

OECD QSAR Toolbox v.3.0

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 filling exercise by the 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 to the user 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 – mechanistic grouping.

Outlook

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

Workflow

- **The Toolbox has six modules which are used in a sequential workflow:**
 - Chemical Input
 - Profiling
 - Endpoints
 - 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 the user with 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
- Chemical IDs such as EC number, Einecs number

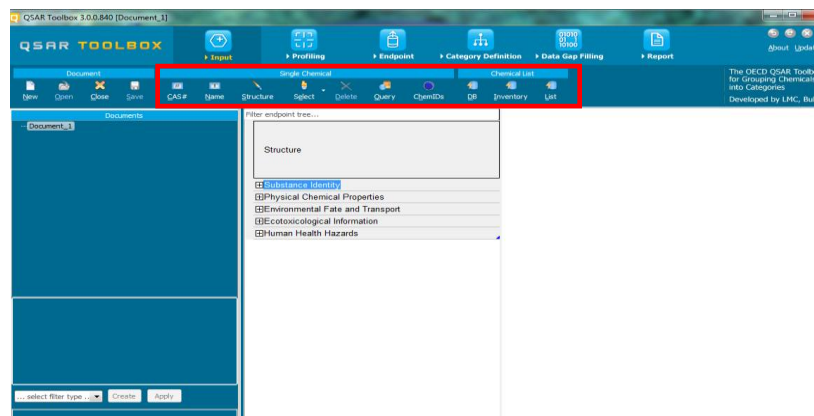
B. Group of chemicals

- User List/Inventory
- Specialized Databases

Getting Started

- Open the Toolbox.
- The six modules in the work flow are seen listed next to "QSAR TOOLBOX".
- **Click** on "Input" (see next screen shot)

Chemical Input Screen Input screen

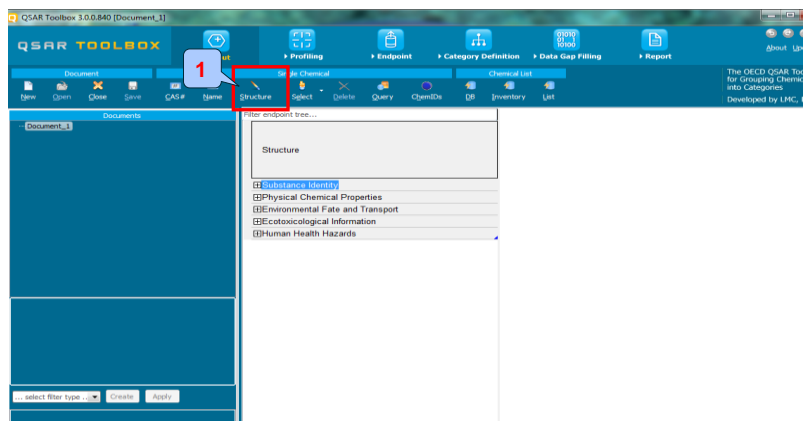


Chemical Input by Drawing

- 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 screen shots will take you through the process for the target chemical.

Chemical Input Screen

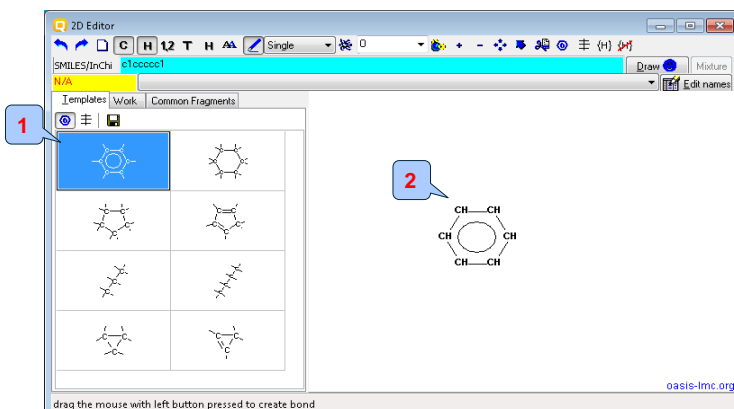
Input target chemical by drawing



1. Click on **Structure**

Chemical Input

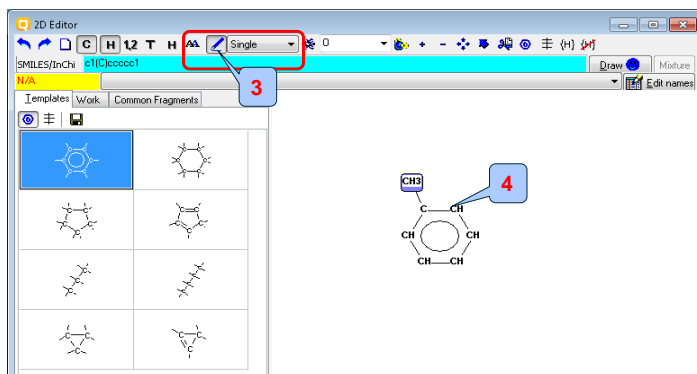
Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor




1. **Left Click** on the appropriate template form from "templates".
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



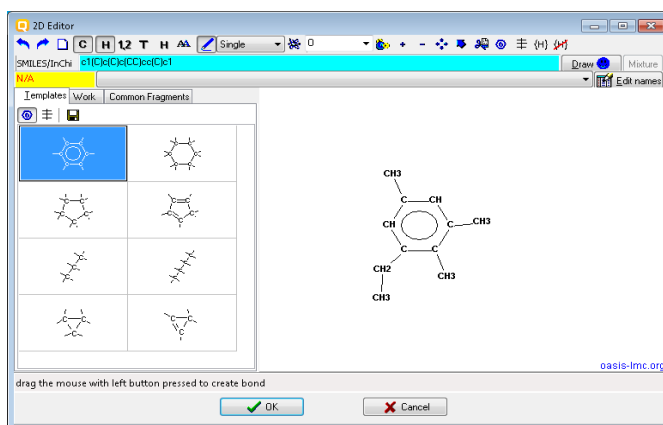
3. Click on  button to add a bond of selected type ("Single" in this case).
4. Drag the mouse (pointing finger) to the appropriate atom and **left click** to create a single bond.

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 on 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).

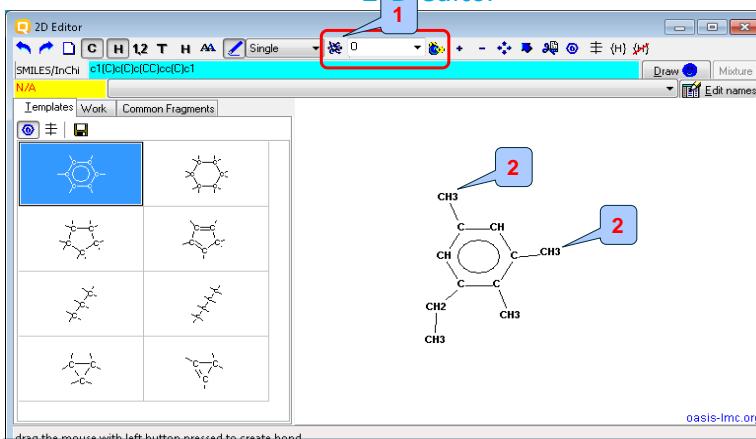
Chemical Input

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

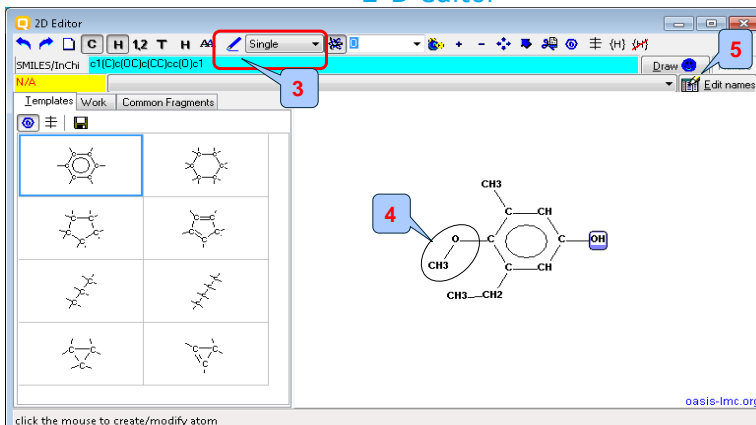


drag the mouse with left button pressed to create bond

1. Click on button to add a hetero atom in this case an oxygen atom.
2. Left click with mouse over the methyl group to insert an oxygen atom.

Chemical Input

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



click the mouse to create/modify atom

3. Click on button
4. Drag the mouse from the O-atom to create a single bond
5. Click **Draw**
6. Click **OK**

QSAR TOOLBOX

Chemical Input

Target chemical identity

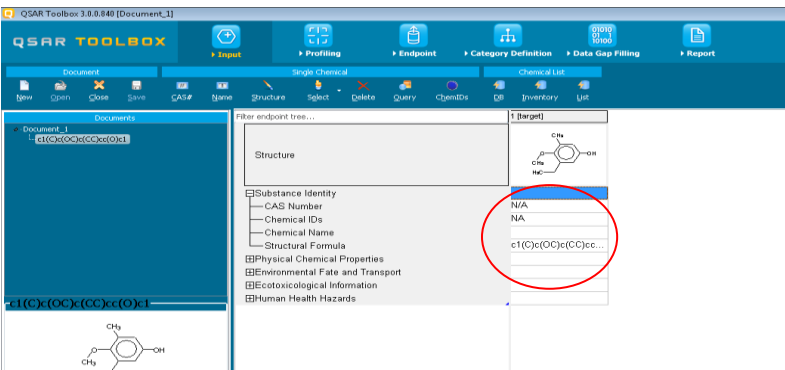
- The already drawn target structure automatically appears on the data matrix
- Note that no CAS number or name is displayed for this chemical. This means the target chemical is not listed in the chemical inventories/databases implemented in the Toolbox.(see next slide).

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 27

QSAR TOOLBOX

Chemical Input

Target chemical identity



The workflow on the first module is now complete, and the user can proceed to the next module. Click on "Profiling".

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 28

Outlook

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

Profiling Overview

- “Profiling” 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 Side-Bar to 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, **click** on "close".

The OECD QSAR Toolbox for Grouping Chemicals into Categories

31

Profiling Side-Bar to 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 on "Show Boundaries" (see next screen shot).

The OECD QSAR Toolbox for Grouping Chemicals into Categories

05.10.2012

32

Profiling Side-Bar to Profiling

1. Highlight the profiler

2. Click **View**

3. Click on one of the Structural alerts (for example Isocyanates)

33

Profiling Side-Bar to Profiling

- The outcome of the profiling determines the most appropriate way to search for analogues (detailed information in Manual for getting started – Toolbox 2.0 (Chapter 4).
<http://www.oecd.org/dataoecd/58/56/46210452.pdf>
- Table 4-1 in chapter 4 (Manual for getting started – Toolbox 2.0) lists a selection of profilers and their relevance for different endpoints of regulatory relevance.
- For this example, the following mechanistic profiling methods are relevant to the aquatic toxicity:
 - EcoSAR – for structural grouping
 - Acute aquatic toxicity MOA by OASIS – mechanistic grouping
 - Protein binding by OASIS v.1.1– mechanistic grouping
 - Acute aquatic toxicity classification by Verhaar – grouping by reactivity

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 34

Profiling

Profiling the target chemical

- Select the “Profiling methods” related to the target endpoint by clicking on the box next to the profilers name.
- This selects (a **green** check mark appears) or deselects(**green** check disappears) profilers.
- For this example, select the following profilers which are relevant to the aquatic toxicity (see next screen shot):
 - EcoSAR – for structural grouping
 - Acute aquatic toxicity MOA by OASIS – mechanistic grouping
 - Protein binding by OASIS v.1.1 – mechanistic grouping
 - Acute aquatic toxicity classification by Verhaar – grouping by reactivity

Profiling

Profiling the target chemical

The screenshot shows the QSAR Toolbox 3.0.0.840 interface. The 'Profiling methods' list is expanded, and the 'Acute aquatic toxicity classification by Verhaar' method is selected, indicated by a green checkmark. The 'Apply' button is highlighted with a red circle and labeled '2'. A red arrow labeled '1' points to the selected method. The 'Filter endpoint tree...' window shows the chemical structure of the target and various property categories.

1. Place a **green** check in the box by the profilers related to the target endpoint.
2. Click **Apply**

QSAR TOOLBOX

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 by 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.

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 37

QSAR TOOLBOX

Profiling

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

1. Double click on the box  to open the nodes of the tree

Grouping Chemicals into Categories 05.10.2012 38

Outlook

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

Endpoint

- “Endpoint” refer to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox database.
- Data gathering can be executed in a global fashion (i.e., collecting all data of all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).
- 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, Aquatic USEPA ECOTOX, and Aquatic OASIS).

Endpoint

1. Expand the **Ecotoxicological Information** section

2. Select databases related to the target endpoint by adding a **green** check in the box before the database name.

3. Click **Gather**

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 41

Endpoint Process of collecting data

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

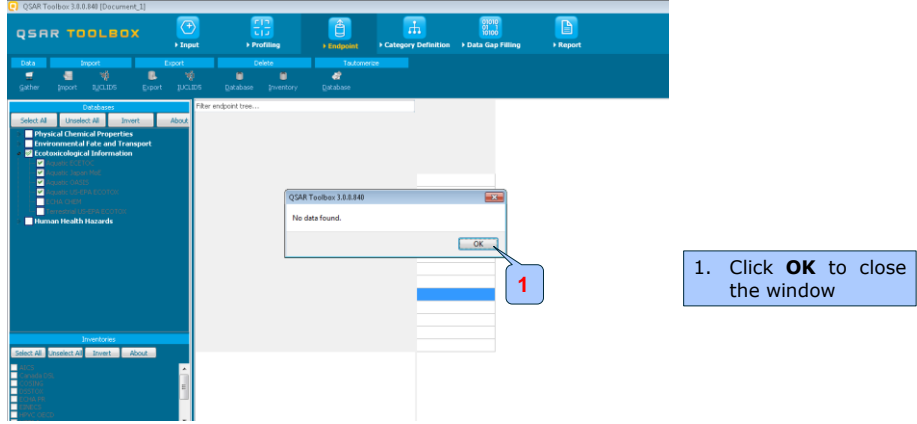
A window with "Read data?" appears. Now the user could choose to collect "all" or "endpoint specific" data.

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 42

Endpoint Process of collecting data

In this example, an insert window appears stating there was "No data found" for the target chemical.



The screenshot shows the QSAR Toolbox 3.0.0.0.000 interface. The main window displays a tree view of chemical categories on the left and a search results area on the right. A small dialog box titled 'QSAR Toolbox 3.0.0.0.000' is open in the center, displaying the message 'No data found.' and an 'OK' button. A red circle with the number '1' is placed over the 'OK' button, and a callout box points to it with the text '1. Click OK to close the window'.

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 43

Outlook

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 44

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.
- Detailed information about grouping chemical (Chapter 4) could be downloaded from:
<http://www.oecd.org/dataoecd/58/56/46210452.pdf>
- 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

Side-bar of 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.
- "Classification by ECOSAR" 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

1. Highlight "Aquatic toxicity classification by ECOSAR"

2. Click **Define**

3. Confirm the category **Phenols** and 4. Click **OK**

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 49

Category Definition Defining ECOSAR category

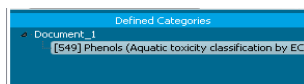
1. Click **OK** to confirm the name of the category

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 50

Category Definition

Analogues

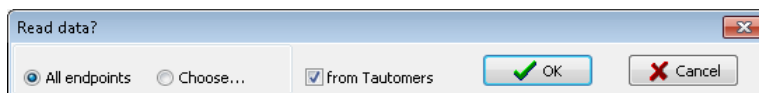
- The Toolbox now identifies all chemicals corresponding to the ECOSAR classification of "phenols" which are listed in the databases selected under "Endpoint".
- 549 analogues are identified. Along with the target they form a category (Phenols) which can be used for data gap filling.
- The name of the category appear in the "Defined Categories" window, along with the number of substances belonging to the category.



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 ecotoxicological endpoints are selected, both options give the same results.
- As the Toolbox must search the database, this may take some time.

Category Definition

Read data for Analogues

Due to overlap between the Toolbox databases for intersecting chemicals the same data may be found simultaneously. data redundancies are identified and the user has the opportunity to select either a single data value or all data values.

Repeated values for: 4517 data-points, 1597 groups, 675 chemicals

| Endpoint | CAS | Structure | Value | Age |
|-------------------------------------|---------|----------------------------------|--|-------------|
| <input checked="" type="checkbox"/> | 59-02-9 | | 50 milligrams per kilogram body weight | 21 month(s) |
| <input checked="" type="checkbox"/> | 59-02-9 | | 50 milligrams per kilogram body weight | 21 month(s) |
| <input checked="" type="checkbox"/> | 59-02-9 | <chem>CC1=CC=C(C=C1)N(C)C</chem> | 50 milligrams per kilogram body weight | 21 month(s) |
| <input checked="" type="checkbox"/> | 59-02-9 | | 50 milligrams per kilogram body weight | 21 month(s) |
| <input checked="" type="checkbox"/> | 59-02-9 | | 50 milligrams per kilogram body weight | 21 month(s) |
| <input checked="" type="checkbox"/> | 59-02-9 | | 50 milligrams per kilogram body weight | 21 month(s) |

1. Click **Select one** and then
2. Click **OK**

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 53

Category Definition

Summary of Analogues

QSAR Toolbox 3.0.0.0 [Document_1]

Categories: Toxic hazard classification by OASIS (with extra), Ultimate biodeg., Endpoint Specific, Acute aquatic toxicity (PNEC), Acute aquatic toxicity (PNEC), Aquatic toxicity classification by ECOSAR, Bioaccumulation - metabolism half-life, Biodegradation fragments (OWS/MTD), Carcinogenicity (genotoxic and non-genotoxic) alerts, Chemical alerts (ARCS, HS and CA by OASIS v1.1), CMRA Cytosine peptide depletion, CMRA (new peptide depletion), Eye Irritation/Corrosion Exclusion rules by BR, Eye Irritation/Corrosion Exclusion rules by BR, In vivo mutagenicity (Gene Tox) alerts by ISS, In vivo mutagenicity (Microtox) alerts by ISS, In vitro gene expression, Oncologic Primary Classification, Skin Irritation/Corrosion Exclusion rules by BR, Skin Irritation/Corrosion Exclusion rules by BR, US EPA ED₀₁ binding expert system, US EPA ED₀₁ binding expert system - dermal, Endpoint

Defined Categories: Document_1, [54] Phenols (Aquatic toxicity classification by ECOSAR), [55] Phenols (Aquatic toxicity classification by ECOSAR)

Filter endpoint tree...

| Structure | 1 target | 2 | 3 | 4 | 5 |
|----------------------------------|-----------|-----------------------|----------------------|-----------------------|-----------------|
| <chem>CC1=CC=C(C=C1)N(C)C</chem> | 646/27262 | M: 50 mg/kg bw/d, ... | M: 102 mg/L, 230 ... | M: 3.21 mg/L, 102 ... | M: 0.868 mg/... |

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 54

Category Definition

Summary information of Analogues

The screenshot shows the QSAR Toolbox interface. On the left, a list of 'Endpoint Specific' methods is visible, with 'Aquatic toxicity classification by ECOTOX' selected. The main workspace shows a table with columns for chemical structures and data. A red circle highlights the cell containing the value '546(27262)'. A callout box points to this cell with the text: 'Chemical statistics presenting the number of chemicals and the available experimental data. This is statistics for the current row on data matrix.'

Category Definition

Side-Bar of experimental data

The screenshot shows the QSAR Toolbox interface. The 'Data points' dropdown menu is open, showing a list of experimental data points. A red circle highlights the first data point. A callout box points to this data point with the text: '1. Double-click on the cell with measured data provides a dropdown box ("Data points") which provides detailed information.'

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 549 analogues have been collected from the selected databases (Aquatic ECETOC, Aquatic Japan MoE, Aquatic USEPA 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 screen shots)

Category Definition

Navigation through the endpoint tree

The screenshot shows the QSAR Toolbox interface with the following data in the endpoint table:

| Endpoint | Count | 1 (target) | 2 | 3 | 4 | 5 |
|-------------------|---------|-----------------------|---|---|------------------------|------------|
| Aquatic Toxicity | 37,221 | | | | | |
| Accumulation | 27,658 | M: 1 micromoles p... | | | M: 8 mg/L | |
| Avoidance | 136,665 | | | | M: 2,91 mg/L, 3,31 ... | M: 10 mg/L |
| Behavior | 79,1463 | M: 50 mg/kg bdwt, ... | | | M: 1,500 mg/L, 10 ... | |
| Biochemistry | 20,136 | | | | M: s1,62E6 g, 0,00 ... | |
| Cells | 46,603 | | | | M: 10 mg/L, 0,32 m... | |
| Development | 44,10 | | | | | |
| Ecosystem Process | 44,490 | M: 50 mg/kg bdwt, ... | | | M: 4 mg/L | |
| Enzyme(s) | 29,87 | | | | M: 30 mg/L | |
| Feeding Behavior | | | | | | |

1. Successfully expand the following nodes: Aquatic toxicity; Mortality; LC50; Animalia; Arthropoda (Invertebrates); Branchiopoda (branchiopods)
2. Find *Daphia magna* - this is the species related to target endpoint

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 59

Category Definition

Navigation through the endpoint tree

The screenshot shows the QSAR Toolbox interface with the following data in the endpoint table:

| Endpoint | Count | 1 (target) | 2 | 3 | 4 | 5 |
|---------------|---------|------------|---|---|-------------------------|---|
| Immunological | 7,229 | | | | | |
| Injury | 11,118 | | | | | |
| Intoxication | 103,603 | | | | | |
| Morphology | 34,605 | | | | M: 2,8(2,5,3) mg/L, ... | |
| Mortality | | | | | | |
| EBEC0 | 2,94 | | | | | |
| EBEC10 | 1,7 | | | | | |
| EBEC100 | 1,99 | | | | | |
| EBEC20 | 3,3 | | | | | |
| EBEC25 | 4,1 | | | | | |
| EBEC50 | 66,1195 | | | | M: 2 mg/L, 1,5 mg/L... | |
| EBET0 | 1,25 | | | | | |
| EBET100 | 1,25 | | | | | |
| EBET50 | 4,65 | | | | | |

1. Successfully expand the following nodes: Aquatic toxicity; Mortality; LC50; Animalia; Arthropoda (Invertebrates); Branchiopoda (branchiopods)
2. Find *Daphia magna* - this is the species related to target endpoint

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 60

Category Definition

Navigation through the endpoint tree

The screenshot shows the QSAR Toolbox interface with the following data in the endpoint tree:

| Structure | Count |
|--------------------|---------|
| LC20 | (1/1) |
| LC25 | (2/9) |
| LC50 | (11/77) |
| Undefined Duration | (1/1) |
| Minutes | (1/1) |
| 18 Minutes | (1/2) |
| 24 Minutes | (1/1) |
| 27 Minutes | (1/1) |
| 36 Minutes | (1/3) |
| 1 h | (9/160) |
| 1.5 h | (1/2) |
| 1.8 h | (1/1) |
| 2 h | (15/18) |
| 2.5 h | (13/13) |

- Successively expand the following nodes: Aquatic toxicity; Mortality; LC50; Animalia; Arthropoda (Invertebrates); Branchiopoda (branchiopodos)
- Find *Daphnia magna* - this is the species related to target endpoint

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012
61

Category Definition

Navigation through the endpoint tree

The screenshot shows the QSAR Toolbox interface with the following data in the endpoint tree:

| Structure | Count |
|------------------------------|----------|
| 48 h | (1/1) |
| Animalia | (8/39) |
| Arthropoda (Invertebrates) | (8/39) |
| Branchiopoda (branchiopodos) | (7/8) |
| Artemia salina | (2/12) |
| Bosmina coregoni | (1/1) |
| Ceriodaphnia dubia | (8/36) |
| Ceriodaphnia pulchella | (1/1) |
| Ceriodaphnia reticulata | (2/5) |
| Cyodorus sphaericus | |
| Daphnia carinata | |
| Daphnia cucullata | |
| Daphnia longispina | (1/1) |
| Daphnia magna | (49/226) |

- Successively expand the following nodes: Aquatic toxicity; Mortality; LC50; Animalia; Arthropoda (Invertebrates); Branchiopoda (branchiopodos)
- Find *Daphnia magna* - this is the species related to target endpoint

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012
62

QSAR TOOLBOX

Category Definition

Navigation through the endpoint tree

- Navigation through the tree by "Filtering"

1. Type **"Daphnia magna"** in the filter box, then press **Enter**

2. Open the tree to the target endpoint by double left clicking.

63

QSAR TOOLBOX

Recap

- You have now retrieved the available experimental data on aquatic toxicity for 549 chemicals classified as "phenols" by the "Aquatic toxicity classification by ECOSAR" profiler found in the databases Aquatic ECETOC, Aquatic Japan MoE, Aquatic USEPA 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 on "Data Gap Filling" (see next screen shot).

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012

64

Outlook

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

Data Gap Filling Overview

- "Data Gap Filling" module give access to three different data gap filling tools:
 - Read-across
 - Trend analysis
 - Q)SAR models
- Depending on the situation, the most relevant data gap mechanism should be chosen, 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.
- In this example, we use trend analysis.

Data Gap Filling Data Gap window

The screenshot shows the QSAR Toolbox interface with the 'Data Gap Filling' window open. The left sidebar contains a 'Data Gap Filling Method' section with options: 'Read-across', 'Trend analysis', and 'QSAR models'. The 'Target Endpoint' section lists various endpoints, with 'LC50' expanded to show 'LC50/48h' selected. The main table displays data for 'Daphnia magna' and other target chemicals. The table has columns for 'Target', 'LC50', and 'M'. The 'LC50/48h' row for 'Daphnia magna' is highlighted in blue.

| Target | LC50 | M |
|------------|------|-----------------------|
| 1 Target | | |
| 2 | | |
| 3 | | |
| 4 | | M: 1.7 mg/L, 0.21 ... |
| 5 | | |
| 6 | | M: 2,8(2,5,3) mg/L... |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |
| 17 | | |
| 18 | | |
| 19 | | |
| 20 | | |
| 21 | | |
| 22 | | |
| 23 | | |
| 24 | | |
| 25 | | |
| 26 | | |
| 27 | | |
| 28 | | |
| 29 | | |
| 30 | | |
| 31 | | |
| 32 | | |
| 33 | | |
| 34 | | |
| 35 | | |
| 36 | | |
| 37 | | |
| 38 | | |
| 39 | | |
| 40 | | |
| 41 | | |
| 42 | | |
| 43 | | |
| 44 | | |
| 45 | | |
| 46 | | |
| 47 | | |
| 48 | | |
| 49 | | |
| 50 | | |
| 51 | | |
| 52 | | |
| 53 | | |
| 54 | | |
| 55 | | |
| 56 | | |
| 57 | | |
| 58 | | |
| 59 | | |
| 60 | | |
| 61 | | |
| 62 | | |
| 63 | | |
| 64 | | |
| 65 | | |
| 66 | | |
| 67 | | |
| 68 | | |
| 69 | | |
| 70 | | |
| 71 | | |
| 72 | | |
| 73 | | |
| 74 | | |
| 75 | | |
| 76 | | |
| 77 | | |
| 78 | | |
| 79 | | |
| 80 | | |
| 81 | | |
| 82 | | |
| 83 | | |
| 84 | | |
| 85 | | |
| 86 | | |
| 87 | | |
| 88 | | |
| 89 | | |
| 90 | | |
| 91 | | |
| 92 | | |
| 93 | | |
| 94 | | |
| 95 | | |
| 96 | | |
| 97 | | |
| 98 | | |
| 99 | | |
| 100 | | |

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012 67

Data Gap Filling Apply Trend analysis

The screenshot shows the same QSAR Toolbox interface as above, but with annotations. A red circle labeled '3' highlights the 'Apply' button in the 'Data Gap Filling Method' section. A red circle labeled '2' highlights the 'Trend analysis' option. A red circle labeled '1' highlights the 'LC50/48h' endpoint for 'Daphnia magna' in the tree view. A red arrow points from the 'Apply' button to the 'LC50/48h' endpoint.

1. Highlight the **data endpoint box** corresponding to Daphnia magna/LC50/48h under the target chemical. It will be empty.
2. Select **Trend analysis**
3. Click **Apply**

QSAR TOOLBOX

Data Gap Filling Results of Trend analysis

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 69

QSAR TOOLBOX

Data Gap Filling Integrating 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** dot represents the predicted value for the target chemical.
- The **BLUE** dots represent the experimental results available for the analogues used in the trend analysis.
- Before accepting the estimated result for the target chemical, the trend analysis should be further refined by subcategorisation (see following slides).

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 70

Data Gap Filling

Side-Bar of Subcategorisation

- 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".
- Subcategorisation refers to the process of applying additional profilers to the previously defined category; subcategorisation identifies chemicals which have differing profiling results and eventually eliminating these chemicals from the final category.

Data Gap Filling

Side-Bar of Subcategorisation

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

- Substance type (mixtures and hydrolizing chemicals)

The categorisation based on substance type allows keeping among the analogues only those that are of the same chemical type: discrete chemicals, mixtures, polymers, inorganics, organometalics. The current target is a discrete chemical 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.

Subcategorisation is demonstrated in the next 4 screen shots.

Data Gap Filling Side-Bar of Subcategorisation

The screenshot displays the QSAR Toolbox interface. On the left, the 'Subcategorisation' side-bar is open, showing various methods and a list of subcategories. The 'Discrete chemical' and 'Ecotoxicology' categories are selected. The main window shows a scatter plot of log Kow vs log Kow with a red regression line. A callout box '1' points to the 'Subcategorize' button in the side-bar, and a callout box '2' points to the 'Select/filter data' button. Below the plot, a table shows predicted values for Daphnia magna: M: 5.21 mg/L, 3.10; M: 0.24 mg/L, 0.24; M: 2.02 mg/L, 2.01; M: 2.97.

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012 73

Data Gap Filling Side-Bar of Subcategorisation

The screenshot displays the QSAR Toolbox interface. On the left, the 'Subcategorisation' side-bar is open, showing various methods and a list of subcategories. The 'Discrete chemical' and 'Ecotoxicology' categories are selected. The main window shows a scatter plot of log Kow vs log Kow with a red regression line. A callout box '1' points to an outlier point on the plot, a callout box '2' points to the 'Substance Type' dialog box, and a callout box '3' points to the 'Remove' button in the side-bar. Below the plot, a table shows predicted values for Daphnia magna: M: 1.5164 mg/L, 1.5164; M: 0.1889.

1. **Double click above the outlier** to see why this chemical is different to the target
The chemical is dissociating chemical and has to be eliminated being different substance type compared to the target (discrete chemical)
2. **Close**
3. Click **Remove** to eliminate dissimilar chemical

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012 73

Data Gap Filling

Subcategorisation by Acute-aquatic toxicity MOA

The screenshot shows the QSAR Toolbox interface during the 'Data Gap Filling' process. The 'Subcategorisation' window is open, with the 'Target' set to 'Phenols and Anilines'. The 'Data Gap Filling' tab is active, displaying a scatter plot of log Kow (x-axis, 2.00 to 6.00) versus Ecotoxicological Information#Aquatic Toxicity (y-axis, 0.00 to 2.00). A red regression line is shown with the equation: $\text{Ecotoxicological Information\#Aquatic Toxicity} = +2.97 + 0.630 * \log \text{Kow}$. The observed target value is N/A, and the predicted target value is 1.78 mg/L. A chemical structure of Daphnia magna is highlighted with a red box. A list of subcategories is visible on the left, with 'Acute aquatic toxicity class' selected. A red box highlights the 'Remove' button.

1. Select **Acute aquatic toxicity MOA by OASIS**
2. Click **Remove** to eliminate dissimilar chemical

75

Data Gap Filling

Results

The screenshot shows the QSAR Toolbox interface during the 'Results' step. The 'Data Gap Filling Method' is 'Read-across'. The 'Target Endpoint' is 'Ecotoxicological Information Aquatic Toxicity Mortality LC50 48h Annelids Amphipods Invertebrates Insect Deposits/Insects/Spiders Daphnia magna'. The 'Data Gap Filling Results' tab is active, displaying a scatter plot of log Kow (x-axis, 2.00 to 6.00) versus Ecotoxicological Information#Aquatic Toxicity (y-axis, 0.00 to 2.00). A red regression line is shown with the equation: $\text{Ecotoxicological Information\#Aquatic Toxicity} = +2.85 + 0.662 * \log \text{Kow}$. The observed target value is N/A, and the predicted target value is 1.65 mg/L. A chemical structure of Daphnia magna is highlighted with a red box. A list of subcategories is visible on the left, with 'Acute aquatic toxicity class' selected. A red box highlights the 'Remove' button.

76

QSAR TOOLBOX

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 and MOA by OASIS).
- By **accepting the prediction** the data gap is filled (see next screen shot).
- By **clicking** on Return to Matrix, the user can close the read-across and proceed with the workflow (see next screen shot).

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 77

QSAR TOOLBOX

Data Gap Filling Accepting prediction results

1. Click **Accept prediction**

2. Click **Yes** if you want to save the model based on the results from trend analysis, otherwise click **NO**

3. Click **Return to matrix**

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05-10-2012 78

Q SAR TOOLBOX

Data Gap Filling

Accepting prediction results

3. Click **Yes** if you want additional data for the analogues to be presented in the report, otherwise click **NO**

4. Click **Return to matrix**

The OECD QSAR Toolbox for Grouping Chemicals into Categories 79

Q SAR TOOLBOX

Outlook

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 80

Export prediction to the IUCLID5

Overview

- The OECD QSAR Toolbox allows the users to export predicted data (by means of the Filling Data Gap tools) to IUCLID 5.
- There are two ways of exporting:
 - by creating an *.i5z file which can then be imported into an IUCLID 5 database.
 - Or by directly connecting to an IUCLID 5 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 screen shot).
- More detailed information could be found in the following link: <http://www.oecd.org/dataoecd/54/27/47136326.pdf>

Exporting the prediction to IUCLID5

Case study

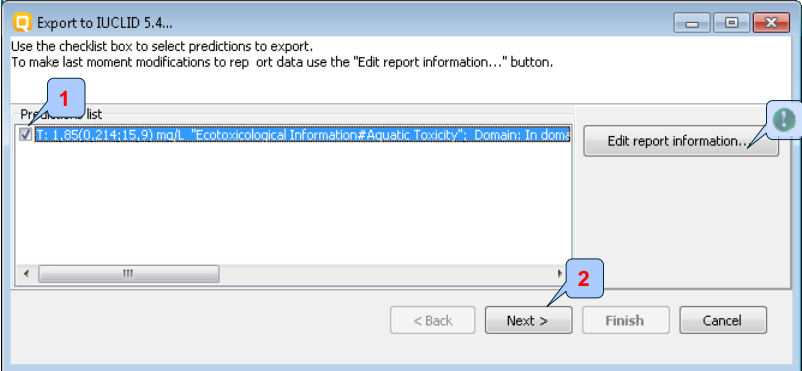
The screenshot shows the QSAR Toolbox 3.8.8.846 interface. The main window displays a table with columns for 'Target', '1', '2', '3', '4', and '5'. The 'Target' column shows chemical structures. The table contains predicted values for various endpoints. A context menu is open over the 'IUCLID5' option in the table, with a red '1' pointing to the 'IUCLID5' option and a red '2' pointing to the 'IUCLID5' option in the menu.

1. Move the mouse in the column of the target substance and click the right mouse button
2. Select IUCLID


Q SAR TOOLBOX

Exporting the prediction to IUCLID5

Case study



1. Select the **prediction** to export (green tick appears)
2. Click **Next** to move through the next step of the export

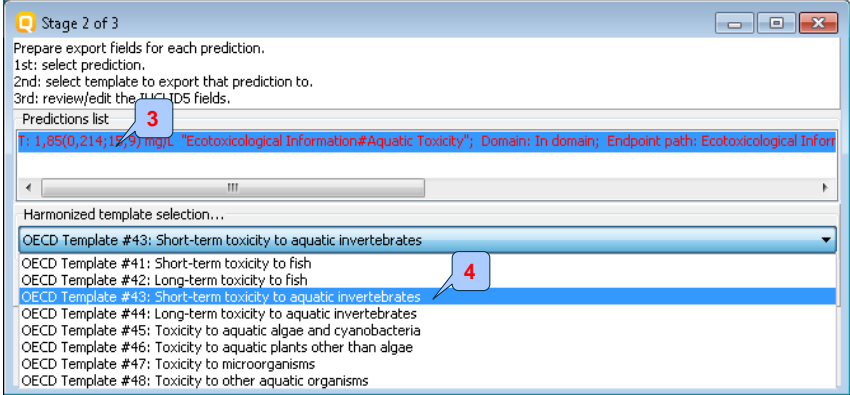
 The user could also edit the report information

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 83

Q SAR TOOLBOX

Exporting the prediction to IUCLID5

Case study



3. Select **prediction**
4. Select **template** to export the prediction

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 84

QSAR TOOLBOX

Exporting the prediction to IUCLID5

Case study

Stage 2 of 3

Prepare export fields for each prediction.
1st: select prediction.
2nd: select template to export that prediction to.
3rd: review/edit the IUCLID5 fields.

Predictions list

1,85(0,214;15,9) mg/L "Ecotoxicological Information#Aquatic Toxicity"; Domain: In domain; Endpoint path: Ecotoxicological Infor...

Harmonized template selection...

OECD Template #43: Short-term t to aquatic invertebrates

Review export data...

< Back Next > Finish Cancel

5. Review/edit the IUCLID5 fields
6. Click Next

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 85

QSAR TOOLBOX

Exporting the prediction to IUCLID5

Case study

Export to...

Specify the export file.

Save as

IUCLID5 medium

Export to IUCLID5 File (*...)

Specify the export file:

C:\Users\Ioanna\Desktop\te...

Inherit data

Information

Export completed!

OK

< Back Next > Finish Cancel

7. Select medium to export, i5z file or export via WebServices
8. Specify the export file
9. Click OK
10. Click Finish

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 86

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Endpoint
 - 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 as well as a template editor with which users can define their own user defined templates.
- The report can then be printed or saved in different formats.

Report Overview

The screenshot shows the QSAR Toolbox software interface. The main window displays the 'Report Overview' for a single chemical prediction. The interface includes a menu bar with options like 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. A sidebar on the left shows 'Available data to report' and 'Available report templates'. A red circle highlights the 'Create' button in the 'Available data to report' section, and a red square highlights the 'Available data to report' section header. A blue box at the bottom contains instructions: '1. Select the prediction in the window Available data to report, then 2. Click Create'.

1. Select the prediction in the window **Available data to report**, then
2. Click **Create**

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 89

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 5.
- 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.

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 90