

OECD QSAR Toolbox v.3.4

Step-by-step example of how to build a category for more than one target chemicals and predict acute toxicity to fish

Outlook

- **Background**
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise 1
- Workflow of the exercise 2

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for evaluating an ad-hoc analogue approach.

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for evaluating an ad-hoc analogue approach.
- By now you are experienced in using the Toolbox so there will be multiple key strokes between screen shots.

Outlook

- Background
- **Objectives**
- Specific Aims
- The exercise
- Workflow of the exercise 1
- Workflow of the exercise 2

Objectives

- To demonstrate how to use the Toolbox to evaluate whether a data gap filling with read-across from potential analogues of target chemicals is robust.

Outlook

- Background
- Objectives
- **Specific Aims**
- The exercise
- Workflow of the exercise 1
- Workflow of the exercise 2

Specific Aims

- To examine the workflow of evaluating an analogue approach.
- To introduce the user to new functionalities within selected modules.
- To explain the rationale behind each step of the exercises.
- To demonstrate with two practical examples how to use the Toolbox to evaluate whether a read-across from a potential analogue to a target chemical is robust.

Outlook

- Background
- Objectives
- Specific Aims
- **The exercise**
- Workflow of the exercise 1
- Workflow of the exercise 2

Exercise

- In this exercise we will perform two examples of ad-hoc read-across for acute toxicity to fish.
- We will do this by first entering the source and target chemicals and analysing the available data for the source chemicals.
- We will then profile the source and target chemicals and evaluate whether the read-across is robust.

Exercise

Side-Bar on the Robustness of a Potential Analogue

- **According to the OECD Guidance on Grouping of Chemicals, the following issues should be taken into account when evaluating the robustness of an analogue approach:**
 - Quality of the experimental result of the source chemical
 - Differences in functionalities in the molecules of the source and target compound (*)
 - Purity and impurity profiles
 - Differences in physical chemical properties
 - Differences in experimental results for other (eco)toxicological endpoints
 - Differences in mode of action (*)
 - Differences in toxicokinetics
- **Some of the issues above (those marked with an *) will be addressed in the current examples with the help of the Toolbox.**

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise 1**
- Workflow of the exercise 2

Workflow of the Exercise 1

- **As you know the Toolbox has 6 modules which are typically used in sequence:**
 - Chemical Input
 - Profiling
 - **Endpoint**
 - Category Definition
 - **Data Gap Filling**
 - Report
- **In this example we will use the modules in a different order, tailored to the aims of the example.**

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise 1**
 - **Chemical Input**
 - Profiling
 - **Endpoint**

Chemical Input Overview

- As you know this module provides the user with several means of entering the chemical of interest or the target chemical.
- It is essential to remember that 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

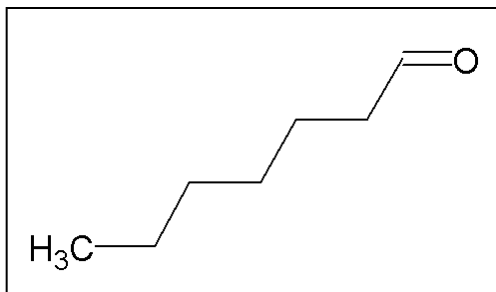
- **Remember there are several ways to enter a target chemical and the most often used are:**
 - CAS#,
 - SMILES (simplified molecular information line entry system) notation, and
 - Drawing the structure.

Chemical Input

Exercise 1

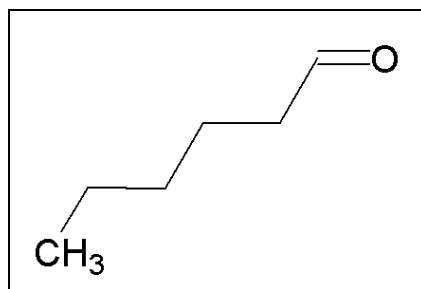
Read-across of acute toxicity to fish from 1-heptanal 1-hexanal and to 3-ethyl-1-pentanal.

1-heptanal

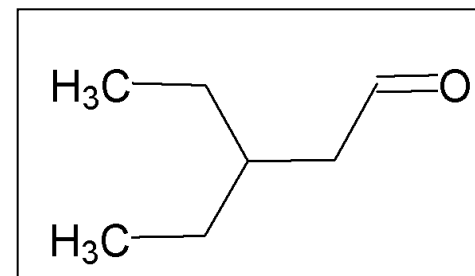


+

1-hexanal



3-ethyl-1-pentanal



Chemical Input

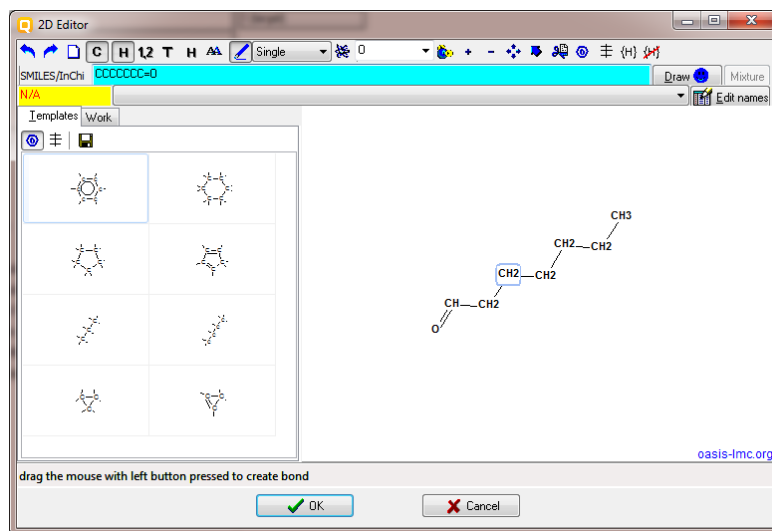
Input target chemical#1 by SMILES

- In this example, we are entering the structure using the SMILES notation.
- **Click** on Structure, then
- **Enter** CCCCCCC=O for n-heptanal **on** "SMILES/InChi" window.
- The structure is drawn simultaneously while entering the SMILES (see next screen shot).

Chemical Input

Target chemical identity

The Toolbox now searches the Toolbox databases and inventories for the presence of the chemical with structure related to the current SMILES notation. It is displayed as a 2D image.



In this case Toolbox found two chemicals answering the required SMILES. This panel displays QA information for presented chemicals. The user can decide which substance is to be retained for the subsequent workflow.

Chemical Input

Input target chemical# 1 by SMILES

The screenshot shows the QSAR Toolbox software interface. The main window is titled '2D Editor' and contains a 'SMILES/InChi' input field with the text 'CCCCC=O' highlighted by a red circle and a callout labeled '4'. Below the input field is a grid of chemical templates. To the right of the templates is a 2D chemical structure of hexanal, also circled in red. At the bottom of the 2D Editor window, there is an 'OK' button highlighted by a red circle and a callout labeled '5', and a 'Cancel' button. The top of the interface shows various toolbars and a menu bar with options like 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The bottom status bar indicates '1 Document' and 'Add chemical by drawing'.

4. Type CCCCC=O in SMILES/InChi window; 5. Click OK.

Chemical Input

Input target chemical#1 by SMILES

- The Toolbox now consults its chemical ID database and finds all chemicals with the structure CCCCCCC=O.
- The Toolbox finds three chemicals with the same structure for 1-heptanal but with different CAS numbers and chemical names. Therefore, the Toolbox find two chemicals with different QA relations (CAS-Name; 2D - Name; CAS-2D (see next screen shot)).

Chemical Input

Input target chemical#1 by SMILES

- The Toolbox finds two chemicals with the same structure and with different QA relations (CAS-Name; 2D -Name; CAS-2D).

The screenshot shows a window titled "Select chemicals" with a table of results. Two entries are selected, both representing the same chemical structure (hexanal, SMILES: CCCCCC=O).

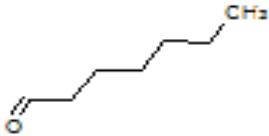
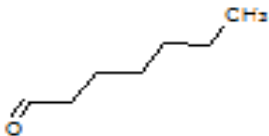
Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS
1. Yes	111-71-7	CCCCCC=O			1:: Low (1:: Low (: Hi 1:: Al 1:: U: 2: 2:: Es 2:: Al 3: 2:: Low (2:: Low (: Hi 4: 1:: Br 1:: U: 2:: Gr 2:: Gr 3:: H		
2. Yes	70955-11	CCCCCC=O			1:: High 1:: Low (: Lo 1: 1: C 1: T: 2: 2: D: 2: Et 3: 3: Et 3: C: 4: 4: HI 4: D:		

At the bottom of the dialog, there are buttons for "Tautomeric sets", "Search", "OK", and "Cancel".

Chemical Input QA

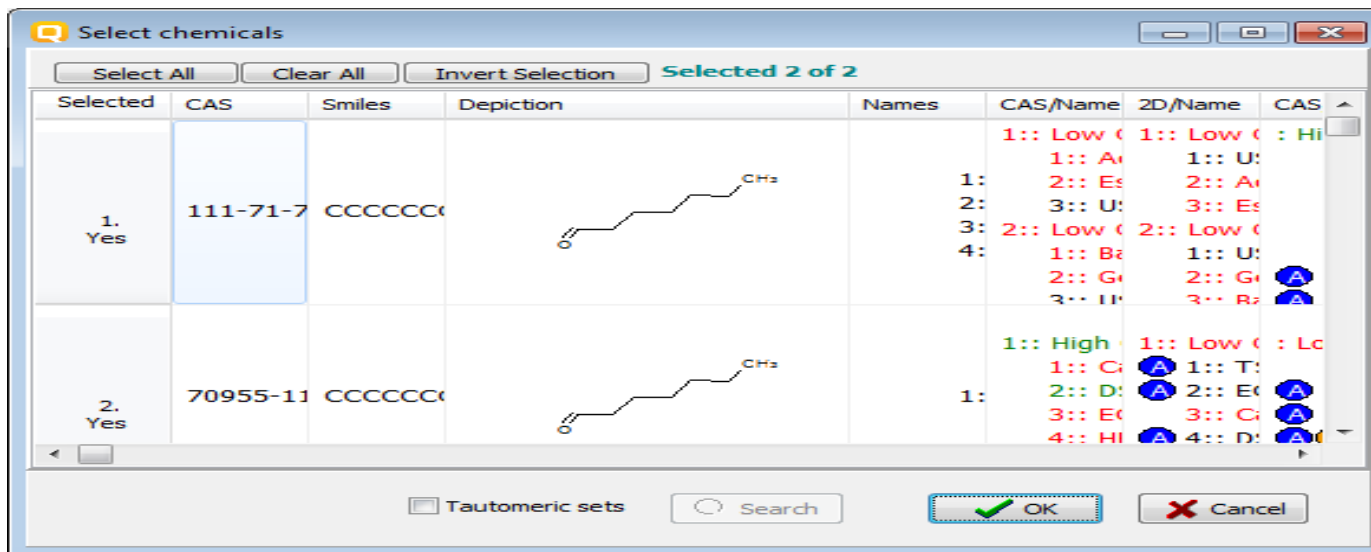
Select chemicals

Select All Clear All Invert Selection Selected 2 of 2

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS
1. Yes	111-71-7	CCCCCCC			1:: Low C 1:: Av 2:: Es 3:: U 4:: 2:: Low C 1:: Ba 2:: G 3:: H	1:: Low C 1:: U 2:: A 3:: Es 2:: Low C 1:: U 2:: G 3:: B	Hi A A
2. Yes	70955-11	CCCCCCC			1:: High 1:: C 2:: D 3:: E 4:: H	1:: Low C 1:: T 2:: E 3:: C 4:: D	Lc A A A A

Note: The last 3 columns represent the chemical identification relations: CAS/Name, 2D/Name, and CAS/2D.

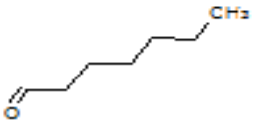
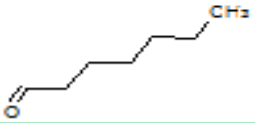
Chemical Input QA



- The columns represent chemical relations
- The colors represent the quality of relation

Text Color	Evaluated Q
Black	N/A Quality
Red	Low Q
Orange	Moderate Q
Green	High Q
Blue	Conflict

Chemical Input QA

Clear All		Invert Selection		Selected 2 of 2		
CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
111-71-7	CCCCCCC=O		1: hept 2: n-he 3: hept 4: hept	1:: Low Quality 1:: Aquatic OAS 2:: Estrogen Re 3:: USER DEFIN 2:: Low Quality 1:: Bacterial mu 2:: Genotoxicity 3:: USER DEFIN	1:: Low Quality : High Qualit 1:: USER D 2:: Aquatic 3:: Estroge 2:: Low Quality 1:: USER D 2:: Genoto 3:: Bacteria	1:: A 2:: B 3:: C 4:: D 5:: E 6:: E 7:: F
70955-11-2	CCCCCCC=O		1: hexe	1:: High Quality 1:: Canada DSL 2:: DSSTOX 3:: ECHA PR 4:: HPVC OFCD	1:: Low Quality : Low Qualit 1:: TSCA 2:: ECHA P 3:: Canada 4:: DSSTO	1:: C 2:: D 3:: E 4:: H

- The columns represent chemical relations
- The colors represent the quality of relation
- Each row represents the quality of relation
- CAS/Name, Name/2D, and CAS/2D** where number corresponds to name numbering

Text Color	Evaluated Q
Black	N/A Quality
Red	Low Q
Orange	Moderate Q
Green	High Q
Blue	Conflict

Chemical Input QA

The screenshot shows the 'Select chemicals' window with a table of chemical data. Two callout boxes provide instructions:

- Callout 1:** Double click on the column to see sources (2) of the representing names chemical presented in the relation (CAS/Name in this case).
- Callout 2:** Points to a detailed 'Relation CAS/Name' window for CAS 111717, which lists various quality relations such as 'heptaldehyde', 'n-heptanal', and 'heptanal' with their respective quality levels and sources.

Selected	CAS	Smiles	Depiction	Name	CAS/Name	2D/Name	CAS/2D
1. Yes	111-71-7	CCCCCCC=O		1: heptaldehyde 2: n-heptanal 3: heptanal 4: heptanal	1:: Low Quality 1:: Aquatic OASIS 2:: Estrogen Receptor 3:: USER DEFINED 2:: Low Quality 1:: Bacterial mutagenicity OASIS 2:: Genotoxicity OASIS 2:: USER DEFINED	1:: Low Quality 1:: USER DEFINED 2:: Aquatic OASIS 3:: Estrogen Receptor Binding Affinity OASIS 2:: Low Quality 1:: USER DEFINED 2:: Genotoxicity OASIS 3:: Bacterial mutagenicity OASIS	: High Quality 1:: Aquatic OASIS 2:: Bacterial mutagenicity OASIS 3:: Canada DSL 4:: DSSTOX 5:: ECHA CHEM 6:: ECHA PR 7:: ECOTOX
2. Yes	70955-11-2	CCCCCCC=O		1: hexanal	1:: High Quality 1:: Canada DSL 2:: DSSTOX 3:: ECHA PR 4:: HPVC OECD 5:: TSCA	1:: Low Quality 1:: TSCA	: Low Quality 1:: Canada DSL

Relation CAS/Name
CAS: 111717

Relations Quality

- heptaldehyde
 - Aquatic OASIS
 - Estrogen Receptor Binding Affinity OASIS
 - USER DEFINED
- n-heptanal
 - Bacterial mutagenicity ISSSTY
 - Genotoxicity OASIS
 - USER DEFINED
- heptanal
 - Canada DSL
 - DSSTOX
 - ECHA CHEM
 - ECHA PR
 - ECOTOX
 - EINECS
 - HPVC OECD
 - METI Japan
 - NICNAS

Chemical Input

Input target chemical#1 by SMILES

Back to our target chemical, the first one is the actual 1-heptanal while the second one is a mixture containing 1-heptanal. As we are not interested in the mixture this chemical can be removed from the exercise (see next screenshot).

The screenshot shows a 'Select chemicals' dialog box with the following data:

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
1. Yes	111-71-7	CCCCCCC=O		1: heptaldehyde 2: n-heptanal 3: heptanal 4: heptanal (n-heptanal) (heptaldehyde)	1:: Low Quality 1:: Aquatic OASIS 2:: Estrogen Receptor 3:: USER DEFINED 2:: Low Quality 1:: Bacterial mutagenicity 2:: Genotoxicity OASIS 2:: USER DEFINED	1:: Low Quality 1:: USER DEFINED 2:: Aquatic OASIS 3:: Estrogen Receptor 2:: Low Quality 1:: USER DEFINED 2:: Genotoxicity OASIS 3:: Bacterial n	: High Quality 1:: 2:: 3:: 4:: 5:: 6:: 7::
2. Yes	70955-11-2	CCCCCCC=O		1: hexene, hydroformylation products	1:: High Quality 1:: Canada DSL 2:: DSSTOX 3:: ECHA PR 4:: HPVC OECD 5:: TSCA	1:: Low Quality 1:: TSCA 2:: ECHA PR 3:: Canada DSL 4:: DSSTOX 5:: HPVC OEC	: Low Quality 1:: 2:: 3:: 4:: 5::

The second row (CAS 70955-11-2) is circled in red in the original image.

Chemical Input

Input target chemical# 1 by SMILES

Select chemicals

Select All Clear All Invert Selection Selected 1 of 2

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
1. Yes	111-71-7	CCCCCCC=O		1: heptaldehyde 2: n-heptanal 3: heptanal 4: heptanal (n-heptanal) (heptaldehyde)	1:: Low Quality 1:: Aquatic OASIS 2:: Estrogen Receptor 3:: USER DEFINED 2:: Low Quality 1:: Bacterial mutagenicity 2:: Genotoxicity OASIS 2:: USER DEFINED	1:: Low Quality 1:: USER DEFINED 2:: Aquatic OASIS 3:: Estrogen Receptor 2:: Low Quality 1:: USER DEFINED 2:: Genotoxicity OASIS 3:: Bacterial n	: High Quality 1:: 2:: 3:: 4:: 5:: 6:: 7::
2. No	0955-11-2	CCCCCCC=O		1: hexene, hydroformylation products	1:: High Quality 1:: Canada DSL 2:: DSSTOX 3:: ECHA PR 4:: HPVC OECD 5:: TSCA	1:: Low Quality 1:: TSCA 2:: ECHA PR 3:: Canada DSL 4:: DSSTOX 5:: HPVC OEC	: Low Quality 1:: 2:: 3:: 4:: 5::

Tautomeric sets

OK Cancel

1. Click over the first column with label Yes, then the column become unmarked (labeled with No); **2. Click** OK

Chemical Input

Target chemical identity

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options for Document, Single Chemical, and Chemical List. The main workspace is divided into several panels:

- Documents Panel (Left):** Shows a document named "Document_1" with the SMILES string CCCCCCC=O.
- Filter endpoint tree... (Top Center):** A dropdown menu showing "1 [target]".
- Structure Panel (Middle Left):** Displays the chemical structure of heptanal.
- Substance Identity Panel (Middle Right):** A tree view showing the following information:
 - CAS Number: 111-71-7
 - Chemical IDs: EINECS:2038984
 - Chemical Name: heptaldehyde, n-heptanal, heptanal, heptanal (n-heptan...)
 - Molecular Formula: C7H14O
 - Structural Formula: CCCCCC=O
- Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards:** These sections are currently collapsed.

At the bottom left, there is a filter selection dropdown with "Create" and "Apply" buttons.

Chemical Input

Exercise 2

- To add additional chemicals by hand into the matrix, **right-click** above the structure and **select** "Add target" and then "Structure".
- **Enter** the SMILES for 1-hexanal: **CCCCC=O** and **click** "OK" (see next screen shot).

Chemical Input

Input target chemical#2 by SMILES

The screenshot shows the QSAR Toolbox interface with the 'Chemical List' tab active. The 'Documents' panel on the left shows a document with the SMILES string CCCCCCC=O. The main workspace displays a chemical structure and a table of chemical properties. A context menu is open over the structure, with three numbered callouts: 1 points to the right-click action, 2 points to 'Add in category', and 3 points to 'Drawing' in the sub-menu.

Property	Value
CAS Number	111-71-7
EINECS	20389
Chemical Name	heptaldehyde n-heptanal heptanal heptanal (n-hep
Molecular Formula	C7H14O
Structural Formula	CCCCCCC=O

1. **Right-click** in the space above the structure; 2. **Select** Add in category;
3. **Select** Drawing.

Chemical Input

Input target chemical#2 by SMILES

The screenshot displays the QSAR Toolbox software interface. The 'Structure' button in the toolbar is circled in red and labeled with a '1'. The 'SMILES/InChi' input field in the '2D Editor' window contains the text 'CCCCCC=O', which is also circled in red and labeled with a '2'. The 'OK' button at the bottom of the '2D Editor' window is circled in red and labeled with a '3'. The chemical structure of hexanal is shown in the 2D editor window, also circled in red.

1. **Click** on Structure; 2. **Type** CCCCCCC=O in SMILES/InChi window; 3. **Click** OK.

Chemical Input

Input target chemical#2 by SMILES

- The Toolbox finds two chemicals with the same structure and with different QA relations (CAS-Name; 2D -Name; CAS-2D).

The screenshot shows a 'Select chemicals' dialog box with a table of results. The table has columns for 'Selected', 'CAS', 'Smiles', 'Depiction', 'Names', 'CAS/Name', '2D/Name', and 'CAS'. Two entries are shown, both for the chemical hexanal (SMILES: CCCCCC=O).

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS
1. Yes	66-25-1	CCCCCC=O			1:: High 1:: Ar 2:: Bi 3:: Ci 4:: D 5:: E 6:: E 7:: F	1:: High 1:: U 2:: T 3:: E 4:: A 5:: R 6:: N 7:: P	
2. No	110-62-3	CCCCCC=O			1:: High 1:: Ar 2:: Ar 3:: E 4:: E 5:: G	1:: Low 1:: G	

At the bottom of the dialog, there are buttons for 'Tautomeric sets', 'Search', 'OK', and 'Cancel'.

Chemical Input

Target chemical identity

- Click OK to add your target to data matrix
- **Click** on the box next to "Substance Identity"; this displays the chemical identification information. (see next screen shot).

Chemical Input

Target chemical identity

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Document, Single Chemical, and Chemical List. The main workspace is divided into several sections:

- Documents:** Shows a document named 'Document_1' with the SMILES string CCCCCCC=O.
- Structure:** Displays the chemical structure of heptanal.
- Substance Identity:** A tree view showing various identification fields. The 'Chemical Name' field is expanded, showing a list of names: heptaldehyde, n-heptanal, heptanal, and heptanal (n-heptan...). The 'EINECS' field is also expanded, showing two entries: 111-71-7 and 66-25-1. The 'EINECS:2006245' entry is circled in red.
- Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards:** These sections are currently collapsed.

At the bottom of the interface, there is a filter selection dropdown menu with 'select filter type...' and buttons for 'Create' and 'Apply'.

Chemical Input

Input target chemical#3 by SMILES

- To add the third chemical by hand into the matrix, **right-click** above the structure and **select** "Add in category" and then "Drawing".
- **Enter** the SMILES for 3-ethyl-1-pentanal: CCC(CC)CC=O and **click** "OK".
- Your data matrix should now contain your three chemicals (see next screen shot).

Chemical Input

Input target chemical#3 by SMILES

The screenshot shows the QSAR Toolbox interface. The main window displays a list of chemicals with the following details for the selected entry:

- 111-71-7
- EINECS:2038984
- heptaldehyde
- n-heptanal
- heptanal
- heptanal
- C7H14O
- CCCCC

The '2D Editor' window is open, showing the SMILES string CCC(CC)CC=O entered in the 'SMILES/InChI' field. A red circle highlights the SMILES string, and another red circle highlights the 'OK' button. A blue callout box with the number '1' points to the SMILES field, and another with the number '2' points to the 'OK' button.

1. **Type** CCC(CC)CC=O in SMILES/InChI window; 2. **Click** OK.

Chemical Input

Input target chemical#3 by SMILES

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Document, Single Chemical, and Chemical List. The main workspace is divided into a left sidebar for document management and a central area for chemical input and data viewing.

In the left sidebar, under 'Documents', a document named 'Document_1' is shown with the SMILES string CCCCCCC=O. Below this, the SMILES string is repeated, and a chemical structure of heptanal is displayed.

The central area shows a 'Filter endpoint tree...' with three target columns: '1 [target]', '2 [target]', and '3 [target]'. Each column contains a chemical structure of heptanal. Below the structures, a table lists various chemical and physical properties for each target.

	1 [target]	2 [target]	3 [target]
Structure			
Substance Identity			
— CAS Number	111-71-7	66-25-1	N/A
— Chemical IDs	EINECS:2038984	EINECS:2006245	NA
— Chemical Name	heptaldehyde n-heptanal heptanal	hexanal hexaldehyde hexylaldehyde	
— Molecular Formula	C7H14O	C6H12O	C7H14O
— Structural Formula	CCCCCCC=O	CCCCC=O	CCC(CC)CC=O
Physical Chemical Properties			
Environmental Fate and Transport			
Ecotoxicological Information			
Human Health Hazards			

At the bottom of the interface, there is a filter selection dropdown menu with 'select filter type...' and buttons for 'Create' and 'Apply'.

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise 1**
 - Chemical Input
 - Profiling
 - **Endpoints**

Endpoints

- Move directly to the module “Endpoints”.
- Remember, “Endpoints” refer to the electronic process of retrieving fate and toxicity data stored in the Toolbox and it can be gathering in a global fashion or on a more defined basis.
- In this example we only want to retrieve data on toxicity to fish so **select** the following databases containing information on aquatic toxicity:
 - Aquatic ECETOC
 - Aquatic Japan MoE
 - Aquatic OASIS
 - Aquatic US-EPA ECOTOX
- **Click** “Gather Data” (see next screen shot).

Endpoints

Gather data

The screenshot shows the QSAR Toolbox interface. The top toolbar contains buttons for 'Gather', 'Import', 'Export', 'Delete', and 'Tautomerize'. The left sidebar shows a tree view of databases, with 'Ecotoxicological Information' selected. The main window displays a table of chemical data for three targets. A dialog box titled 'Read data?' is open, with the 'OK' button highlighted.

Filter endpoint tree...	1 [target]	2 [target]	3 [target]
Structure			
Substance Identity	111-71-7	66-25-1	N/A
CAS Number	EINECS:2038984	EINECS:2006245	NA
Chemical IDs	heptaldehyde	hexanal	
Chemical Name	n-heptanal	hexaldehyde	
	heptanal	hexylaldehyde	
Molecular Formula	heptanal (n-heptan...	C6H12O	C7H14O
Structural Formula	C7H14O	CCCCC=O	CCC(CC)CC=O
Physical Chemical Properties			
Environmental Fate and Transport			
Ecotoxicological Information			
Human Health Hazards			

1. Select databases related to the target endpoint; **2. Click** Gather; **3. Click** OK.

Endpoints

Available experimental data

- Results are available for two effects:
 - **Growth** for *Tetrahymena pyriformis* for both n-hexanal and n-heptanal.
 - **Mortality** for two species: *Pimephales promelas* and *Poecilia reticulata* for both n-hexanal and n-heptanal (see next screen shot).
- These can potentially be used for read-across to fill in the data gap for the third target: **3-ethyl-1-pentanal** (e.g. using the lowest available LC50 result).

Endpoints

Available experimental data

The screenshot displays the QSAR Toolbox software interface. The main window shows a table of experimental data for various endpoints. The table has four columns: 'Structure', '1 [target]', '2 [target]', and '3 [target]'. The 'Structure' column contains a hierarchical tree of endpoints, and the other columns contain numerical data and chemical structures. Two rows are circled in red:

Structure	1 [target]	2 [target]	3 [target]
Structure			
Structure			
48 h			
Protozoa			
Ciliophora			
Ciliatea			
<i>Tetrahymena pyriformis</i>	(2/2)	M: 114 mg/L	M: 152 mg/L
Immobilisation			
Intoxication	(1/12)		M: 9 mg/L, 5.9 mg...
Mortality			
EC50			
LC50			
3 h	(1/2)		M: >10 mg/L, >22 ...
12 h			
24 h	(1/6)		M: >18 mg/L, >10 ...
48 h	(1/6)		M: 5.5(4.5;6.7) mg...
72 h	(1/4)		M: ≈4.3 mg/L, 5.5(...)
96 h			
Animalia			
Arthropoda (Invertebrates)	(1/2)		M: 5.3(4.3;6.7) mg...
Chordata (Vertebrates)			
Actinopterygii (Fish)			
<i>Lepomis macrochirus</i>			
<i>Leuciscus idus</i>			
<i>Pimephales promelas</i>	(2/6)	M: 12 mg/L	M: 17.8 mg/L, 22(2...)
<i>Poecilia reticulata</i>	(2/2)	M: 8.86 mg/L	M: 9.79 mg/L
Undefined Test Organisms (species)			
Mollusca (Invertebrates)			

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise 1**
 - Chemical Input
 - **Profiling**
 - Endpoints

Profiling

- Click on “Profiling” to move back (**yes back**) to the previous module.
- Remember that “Profiling” refers to the process of retrieving information on the target compounds, other than and toxicity data.
- Available information includes likely mechanism(s) of action.
- In this exercise we will use the profiling results to evaluate the robustness of the analogue approach.

Profiling

Profiling the target chemical

- As you remember, the outcome of the profiling determines the most appropriate way to search for analogues.
- For this example the following mechanistic and endpoint specific profiling methods should be selected:
 - Aquatic toxicity classification by ECOSAR
 - Acute aquatic toxicity MOA by OASIS
 - Acute aquatic toxicity classification by Verhaar(Modified)
- Select those 3 “profiling methods” by **clicking** on the boxes before the names of the profilers before **clicking** “Apply” (see next screen shot).

Profiling

Profiling the target chemical

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is open, showing 'Profiling Schemes' with an 'Apply' button circled in red. Below the menu, the 'Profiling methods' list is visible, with 'Endpoint Specific' circled in red. The main window displays a tree view of endpoints and a table of results for three target chemicals.

Structure	1 [target]	2 [target]	3 [target]
Structure	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>
Immobilisation	(1/12)		
Intoxication		M: 9 mg/L, 5.9 mg...	
Mobility			
Mortality			
EC50			
EL50			
LC50			
3 h	(1/2)	M: >10 mg/L, >22 ...	
12 h			
24 h	(1/6)	M: >18 mg/L, >10 ...	
48 h	(1/6)	M: 5.5(4.5;6.7) mg...	
72 h	(1/4)	M: ≈4.3 mg/L, 5.5(...)	
96 h			
Animalia			
Arthropoda (Invertebrates)	(1/2)	M: 5.3(4.3;6.7) mg...	
Chordata (Vertebrates)			
Actinopterygii (Fish)			
Lepomis macrochirus			
Leuciscus idus			
Oncorhynchus mykiss			
Pimephales promelas	(2/6)	M: 12 mg/L	M: 17.8 mg/L, 22(2)
Poecilia reticulata	(2/2)	M: 8.86 mg/L	M: 9.79 mg/L
Salmo gairdneri (new name: oncorh...)			
Undefined Test organisms (species)			
Mollusca (Invertebrates)			
Rhynchonitiformes (Mollusca)			
Rhynchonitiformes (Mollusca)			
Rhynchonitiformes (Mollusca)			

1. Select the profilers related to the target endpoint; **2. 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 appeared as a dropdown box under the target chemical.
- The target and source chemicals have the same mechanisms or modes of action relevant for acute aquatic toxicity.
- The Toolbox does not provide any arguments against read-across (see next screen shot).

Profiling

Profiles of the targets

The screenshot shows the QSAR Toolbox interface. On the left, the 'Endpoint Specific' list is expanded to show 'Acute aquatic toxicity classification by ECOSAR'. The main window displays a table with three columns for different targets. A red circle highlights the classification results for 'Class 3 (unspecific... Aldehydes (Mono))' in the third column, with a callout box labeled '1' pointing to it. A text box at the bottom explains that a right-click is used to see why this target is classified by ECOSAR.

1. Right click to see why this target is classification by ECOSAR

- In this case the target and source chemicals have the same mechanisms and modes of action.
- So the Toolbox does not provide any arguments against read-across.
- This step is critical for next grouping of analogues.

Aquatic toxicity classification by ECOSAR of "n-heptanal"

Category definitions

- ▲ Aquatic toxicity classification by ECOSAR
 - Acid Halides
 - Acid moiety
 - Acrylamides
 - Acrylates
 - Aldehydes (Mono)
 - Aldehydes (Poly)
 - Aliphatic Amines
 - Alkoxy Silanes
 - Amides
 - Anilines (amino-meta)
 - Anilines (amino-ortho)
 - Anilines (amino-para)
 - Anilines (Hindered)
 - Anilines (Unhindered)
 - Aziridines
 - Benzodioxoles
 - Benzotriazoles
 - Benzoylcyclohexanedione
 - Benzyl Alcohols
 - Benzyl Halides
 - Benzyl Imines
 - Benzyl Nitriles
 - Carbamate Esters
 - Carbamate Esters, Phenyl
 - Carbonyl Ureas
 - Diazoniums, Aromatic
 - Diketones
 - Epoxides, mono acid subst
 - Epoxides, mono
 - Epoxides, Poly
 - Esters (phosphate)
 - Esters (Phosphinates)
 - Esters
 - Esters, Dithiophosphates
 - Esters, Monothiophosphates
 - Halo Acids
 - Halo Alcohols
 - Halo Epoxides
 - Halo Ester
 - Halo Ethers
 - Halo Ketones (2 free H)
 - Halo Nitriles
 - Haloacetamides
 - Haloimides
 - Halopyridines
 - Hydrazines
 - Hydroquinones
 - Imidazoles
 - Imides
 - Inorganic Compound

Profile Description

ECOSAR Class Definition: Aldehydes (Mono)

The Aldehydes (Mono) class is identified by the following structure:

R1 - attachment must be either an alkyl carbon, aromatic carbon, carbonyl or hydrogen.

The structure can contain only one aldehyde functional group to be classified as Aldehydes (Mono). If a structure contains more than one aldehyde group, it will be classified as Aldehydes (Poly).

If the R1 attachment is an olefinic carbon, acetylenic carbon or allyl group (-CC=C), the structure will be classified as a Vinyl/Allyl Aldehyde. In the current ECOSAR program, structures classified as Vinyl/Allyl Aldehydes are not additionally classified as Aldehydes (Mono).

SMILES String Identifications:
O=C or O=C[H] or O=CH
O=CC ... (no other attachments to carbonyl)
O=Cc1cccc1 ... (no other attachments to carbonyl)

Associated ECOSAR Class(s):

Aldehydes (Poly) - If a structure contains more than one aldehyde group, it will be classified as Aldehydes (Poly) instead of Aldehydes (Mono)

Vinyl/Allyl Aldehydes - If the R1 attachment is an olefinic carbon, acetylenic carbon or allyl group (-CC=C), the structure will be classified as a Vinyl/Allyl Aldehyde. In the current ECOSAR program, structures classified as Vinyl/Allyl Aldehydes are not additionally classified as Aldehydes (Mono).

Example Aldehydes (Mono):

CAS No.	Name	SMILES Notation
75-07-0	Acetaldehyde	O=CC
123-72-8	Butanal	O=CCCC
555-16-8	Benzaldehyde, 4-nitro-	O=Cc(ccc(N(=O)(=O))c1)c1
148-53-8	o-Vanillin	O=Cc(c(O)c(OC)cc1)c1
454-89-7	Benzaldehyde, 3-(trifluoromethyl)-	O=C(ccc1C(F)(F)F)c1

Profiling Recap

- You have entered the source and target chemicals being sure of the correct structures.
- You have checked the relevant databases for available experimental results.
- You have profiled the source and target chemicals.
- You have evaluated the robustness of the analogue approach and concluded that the read-across may be acceptable.

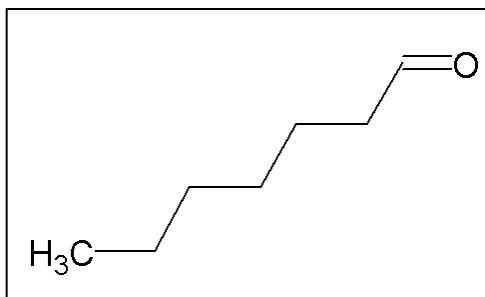
Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise 1
- **Workflow of the exercise 2**
 - **Chemical Input**
 - Profiling
 - Endpoints

Chemical Input Exercise 2

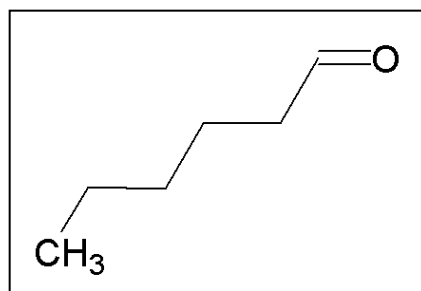
Read-across of acute toxicity to fish from 1-heptanal and 1-hexanal to 2,5-diene-4-methyl-hexan-1-al.

1-heptanal

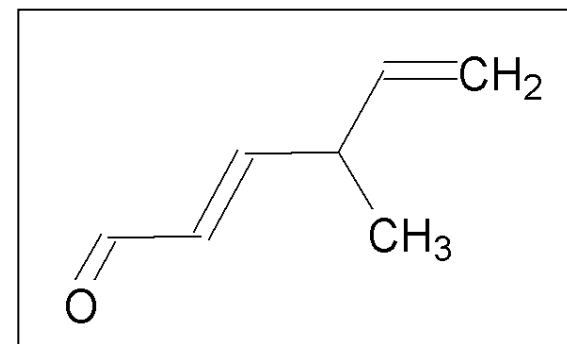


+

1-hexanal



2,5-diene-4-methyl-hexan-1-al



Chemical Input

Exercise 2

- In the second example, we use the same source chemicals and a different target chemical.
- We can therefore simply delete the previous target chemical and enter the identity of the new target chemical.
- **Right-click** above the structure of chemical 3-ethyl-1-pentanal and **select** "Delete chemical" (see next screen shot).

Chemical Input Exercise 2

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

- Top Bar:** Navigation icons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report.
- Left Panel:** Profiling methods and Metabolism/Transformations sections with various checkboxes.
- Table:** A table with columns for target IDs (1, 2, 3) and rows for Substance Identity, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, Human Health Hazards, Profile, and Endpoint Specific.
- Context Menu:** A menu is open over the first target, listing actions such as 'Remove this target', 'Set AOP target', 'Add to Study Pad', 'Delete', and 'Export CAS list'. The 'Delete' option is circled in red.
- Callouts:** A blue callout '1' points to the first target, and another blue callout '2' points to the 'Delete' option in the menu.

1. Right click on the previous target; 2. Select Delete chemical.

Chemical Input

Exercise 2

- Add the new target chemical as in the previous exercise.
- To add the third chemical by hand into the matrix, **right-click** above the structure and **select** "Add category" and then "Drawing".
- **Enter** the SMILES for 2,5-diene-4-methyl-hexan-1-al: O=CC=CC(C)C=C and **click** "OK".

Chemical Input Exercise 2

The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes tabs for Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The 'Input' tab is active, showing a 'Filter endpoint tree...' panel on the left with various endpoints like 'Acute aquatic toxicity classification by Verhaar (Modified)'. The main workspace displays a table with two columns: '1 [target]' and '2 [target]'. The table contains chemical data for two targets, including CAS numbers, EINECS codes, chemical names (e.g., heptaldehyde, hexanal), molecular formulas, and structural formulas. A context menu is open over the structure column, listing actions such as 'Focus', 'Remove this target', 'Add category', and 'Drawing'. Three callouts are present: '1' points to the right-click action above the structure; '2' points to the 'Add category' option in the menu; '3' points to the 'Drawing' option in the sub-menu.

1. Right-click above the structure; 2. Select Add category; 3. Select Drawing

Chemical Input Exercise 2

The screenshot shows the QSAR Toolbox software interface. The main window is divided into several panels:

- Top Panel:** Navigation tabs for Profiling, Endpoint, Category Definition, Data Gap Filling, and Report.
- Left Panel:** Profiling methods and Metabolism/Transformations sections with various checkboxes.
- Center Panel:** Filter endpoint tree... showing a tree structure of endpoints.
- Right Panel:** 2D Editor window with a SMILES/InChI input field containing O=CC=CC(C)C=C (circled in red with a blue callout '4'). Below the input field is a grid of chemical templates. A red circle highlights a specific template in the grid, and a blue callout '5' points to the 'OK' button at the bottom of the 2D editor window.

4. Type O=CC=CC(C)C=C in SMILES/InChI window; 5. Click OK.

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise 1
- **Workflow of the exercise 2**
 - Chemical Input
 - **Profiling**
 - Endpoints

Profiling

- In the module profiling, profile the new target chemical with the 3 profilers relevant for aquatic toxicity, in the same way as for the previous example.

Profiling

Profiles of the targets

The screenshot shows the QSAR Toolbox interface with the 'Profiling' tab selected. The 'Endpoint Specific' list is checked, and the 'ECOSAR' classification is highlighted in red. A callout box points to the ECOSAR classification with the text '1. Right click to see why this target is classification by ECOSAR'.

Structure	1 [target]	2 [target]	3 [target]
Substance Identity			
— CAS Number	66-25-1	66-25-1	N/A
— Chemical IDs	EINECS:2038984	EINECS:2006245	NA
— Chemical Name	heptaldehyde n-heptanal heptanal	hexanal hexaldehyde hexylaldehyde	
— Molecular Formula	heptanal (n-heptan... C7H14O	C6H12O	C7H10O
— Structural Formula	CCCCCCC=O	CCCCC=O	CC(C=C)C=CC=O
Physical Chemical Properties			
Environmental Fate and Transport			
Ecotoxicological Information	(2/49) M: 12 mg/L, 8.86 ...	M: 17.8 mg/L, 9.79...	
Human Health Hazards			
Profile			
— Endpoint Specific			
— Acute aquatic toxicity classification by Verhaar (Modified)	Class 3 (unspecific... Aldehydes	Class 3 (unspecific... Aldehydes	Class 3 (unspecific... Aldehydes
— Acute aquatic toxicity MOA by OASIS	Aldehydes (Mono)	Aldehydes (Mono)	Vinyl/Allyl Aldehydes
— Aquatic toxicity classification by ECOSAR			

- In this case the target and analogue (source) chemicals do not have same mechanism and modes of action, regarding ECOSAR classification
- So the read-across is questionable in this case with this particular analogues

1. Right click to see why this target is classification by ECOSAR

Recap

- You have replaced a target chemical with another target chemical in the data matrix.
- You have profiled the new target chemical.
- You have evaluated the robustness of the analogue approach and concluded that the read-across may not be acceptable by using the current analogue chemicals (source)
- The further workflow is to search for more suitable target analogues

Searching for More Suitable Analogues

- Before searching for more suitable analogues, delete n-hexanal and n-heptanal from the data matrix by **right-clicking** above each of them and **select** “Delete chemical” or **right-clicking** above the target (2,5-diene-4-methyl-hexan-1-al) and **select** “Delete all except current” (see next screen shot).
- The aim of the next part of the exercise will be to find analogues which have the same profiling results as the target chemical.

Searching for More Suitable Analogues

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' section is active, showing various profiling methods and metabolism/transformation options. The main workspace is divided into a 'Filter endpoint tree...' on the left and a table of targets on the right. The table has three columns labeled '1 [target]', '2 [target]', and '3 [target]'. A context menu is open over the table, with a red '1' pointing to the target space and a red '2' pointing to the 'Delete all except current' option.

Structure	1 [target]	2 [target]	3 [target]
Substance Identity			
CAS Number	111-71-7	66-25-1	N/A
Chemical IDs	EINECS:2038984	EINECS:2006245	NA
Chemical Name	heptaldehyde n-heptanal heptanal heptanal (n-heptan...	hexanal hexaldehyde hexylaldehyde	
Molecular Formula	C7H14O	C6H12O	C7H10O
Structural Formula	CCCCC=O	CCCCC=O	CC(C=C)C=O
Physical Chemical Properties	(2/49) M: 12 mg/L, 8.86 ...	M: 17.8 mg/L, 9.79...	
Profile			
Endpoint Specific			
Acute aquatic toxicity classification by Verhaar (M...	Class 3 (unspecific...	Class 3 (unspecific...	Class 3 (uns...
Acute aquatic toxicity MOA by OASIS	Aldehydes	Aldehydes	Aldehydes
Aquatic toxicity classification by ECOSAR	Aldehydes (Mono)	Aldehydes (Mono)	Vinyl/Allyl Al...

1. Right click in the space above the target chemical; **2. Select** Delete All except current.

Searching for More Suitable Analogues

The screenshot displays the QSAR Toolbox software interface. The 'Category Definition' module is highlighted with a red circle and a callout box containing the number '3'. The interface shows a sidebar with 'Grouping methods' and 'Defined Categories', a central 'Filter endpoint tree...' panel, and a right-hand table with chemical data. The table includes fields for CAS Number, Chemical IDs, Chemical Name, Molecular Formula (C7H10O), and Structural Formula (CC(C=C)C=CC=O). Under the 'Profile' section, it lists classification results: 'Class 3 (unspecific... Aldehydes)' and 'Vinyl/Allyl Aldehydes'.

3. Move to the module Category Definition to launch a search for more suitable analogues.

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise 1
- **Workflow of the exercise 2**
 - Chemical Input
 - Profiling
 - Endpoints
 - **Category definition**

Category definition

Searching for More Suitable Analogues

- Currently it is not possible to query directly by several profiling results in parallel. The user has first to query according to one profiler and then subcategorise the results step-by-step according to other profilers.
- For this example, the user could first select the ECOSAR profiler of the target chemical and query for all the chemicals with the same structural feature in the selected databases (see next screen shot).

Category definition

Searching for More Suitable Analogues

The screenshot displays the QSAR Toolbox interface during the 'Category Definition' process. The top toolbar contains icons for 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Category Definition' toolbar includes 'Define', 'Define with metabolism', 'Subcategorize', 'Combine', 'Clustering', 'Delete', and 'Delete All'. The 'Define' button is circled in red and labeled '2'. The 'Grouping methods' list on the left has 'Aquatic toxicity classification by ECOSAR' circled in red and labeled '1'. A dialog box titled 'Aquatic toxicity classification by ECOSAR' is open, showing a list of profiles with 'Vinyl/Allyl Aldehydes' selected and the 'OK' button highlighted in red and labeled '3'. The main window shows a 'Filter endpoint tree...' with a chemical structure and a table of properties including CAS Number (N/A), Chemical IDs (NA), Molecular Formula (C7H10O), and Structural Formula (CC(C=C)C=CC=O). The 'Profile' section lists 'Acute aquatic toxicity classification by Verhaar (M...)' as 'Class 3 (unspecific... Aldehydes)' and 'Aquatic toxicity classification by ECOSAR' as 'Vinyl/Allyl Aldehydes'.

1. Highlight "Aquatic toxicity classification by ECOSAR"; **2. Click** Define and confirm the category from classification by ECOSAR profiler; **3. Click** OK.

Category definition

Searching for More Suitable Analogues

The screenshot shows the QSAR Toolbox software interface. The main window is titled 'Category Definition' and displays a 'Filter endpoint tree...' window. The tree structure is as follows:

- Structure
- Substance Identity
 - CAS Number: N/A
 - Chemical IDs: NA
 - Chemical Name
 - Molecular Formula: C7H10O
 - Structural Formula: CC(C=C)C=CC=O
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards
- Profile
 - Endpoint Specific
 - Acute aquatic toxicity classification by Verhaar (M...): Class 3 (unspecific...)
 - Acute aquatic toxicity MOA by OASIS: Aldehydes
 - Aquatic toxicity classification by ECOSAR: Vinyl/Allyl Aldehydes

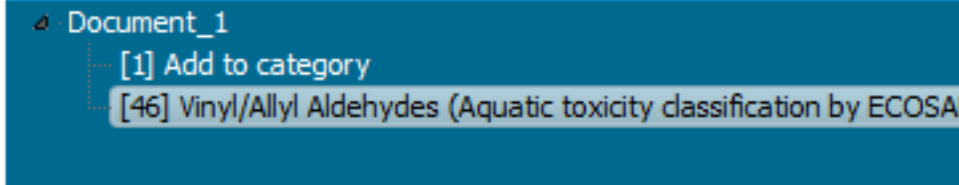
A dialog box titled 'Define category name' is open, showing the category name 'Aldehydes (Aquatic toxicity classification by ECOSAR)' for 46 chemicals. The 'OK' button is highlighted with a blue callout box containing the number '1'.

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

Category definition

Searching for More Suitable Analogues

- The Toolbox now identifies all chemicals corresponding to mechanism “Vinyl/Allyl Aldehydes” by Aquatic toxicity classification by ECOSAR listed in the databases selected under “Endpoints”.
- 45 analogues are identified. Along with the target they form a mechanistic category used for gap filling.
- The name of the category appears in the “Defined Categories” window, indicating the number of substances belonging to the category.

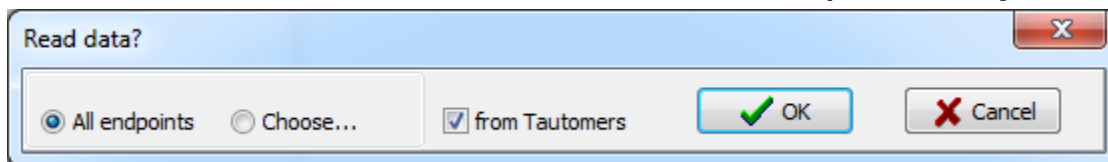


```
Document_1
  [1] Add to category
  [46] Vinyl/Allyl Aldehydes (Aquatic toxicity classification by ECOSA)
```

Category definition

Reading data for Analogues

- The Toolbox will now retrieve those chemicals that have the same structural functionality as the target chemical based on ECOSAR profiler (Vinyl/Allyl aldehydes).
- 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, as only databases are selected that contain information for aquatic toxicity endpoints, both options give the same results.

Category definition

Reading data for Analogues

Due to the overlap between the Toolbox databases same data for intersecting chemicals is found simultaneously in more than one database. The data redundancy is identified and the user has the opportunity to select either a single data value or all data values.

Repeated values for: 90 data-points, 31 groups, 14 chemicals

Data points...

	Endpoint	CAS	Structure	Value	additional_comm
<input checked="" type="checkbox"/>	LC50	107-02-8	<chem>CC=CC(F)=O</chem>	27(24;30) micrograms per liter	
<input checked="" type="checkbox"/>	LC50	107-02-8	<chem>CC=CC(F)=O</chem>	27(24;30) micrograms per liter	
<input checked="" type="checkbox"/>	IGC50	N/A	<chem>CC1=CC=C(C=C1)C2=CC=CC=C2</chem>	0.000832 mol/L	
<input checked="" type="checkbox"/>	IGC50	N/A	<chem>CC1=CC=C(C=C1)C2=CC=CC=C2</chem>	0.000832 mol/L	
<input checked="" type="checkbox"/>	LC50	4170-30-3	<chem>CC=CC=C</chem>	0.71 miligram per liter	TIME/OTHER DURATIONS ALL REPORTED//SO TO 47.00* mg/L//CL/21.00*

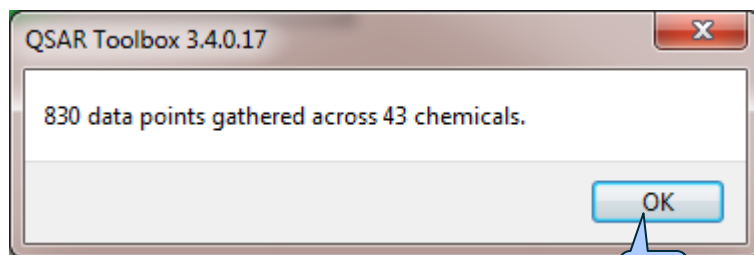
Buttons: Select one, Invert, Check All, Uncheck All, OK, Cancel

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

Category definition

Reading data for Analogues

The system automatically gives indication for the number of gather experimental data points



1. Click OK

Category Definition

Defined category

The screenshot shows the QSAR Toolbox interface with the 'Category Definition' workflow selected. The 'Defined Categories' list on the left includes 'Vinyl/Allyl Aldehydes (Aquatic toxicity classification by ECOSAR)'. The main table displays a filter endpoint tree with a red circle highlighting the 'Aquatic toxicity classification by ECOSAR' row, which is associated with the 'Class 3 (unspecific... Aldehydes Vinyl/Allyl Aldehydes)' category.

Filter endpoint tree...	1 [target]	2	3	4	5	6	7
Structure							
Substance Identity							
Physical Chemical Properties							
Environmental Fate and Transport							
Ecotoxicological Information							
Aquatic Toxicity	(38/392)	M: 0.0114 mg/L, 0...	M: 14 mg/L, 14.6 ...	M: 5.94 mg/L, 52.9...	M: 3.4 mg/L, 7.05(...	M: 0.014 mg/L, 0.0...	M: 103 mg...
Sediment Toxicity							
Terrestrial Toxicity	(12/430)		M: 5.61E-5 g, 1.74...	M: 24.1 mg/L		M: 1.2E3 milligram...	
Human Health Hazards							
Profile							
Endpoint Specific							
Acute aquatic toxicity classification by Verhaar (M...							
Acute aquatic toxicity MOA by OASIS							
Aquatic toxicity classification by ECOSAR							

Category Definition Subcategorisation

- After the available data has been retrieved, the user can then further subcategorize the results according to the following subcategorisations:
 - MOA of action
 - Verhaar classification
- These steps are summarized in the next screen shots.

Category Definition

Subcategorisation by Acute aquatic toxicity MOA by OASIS

The screenshot displays the QSAR Toolbox software interface during the 'Category Definition' process. The main window shows a 'Filter endpoint tree...' table with columns for 'Structure', '1 [target]', '2', '3', and '4'. The 'Acute aquatic toxicity MOA by OASIS' category is selected in the tree. A 'Define category name' dialog box is open, showing the category name 'Categorized: Acute aquatic toxicity MOA by OASIS'. On the right, a 'Grouping methods' list is visible, with 'Aldehydes' and 'Acute aquatic toxicity MOA by OASIS' circled in red. A 'Metabolism/Transformations' panel is also visible at the bottom right. Five numbered callouts (1-5) highlight key steps: 1. Selecting the current category in the 'Defined Categories' list; 2. Clicking the 'Subcategorize' button in the toolbar; 3. Selecting 'Acute aquatic toxicity MOA by OASIS' in the 'Grouping methods' list; 4. Removing dissimilar chemicals in the 'Metabolism/Transformations' panel; 5. Clicking 'OK' in the 'Define category name' dialog box.

1. Select current category; 2. Click Subcategorize; 3. Select Acute aquatic toxicity MOA by OASIS profiler; 4. Remove dissimilar chemicals and 5. Confirm new category by clicking OK.

Category Definition

Subcategorisation by Acute aquatic classification by Verhaar(Modified)

The screenshot displays the QSAR TOOLBOX interface during a subcategorization task. The main window shows a 'Filter endpoint tree' table with columns for target and analogues. The 'Subcategorization' dialog box is open on the right, showing a list of grouping methods and a 'Target' field. The 'Defined Categories' list at the bottom left shows a category with 39 analogues. The 'Subcategorization' dialog box has a 'Target' field containing 'Class 3 (unspecific reactivity)' and an 'Analogues' field containing '39 Class 3 (unspecific reactivity)'. The 'Subcategorization' dialog box also has a 'Differ from target by' section with radio buttons for 'At least one category' and 'All categories'. The 'Subcategorization' dialog box has a 'Selected 0 (39/39)' indicator and 'Select different' and 'Remove' buttons.

Filter endpoint tree...	1 [target]	2	3
Structure	<chem>C=CC=C</chem>	<chem>C=CC=C</chem>	<chem>C=CC=C</chem>
Substance Identity			
Physical Chemical Properties			
Environmental Fate and Transport			
Ecotoxicological Information			
Aquatic Toxicity (32/78)		M: 5.94 mg/L, 52.9...	M: 3.4 mg/L, 7.05(...
Sediment Toxicity			
Terrestrial Toxicity (9/84)			M: 1.2E3 milligram...
Human Health Hazards			
Profile			
Endpoint Specific			
Acute aquatic toxicity classification by Verhaar (M...	Class 3 (unspecific...	Class 3 (unspecific...	Class 3 (unspecific...
Acute aquatic toxicity MOA by OASIS	Aldehydes		
Aquatic toxicity classification by ECOSAR	Vinyl/Allyl Aldehydes		

1. Select category with 39 analogues; 2. Click Subcategorise; 3. Select Verhaar profiler; Note all analogues are in the same category as the target chemical so no further action is required.

Category Definition

Results after subcategorisation

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition' (selected), 'Data Gap Filling', and 'Report'. Below this, a secondary bar shows 'Categorize' and 'Delete' options. The main workspace is divided into several panels:

- Left Panel:** A tree view of 'Grouping methods' and 'Endpoint Specific' methods. Under 'Endpoint Specific', 'Aquatic toxicity classification by ECOSAR' is selected, and it is further subcategorized into 'Acute aquatic toxicity MOA by OASIS'.
- Top Center Panel:** A 'Filter endpoint tree...' section with a 'Structure' tab and a 'Substance Identity' tree. The 'Aquatic Toxicity' node is expanded, showing counts for various sub-categories.
- Table:** A table with 7 columns representing different chemical classes. The first column is labeled '1 [target]'. The table contains chemical structures in the first row and numerical data in subsequent rows. The data is as follows:

1 [target]	2	3	4	5	6	7
(32/78)	M: 5.94 mg/L, 52.9...	M: 3.4 mg/L, 7.05(...	M: 0.014 mg/L, 0.0...	M: 103 mg/L, 569 ...	M: 138(123;165) m...	M: 3.92 mg...
(9/84)		M: 1.2E3 milligram...			M: 17.5 mg/L, 0.8 ...	
Class 3 (unspecific... Aldehydes	Class 3 (unspecific... Aldehydes	Class 3 (unspecific... Aldehydes	Class 3 (unspecific... Aldehydes	Class 3 (unspecific... Aldehydes	Class 3 (unspecific... Aldehydes	Class 3 (unspecific... Aldehydes
Vinyl/Allyl Aldehydes						
- Bottom Left Panel:** 'Defined Categories' showing a hierarchy: Document_1 > [1] Add to category > [46] Vinyl/Allyl Aldehydes (Aquatic toxicity classification by ECOSAR) > [40] Subcategorized: Acute aquatic toxicity MOA by OASIS > [40] Subcategorized: Acute aquatic toxicity classification by ECOSAR.

Category Definition

Interpretation of the results

- Following the above-described subcategorisation exercise, 39 chemicals are left in the category. All have same mechanisms of action.
- The result is a group of chemicals which are classified as Vinyl/Allyl class by ECOSAR category and have same mode of action according to the MOA profiler.
- For 4 chemicals, experimental results for acute toxicity to fish are available- 4 chemicals have 96h-LC50 results from 3.4 to 7.29 mg/l for *Pimephales promelas*; 2 chemicals have 96h-LC50 results from 7.62 to 9.81mg/l for *Poecilia reticulata*; 1 chemical has 96h-LC50 0.91mg/l for *Oryzias latipes*.

(see next two screen shots)

Category Definition

Interpretation of the results

96h-LC50 for *Pimephales promelas*

The screenshot shows the QSAR Toolbox interface with the 'Category Definition' workflow active. The 'Filter endpoint tree...' panel is open, showing a hierarchical tree of endpoints. A context menu is open over the 'Pimephales promelas' entry, with 'Sort (targets priority)' selected and 'Descending' chosen. Three callout boxes (1, 2, 3) highlight the right-click, the 'Sort' option, and the 'Descending' option respectively. The main table shows chemical structures and associated data for various endpoints.

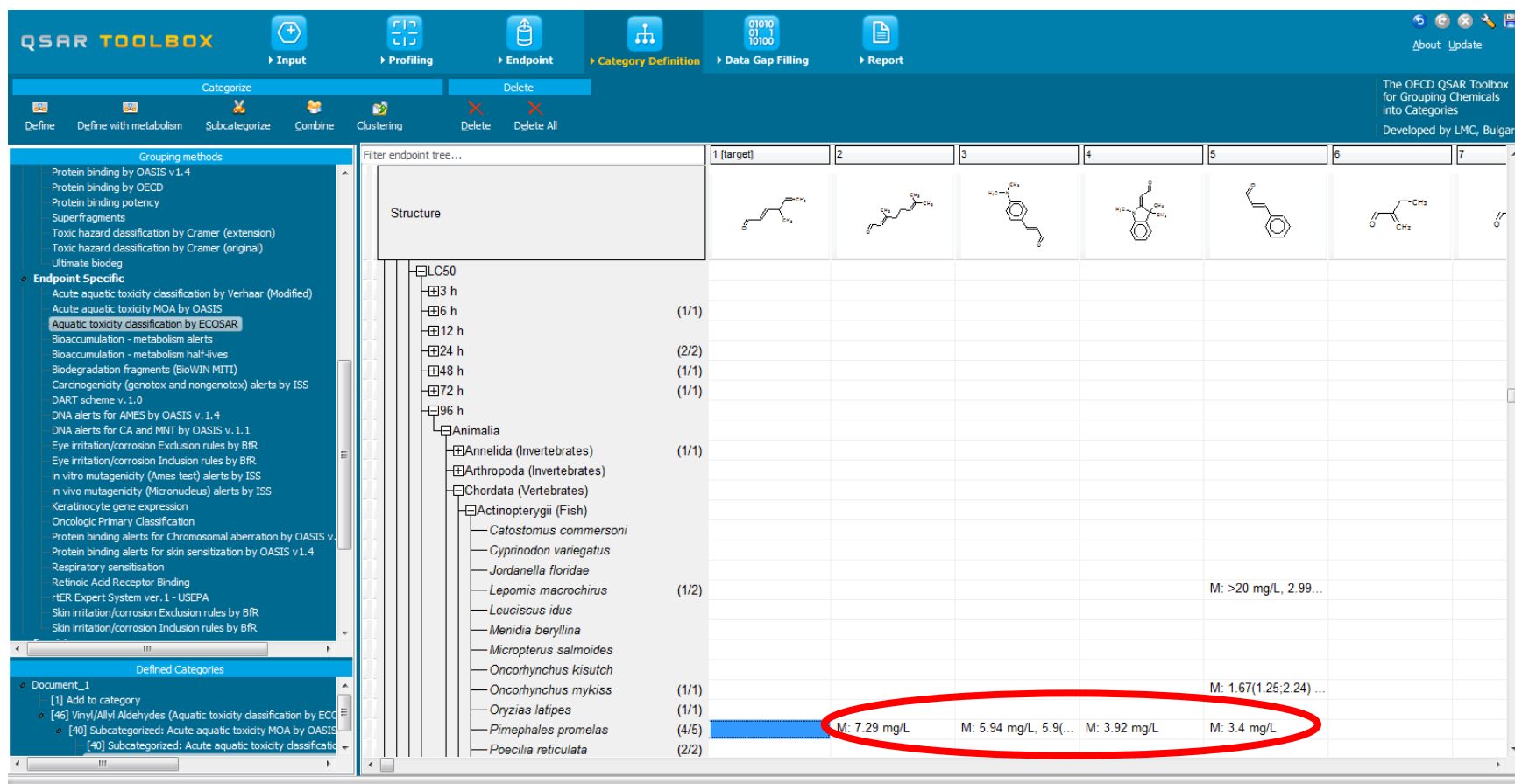
Endpoint	Structure	1 [target]	2	3	4	5	6	7
LC50								
3 h								
6 h		(1/1)					M: 500 mg/L	
12 h								
24 h		(2/2)					M: 78.3(71.9;84.8) ...	
48 h		(1/1)						
72 h		(1/1)						
96 h								
Animalia		(1/1)						
Arthropoda (Invertebrates)								
Chordata (Vertebrates)								
Actinopterygii (Fish)								
Catostomus commersoni								
Cyprinodon variegatus								
Jordanella floridae								
Lepomis gibbosus		(1/1)					M: >20 mg/L, 2.99...	
Leuciscus leuciscus								
Melanostictus albertanus								
Micropterus salmoides								
Oncorhynchus kisutch								
Oncorhynchus mykiss		(1/1)					M: 1.67(1.25;2.24) ...	
Oryzias latipes		(1/1)						
Pimephales promelas		(4/5)					M: 0.91 mg/L	
Poecilia reticulata		(2/2)						

1. Right click above the *Pimephales promelas*; **2. Select Sort (targets priority)**, then **3. Descending**.

Category Definition

Interpretation of the results

96h-LC50 from 3.4 to 7.29 mg/l for *Pimephales promelas*



The screenshot displays the QSAR Toolbox interface during the 'Category Definition' process. The 'Filter endpoint tree...' panel is expanded to show a list of endpoints, with '96 h' selected. The 'Structure' panel shows chemical structures for various endpoints. The 'Defined Categories' panel shows a list of categories, with '40] Subcategorized: Acute aquatic toxicity MOA by OASIS' selected. The 'Endpoint' table shows the results for the selected endpoint, with a red circle highlighting the LC50 values for 'Pimephales promelas' (7.29 mg/L) and 'Pimephales promelas' (3.4 mg/L).

Endpoint	1 [target]	2	3	4	5	6	7
LC50							
3 h							
6 h	(1/1)						
12 h	(2/2)						
24 h	(1/1)						
48 h	(1/1)						
72 h	(1/1)						
96 h							
Animalia	(1/1)						
Annelida (Invertebrates)							
Arthropoda (Invertebrates)							
Chordata (Vertebrates)							
Actinopterygii (Fish)							
<i>Catostomus commersoni</i>							
<i>Cyprinodon variegatus</i>							
<i>Jordanella floridae</i>							
<i>Lepomis macrochirus</i>	(1/2)					M: >20 mg/L, 2.99...	
<i>Leuciscus idus</i>							
<i>Menidia beryllina</i>							
<i>Micropterus salmoides</i>							
<i>Oncorhynchus kisutch</i>							
<i>Oncorhynchus mykiss</i>	(1/1)					M: 1.67(1.25;2.24) ...	
<i>Oryzias latipes</i>	(1/1)						
<i>Pimephales promelas</i>	(4/5)	M: 7.29 mg/L	M: 5.94 mg/L, 5.9(...	M: 3.92 mg/L			M: 3.4 mg/L
<i>Poecilia reticulata</i>	(2/2)						

Category Definition

Interpretation of the results

96h-LC50 results for *Poecilia reticulata*

The screenshot shows the QSAR Toolbox software interface. The main window displays a table of results for 96h-LC50 tests. The 'Filter endpoint tree...' panel is open, showing a hierarchical tree of endpoints and taxonomic groups. A context menu is open over the 'Poecilia reticulata' entry, with 'Sort (targets priority)' selected and 'Descending' chosen. Red callouts 1, 2, and 3 highlight the right-click, the 'Sort' menu, and the 'Descending' option respectively.

Structure	1 [target]	2	3	4	5	6	7
Structure							
3 h							
6 h	(1/1)						
12 h							
24 h	(2/2)						
48 h	(1/1)						
72 h	(1/1)						
96 h							
Animalia							
Arthropoda (Invertebrates)	(1/1)						
Arthropoda (Invertebrates)							
Chordata (Vertebrates)							
Actinopterygii (Fish)							
Catostomus commersoni							
Cyprinodon variegatus							
Jordanella floridae							
Lepomis macrochirus							
Leuciscus idus							M: >20 mg/L, 2.99...
Menidia beryllina							
Micropterus salmoides							
Oreochromis mossambicus							
Oreochromis mykiss							M: 1.67(1.25;2.24) ...
Oreochromis niloticus							
Pimephales promelas							
Poecilia reticulata		7.29 mg/L	M: 5.94 mg/L, 5.9(...	M: 3.92 mg/L			M: 3.4 mg/L
Salmo gairdneri (new name: oncorh...							

1. Right click above the *Poecilia reticulata*; **2. Select Sort (targets priority)**, then **3. Descending**.

Category Definition

Interpretation of the results

96h-LC50 results from 7.6 to 9.8 mg/l for *Poecilia reticulata*.

The screenshot displays the QSAR Toolbox software interface, specifically the 'Category Definition' workflow. The 'Filter endpoint tree...' panel is expanded to show a list of species and their associated LC50 values. The row for 'Poecilia reticulata' is highlighted, with its LC50 values of 9.81 mg/L and 7.62 mg/L circled in red.

Species	LC50 (mg/L)
3 h	
6 h (1/1)	M: 500 mg/L
12 h	
24 h (2/2)	M: 78.3(71.9;84.8) ...
48 h (1/1)	
72 h (1/1)	
96 h	
Animalia	
Annelida (Invertebrates) (1/1)	
Arthropoda (Invertebrates)	
Chordata (Vertebrates)	
Actinopterygii (Fish)	
<i>Catostomus commersoni</i>	
<i>Cyprinodon variegatus</i>	
<i>Jordanella floridae</i>	
<i>Lepomis macrochirus</i> (1/2)	
<i>Leuciscus idus</i>	
<i>Menidia beryllina</i>	
<i>Micropterus salmoides</i>	
<i>Oncorhynchus kisutch</i>	
<i>Oncorhynchus mykiss</i> (1/1)	
<i>Oryzias latipes</i> (1/1)	
<i>Pimephales promelas</i> (4/5)	
<i>Poecilia reticulata</i> (2/2)	M: 9.81 mg/L M: 7.62 mg/L
<i>Salmo gairdneri</i> (new name: oncorh...)	

Category Definition

Interpretation of the results

96h-LC50 results for *Oryzias latipes*

The screenshot shows the QSAR Toolbox interface with the 'Category Definition' workflow. The 'Endpoint' tab is active, displaying a list of endpoints and a tree view of taxonomic categories. A context menu is open over the 'Oryzias latipes' entry, with 'Sort (targets priority)' selected and 'Descending' chosen. Red callouts 1, 2, and 3 highlight the right-click, the 'Sort' option, and the 'Descending' option respectively.

1. Right click above the current fish; **2. Select** Sort (targets priority), then **3. Descending**.

Category Definition

Interpretation of the results

- Further visual analysis of the structures (see next two screen shots) could indicate that the following results are most suitable for read-across:
- *Pimephales promelas* : (E)-3,7-Dimethyl-2,6-octadienal - 96hLC50 = 7.3 mg/l
- *Poecilia reticulata* : 2-Ethyl-2-butenal - 96hLC50 = 7.6 mg/l
- Indeed those chemical are structurally most similar based on branching and functional groups in the molecule.

Category Definition

Interpretation of the results

The screenshot shows the QSAR Toolbox software interface. The main window is titled 'Category Definition' and contains several panels:

- Input Panel:** Includes 'Define', 'Define with metabolism', 'Subcategorize', 'Combine', 'Clustering', 'Delete', and 'Delete All' buttons.
- Grouping methods Panel:** Lists various methods such as 'Ionization at pH = 7,4', 'Protein binding by OASIS v1.3', and 'Toxic hazard classification by Cramer (original)'. The 'Endpoint Specific' section is expanded, showing 'Aquatic toxicity classification by ECOSAR' as the selected method.
- Structure Panel:** Displays a tree view of chemical structures. A red circle highlights 'Chemical 2', which is the most structurally similar to the target analogue. A callout box with the number '1' points to this structure.
- Endpoint Specific Panel:** Lists various toxicity endpoints, including 'Acute aquatic toxicity classification by Verhaar (Modified)', 'Acute aquatic toxicity MOA by OASIS', and 'Aquatic toxicity classification by ECOSAR'.
- Defined Categories Panel:** Shows subcategorized results, including 'Acute aquatic toxicity MOA by OASIS' and 'Acute aquatic toxicity classification by ECOSAR'.

The 'Structure' panel also includes a table of results for various species and endpoints. The table shows the following data:

Species	Endpoint 1	Endpoint 2	Endpoint 3	Endpoint 4	Endpoint 5
<i>Catostomus commersoni</i>					
<i>Cyprinodon variegatus</i>					
<i>Jordanella floridae</i>					
<i>Lepomis macrochirus</i> (1/2)					M: >20 mg/L, 2.99...
<i>Leuciscus idus</i>					
<i>Menidia beryllina</i>					
<i>Micropterus salmoides</i>					
<i>Oncorhynchus kisutch</i>					
<i>Oncorhynchus mykiss</i> (1/1)					M: 1.67(1.25;2.24) ...
<i>Onyzias latipes</i> (1/1)					
<i>Pimephales promelas</i> (4/5)		M: 7.29 mg/L	M: 5.94 mg/L, 5.9(...	M: 3.92 mg/L	M: 3.4 mg/L
<i>Poecilia reticulata</i> (2/2)					
Undefined Test organism...					
Amphibia (Amphibians, Frog)					
Mollusca (Invertebrates)					
Rotifera					

1. Chemical 2 is most structurally similar to the **target** analogue.

Category Definition

Interpretation of the results

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Category Definition' workflow is active, showing a 'Filter endpoint tree...' panel. The 'Structure' column displays chemical structures, with chemical 3 circled in red and highlighted by a callout box labeled '1'. The 'Filter endpoint tree...' panel shows a list of taxonomic groups and endpoints, including 'Chordata (Vertebrates)', 'Actinopterygii (Fish)', and 'Rotifera'. The 'Structure' column shows chemical structures, with chemical 3 circled in red and highlighted by a callout box labeled '1'. The 'Filter endpoint tree...' panel shows a list of taxonomic groups and endpoints, including 'Chordata (Vertebrates)', 'Actinopterygii (Fish)', and 'Rotifera'. The 'Structure' column shows chemical structures, with chemical 3 circled in red and highlighted by a callout box labeled '1'. The 'Filter endpoint tree...' panel shows a list of taxonomic groups and endpoints, including 'Chordata (Vertebrates)', 'Actinopterygii (Fish)', and 'Rotifera'. The 'Structure' column shows chemical structures, with chemical 3 circled in red and highlighted by a callout box labeled '1'.

1. Chemical 3 is most structurally similar to the **target** analogue.

Category Definition

Recap

- You have searched for suitable analogues having the same profile than the target compound by successive subcategorisation with 3 profilers.
- You have chosen the most suitable candidates to be used for read-across based on a visual analysis of their molecular structure.

Report

- Remember the report module (not reviewed in this exercise) allows you to generate a report on the predictions performed with the Toolbox. This 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.
- The obtained prediction could be saved as a file and loaded later on in the system (see Tutorial 5).

Congratulation

- You have used some more functions of the Toolbox and changed up the workflow to address new issues.
- By now you should feel comfortable moving the curser around the basic screens for each one the modules.
- Continue to using the Toolbox and you speed and confidence will increase sharply.