

OECD QSAR Toolbox v.4.1

Tutorial of how to use Standardized workflow for
ecotoxicological prediction

Aim

This is a step-by-step presentation designed to take the user of Toolbox through the Standardized workflow for ecotoxicity prediction.

Outlook

- Aim
- Ecotoxicological prediction
- Workflow of the exercise
- Report

Outlook

- Aim
- **Ecotoxicological prediction**
- Workflow of the prediction
- Report

Outlook

Ecotoxicological prediction:

- Acute aquatic toxicity
- Fish, LC50 (EC50), 96h, mortality
- or*
- Invertebrates, EC50(LC50), 48h, mortality, immobilization, intoxication
- or*
- Algae, LC50 (EC50), 72-96h, population or growth

Outlook

- Aims
- Ecotoxicological prediction
- **Workflow of the prediction**
- Report

Workflow: Steps

- Input of target chemical(s)
- Data gap filling: Standardized workflow
- Report

Outlook

- Aims
- Ecotoxicological prediction
- **Workflow of the prediction**
 - **Chemical input**
 - Standardized workflow in Data-gap filling section
- Report

Chemical Input

- 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

I. Single target chemical:

- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation
- Chemical with defined composition
- Drawing chemical structure
- Select from User List/Inventory/Databases

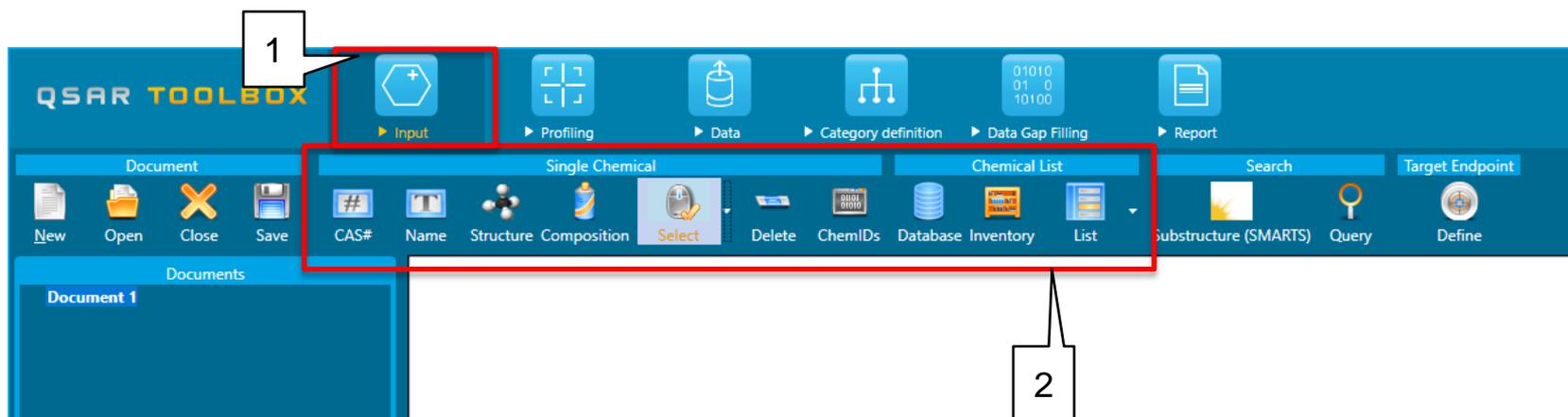
II. Group of chemicals:

- User's List
- Inventory/Database

Chemical Input: Single chemical

- Open the Toolbox.
- Click on “Input” (see next screen shot).

Chemical Input Single chemical



1. Click on Input (1) to display the main Input section (2).

Chemical Input

Single chemical: CAS RN

The screenshot shows the QSAR Toolbox software interface. The 'Single Chemical' menu is open, and the 'Search by CAS #' dialog box is displayed. The dialog box has a text input field containing '111864', a 'Search' button, and 'OK' and 'Cancel' buttons. A table of search results is shown, with the following data:

Property	Value
CAS	11186-4
SMILES	CCCCCCCN
CS Relation	High
Substance	Mono constituent
Composition	
Name	1-aminooctane 1-Octanamine 1-octylamine

On the right side of the dialog box, the chemical structure of 1-octanamine is displayed: CCCCCCCCN.

1. Press CAS# (1); 2. Type in the CAS # (2) ; 3. Click on Search (3);
4. Press OK (5).

Outlook

- Aims
- Ecotoxicological prediction
- **Workflow of the prediction**
 - Chemical input
 - **Standardized workflow in Data gap filling section**
- Report

Data gap filling

An overview

- “Data Gap Filling” module gives 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.

Data gap filling

Standardized workflow: An overview

- Algorithms for standardized data gap filling (SWs) have been developed for skin sensitization (LLNA and GPMT data) and acute aquatic toxicity to fish, invertebrates and algae. Once started, the standardized workflows (SWs) follow an implemented logic.
- The main differences compared to the automated workflows (AWs) are that the domain of application is expanded in the SWs (including other species, durations, etc.) and SWs allow interactions by the user and thus, different selection than those in AW could be done.

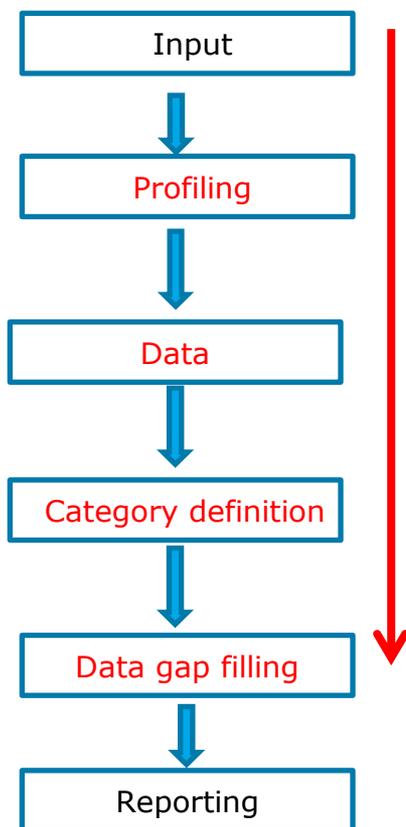
Data gap filling

Standardized workflow: An overview

- The Standardized workflow can be used in a single chemical mode or in a batch mode.
- In this tutorial only single chemical mode for acute aquatic toxicity to fish is discussed.

Data Gap Filling Overview

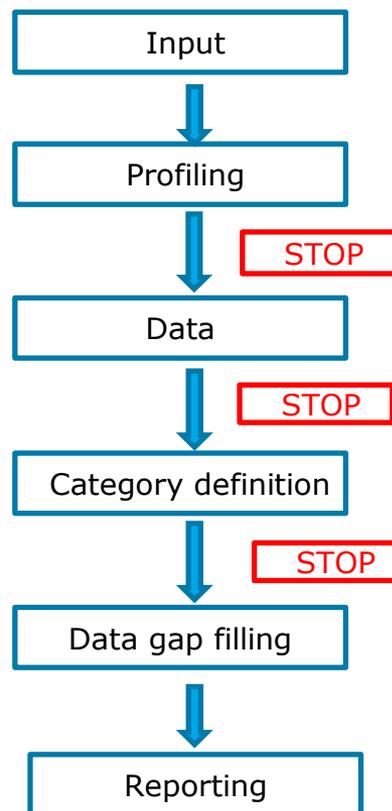
Automated workflow (AW)



Same components as defined in the AW are used in the SW

The SW pauses at each of the stages and user is able to make different selection than those implemented in the AW

Standardized workflow (SW)



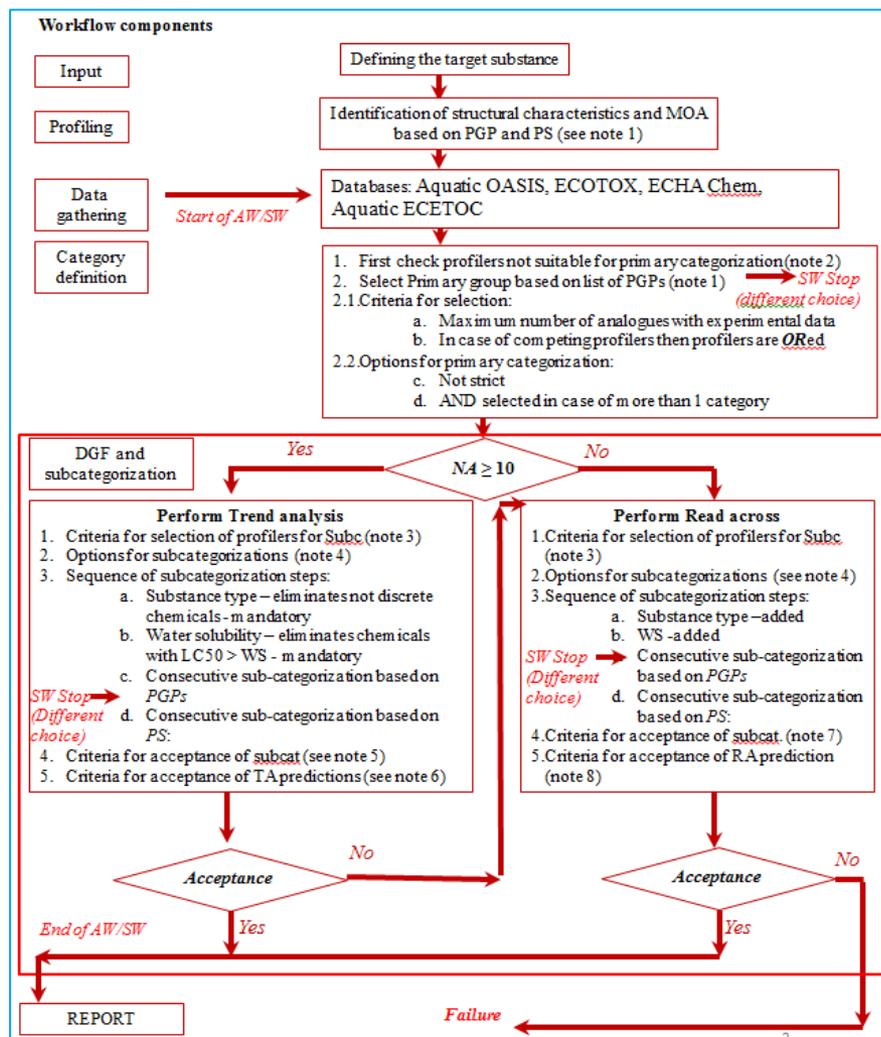
Databases with data for the target endpoint are listed and user select to use all of them or make specific selection

Relevant to the workflow profilers appropriate for DGF are listed and ordered hierarchically based on the population of the group and user is able to select any of them

Additional data filtering could be applied (e.g. different species selection)

Data Gap Filling

Algorithm of Ecotoxicological workflow



Data Gap Filling

Standardized workflow: single chemical

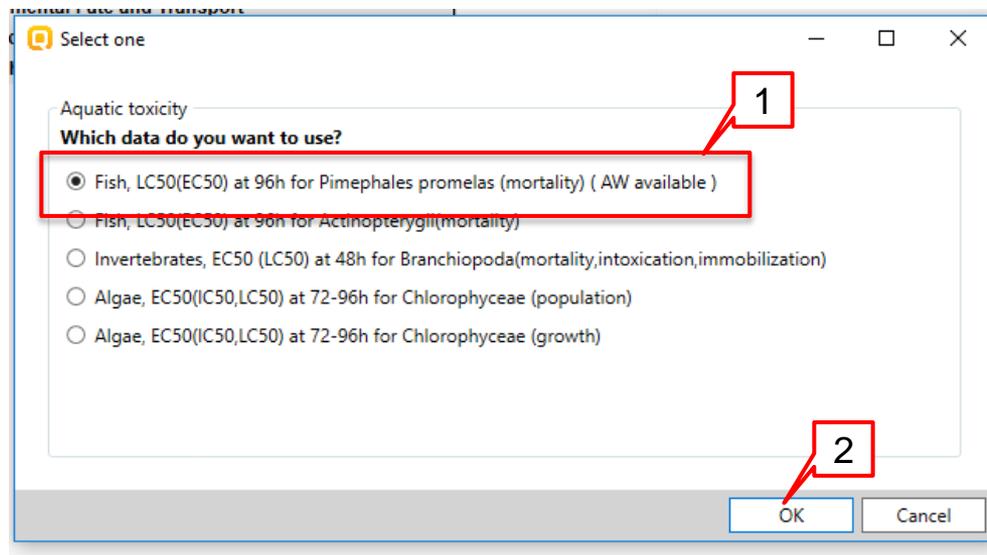
The screenshot displays the QSAR Toolbox interface. The top menu bar contains 'Data Gap Filling' (1), 'Standardized' (2), and 'OK' (5) buttons. A dialog box titled 'Select workflow' is open, showing 'Ecotoxicological Endpoint' (3) selected. The 'Filter endpoint tree...' panel on the right shows a list of categories including 'Structure info', 'Parameters', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', and 'Human health hazards'. The 'Documents' panel shows 'Document 1' with CAS: 120832. The 'Data Gap Filling Settings' panel is visible at the bottom.

1. Select *Data gap filling* tab (1)
2. Press *Standardized* (2)
3. Select *Ecotoxicological endpoint* (3) from the pop-up window (4)
4. Click on *OK* (5)

Data Gap Filling

Standardized workflow: single chemical

- A dialogue window gives the user a choice to select the end-point (1)
- *AW available in brackets* means that the endpoint is used in the automated workflow
- Click on OK (2)

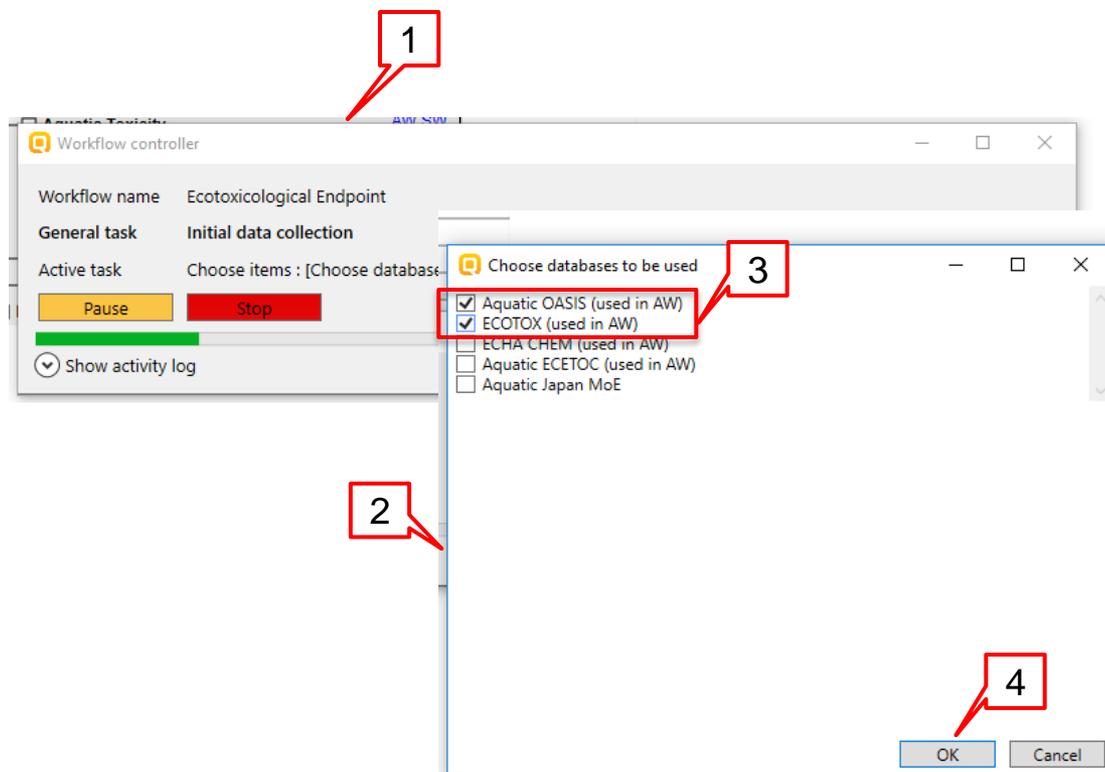


Data Gap Filling

Standardized workflow: single chemical

A workflow controller window (1) is displayed for a second and then a dialogue window for selection (2) of databases, which are recommended for use in the SW is displayed

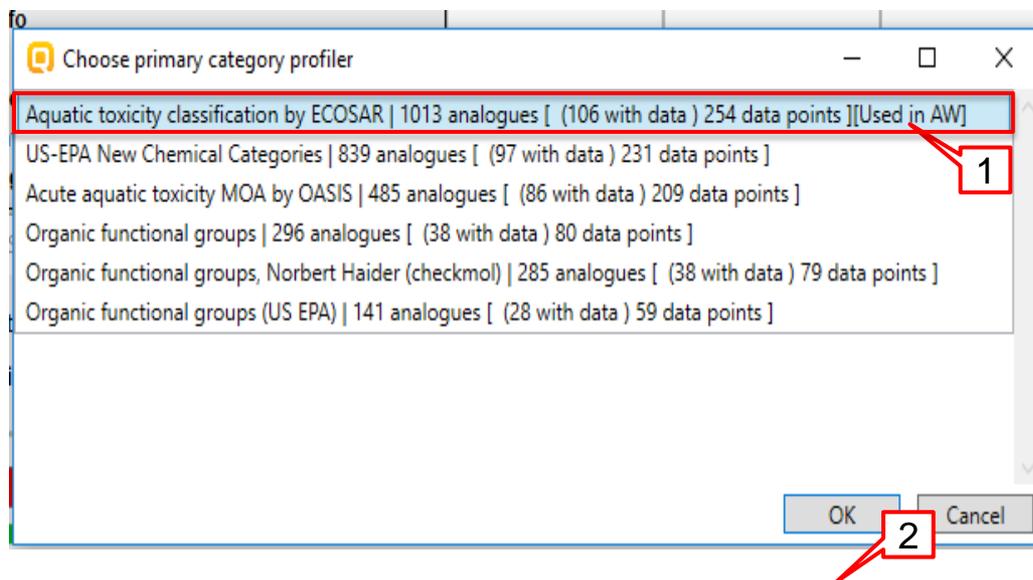
- Indication about which databases are used in the AWs is also included in brackets
- Select the databases (3)
- Click on OK(4)



Data Gap Filling

Standardized workflow: single chemical

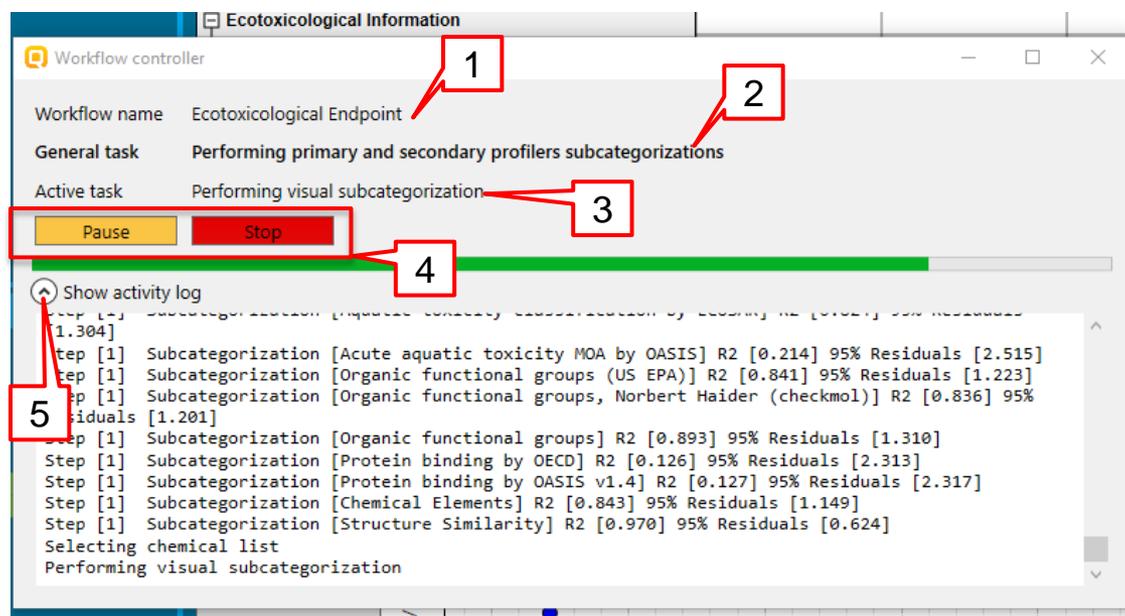
- A list with all primary category profilers is displayed, where you have to select the most suitable one
- Here, select Aquatic toxicity classification by ECOSAR (1) and then click on OK (2)



Data Gap Filling

Standardized workflow: single chemical

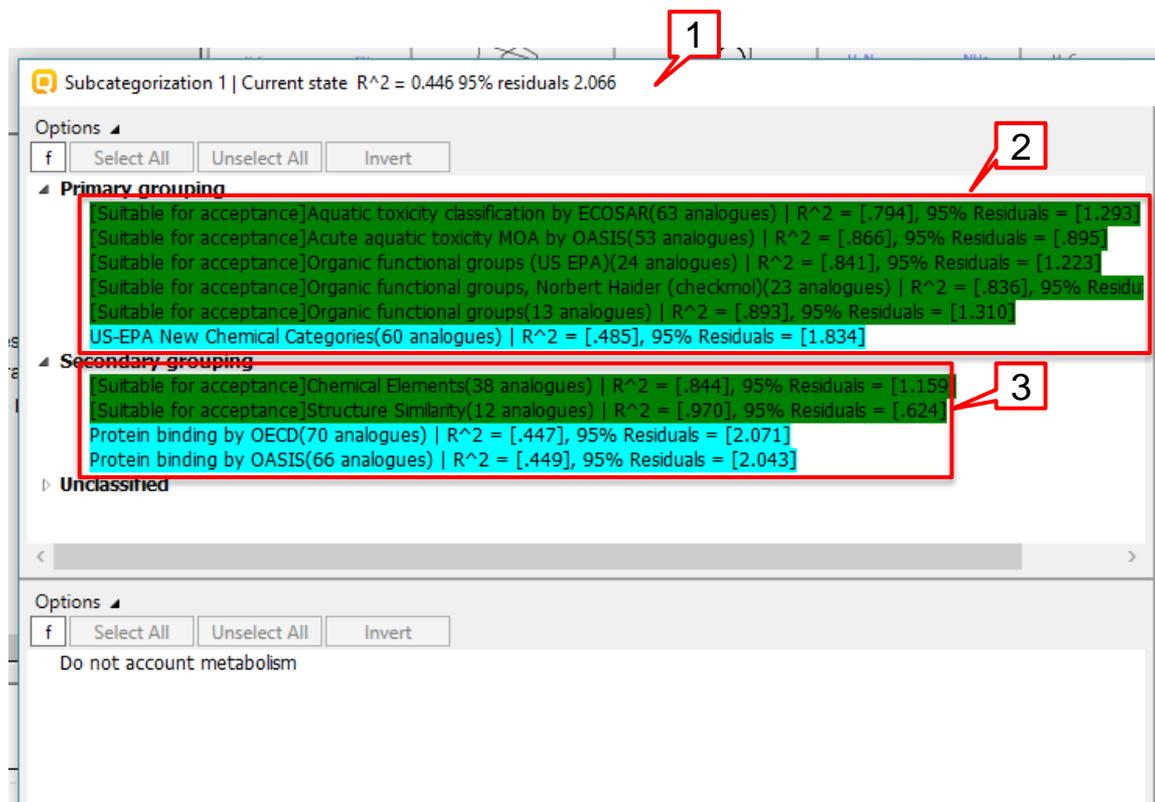
- A workflow controller window is displayed throughout the standardized workflow procedure. It includes:
- Workflow name (1)
 - General task (2)
 - Active task (this is subtask of the general task, which is currently being performed)(3)
 - Navigation options (4)
 - Activity log (5)



Data Gap Filling

Standardized workflow: single chemical

The subcategorisation window (1) is separated into two sections: Primary grouping (2) and secondary grouping (3)



Data Gap Filling

Standardized workflow: single chemical

- Profilers are highlighted in different color (1)
- Click on *Options - Legend* (2) to display the colors' meaning (3)
- "Priority" factor along with criteria for acceptance of the prediction (4) as compared with the statistics of *the current state* (5)
- An indication for *suitability* (6) of the profiler for acceptance of the prediction is included

Subcategorization 1 Current state $R^2 = 0.446$ 95% residuals 2.066

Options

Group by: Priority

Sort by: Relevancy

Color by: Outcome Legend

Filter by: None

Secondary grouping

US-EPA New Chemical Categories(60 analogues) | $R^2 = [.485]$, 95% Residuals = [1.834]

Suitable for acceptance Chemical Elements(38 analogues) | $R^2 = [.844]$, 95% Residuals = [1.159]

Suitable for acceptance Structure Similarity(12 analogues) | $R^2 = [.970]$, 95% Residuals = [.624]

Protein binding by OECD(70 analogues) | $R^2 = [.447]$, 95% Residuals = [2.071]

Protein binding by OASIS(66 analogues) | $R^2 = [.449]$, 95% Residuals = [2.043]

Unclassified

Legend

Outcome

- Converges
- Improves statistic
- Does not improve statistic
- Prediction not suitable
- Applied or Doesn't modify state

Options

Select All Unselect All Invert

Do not account metabolism

Data Gap Filling

Standardized workflow: single chemical

- Clicking on the first profiler (1) shows all categories found in the analogues (2) as the ones colored in blue (3) are the ones that are not applicable to the target chemical and hence can be removed (4)

Subcategorization 1 | Current state: R² = 0.446 95% residuals 2.066

Options ▾
 f Select All Unselect All Invert About Options

▲ Primary grouping

- [Suitable for acceptance]Aquatic toxicity classification by ECOSAR(63 analogues) | R² = [.794], 95% Residuals = [1.293]
- [Suitable for acceptance]Acute aquatic toxicity MOA by OASIS(53 analogues) | R² = [.866], 95% Residuals = [.895]
- [Suitable for acceptance]Organic functional groups (US EPA)(24 analogues) | R² = [.841], 95% Residuals = [1.223]
- [Suitable for acceptance]Organic functional groups, Norbert Haider (checkno)(23 analogues) | R² = [.836], 95% Residuals = [1.310]
- [Suitable for acceptance]Organic functional groups (13 analogues) | R² = [.893], 95% Residuals = [1.310]
- US-EPA New Chemical Categories(60 analogues) | R² = [.485], 95% Residuals = [1.834]

▲ Secondary grouping

- [Suitable for acceptance]Chemical Elements(38 analogues) | R² = [.844], 95% Residuals = [1.159]
- [Suitable for acceptance]Structure Similarity(12 analogues) | R² = [.970], 95% Residuals = [1.624]
- Protein binding by OECD(70 analogues) | R² = [.447], 95% Residuals = [2.071]
- Protein binding by OASIS(66 analogues) | R² = [.449], 95% Residuals = [2.043]

▷ Unclassified

Options ▾
 f Select All Unselect All Invert
 Do not account metabolism

Adjust options

Aliphatic Amines

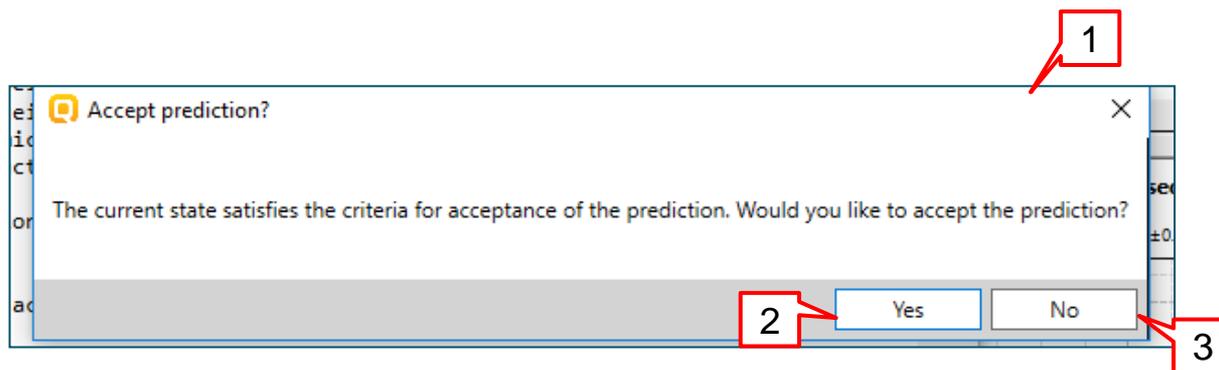
Differ from target by
 At least one category
 All categories [STOP]

(4) Acid moiety
 (71) Aliphatic Amines
 (1) Neonicotinoids
 (1) Phenol Amines
 (1) Phenols
 (1) Triazines, Aliphatic
 (1) Triazines, Aromatic

Selected 8 (63/71)
 Select different
 Remove selected

Data Gap Filling

Standardized workflow: single chemical



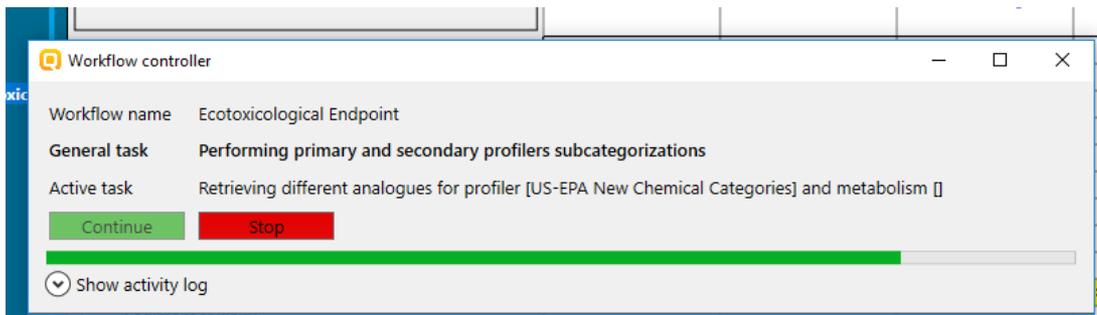
- An Accept prediction? window is visualized (1)
- Press Yes (2) , if you want to accept the prediction (next see slide 30)
- Press No (3) and continue with the workflow if you are not satisfy with the outcome (see slide 29)

Data Gap Filling

Standardized workflow: single chemical

- Press **No (3)** and continue with the workflow if you are not satisfy with the outcome.

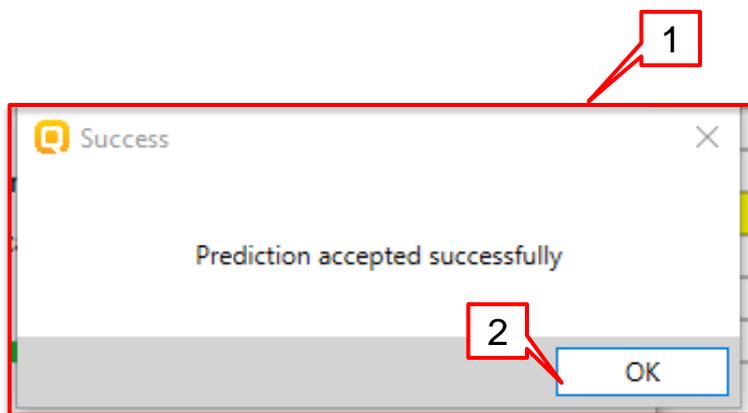
- Press Continue button (1) in the workflow controller
- Then repeat the steps described in slides 25-27



Data Gap Filling

Standardized workflow: single chemical

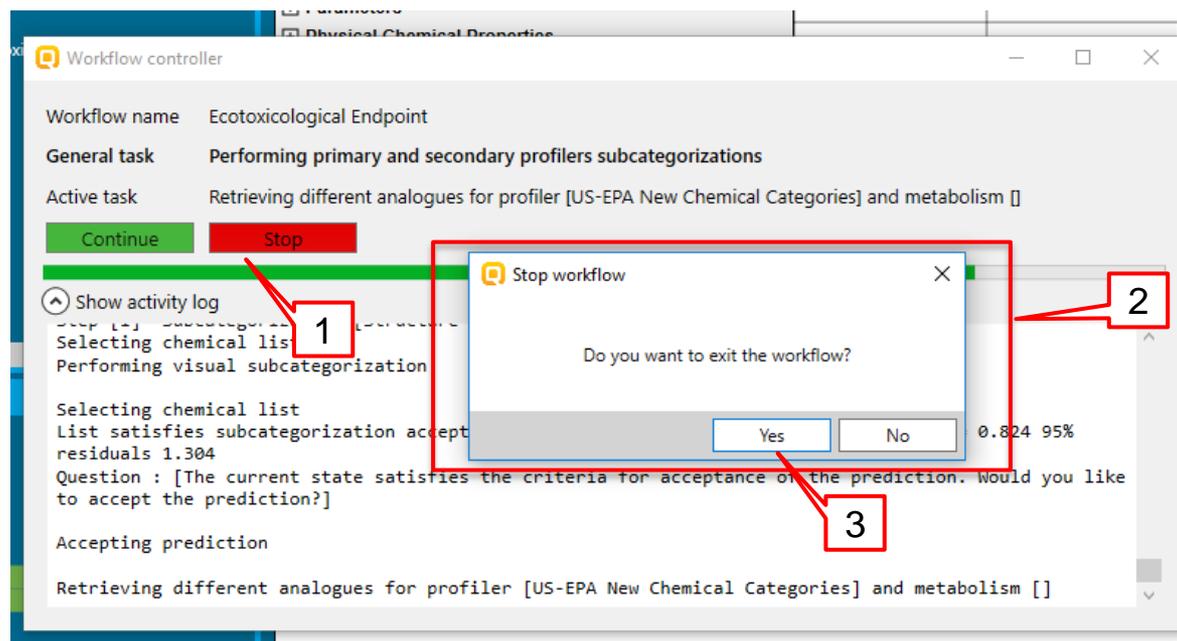
- A pop-up window (1) is displayed informing that the prediction is accepted
- Press OK (2)



Data Gap Filling

Standardized workflow: single chemical

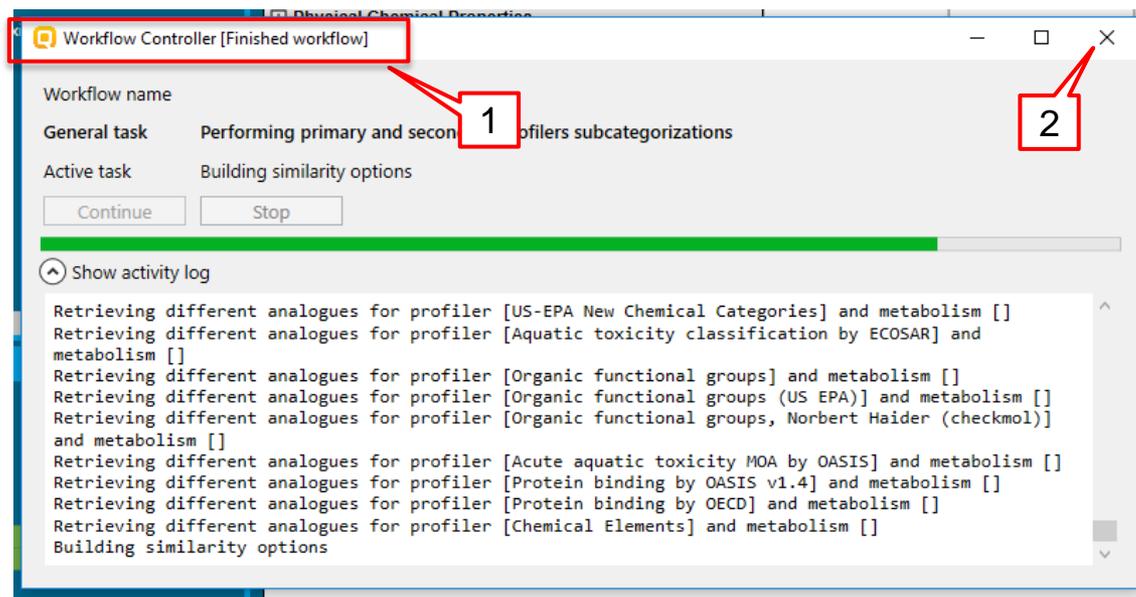
- Press Stop button (1) on the Workflow controller.
- A pop-up window (2) asks to confirm the exit of the workflow
- Press Yes button(3)



Data Gap Filling

Standardized workflow: single chemical

- A message is displayed that the workflow has finished (1)
- Press **X** button (2)



Data Gap Filling

Standardized workflow: single chemical

The screenshot displays the QSAR Toolbox interface for a single chemical workflow. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, a sub-menu offers Trend analysis, Read across, (Q)SAR, and Standardized Automated options. The main workspace is divided into three panels:

- Documents:** A tree view showing the workflow structure. The 'US-EPA New Chemical Categories' node is selected and highlighted with a red box and a '2' in a white box.
- Filter endpoint tree...:** A hierarchical tree of endpoints. The 'Aquatic Toxicity' node is expanded to show 'Mortality' and 'EC50 <OR> LC50'. The 'EC50 <OR> LC50' node is further expanded to show 'Animalia (animals)', 'Chordata (chordates)', 'Actinopterygii (ray-finned fishes, sp ...)', and 'Pimephales promelas'. The 'Actinopterygii' node is highlighted with a red box and a '1' in a white box.
- Data Matrix:** A table with columns for '1 [target]', '2', and '3'. The first row contains chemical structures. The second row contains numerical data for '96 h' exposure, with values for 'N', 'M', and 'T' for each column. The 'T' values are highlighted in yellow.

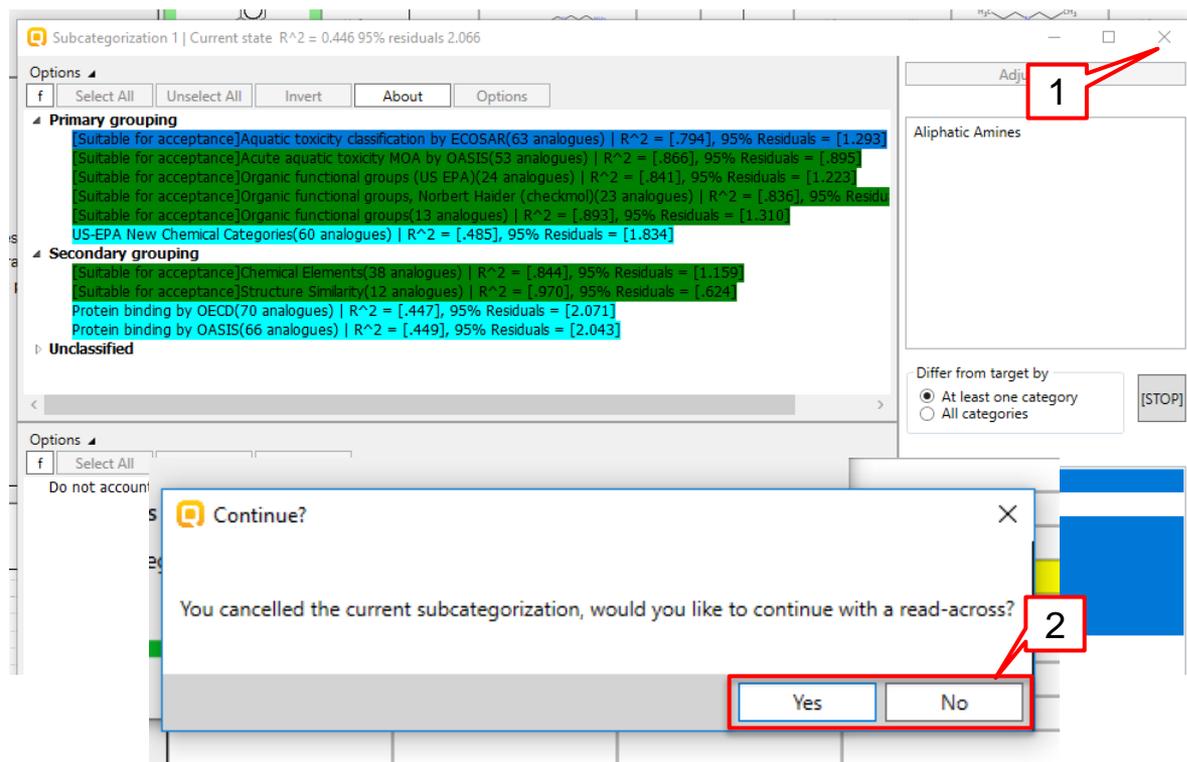
- The result is displayed on the data matrix (1) marked with "T" ("M" stands for measured data)
- The workflow finishes on the document level of the primary grouping(2)

Data Gap Filling

Standardized workflow: single chemical

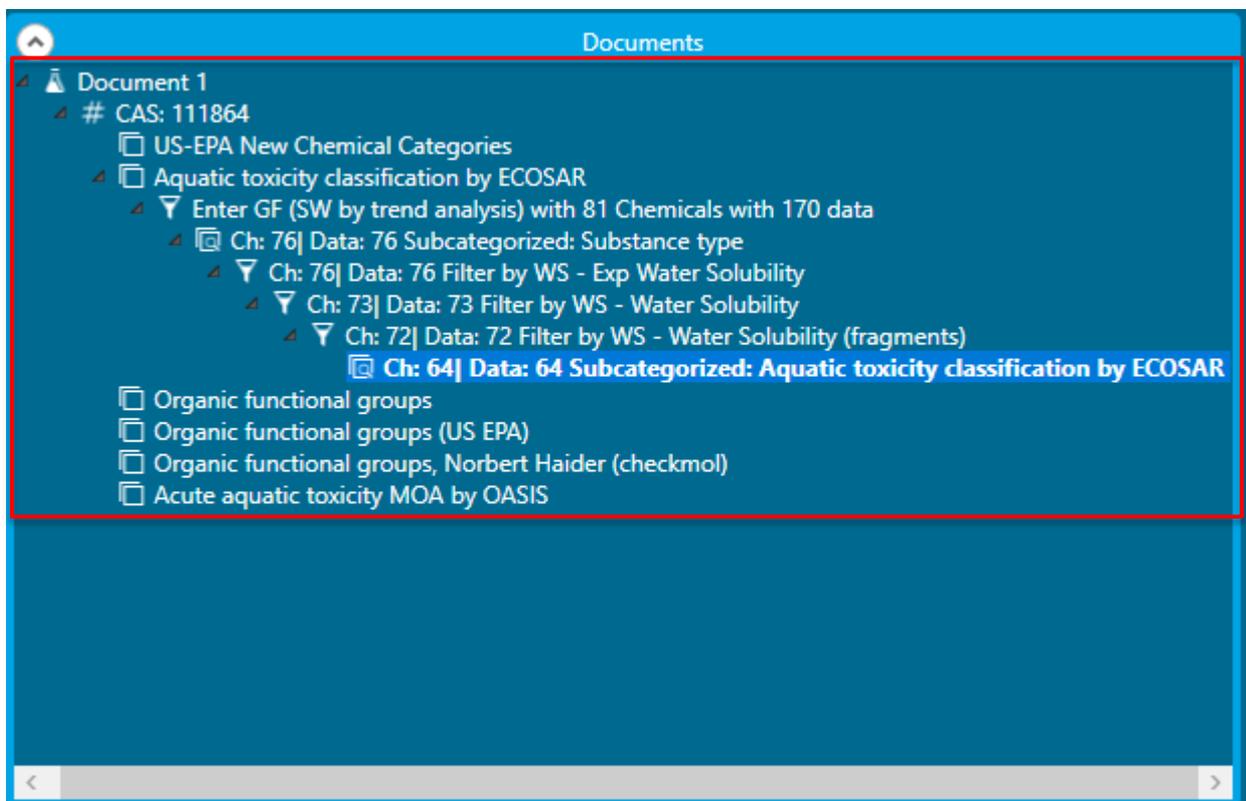
If the subcategorisation window is closed by pressing **X** button (1) without performing any subcategorization, a dialogue window is displayed (2):

- Press Yes if you want to exit the trend analysis and continue with read-across analysis.
- Press No if you want to finish the workflow.



Data Gap Filling

Standardized workflow: single chemical



The steps executed in the SW are listed in the Documents panel.

Outlook

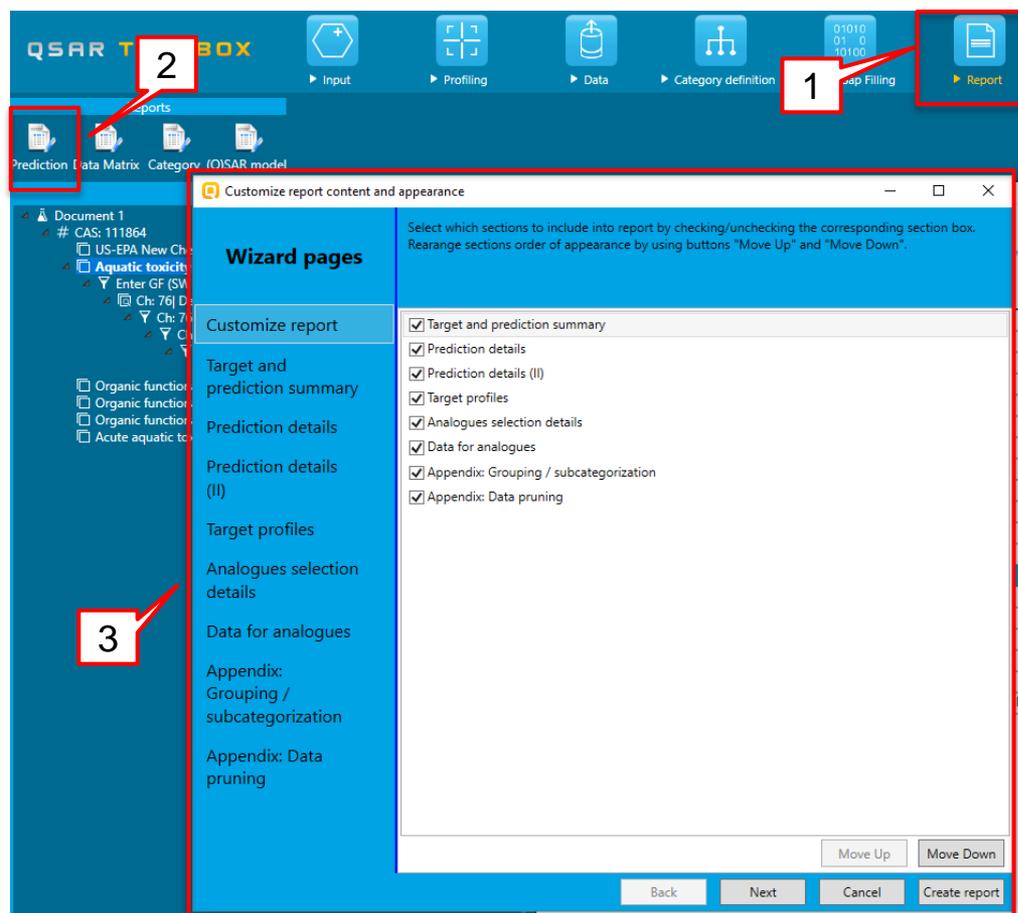
- Aims
- Ecotoxicological prediction
- Workflow of the prediction
 - Chemical input
 - Standardized workflow in Data gap filling section
- **Report**

Report Overview

- Report module could generate report of any of predictions performed with the Toolbox.
- Report module contains Wizard pages which navigate you through predefined and user-editable report templates.

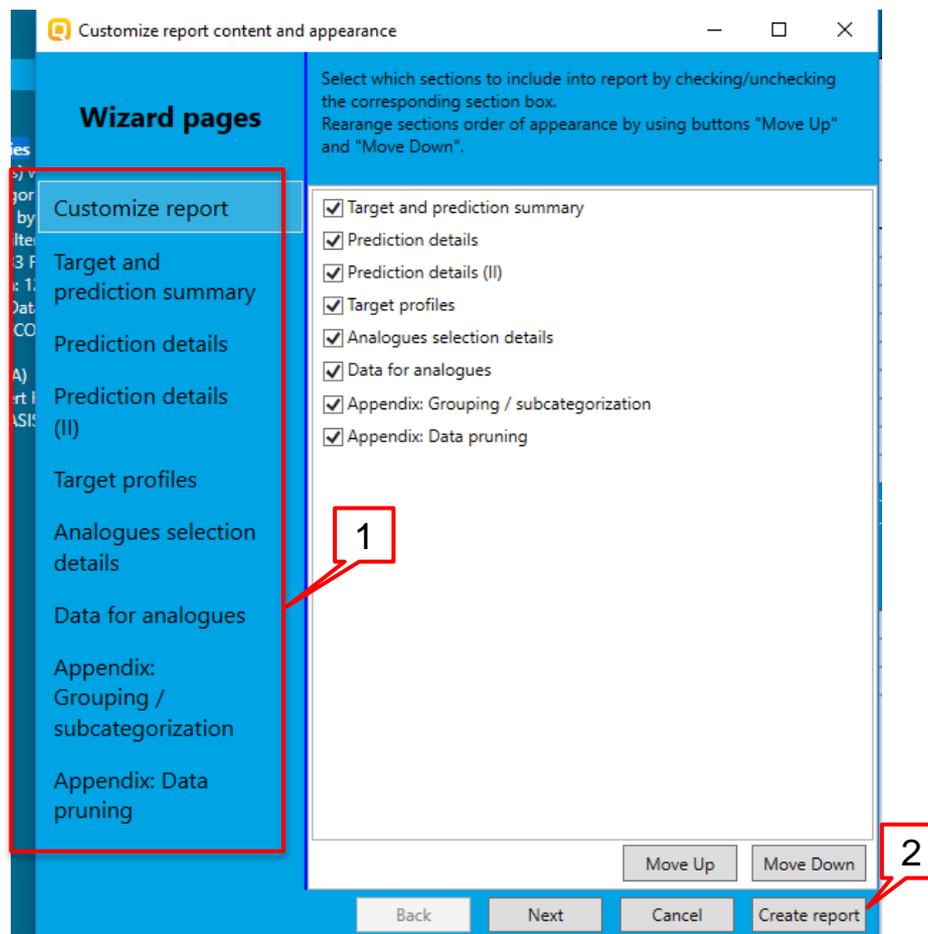
Report Generation report

- Go to Report module (1)
- Select Prediction (2)
- A Wizard pages window is displayed (3)



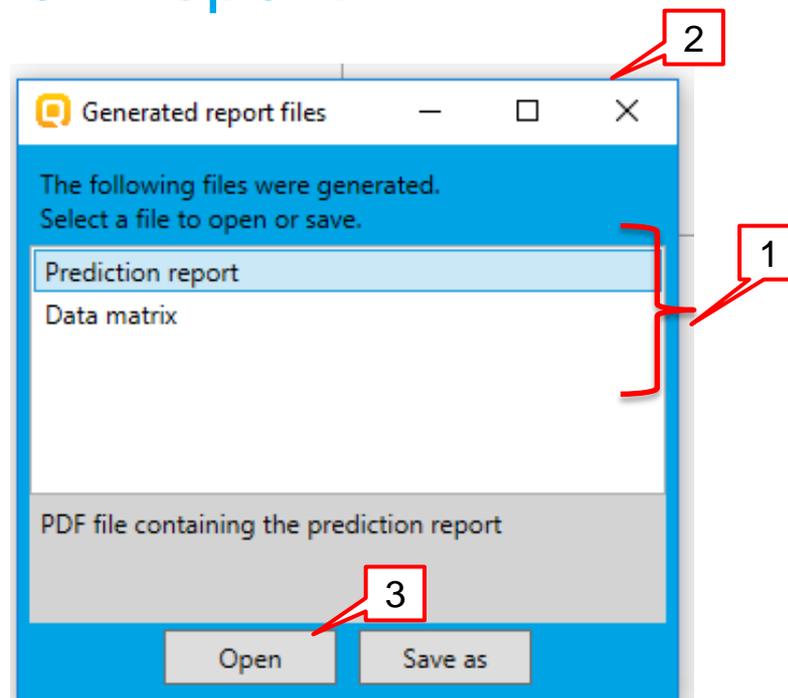
Report Generation report

- Select different levels to customize the information that is going to be shown in the report (1);
- Select Create report (2) to display the report.



Report Generation report

- Two files (1) are generated, which can be selected from the Generated report files window (2) by clicking Open (3)
- Select Create report (2) to display the report



Report Overview

- The prediction report (1) is a PDF file;
- The execution of SW "Ecotoxicological Endpoint" is included in the Prediction summary (2)

Prediction of LC50 for octylamine

1 / 49

QSAR Toolbox prediction for single chemical

Date: 10 Jul 2017
 Author(s):
 Contact details:



Target information		
Structural information	Numerical identifiers	Chemical names
SMILES: CCCCCCCCN	EC#: N/A CAS#: 111-86-4 Other: N/A	1-aminooctane 1-Octanamine 1-octylamine
Structure		
		

Prediction summary
Predicted endpoint: LC50; Mortality; Pimephales promelas; 96h; No guideline specified
Predicted value: 9.77 (from 0.567 to 169)
Unit/scale: mg/L
Data gap filling method: Trend analysis, executed via SW "Ecotoxicological Endpoint"
Summary: manually editable field Not provided by the user



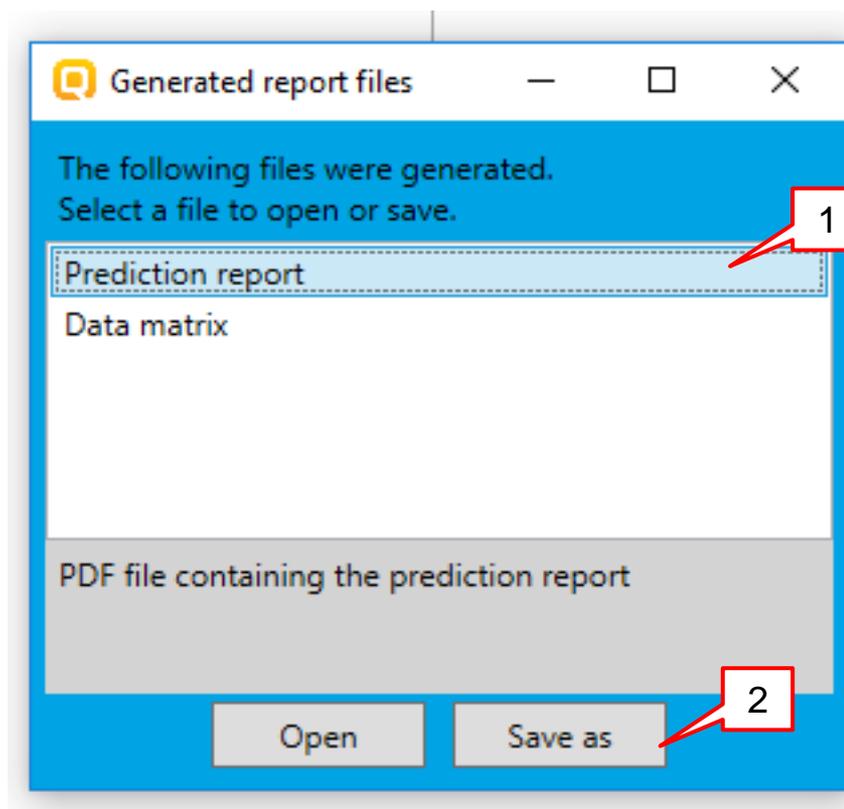
Report Overview

	Target chemical	Analogue #1	Analogue #2	Analogue #3
Substance Identity				
Structure	<chem>CCCCCCCCN</chem>	<chem>NC12CC3CC(C(C3)C1)C2</chem>	<chem>Nc1ccccc1</chem>	<chem>NCCCN</chem>
CAS number	111-86-4	768-94-5	100-46-9	109-76-2
Chemical name	octylamine	Amantadine	Benzylamine	1,3-Diaminopropane
Other identifier				
SMILES	CCCCCCCN	NC12CC3CC(C(C3)C1)C2	Nc1ccccc1	NCCCN
Parameters				
Profilers				
Profiles used for grouping/subcategorization				
Aquatic toxicity classification by ECOSAR	Aliphatic Amines Discrete chemical;	Aliphatic Amines Discrete chemical;	Aliphatic Amines Discrete chemical;	Aliphatic Amines Discrete chemical;
Substance type (subcategorization)	Mono constituent (predefined); Organic	Mono constituent (predefined); Organic	Mono constituent (predefined); Organic	Mono constituent (predefined); Organic
Aquatic toxicity classification by ECOSAR	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines
Measured and predicted data				
Data used for prediction				

The data matrix (1) is an *Excel* file, which contains information about the analogues.

Report Save

To save any of the two files, select the file (1) and then click on Save as (2)



Report Save

The report is saved as a pdf file (1) while the data matrix is saved as an .xlsx file (2)

