradshaw, J.L. Forte, and G.P. Wright An Evaluation of the Acute Toxicity to Aquatic Biota of a Coal Conversion Effluent and its Major Components Daphnia magna 48 h	3.54	mg/L	Daphnia magna 48 h Trans. Am. Fish. Soc.113(1): 74-85 Millemann,R.E., W.J. Birge, J.A. Black, R.M. Cushman, K.L. Daniels, P.J. Franco, J.M. Giddings, J.F. McCarthy, and A.J. Comparative Acute Toxicity to Aquatic Daphnia magna 48 h	9.93	mg/L	Daphnia magna 48 h 2005 Chem. Res. Toxicol., 18 (3), 536-555. Peter C. von der Ohe, Ralph Kühne, Ralf-Uwe Ebert, Rolf Altenburger, Matthias Liess, and Gerrit Schüürmann Structural Alerts-A New Classification Model to	3.64	mg/L	Daphnia magna 48 h 2005 Chem. Res. Toxicol., 18 (3), 536-555. Peter C. von der Ohe, Ralph Kühne, Ralf-Uwe Ebert, Rolf Altenburger, Matthias Liess, and Gerrit Schüürmann Structural Alerts-A New Classification Model to	25.9	mg/L	Daphnia magna 48 h Chem.-Biol. Interact.9(4): 245-251 Kopperman,H.L., R.M. Carlson, and R. Caple Aqueous Chlorination and Ozonation Studies. I. Structure-Toxicity Correlations of Phenolic Compounds to Daphnia magna 48 h	4.23	mg/L	Daphnia magna 48 h Chem. Res. T (3), 536 Peter C. von Ralph Kühne Ebert, Rolf A Matthias L Gerrit Schü Structural Al			

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- Workflow of the exercise
- **Save the prediction result**

Saving the prediction result

- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc., on the same computer. The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document and later executing these actions in the same sequence in order to get the same result(s).
- Saving/Loading the file with TB prediction is shown on next screenshots

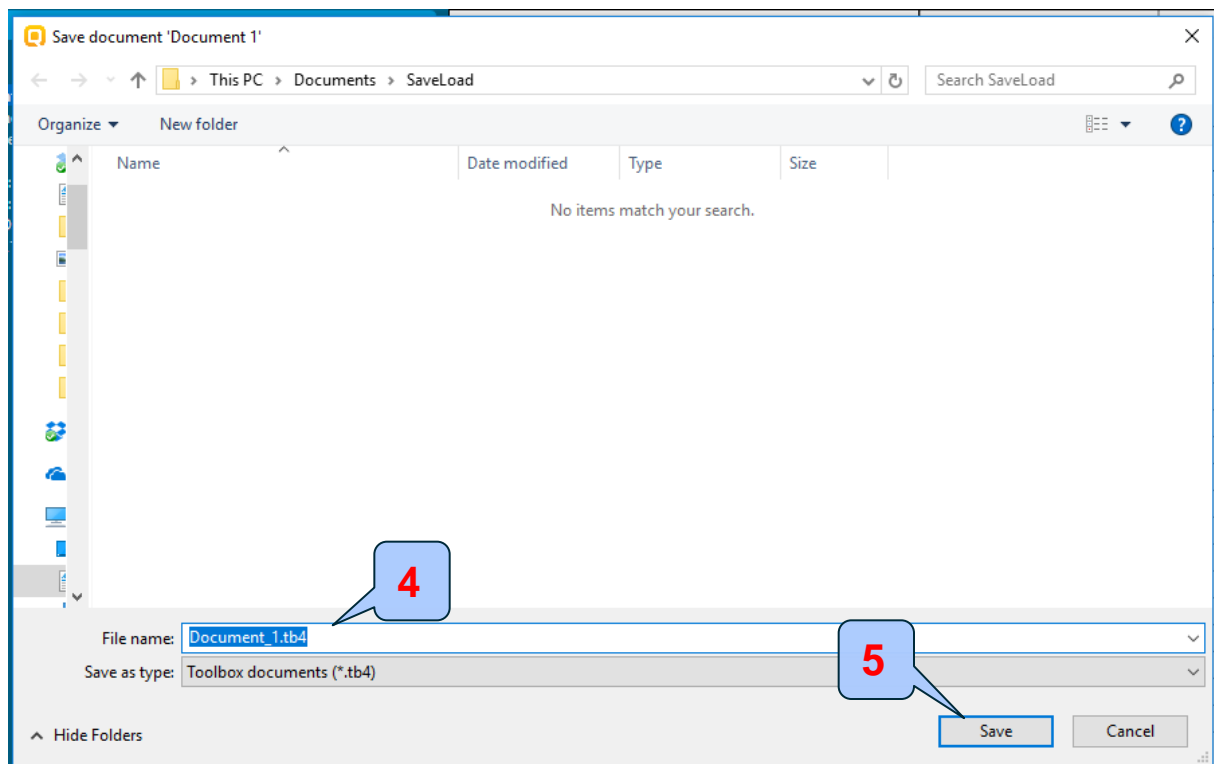
Saving the prediction result

The screenshot shows the QSAR Toolbox software interface. The 'Input' module is selected in the top menu. A red dashed box highlights the 'Save' button in the 'Document' toolbar. A blue callout box with the number '1' points to this button. Another blue callout box with the number '2' points to the 'Save' button in the 'Document' toolbar. A dialog box titled 'Toolbox' is open, asking 'Do you want to save changes to document 'Document 1'?'. A blue callout box with the number '3' points to the 'Yes' button in the dialog box. The background shows a chemical list with various chemical structures and their predicted values.

Structure	1 [target]	2	3	4	5	6	7
Structure	<chem>Oc1ccc(O)cc1</chem>	<chem>Oc1ccc(O)cc1</chem>	<chem>Oc1ccc(O)cc1</chem>	<chem>Oc1ccc(O)cc1</chem>	<chem>Oc1ccc(O)cc1</chem>	<chem>Oc1ccc(O)cc1</chem>	<chem>Oc1ccc(O)cc1</chem>
35 h	(1/1)						
36 h	(3/3)						
44 h	(3/3)						
46 h	(1/1)						
48 h							
Daphnia magna (57/251)			M: 17.1 mg/L				
Daphnia pulex (9/64)			M: 22.7 mgr/L				
Daphnia pulicaria (4/5)			M: 22.7 mgr/L				
Lynceus brachyurus (1/1)							
Moina macrocopa (1/1)							
Sida crystallina (1/1)							
Simocephalus vetulus (2/29)							
Streptocephalus torvicornis(2/2)							

1. Go to the *Input* module; 2. Click **Save** button; 3. Click Yes to confirm;

Saving the prediction result



4. Define name of the file; 5. Click **Save**

Open file

The screenshot shows the QSAR Toolbox software interface. The top toolbar contains several icons for different functions. The 'Open' icon, which is a folder with a plus sign, is highlighted with a red box and a callout labeled '1'. Below the toolbar, a file explorer window is open, showing the contents of a folder named 'SaveLoad'. A file named 'Document_1.tb4' is selected, and a callout labeled '2' points to it. At the bottom of the file explorer window, the 'File name' field contains 'Document_1.tb4', and the 'Open' button is highlighted with a red box and a callout labeled '3'. A blue banner at the bottom of the image contains the text: 'To open a file: 1. Click **Open**; 2. Find and select the file; 3. Click **Open**.'

Congratulations

- You have now been introduced to the work flow of the Toolbox and completed the tutorial on data gap filling by trend analysis and exported the prediction to IUCLID 6.
- You have been introduced to the six modules of the Toolbox, the basic functionalities within each module and the rationale behind each module.
- Remember proficiency comes with practice.