QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD (Q)SAR Toolbox v.4.4.1

Tutorial of how to predict ecotoxicological endpoint of chemicals by standardized workflow

Outlook

- **Background**
- Keywords
- Objectives
- Specific Aims
- Standardized workflow for Ecotoxicity
- The exercise
- Standardized workflow execution

Background

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

Outlook

- Background
- **Keywords**
- Objectives
- Specific Aims
- Standardized workflow for Ecotoxicity
- The exercise
- Standardized workflow execution

Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options (e.g. Profiling)

WORKFLOW – the use, in combination, of the different modules (e.g. prediction workflow: from input to report)

PROFILER - algorithm (rule set) for the identification of specific features of the chemicals. Several types of profilers are available, such as structural (e.g. Organic functional groups), mechanistic (e.g. Protein binding by OECD) and endpoint-specific (e.g. in vitro in vitro mutagenicity (Ames test) alerts by ISS) profilers.

ALERT - the profilers consist of sets of rules or alerts. Each of the rules consists of a set of queries. The queries could be related to the chemical structure, physicochemical properties, experimental data, comparison with the target or list with substances and external queries from other predefined profilers (reference queries).

CATEGORY – "group" of substances sharing same characteristics (e.g. the same functional groups or mode of action). In a typical Toolbox workflow, it consists of the target chemical and its analogues gathered according to the selected profilers

ENDPOINT TREE – Endpoints are structured in a branched scheme, from a broader level (Phys-Chem properties, Environmental Fate and transport, Ecotoxicology, Human health hazard) to a more detailed one (e.g. EC3 in LLNA test under Human health hazard-Skin sensitization)

DATA MATRIX – Table reporting the chemical(s) and data (experimental results, profilers outcomes, predictions). Each chemical is in a different column and each data in a different row

Outlook

- Background
- Keywords
- **Objectives**
- Specific Aims
- Standardized workflow for Ecotoxicity
- The exercise
- Standardized workflow execution

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Identify analogues for a target chemical;
- Retrieve experimental results available for those analogues;
- Color the profiling schemes according to their suitability for subcategorization;
- Fill data gaps by standardized workflow;

Outlook

- Background
- Keywords
- Objectives
- **Specific Aims**
- Standardized workflow for Ecotoxicity
- The exercise
- Standardized workflow execution

Specific Aims

- To introduce to the Toolbox user to the standard workflow for predicting of ecotox endpoint (LC50);
- To familiarize the user with the new Toolbox interface;
- To explain to the user the rationale behind each step of the exercise.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- **Standardized workflow for Ecotoxicity**
- The exercise
- Standardized workflow execution

Standardized workflow for Ecotoxicity endpoints
Overview

Standardized workflow for Ecotox covers the following acute aquatic toxicity endpoints:

- *Fish*, LC50 (EC50), 96h, mortality

  or

- *Invertebrates*, EC50(LC50), 48h, mortality, immobilization, intoxication

  or

- *Algae*, LC50 (EC50), 72-96h, population or growth

Standardized workflow for Ecotox endpoints
Overview

- The standardized workflow (SW) is designed to apply data gap filling for discrete chemicals only
- The SWs has been developed to be applicable for the same endpoints used for application of the AWs (i.e. LC50, Mortality, 96h, *Pimephales promelas*).
- Once started, the SW follows the implemented logic under the user control.
- As opposite to the automated workflow (AW), the domain of application is expanded in the SWs (including other endpoints, effects, species, durations, etc.) and SWs allow interactions by the user.
- In case more than one further application is possible, the workflow stops and waits for the decision of the user.
- SW can be executed for a single chemical only.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Standardized workflow for Ecotoxicity
- **The exercise**
- Standardized workflow execution

The Exercise

- In this exercise we will predict the aqute aquatic toxiity endpoint (LC50) of N-Octylamine [CAS# 111-47-8], which will be the “target” chemical.
- This prediction will be accomplished by using of the developed standardized workflow for ecotoxicity (LC50, Mortality, 96h, *P.promelas*).

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Standardized workflow for Ecotoxicity
- The exercise
- **Standardized workflow execution**

Standardized workflow execution

- Only three of the general Toolbox modules are used in a sequential workflow:
  - Input
  - Data Gap Filling
  - Report

The rest of the modules – *Profiling*, *Data* and *Category definition* are included as a part of the algorithm of the standardized workflow. The workflow stops at them and waits for the decision of the user.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Standardized workflow for Ecotoxicity
- The exercise
- **Standardized workflow execution**
  - 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

I. Single target chemical:

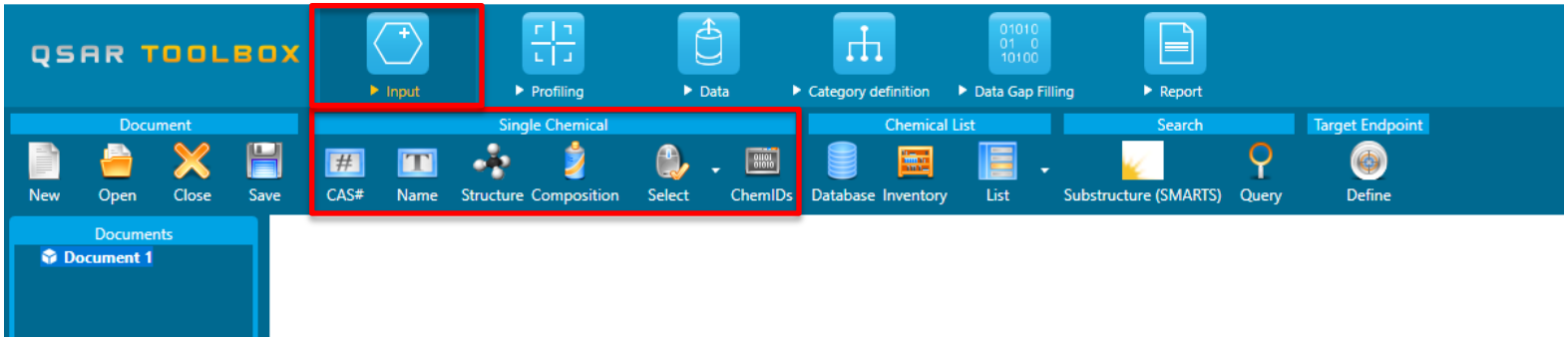
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation
- Chemical with defined composition
- Drawing chemical structure
- Select from User List/Inventory/Databases

II. Group of chemicals:

- User's List
- Inventory/Database

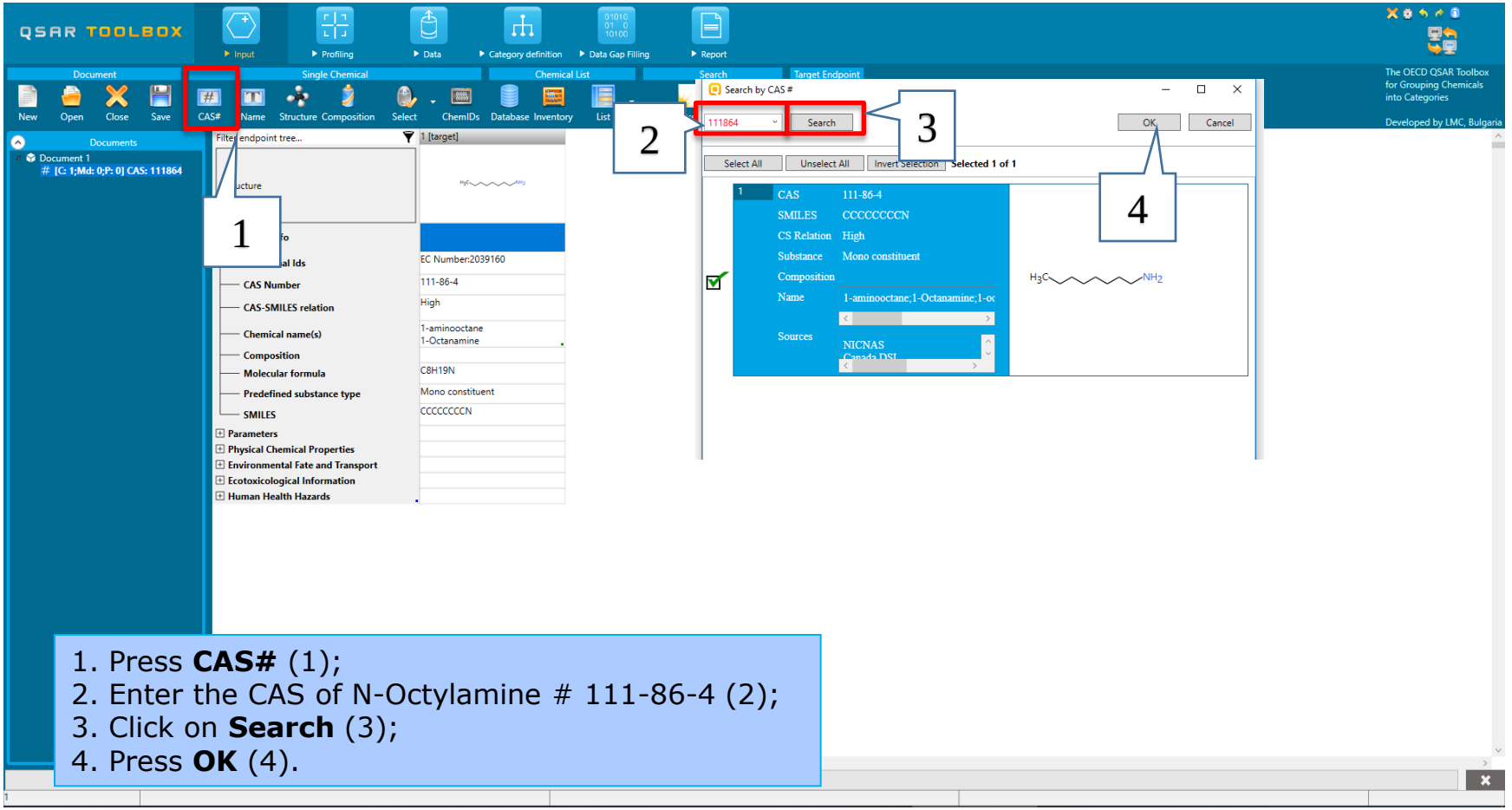
Chemical Input
Single chemical

As mentioned before SW is allowed for single chemical only. In this respect below are provided different ways for entering a single chemical



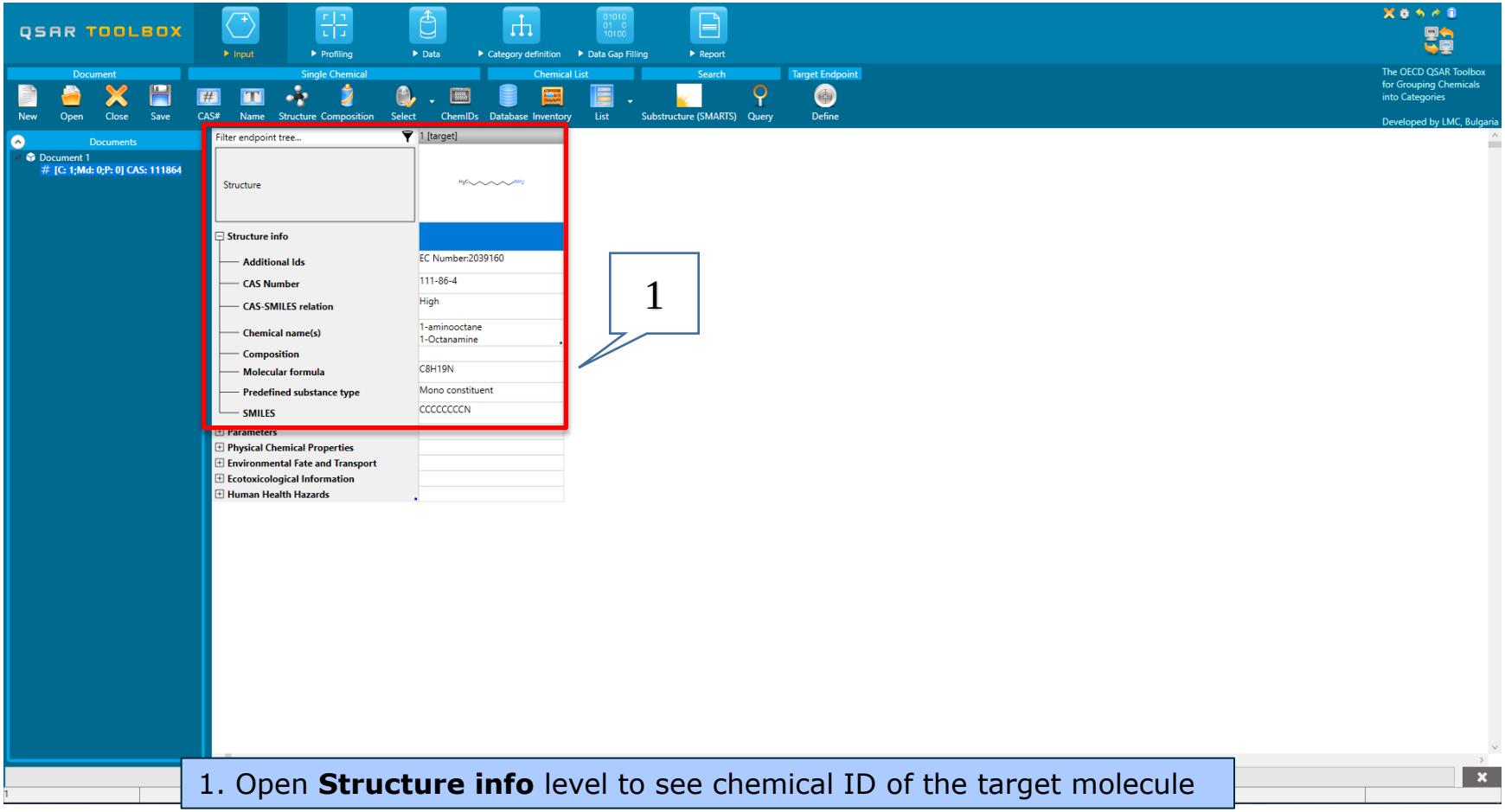
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- Drawing chemical structure
- Select from User List/Inventory/Databases

Chemical Input
Single chemical: CAS RN



1. Press **CAS#** (1);
2. Enter the CAS of N-Octylamine # 111-86-4 (2);
3. Click on **Search** (3);
4. Press **OK** (4).

Chemical Input
Target chemical identity



1. Open **Structure info** level to see chemical ID of the target molecule

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Standardized workflow for Ecotoxicity
- The exercise
- **Standardized workflow execution**
  - Input
  - **Data Gap Filling**

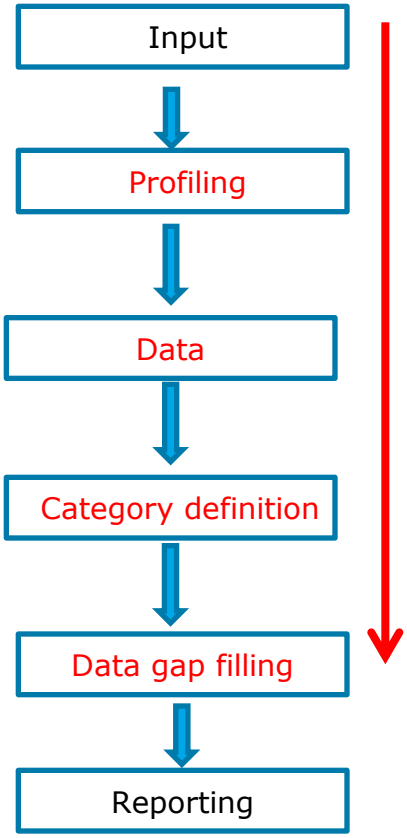
Data gap filling
An overview

- "Data Gap Filling" module gives access to three different data gap filling tools:
  - Read-across
  - Trend analysis
  - (Q)SAR models
  - Automated workflows
  - Standardized workflows

- 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
An overview

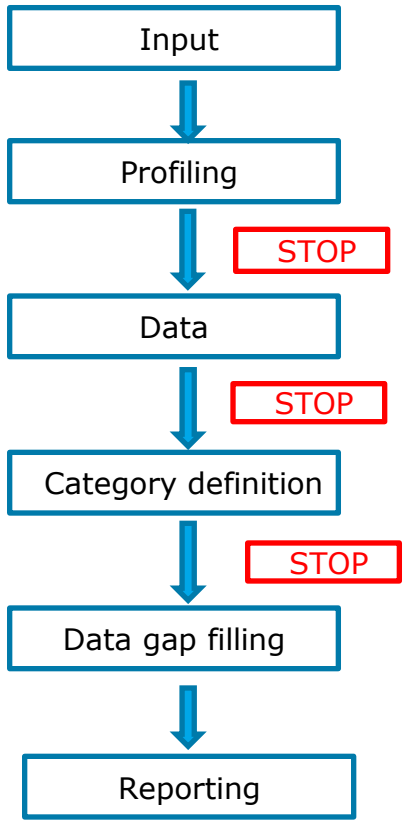
Automated workflow (AW)



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

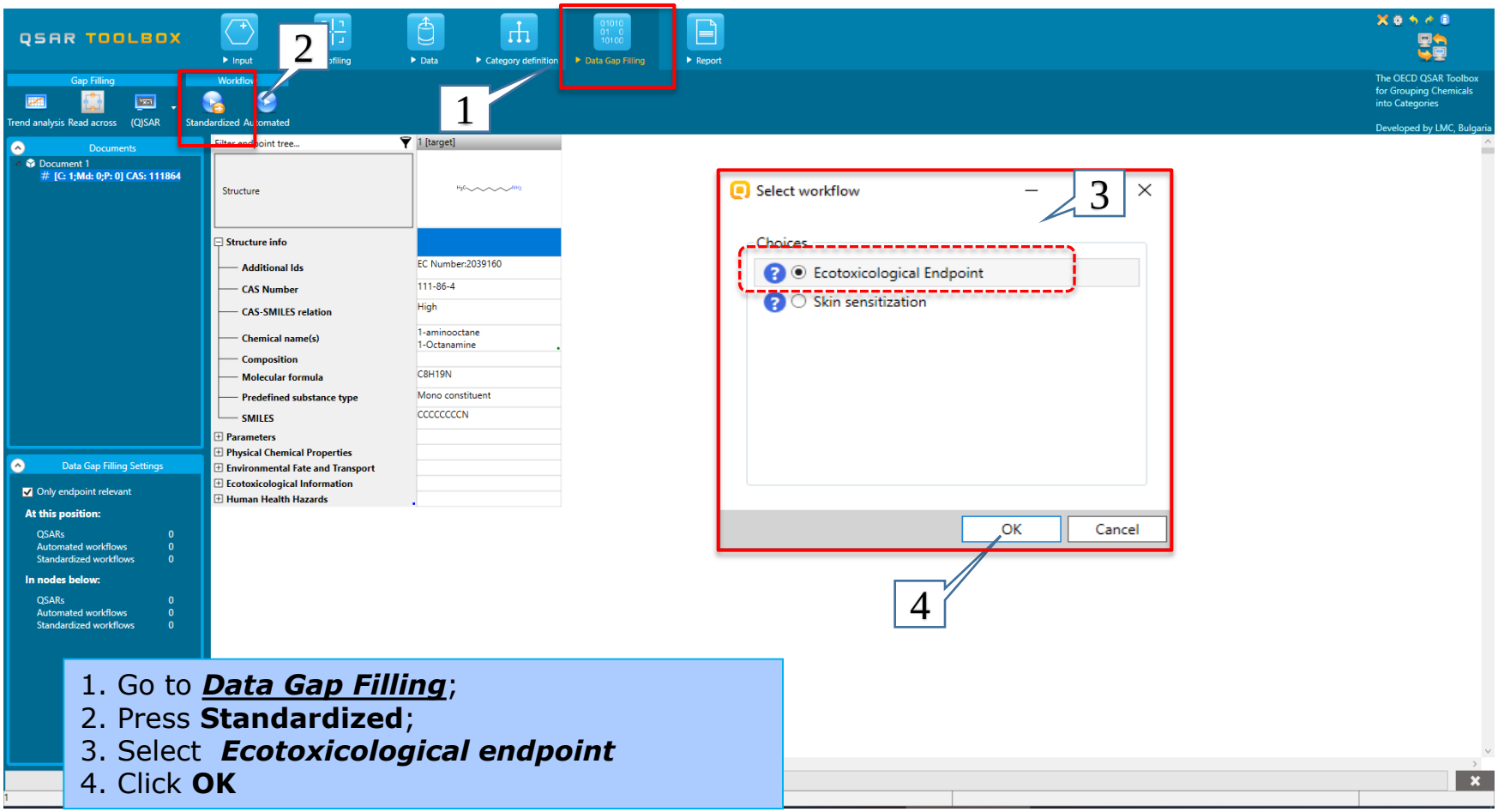
Relevant to the workflow profilers appropriate for DGF are listed and ordered hierarchically based on the population of the group and user is able to select any of them

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

In this example, we will use the Standardized workflow approach.

Data Gap Filling

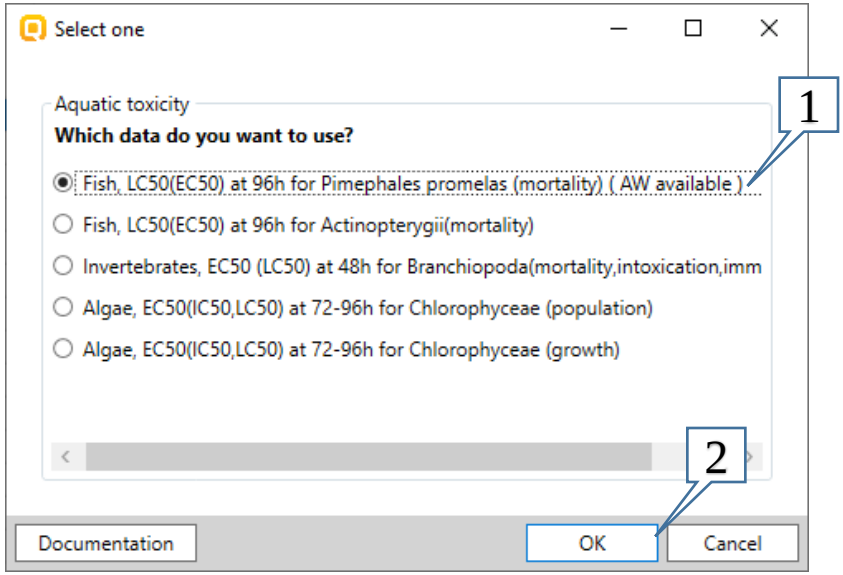
Apply Standardized workflow



1. Go to ***Data Gap Filling***;
2. Press **Standardized**;
3. Select ***Ecotoxicological endpoint***
4. Click **OK**

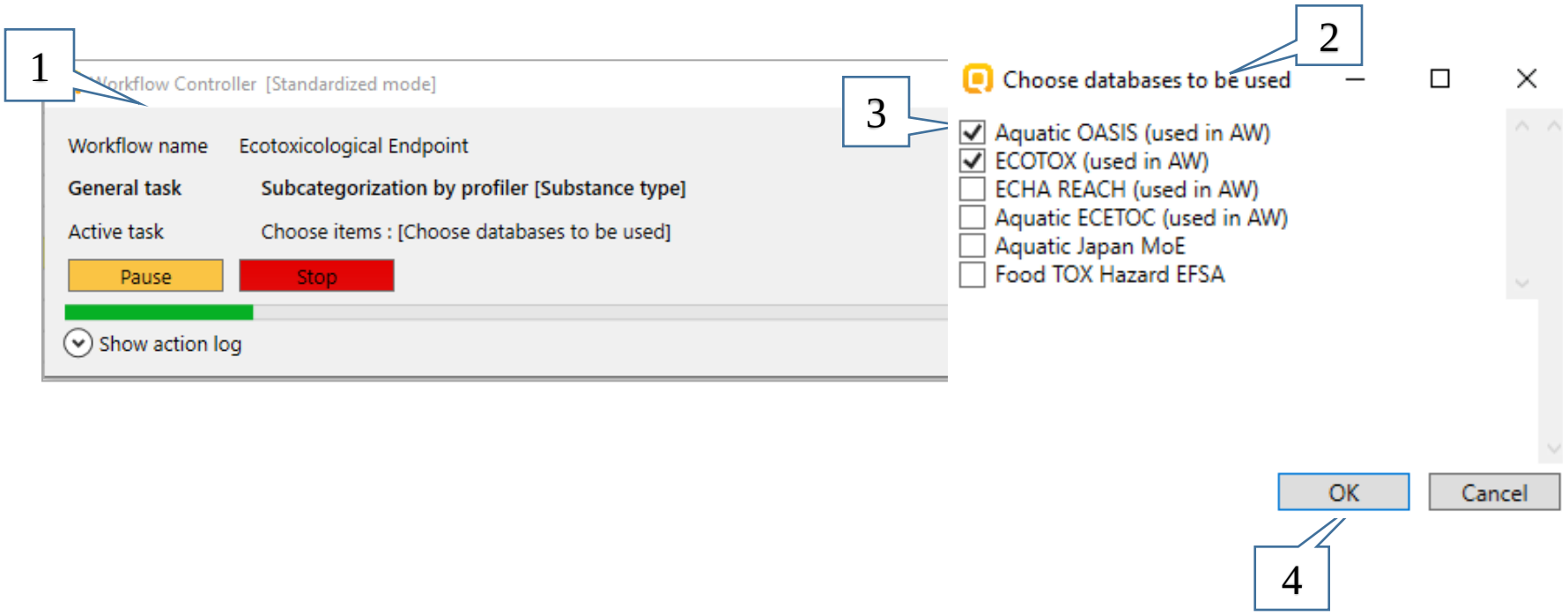
Data Gap Filling
Apply Standardized workflow

There are several options for selection of endpoint and user should select one of them. The first option covers the same endpoint and metadata as the automated workflow, while the others are associated with different endpoints.



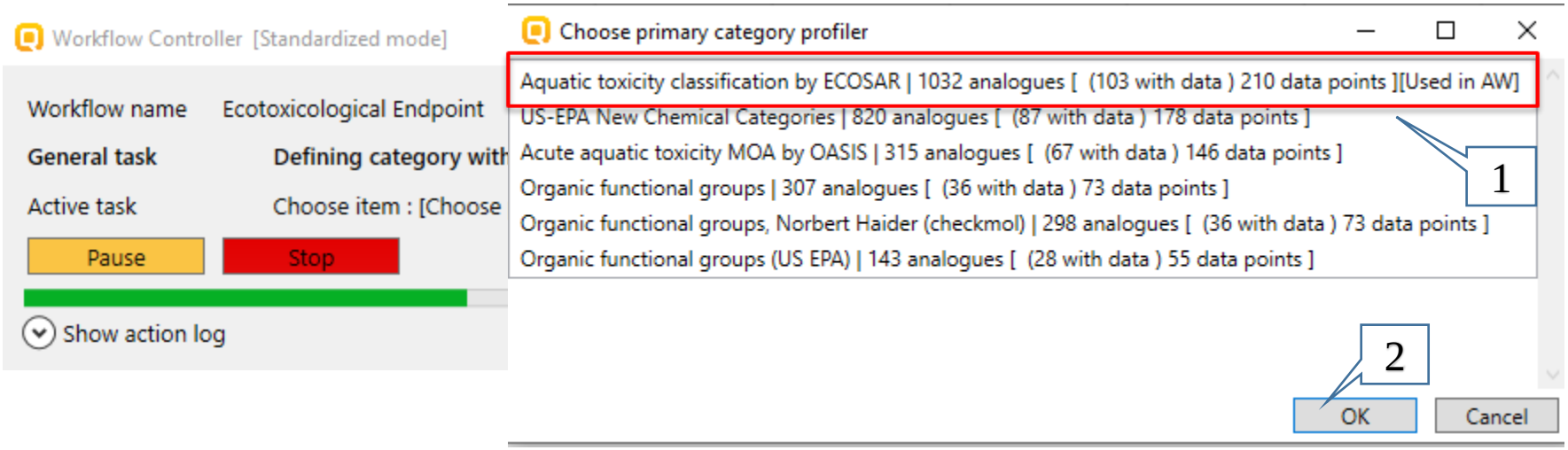
In our case select the first endpoint – **Fish, LC50(EC50) at 96h for Pimephales promelas** *(1).*
The user could make other selection. *AW available in brackets* means that the endpoint is used in the automated workflow;
2. Click on **OK** (2)

Data Gap Filling
Apply Standardized workflow



A workflow controller window (1) is displayed and then a dialogue window for selection (2) of databases appears. Indication about which databases are used in the AWs is also included in brackets. Select the databases which you would like to be used in SW. In our case the first two databases are selected (3); Click on **OK**(4)

Data Gap Filling
Apply Standardized workflow

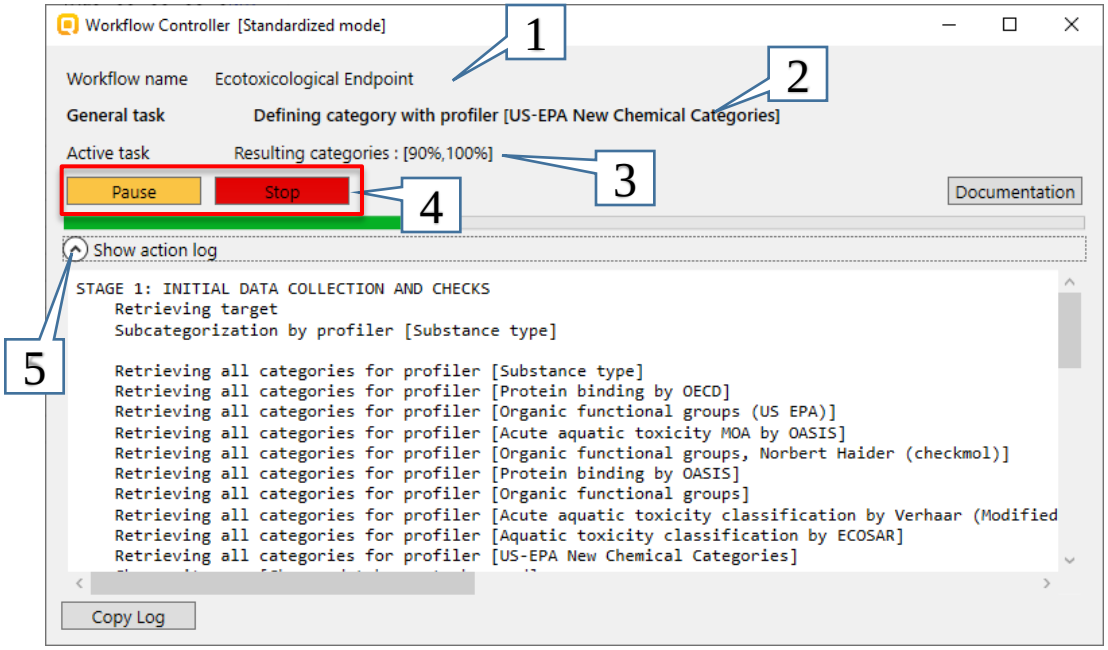


Now the system define groups with analogues and provide a list with the most appropriate for primary grouping profilers starting with the most populated one. Here, we select **Aquatic toxicity classification by ECOSAR** (1) and then click on **OK** (2).

Data Gap Filling
Apply Standardized workflow

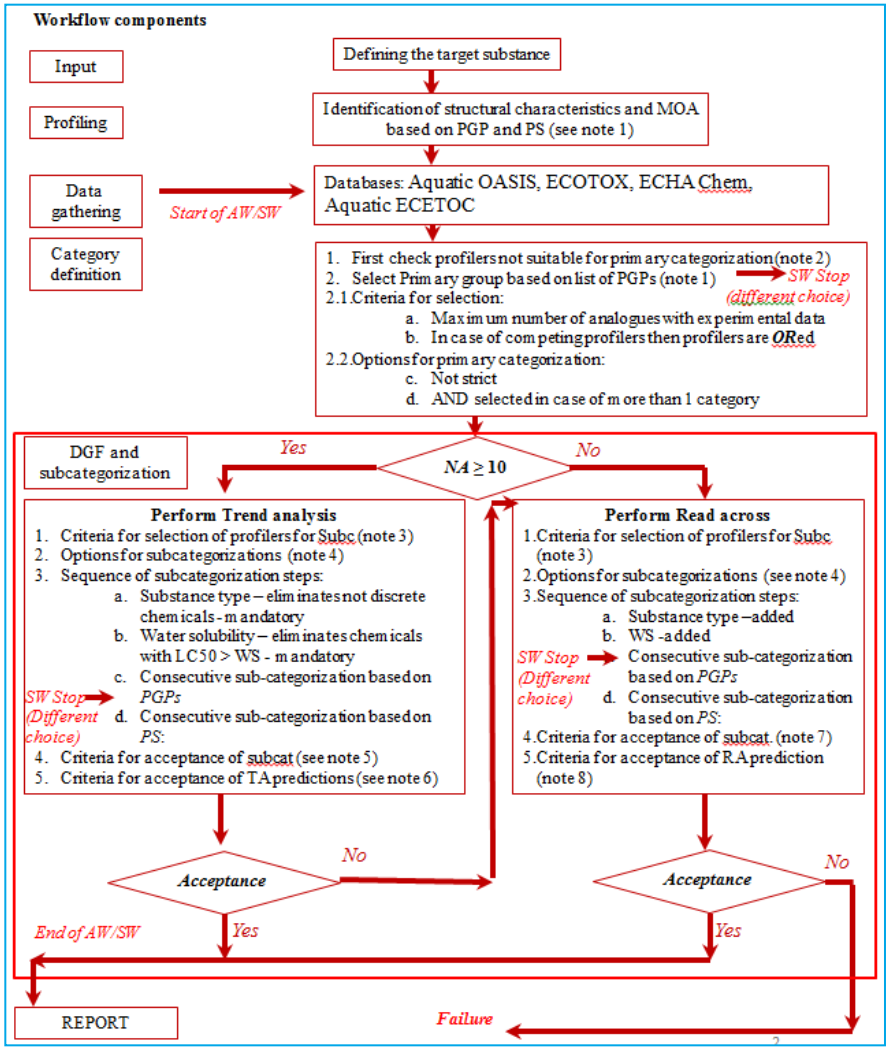
A workflow controller window is displayed throughout the standardized workflow procedure. It includes:

1) Workflow name (1)
2) General task (2)
3) Active task (this is subtask of the general task, which is currently being performed)(3)
4) Navigation options (4)
5) Activity log (5)



Data gap filling

Algorithm of Ecotoxicological workflow



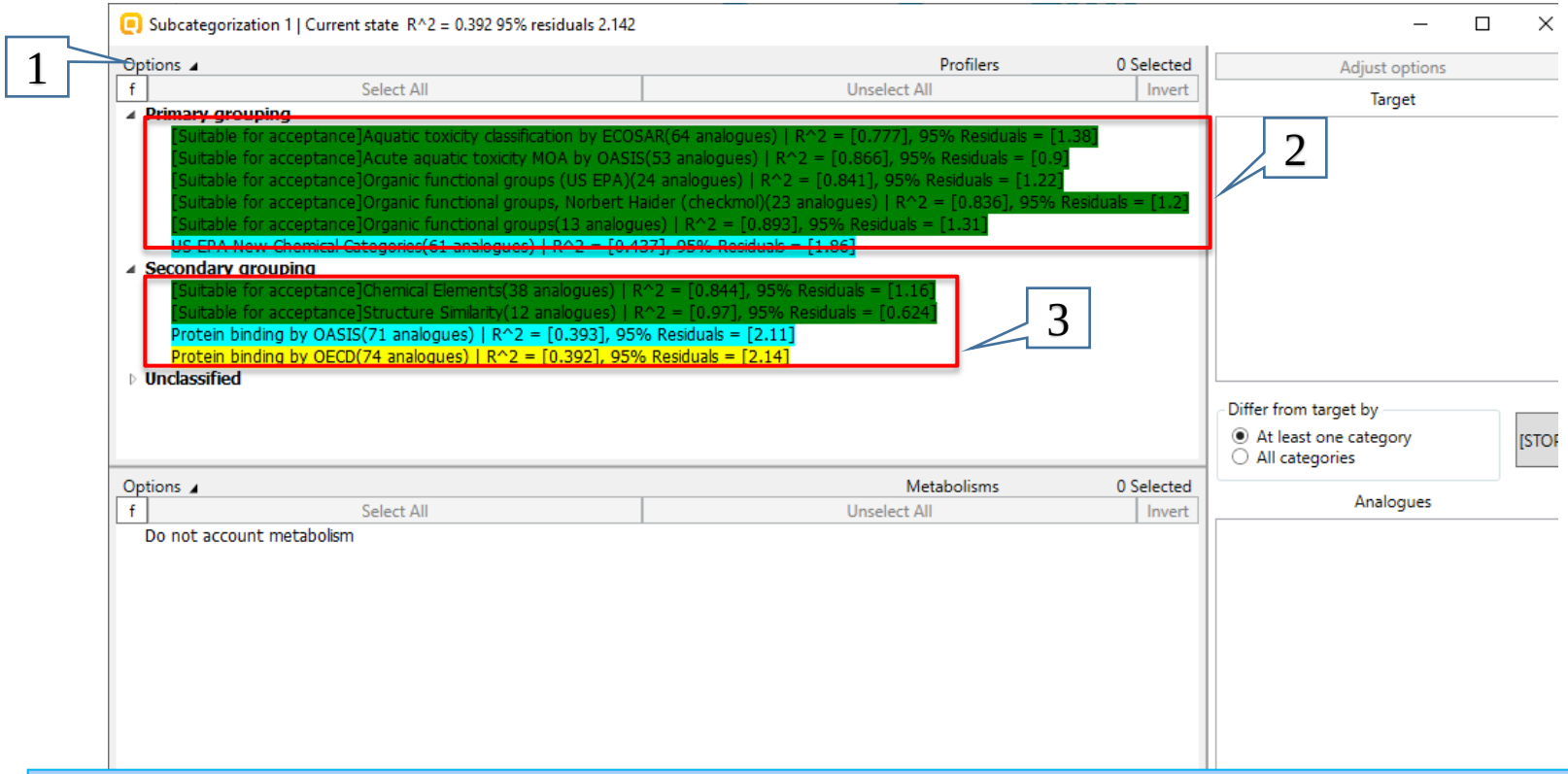
Data Gap Filling
Apply Standardized workflow

When the profiling scheme for primary categorization is selected, the workflow makes a category and enters in Data Gap Filling, where the next step is subcategorization. Depending on the outcome obtained by the applied subcategorization, the profilers are colored as follows:

- **Green** – application of the profiler will satisfy the criteria for acceptance of the prediction
- **Blue** – application of the profiler increase the confidence of the prediction only
- **Yellow** – application of the profiler does not change the current state
- **Red** – criteria for acceptance the subcategorization will be not reached
- **Grey** – already applied profiler

Data Gap Filling

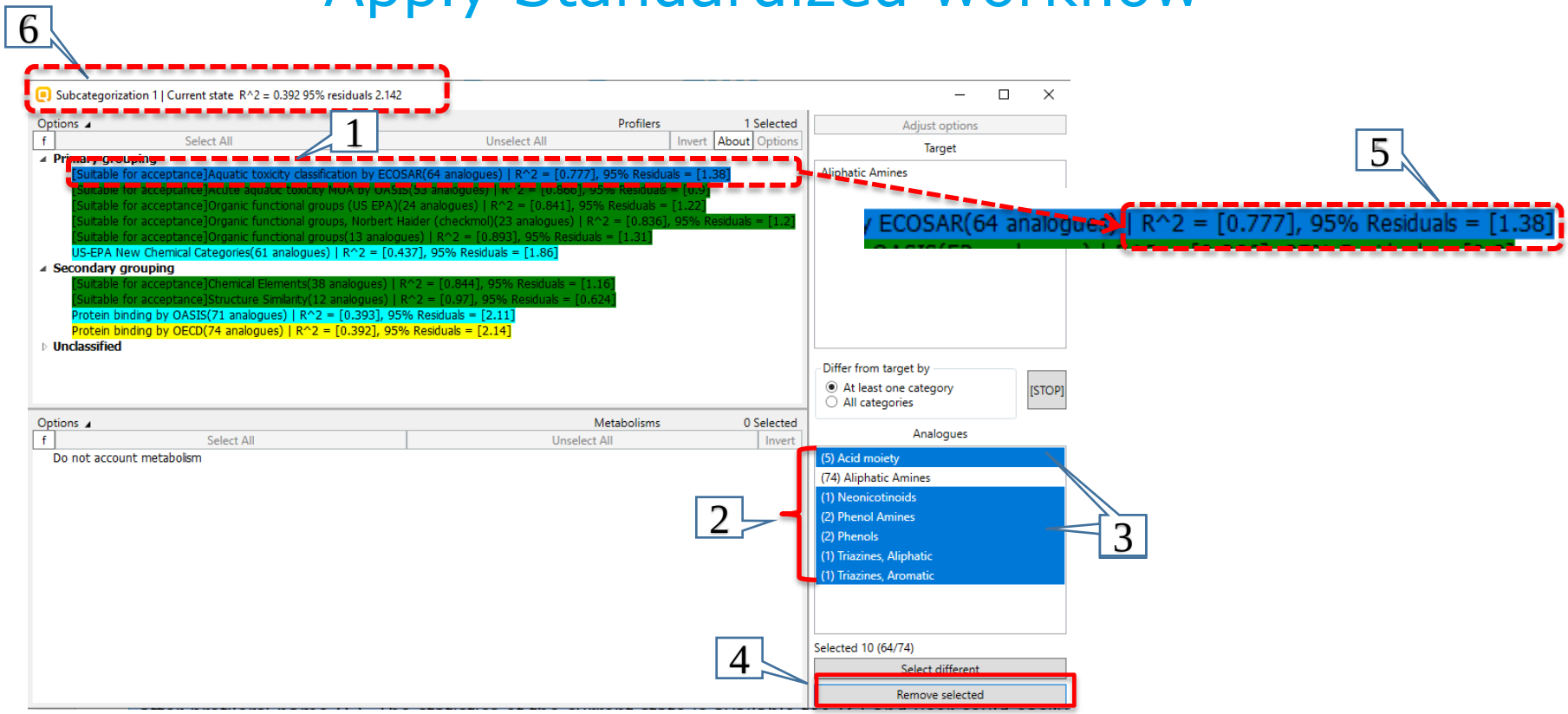
Apply Standardized workflow



Once the primary group is selected, the implemented logic of Ecotoxicological workflow is applied. As result a list with profilers appears highlighted and ordered appropriately. The profilers are separated into two sections in the subcategorization window (1):
Primary grouping (2) and *Secondary grouping* (3).

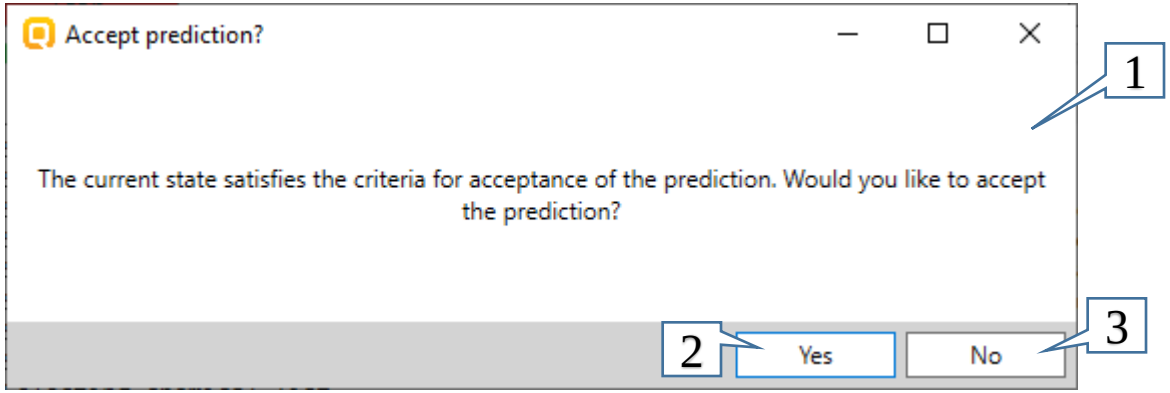
Data Gap Filling

Apply Standardized workflow



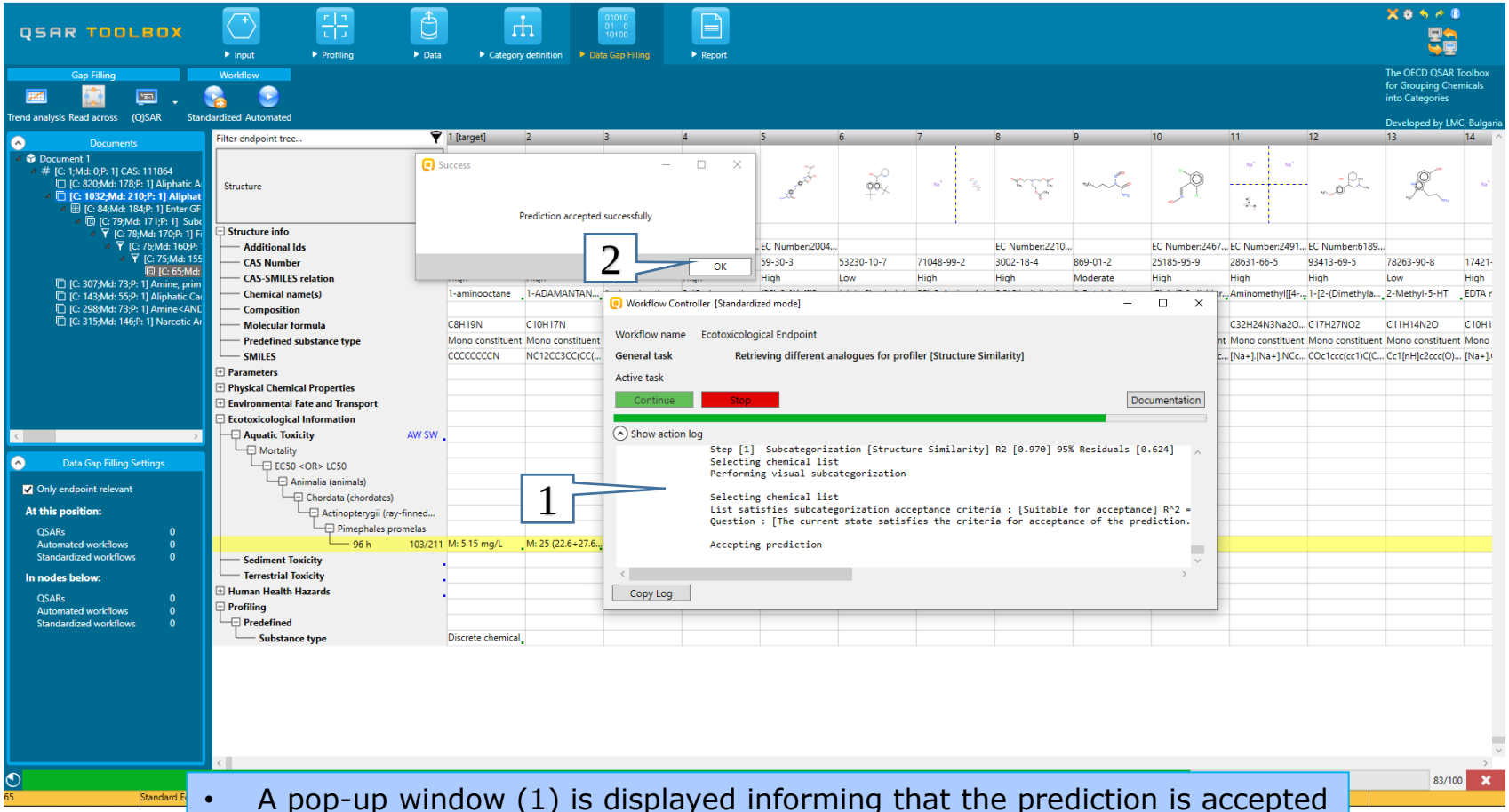
Clicking on the first profiler (1) shows all categories found in the analogues (2) as the ones colored in blue (3) are the ones that are not applicable to the target chemical (different to the target) and hence can be removed (4). Moreover a statistic coefficients which will be obtained if the profiler is applied are provided after profilers' name (5). The statistics of the current state is available too (6) and user could easily compare them.

Data Gap Filling
Apply Standardized workflow



- If obtained results after applied subcategorization satisfy the criteria for acceptance the prediction, then the message appears:
- *The current state satisfy the criteria for acceptance the prediction. Would you like to accept the prediction ?(*1)
- Press Yes (2) , if you want to accept the prediction (next slide)
- Press No (3) and continue with the workflow if you are not satisfy with the outcome and want to continue

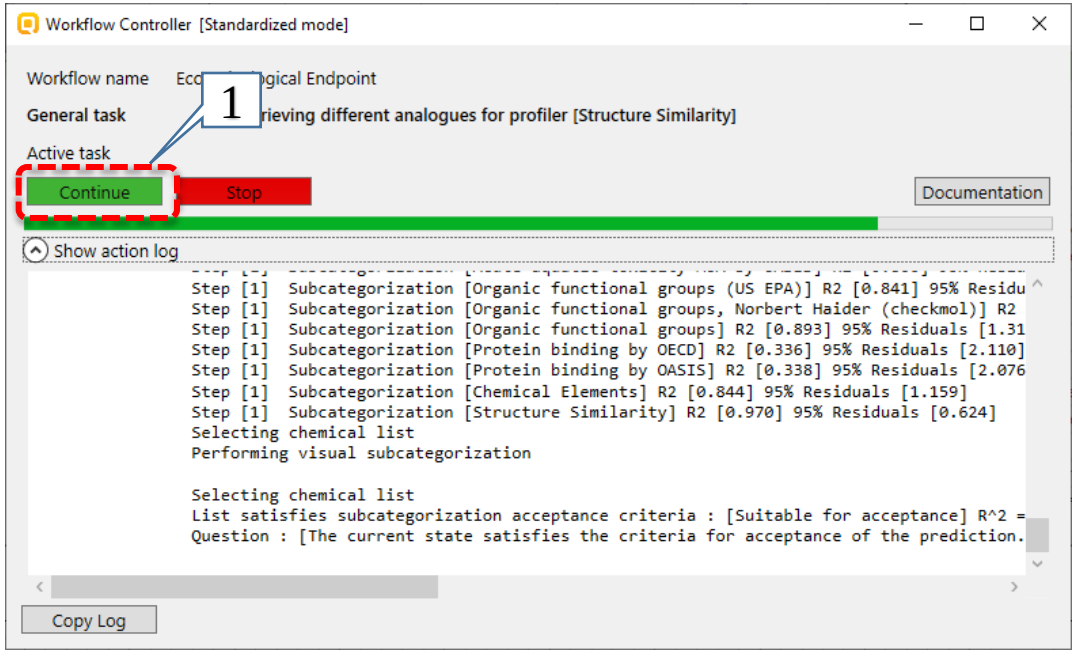
Data Gap Filling
Apply Standardized workflow



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

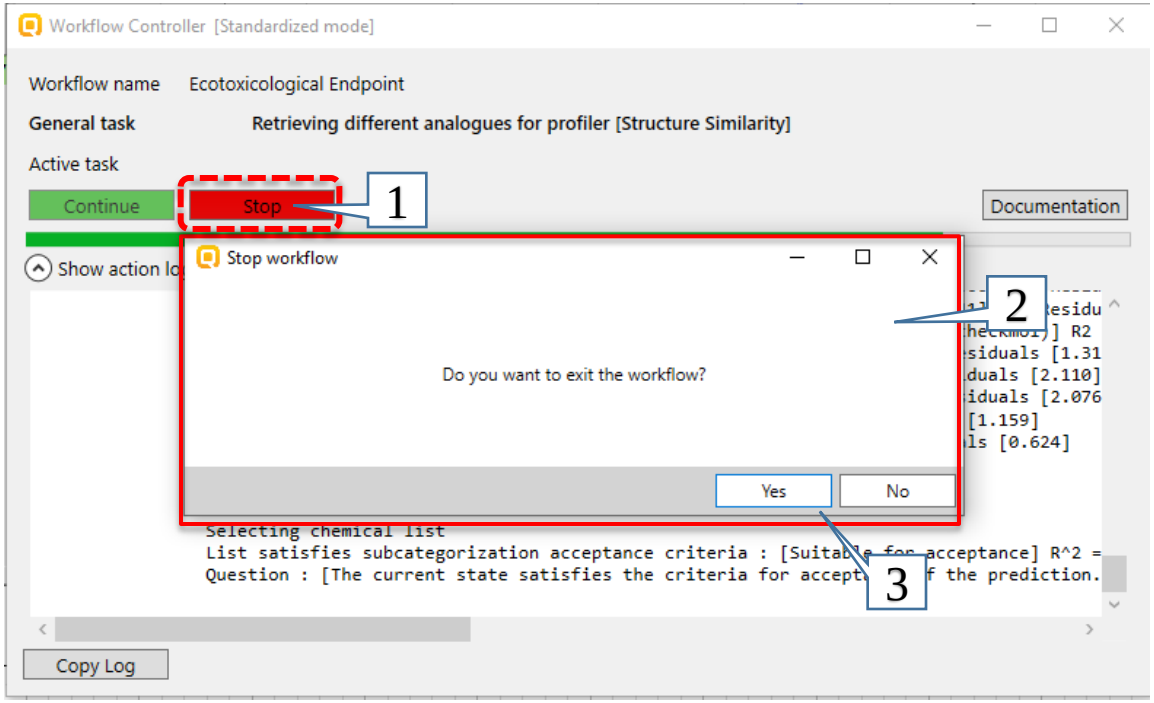
Data Gap Filling
Apply Standardized workflow

- **Continue the workflow if you are not satisfied with the outcome.**



- Press *Continue* button (1) in the workflow controller
- Then repeat the steps described on slide 34

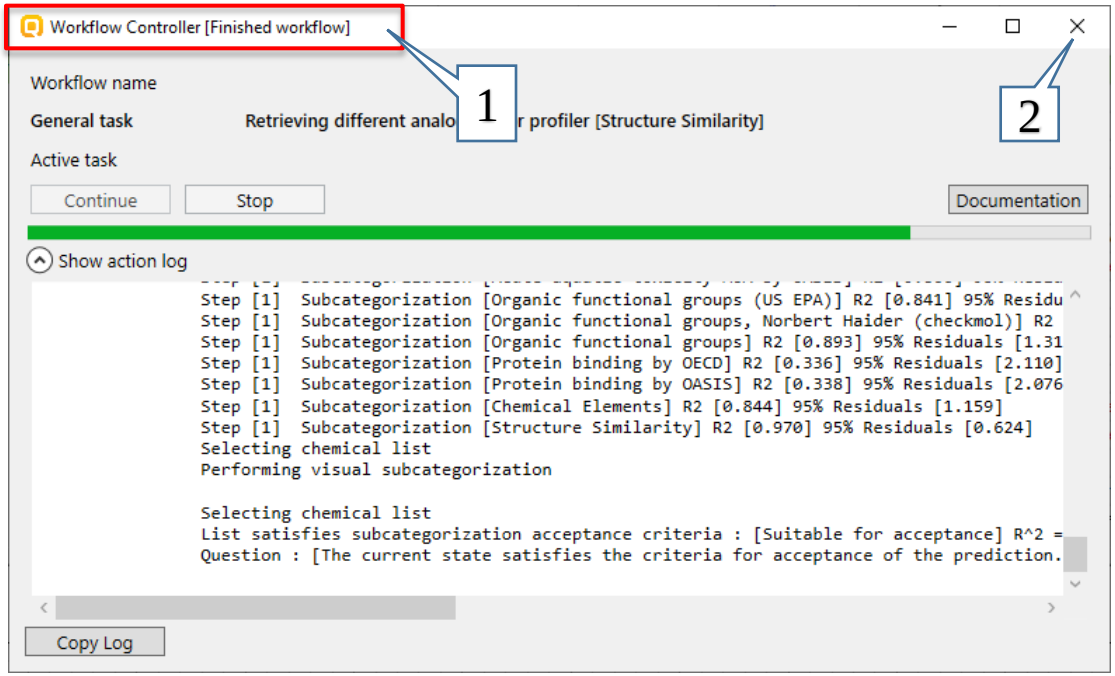
Data Gap Filling
Apply Standardized workflow



- Press *Stop* button (1) on the Workflow controller, once you are satisfied with the prediction
- A pop-up window (2) asks to confirm the exit of the workflow
- Press *Yes button*(3)

Data Gap Filling
Apply Standardized workflow

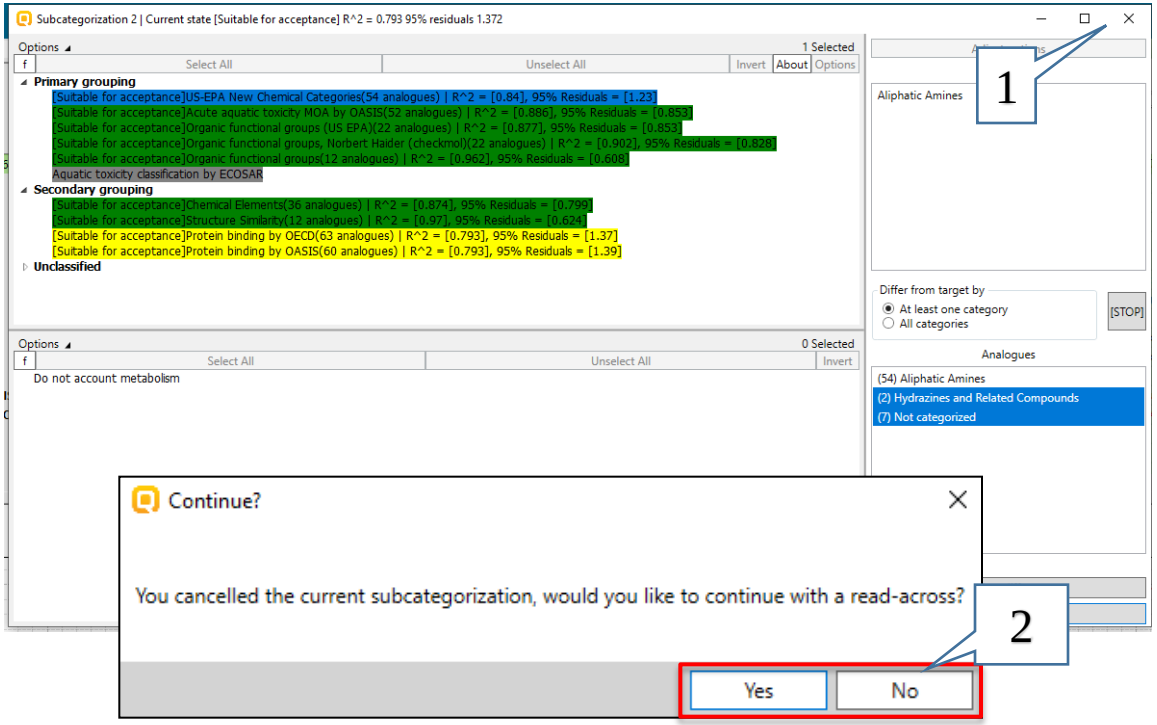
- A message is displayed that the workflow has finished (1)
- Press **X** button (2)



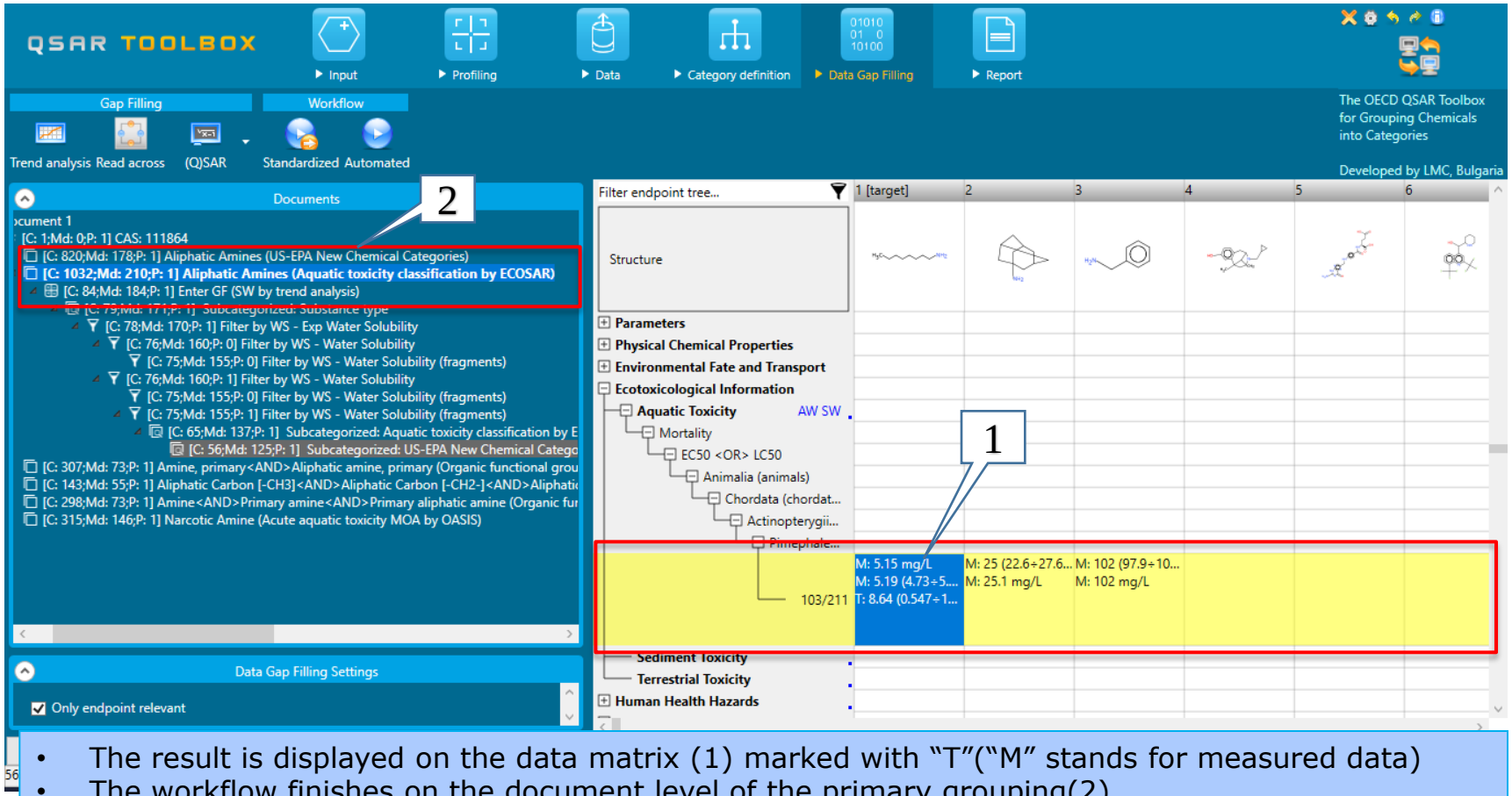
Data Gap Filling
Apply Standardized workflow

If the subcategorization 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.

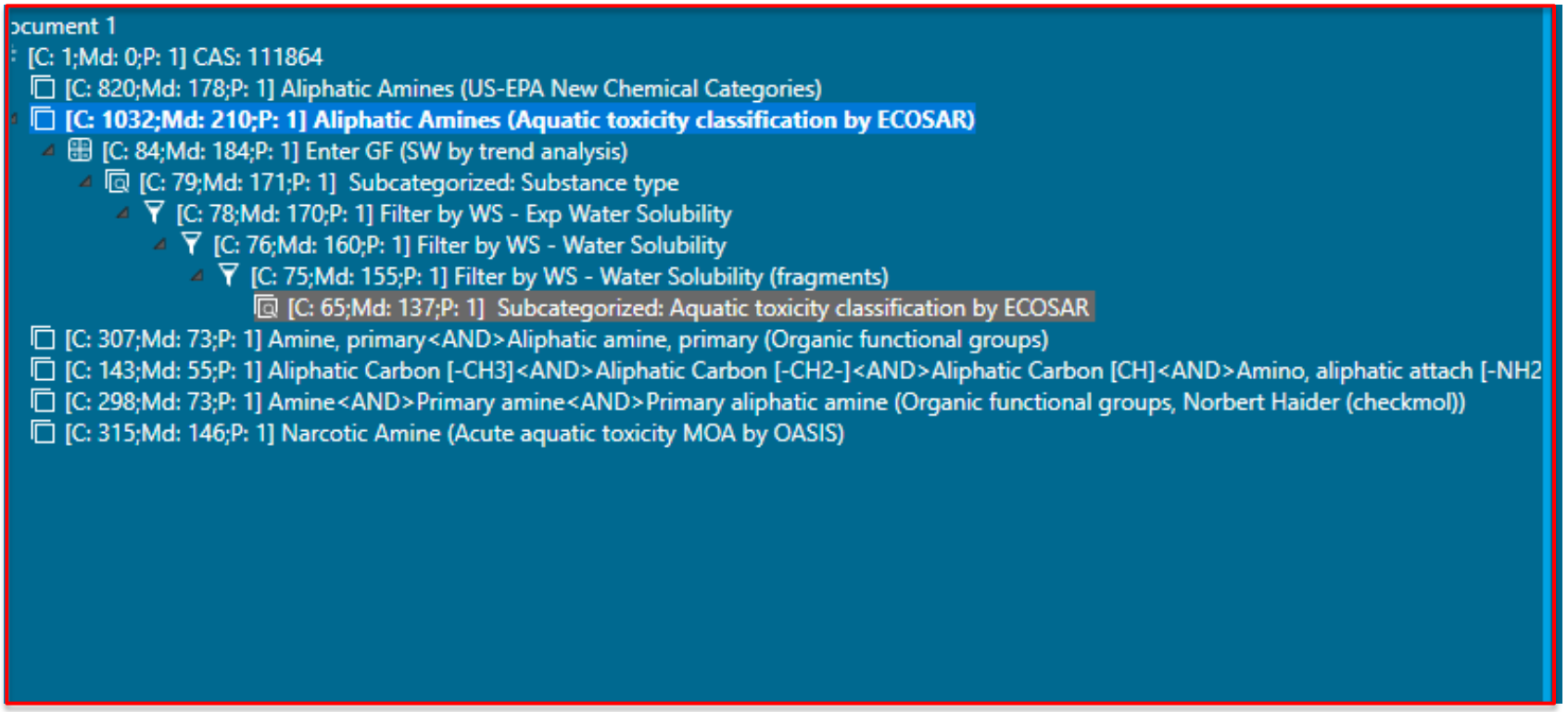


Data Gap Filling
Apply Standardized workflow



- 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
Apply Standardized workflow



All the steps executed in the SW are listed in the Document's panel.
The grey highlighted level(s) of documented tree indicates that a prediction is accepted at this level.

Outlook

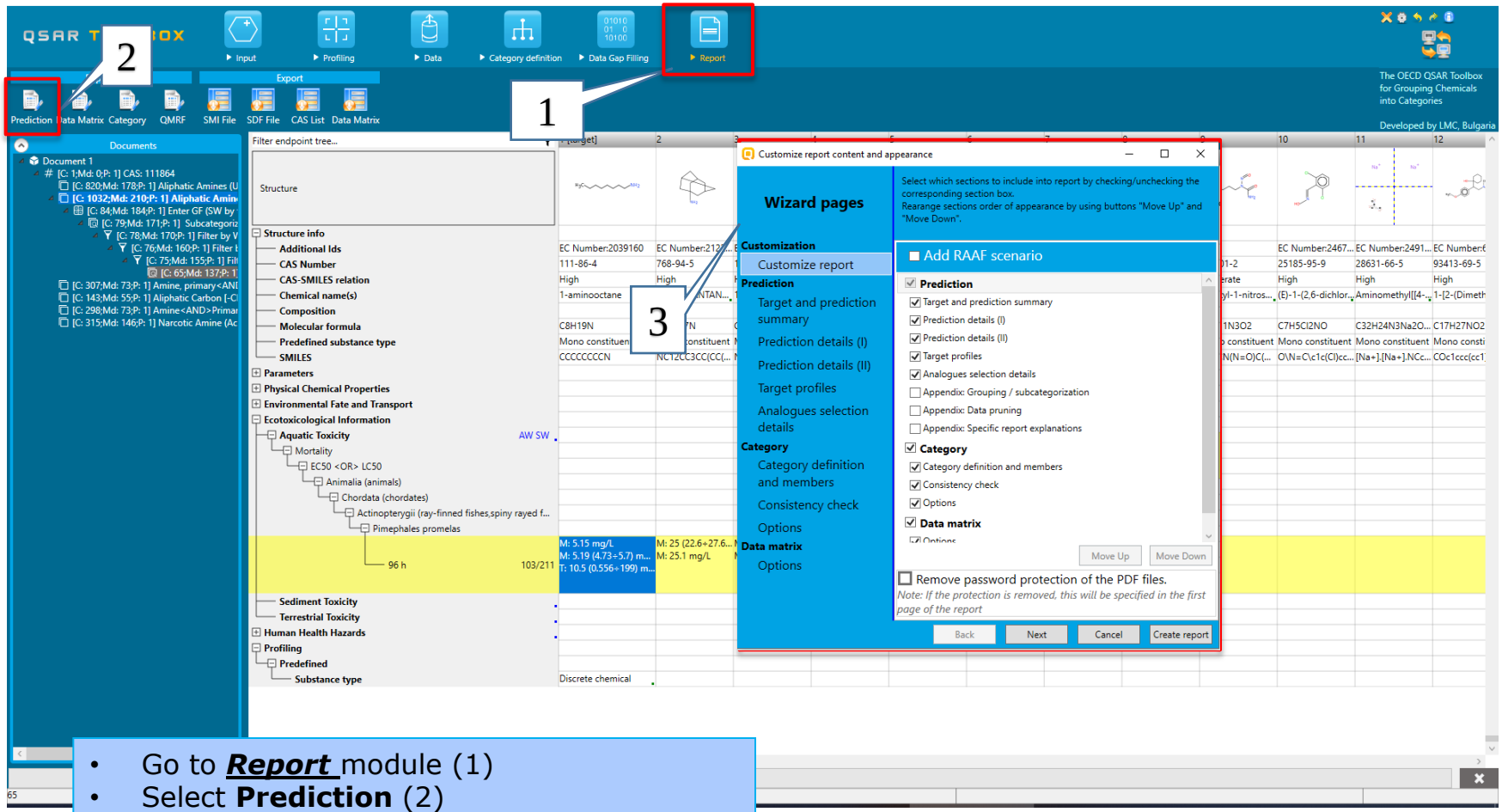
- Background
- Keywords
- Objectives
- Specific Aims
- Standardized workflow for Ecotoxicity
- The exercise
- **Standardized workflow execution**
  - Input
  - Data Gap Filling
  - **Report**

Report
Overview

- The report module can generate reports on predictions performed with the Toolbox.
- The report module contains a predefined report template which users can customized.
- Three types of report files are generated:

  - Prediction report – containing information for the target
  - Category report – containing information for the analogues in the category
  - Data matrix – containing information for the analogues used for the prediction

Report
Generation report

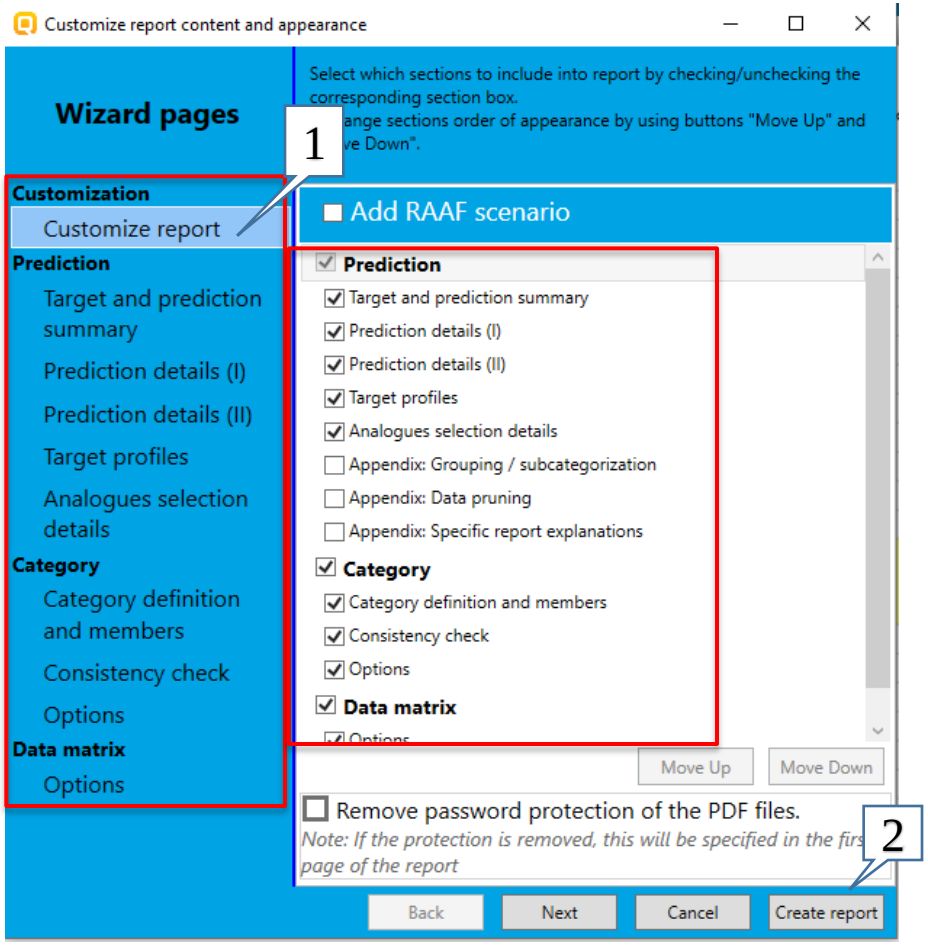


- Go to ***Report*** module (1)
- Select **Prediction** (2)
- A Wizard pages window is displayed (3)

Report
Generation report

The user can customize the report content (1) and appearance.

Generation of the reports happens by click on the **Create report** button (2).



Report
Wizard pages

I. **Customized report** - the user is able to include or exclude the sections in the report;

II. **Prediction report:**

- **Target and prediction summary** – This section includes substance ID of the target chemical and the prediction outcome. Fields which are automatically populated by the system are indicated. Here the user could add information for the author, contact details and summary information;
- **Prediction details** and **Prediction details (II)** – section prediction details provides details about the prediction and its reliability. Prediction details (II) is optional it provides specific information about the prediction depending on the gap filling approach;
- **Target profiles** – this section summarize profiles used for the prediction. Additional profiles could be also included by the user;
- **Analogues selection details** – This section illustrates how analogues were selected. It displays selected databases, category boundaries and applicability domain.

Report
Wizard pages

III. Category report:

- **Category definition and members** – This part includes sections related to list of category members, basic definition of the target endpoint and category hypothesis. Also information for calculated physico-chemical parameters for the category members are provided. Some of the sections are automatically populated while for the others a report items from Report basket could be added manually.
- **Consistency check** – This part includes sections related to the layers of the consistency check: physicochemical similarity; structural similarity, mechanistic similarity and additional endpoint data. Similarly to the previous section some of the sections are automatically populated and for the others items from the Report basket could be added.;
- **Options** – in this section number of the category members used for reporting could be changed;

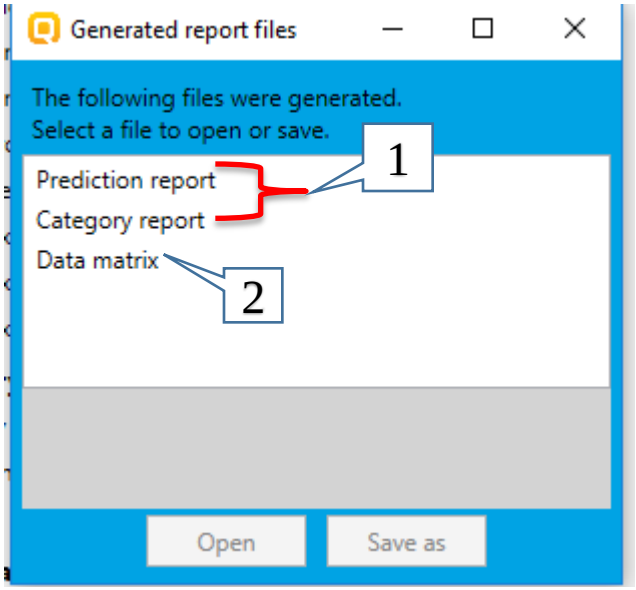
IV. Data matrix report

- Data matrix report gives the possibility to export information for the chemicals in the data matrix including parameters, profilers and experimental data.

Report
Generation report

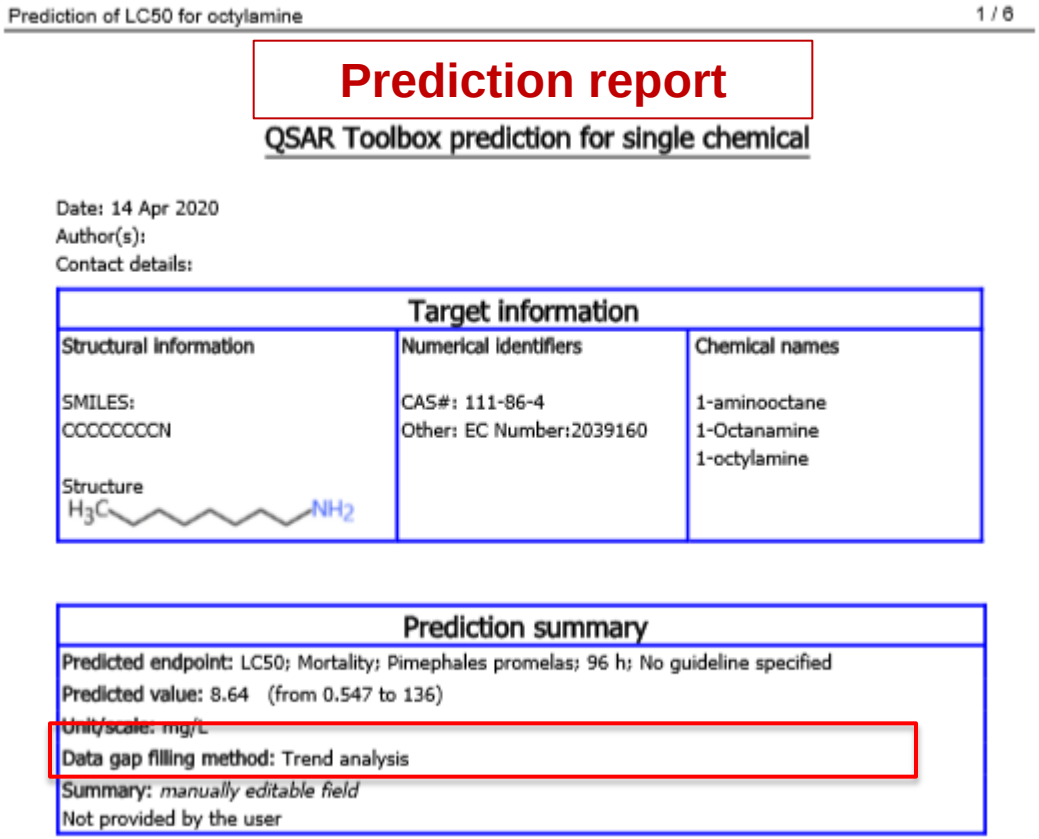
After clicking on the Create report button, *Generated report files* window appears. It contains two types of files:

1) **Prediction report** - a PDF file containing the prediction information related to the target.
2) **Category report -** a PDF file containing information for the consistency of the final category (target plus used analogues).
3) **Data matrix** - a MS Excel file containing chemicals used for prediction along with their data for selected parameters, profiles and endpoint tree positions.



Report
Prediction report

Using of a standardized workflow for predicting of ecotoxicological endpoint is noted in the *Prediction report*.



Prediction of LC50 for octylamine 1 / 6

Prediction report

QSAR Toolbox prediction for single chemical

Date: 14 Apr 2020
Author(s):
Contact details:

Target information

Structural information	Numerical identifiers	Chemical names
SMILES: CCCCCCCCN Structure H_3C ~~~ NH_2	CAS#: 111-86-4 Other: EC Number:2039160	1-aminooctane 1-Octanamine 1-octylamine

Prediction summary

Predicted endpoint: LC50; Mortality; Pimephales promelas; 96 h; No guideline specified
Predicted value: 8.64 (from 0.547 to 136)
Unit/scale: mg/L
Data gap filling method: Trend analysis
Summary: *manually editable field*
Not provided by the user

Report
Category report

Category report

Information for the members of the category obtained as a result of SW application is included in the Category *report*.

QSAR Toolbox report for category

1. Category definition

1.1. Category definition

Category name *manually editable field*

Not provided by the user

Covered (target) endpoint(s)

- Ecotoxicological Information/Aquatic Toxicity: Pimephales promelas, Actinopterygii (ray-finned fishes,spiny rayed fishes), Chordata (chordates), Animalia (animals), EC50 <OR> LC50, Mortality, Duration=96 h

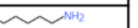
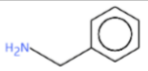
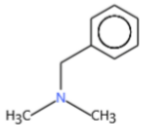
Category hypothesis *manually editable field*

Not provided by the user

1.2. Category members

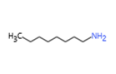
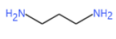
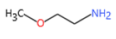
Information of category members

Table of category members

#	CAS	Name	SMILES	Structure
1	111-86-4	octylamine	CCCCCCCCN	
2	768-94-5	Amantadine	NC12CC3CC(CC(C3)C1)C2	
3	100-46-9	Benzylamine	NCc1ccccc1	
4	109-76-2	1,3-Diaminopropane	NCCCN	
5	109-85-3	2-methoxyethanamine	COCCN	
6	103-83-3	Benzyldimethyl	CN(C)Cc1ccccc1	

Report
data matrix report

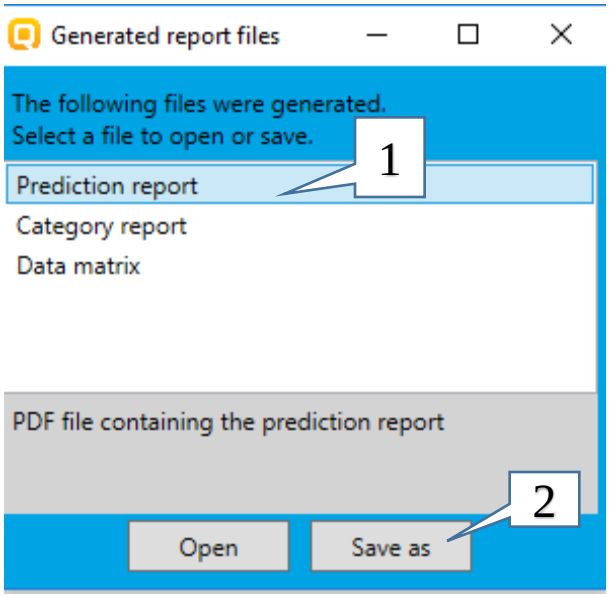
Data matrix report

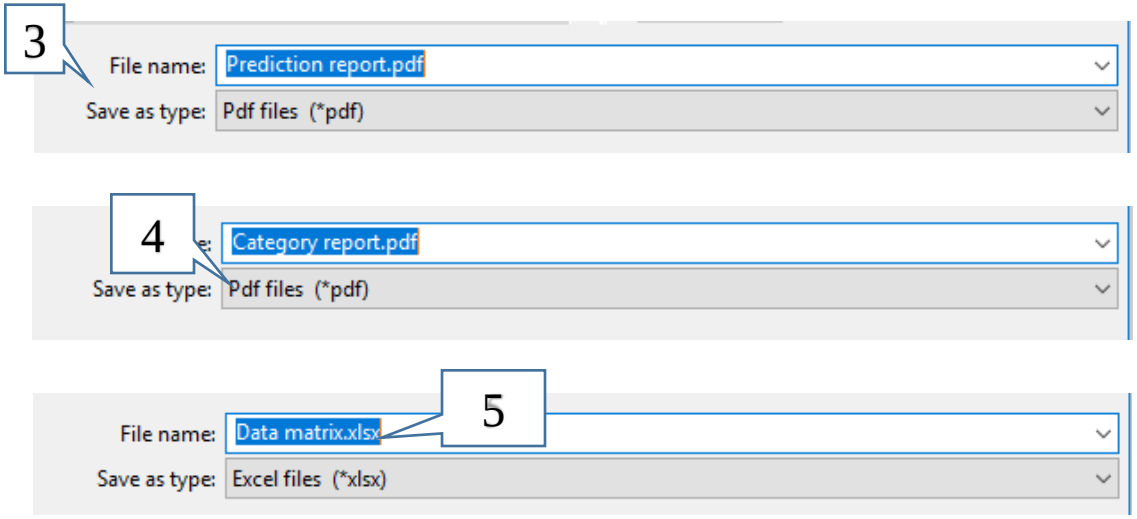
	Target chemical	Analogue #1	Analogue #2	Analogue #3	Analogue #4	Analogue #5	Analogue #6
Substance identity							
Structure							
CAS number	111-86-4	768-94-5	100-46-9	109-76-2	109-85-3	103-83-3	15673-00-4
Chemical name	octylamine	Amantadine	Benzylamine	1,3-Diaminopropane	2-methoxyethanamine	Benzyldimethyl	3,3-Dimethylbutylamine
Other identifier							
SMILES	CCCCCCCCN	NC12CC3CC(CC(C3)C1)C2	NCc1ccccc1	NCCCN	COCCN	CN(C)Cc1ccccc1	CC(C)(C)CCN
Parameters **unit**							
Profilers							
Profiles used for grouping/subcategorization							
Aliphatic Amines (Aquatic toxicity	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines
Substance type (subcategorization)	Discrete chemical; Mono constituent (predefined); Organic	Discrete chemical; Mono constituent (predefined); Organic	Discrete chemical; Mono constituent (predefined); Organic	Discrete chemical; Mono constituent (predefined); Organic	Discrete chemical; Mono constituent (predefined); Organic	Discrete chemical; Mono constituent (predefined); Organic	Discrete chemical; Mono constituent (predefine Organic
Aquatic toxicity classification by ECOSAR	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines
US-EPA New Chemical Categories	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines
General Mechanistic							
Protein binding by OASIS	No alert found	No alert found	No alert found	No alert found	No alert found	No alert found	No alert found
Protein binding by OECD	No alert found	No alert found	No alert found	No alert found	No alert found	No alert found	No alert found
Endpoint Specific							
Acute aquatic toxicity MOA by OASIS	Narcotic Amine	Narcotic Amine	Narcotic Amine	Narcotic Amine	Narcotic Amine	Narcotic Amine	Narcotic Amine
Empiric							
Organic functional groups (US EPA)	Aliphatic Carbon [-CH3]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [CH]; Amino, aliphatic attach [-NH2]; Amino, aliphatic attach [-N<]	Fused Aliphatic ring unit ; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [CH]; Aliphatic Carbon [C]; Amino, aliphatic attach [-NH2]; Amino, aliphatic attach [-N<]; Tertiary Carbon	Aliphatic Carbon [-CH2-]; Aliphatic Carbon [CH]; Amino, aliphatic attach [-NH2]; Amino, aliphatic attach [-N<]; Aromatic Carbon [C]	Aliphatic Carbon [-CH2-]; Aliphatic Carbon [CH]; Amino, aliphatic attach [-NH2]; Amino, aliphatic attach [-N<]	Aliphatic Carbon [-CH3]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [CH]; Oxygen, aliphatic attach [-O-]; Amino, aliphatic attach [-NH2]; Amino, aliphatic attach [-N<]	Aliphatic Carbon [-CH3]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [CH]; Amino, aliphatic attach [-N<]; Aromatic Carbon [C]	Aliphatic Carbon [-CH3]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [CH]; Amino, aliphatic attach [-NH Amino, aliphatic attach [-N<
Organic functional groups	Amine, primary; Aliphatic amine, primary	Amine, primary; Tricyclodecane; Cycloalkane; Bridged-ring carbocycles; Aliphatic amine, primary	Amine, primary; Benzyl; Aryl; Aliphatic amine, primary	Amine, primary; Aliphatic amine, primary	Amine, primary; Ether; Aliphatic amine, primary	Amine, tertiary; Benzyl; Aryl; Aliphatic amine, tertiary	Alkane, branched with quaternary Amine, primary; tert-Butyl; Aliphatic amine, primary
Structure similarity	[90%,100%]	[0%,10%)	[10%,20%)	[40%,50%)	[20%,30%)	[0%,10%)	[20%,30%)
Chemical elements	Group 14 - Carbon C; Group 15 - Nitrogen N	Group 14 - Carbon C; Group 15 - Nitrogen N	Group 14 - Carbon C; Group 15 - Nitrogen N	Group 14 - Carbon C; Group 15 - Nitrogen N	Group 14 - Carbon C; Group 15 - Nitrogen N; Group 16 - Oxygen O	Group 14 - Carbon C; Group 15 - Nitrogen N	Group 14 - Carbon C; Group 15 - Nitrogen N

Analogues used for the target prediction can be seen the **Data matrix** report. Their selected profiling results, experimental data and/or parameters are also shown.

Report
Saving the report files

To save any of the reports, select the **report** (1) and then click on **Save as** (2); The prediction and category reports are saved as a **pdf file** (3, 4) while the **data matrix** is saved as an ***.xlsx* file** (5)





Congratulations!

- You have completed the tutorial on the standardized workflow for ecotoxicological endpoint.

- You have now been introduced to the consecutive steps of the standardized workflow of the (Q)SAR Toolbox and the rationale behind each step.

- Note, proficiency comes with practice!