QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.1

Tutorial of how to use Automated workflow for ecotoxicological prediction

Outlook

- **Aim**
- Automated workflow
- The exercise
- Report

Aim

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

Outlook

- Aim
- **Automated workflow**
- The exercise
- Report

Automated workflow (AW) for ecotoxicity

Mechanistic understanding:

According to **McFarland**, the toxicity to aquatic organisms depends on penetration (log K_{OW}) of the chemical, followed by interaction with cellular biomolecules.

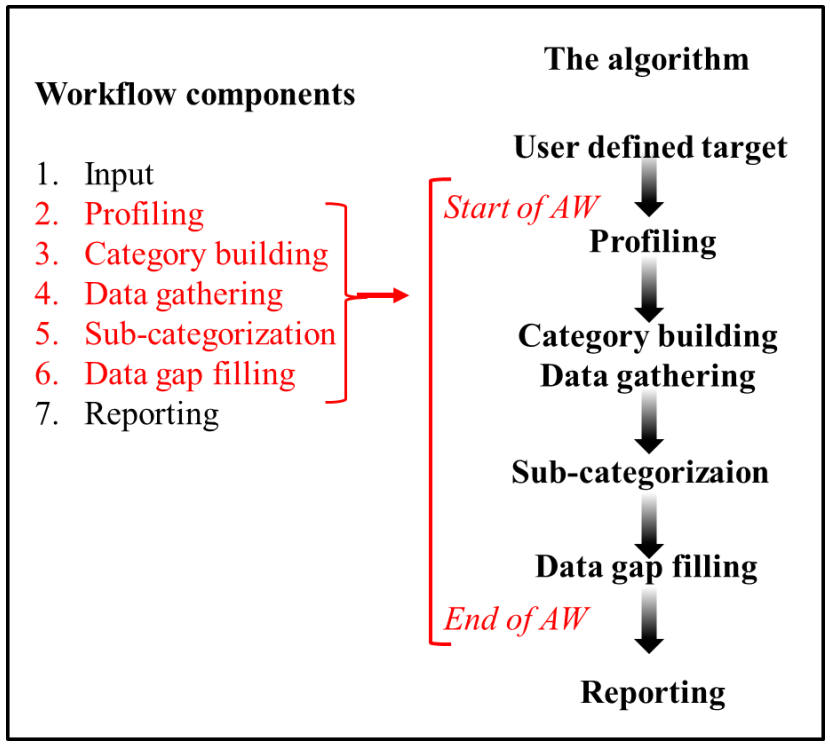
Ecotoxicological endpoint is:

- Sub-hazard – ***Aquatic toxicity***
- Effect – ***Mortality***
- Endpoint – ***LC50 (EC50)***
- Test duration – ***96 h***
- Test species – ***P. promelas***

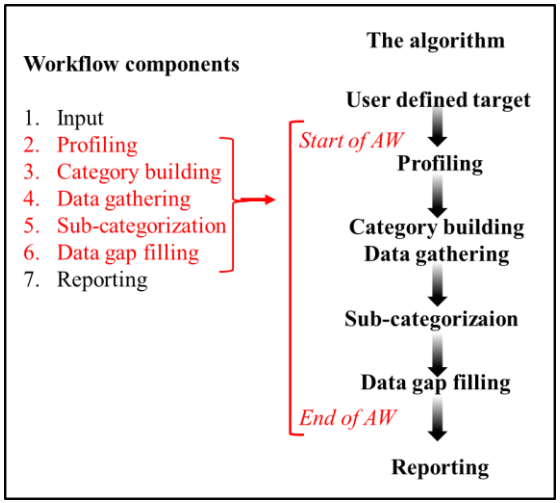
Automated workflow (AW) for ecotoxicity

Workflow components:

- Except *Input* and *Reporting*, the rest of the Toolbox modules are part of the automated workflow (AW).



Automated workflow (AW) for ecotoxicity

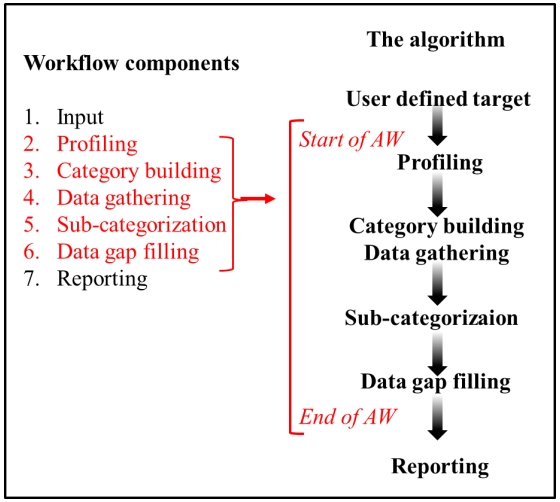


Workflow components: Profiling

- The aim is to collect structural and mechanistic information about the target
- Profilers collecting information for the target are organized in two groups:
  - **Primary grouping profilers (PGPs):**
    US EPA,
    Verhaar,
    MOA,
    ECOSAR,
    OFG (without nested)
  - **Profilers for subcategorization (PS):**
    Substance type,
    Protein binding (OASIS/OECD),
    Chemical elements,
    Str. Similarity (ACF-first neighbors, Dice > 50%)

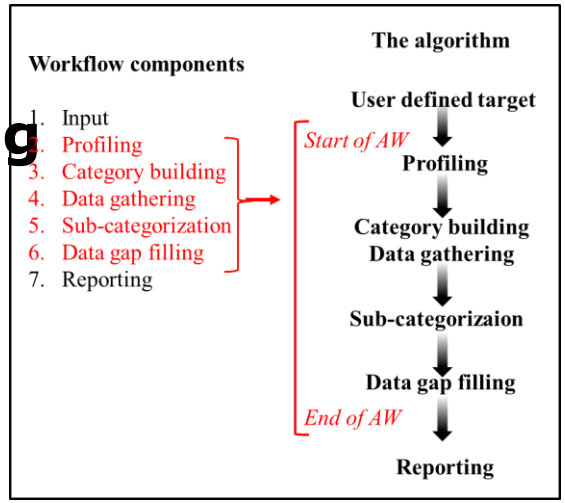
Automated workflow (AW) for ecotoxicity

Workflow components: Data



- Databases
  - Aquatic OASIS
  - ECOTOX
  - ECHA CHEM
- Endpoint data
  - *P. promelas* ~ 5400 data for ~ 2400 chemicals
  - *O. mykiss* ~ 6000 data for ~ 2000 chemicals

Automated workflow (AW) for ecotoxicity

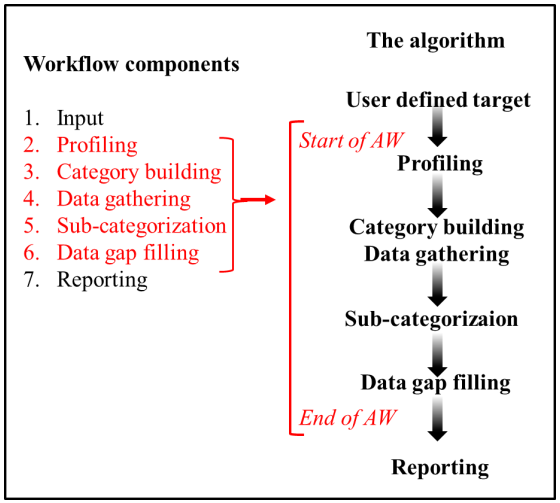


Workflow components: Category building

- The aim is to collect analogues based on global molecular features
- Profilers suitable for primary categorization:
  - US EPA,
  - Verhaar,
  - MOA,
  - ECOSAR,
  - OFG,
  - OFG US-EPA,
  - OFG, Norbert Haider.

Automated workflow (AW) for ecotoxicity

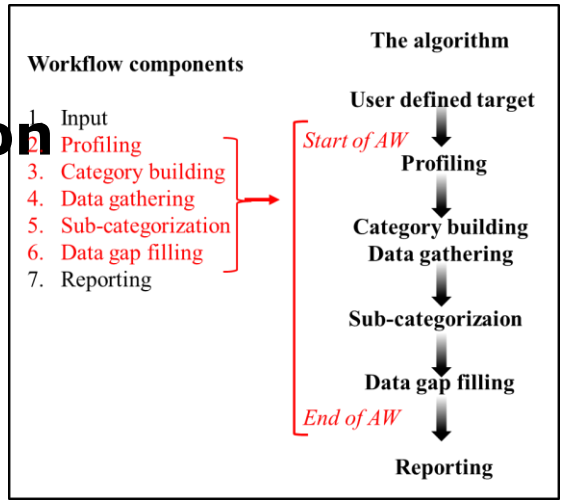
Workflow components: Data Gap Filling



- The aim is to fill in missing data gap by
  - Read-across (RA), or
  - Trend analysis (TA)
- Trend analysis is the default approach
- Read across is applied if
  - Prediction by Trend analysis is not acceptable, or
  - The number of analogues is < 10
- Gap filling and subcategorizations are sequence of logical operations (if, then), combined with criteria for acceptance.

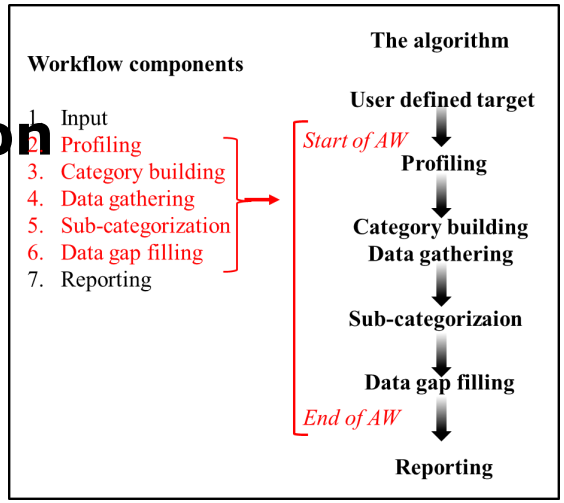
Automated workflow (AW) for ecotoxicity

Workflow components: Subcategorization



- The aim is to increase the similarity of analogues with the target
- It is consecutive process of application of primary grouping profilers (PGPs) and profilers for subcategorization (PS)
- Hierarchy of application of PGPs and PS depends on the number of analogues they have collected
- Sub-categorization process is based on:
  - Sequence of subcategorization steps
  - Criteria for acceptance of subcategorization steps

Automated workflow (AW) for ecotoxicity



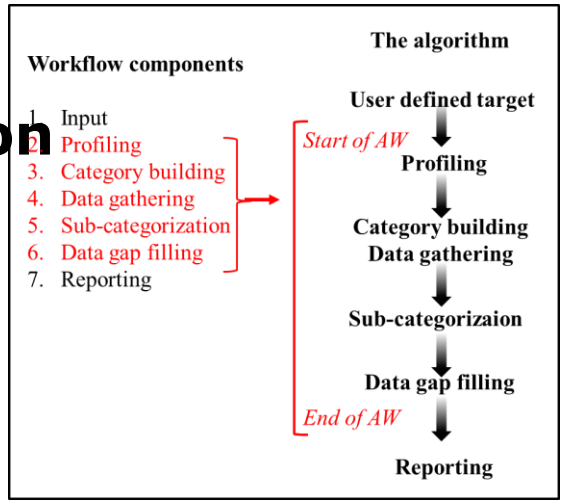
Workflow components: Subcategorization

Sequence of subcategorization steps

1. *Substance type* – eliminates not discrete chemicals
2. *Water solubility* (WSKOWWIN + WATERNT) – eliminates chemicals with *LC50 > WS*
3. Consecutive sub-categorization based on *PGPs*:
   - US EPA,
   - Verhaar,
   - MOA,
   - ECOSAR,
   - OFG (without nested)
4. Consecutive sub-categorization based on *PS*:
   - Substance type,
   - Protein binding (OASIS + OECD),
   - Chemical elements,
   - Str. Similarity

Automated workflow (AW) for ecotoxicity

Workflow components: Subcategorization



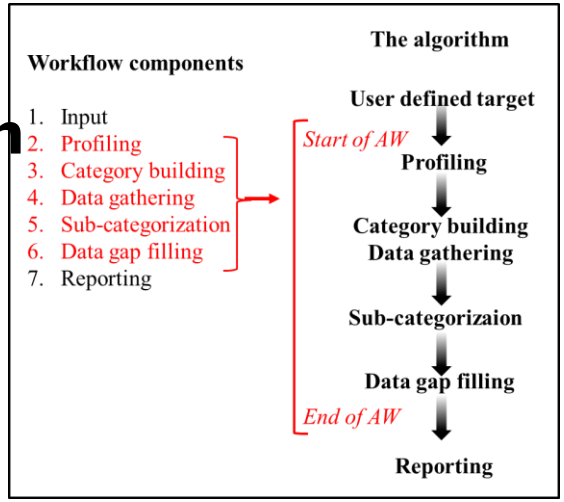
Criteria for acceptance of subcategorization step:

- Depends on the specific statistical and structural criteria (e.g, experimental error, 95% of residuals, log K_{OW}, range of variation of the analogues etc.)

- Criteria are different for RA and TA

Automated workflow (AW) for ecotoxicity

Workflow components: Subcategorization



Criteria for acceptance of subcategorization for **Trend analysis**

1. Sub-categorization by PGPs

IF *Interpolation* **AND** (R^2 ↑ OR *95% of residuals* ↓) **AND** *NA* ≥ 10
THEN accept the subcategorization and continue with the next PGP,
ELSE reject subcategorization and continue with the next PGP

2. Sub-categorization by PSs

IF *Interpolation* **AND** (R^2 ↑ **OR** *95% of residuals* ↓) **AND** *NA* ≥ 10
THEN accept sub-categorization and continue with the next PS,
ELSE continue with the next PS

Definitions:

Interpolation: log K_{OW} of the target should be within the range of log K_{OW} of analogues

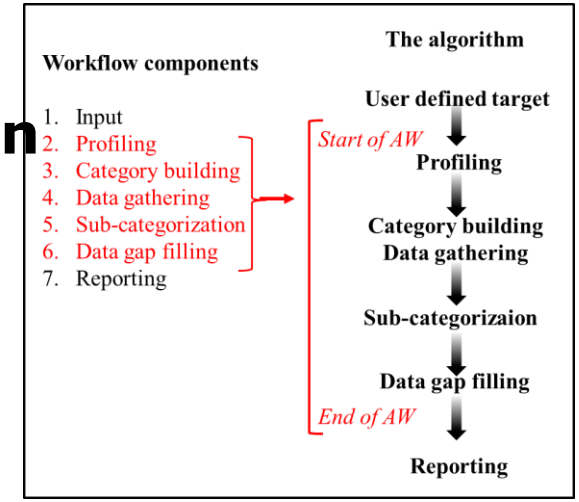
Correlation coefficient R^2 ↑: R^2 increases **OR** R^2 ≥ 0.85

95% of residuals ↓: 95% of residuals decreases **OR** 95% of residuals ≤ 1.0

NA – Number of analogues

Automated workflow (AW) for ecotoxicity

Workflow components: Subcategorization



Criteria for acceptance of subcategorization for **Read across**

1. Sub-categorization by PGPs

IF *Interpolation* **AND** (*LC50* ↓ **OR** log *Kow* ↓) **AND** *NA* ≥ 5
THEN accept sub-categorization and continue with the next profiler
ELSE reject sub-categorization and continue with the next profiler

2. Sub-categorization by PSs
IF *Interpolation* **AND** (*LC50* ↓ **OR** log *Kow* ↓) **AND** *NA* ≥ 5
THEN accept sub-categorization and continue with the next profiler
ELSE reject sub-categorization and continue with the next profiler

Definitions

Interpolation: log K_{OW} of the target should be within the range of log K_{OW} of analogues

LC50 ↓*:* for the 5 closest analogues the range of variation of *LC50* decreases **OR** range of variation is ≤ 2 log units

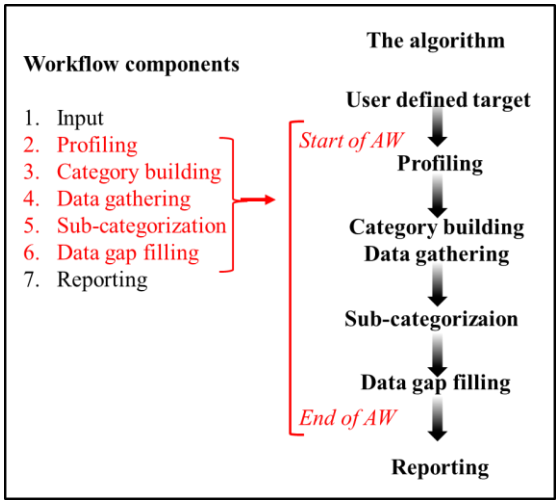
log K_{OW} ↓*:* for the 5 closest analogues the range of variation of log Kow decreases **OR** range of variation is ≤ 2 log units

NA – Number of analogues

Automated workflow (AW) for ecotoxicity

Workflow components

Criteria for acceptance of **prediction for Trend analysis**

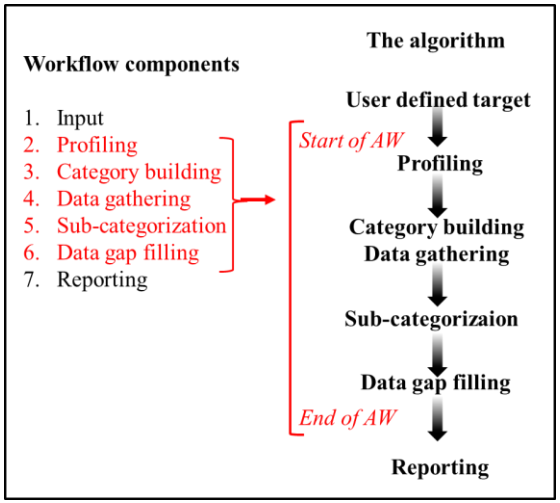


After sub-categorization by all PGPs and PS

IF (R^2 ≥ 0.8 **and** *NA*≥ 5) **OR** (R^2 ≥ 0.7 **and** *NA* ≥ 10)
THEN accept the prediction and generate report,
ELSE switch to Read across

Automated workflow (AW) for ecotoxicity

Workflow components



Criteria for acceptance of **prediction for Read across**

After sub-categorization by all PGPs and PS

IF *Interpolation* **AND** *LC50* ≤ 2 log units **OR** log *Kow* ≤ 2 log units **AND** *NA* ≥ 5 **THEN** accept prediction and proceed with Report

Definitions

Interpolation: log K_{OW} of the target should be within the range of log K_{OW} of analogues

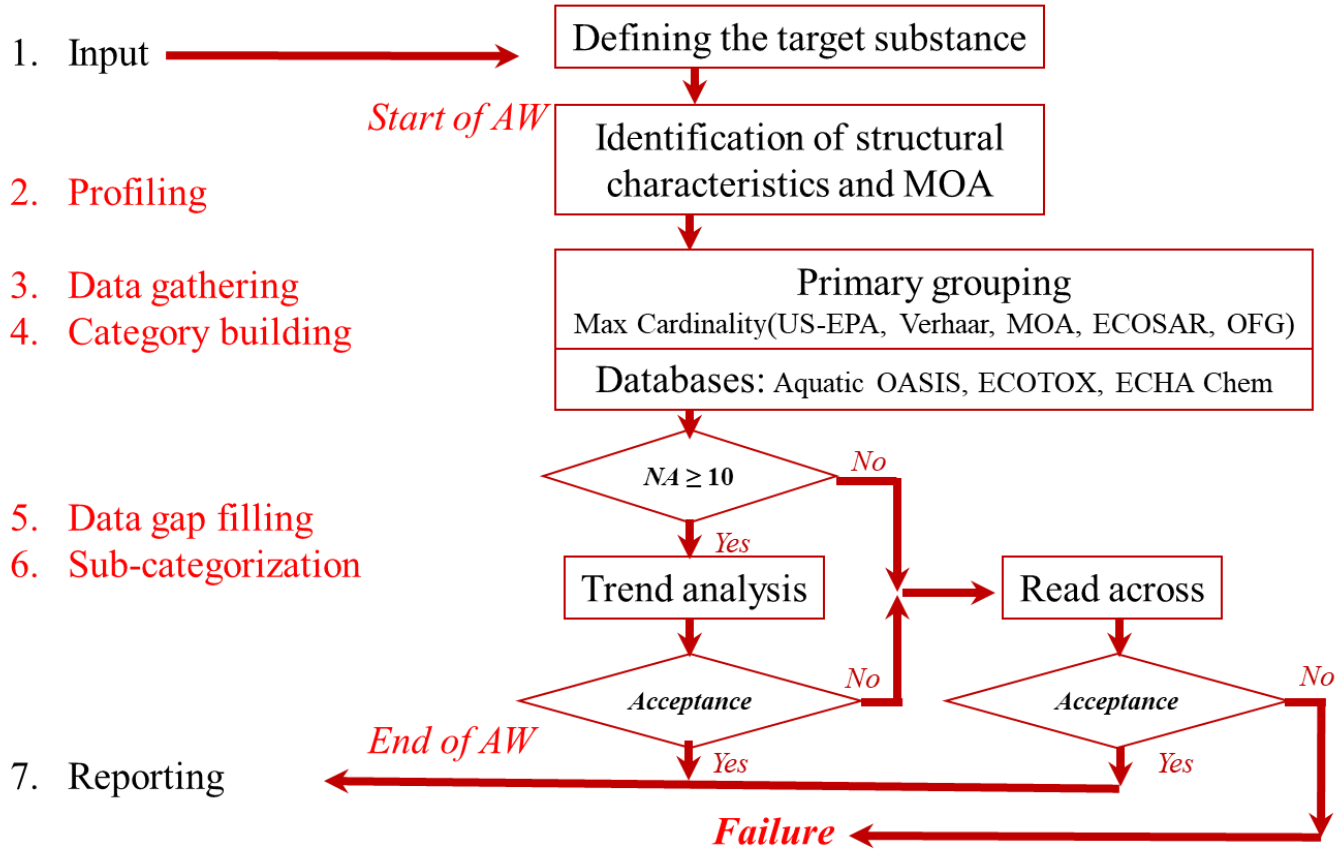
LC50 ≤ 2*:* for the 5 closest analogues the range of variation of *LC50* is ≤ 2 log units

log K_{OW} ≤ 2*:* for the 5 closest analogues the range of variation of log K_{OW} is ≤ 2 log units

NA – Number of analogues

Automated workflow (AW) for ecotoxicity

Workflow algorithm – illustration



Outlook

- Aim
- Automated workflow
- **The exercise**
- Report

The Exercise

- In this exercise we will predict the skin sensitization effect for:
  - **Single chemical – CAS# 111-86-4;**
  - Batch of chemicals

- This prediction will be accomplished by using of the automated workflow for skin sensitization.

Execution of AW for Ecotoxicity

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.

Execution of AW for Ecotoxicity

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's List/Inventory/Databases

II. Group of chemicals

- User's List
- Inventory/Database

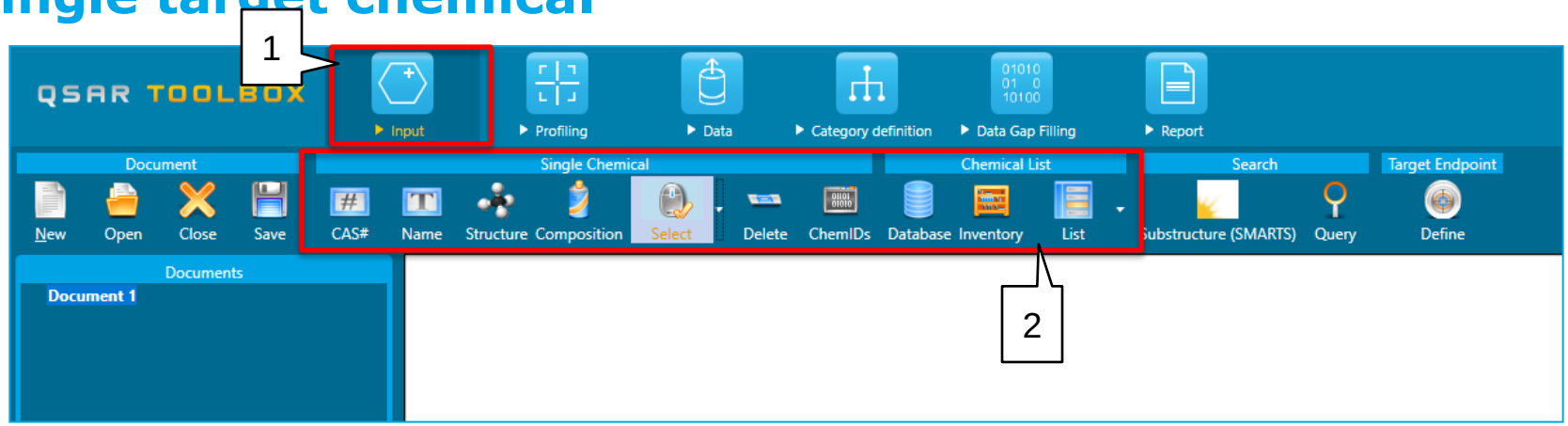
Execution of AW for Ecotoxicity

Chemical Input: Single target chemical

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

Execution of AW for Ecotoxicity

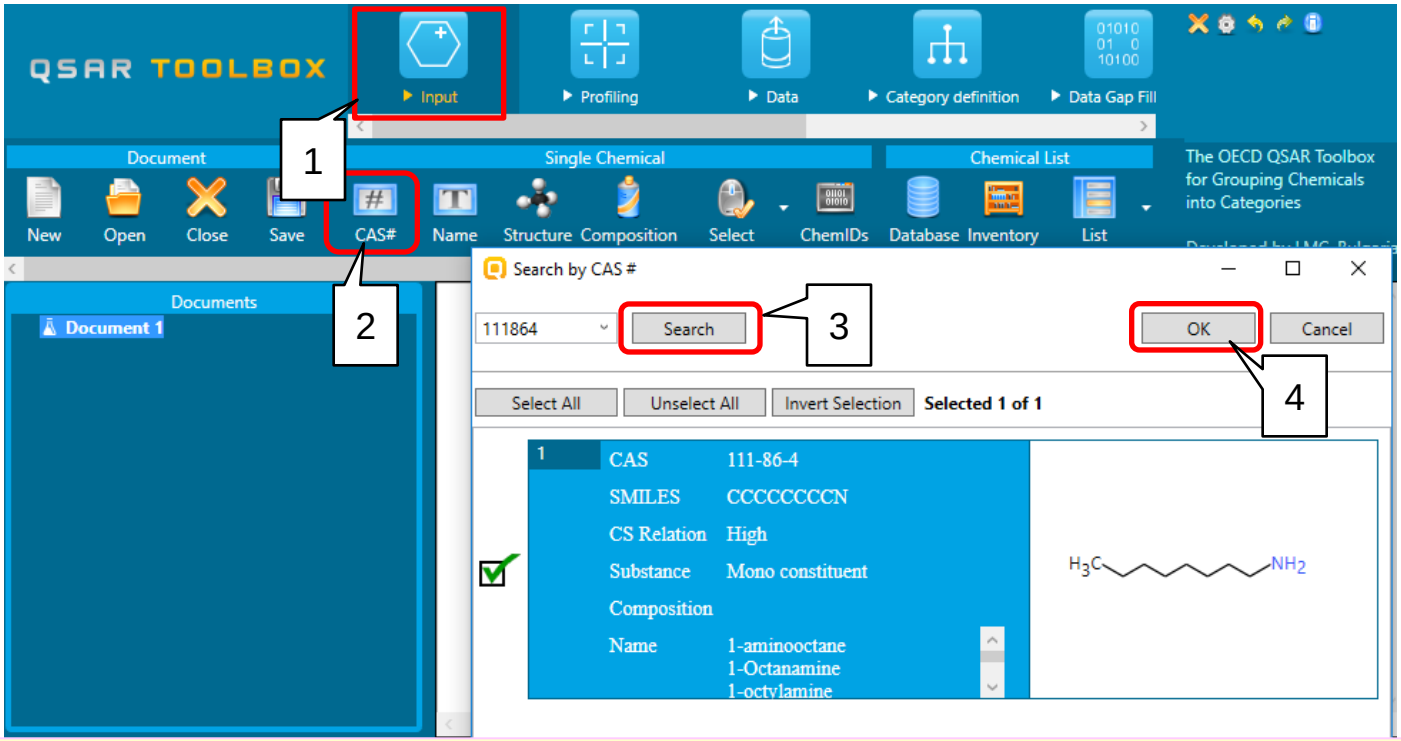
Chemical Input
Single target chemical



1. Click on *Input*
2. Main *Input* section

Execution of AW for Ecotoxicity

Input Single target chemical
by CAS RN



1. Go to *Input*; 2. Click and type CAS # 111864; 3. Press *Search*; 4. Confirm by *OK*.

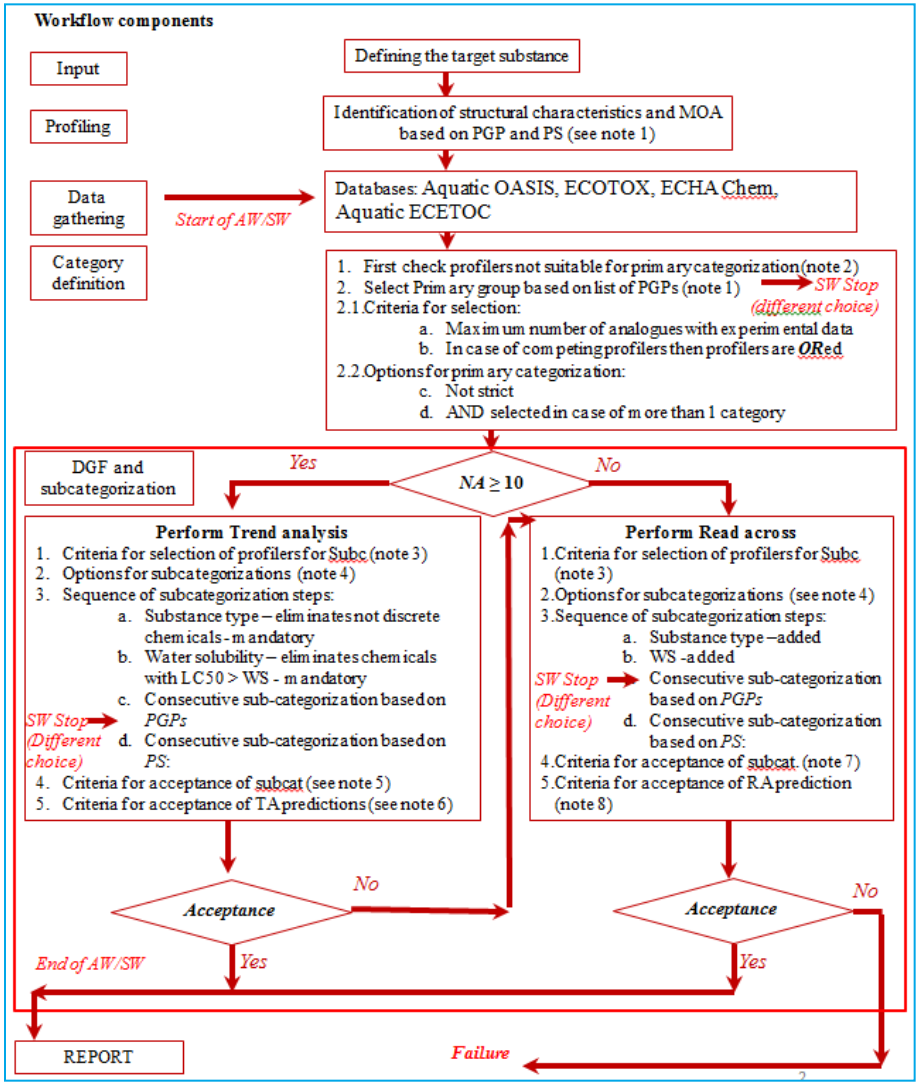
Execution of AW for Ecotoxicity

Data gap filling
Automated workflow: An overview

- Algorithms for automated and standardized data gap filling have been developed for skin sensitization (LLNA and GPMT data) and acute aquatic toxicity to fish (*Pimephales promelas*, Mortality, LC50, 96 h).
- Once started, the automated workflows (AWs) follow the implemented logic and finished with prediction without interaction by the user.
- In this tutorial only the acute aquatic toxicity to fish is discussed.
- The Automated workflow can be used for a single chemical mode or in a batch mode.

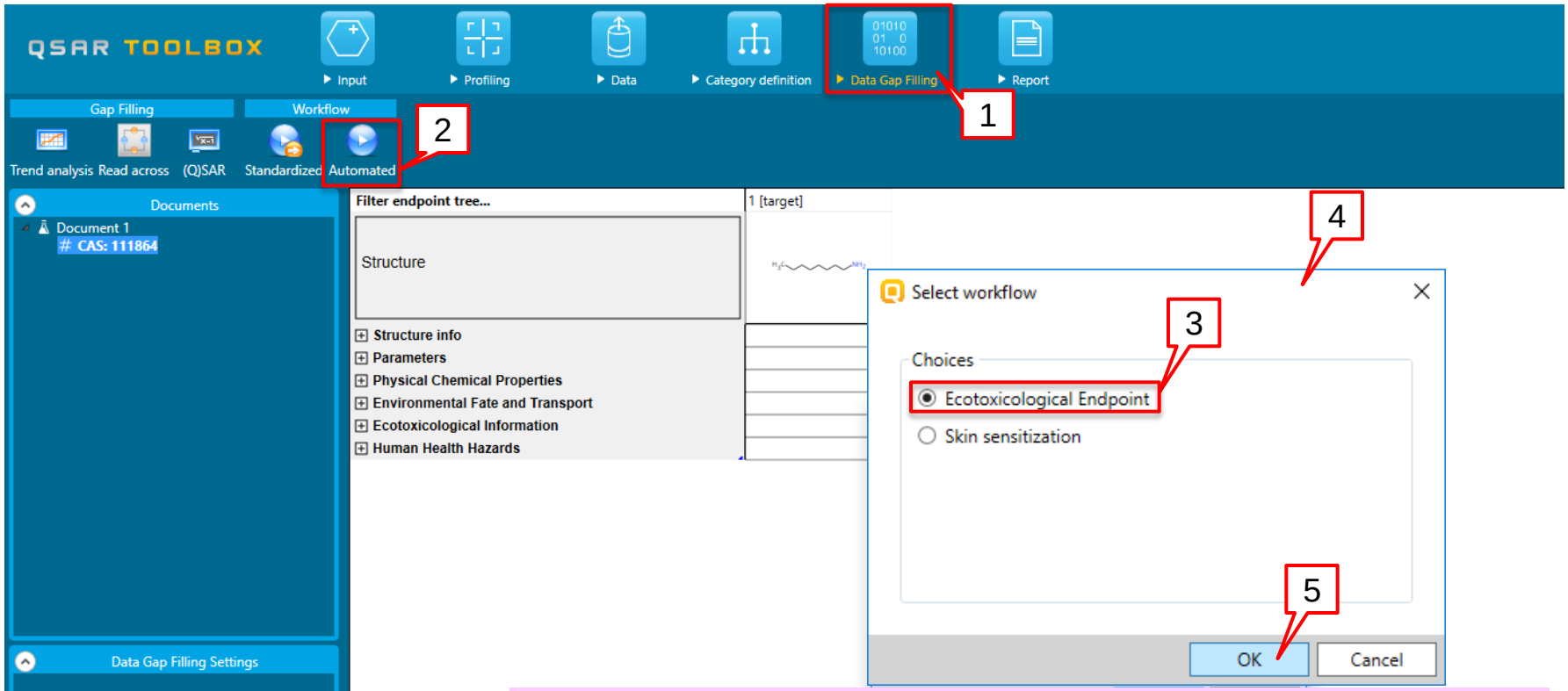
Execution of AW for Ecotoxicity

Algorithm of Ecotoxicological workflow



Execution of AW for Ecotoxicity

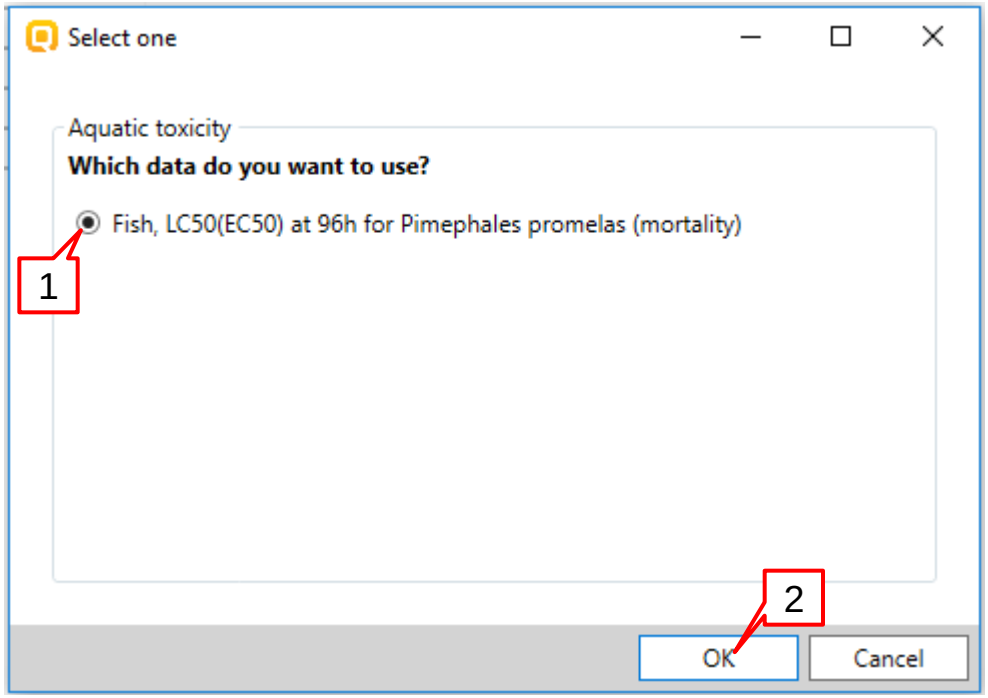
Location of the Automated workflow for Ecotoxicity



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

Execution of AW for Ecotoxicity

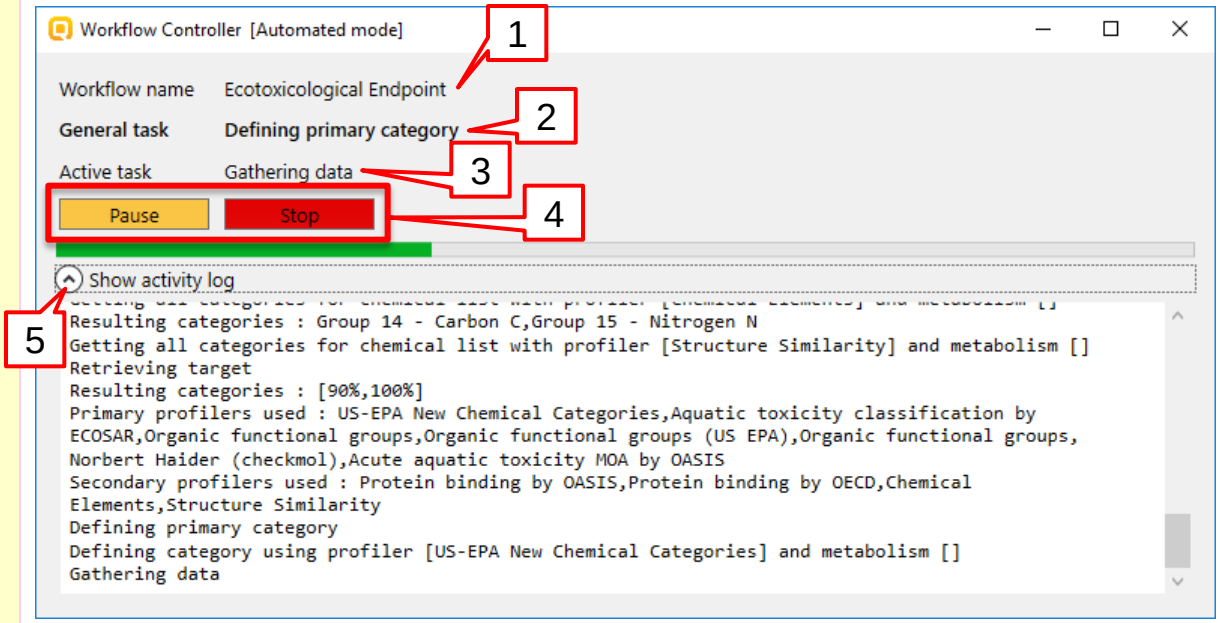
- A dialogue window gives the user a choice to select the end-point (1);
- Then select *OK* (2).



Execution of AW for Ecotoxicity

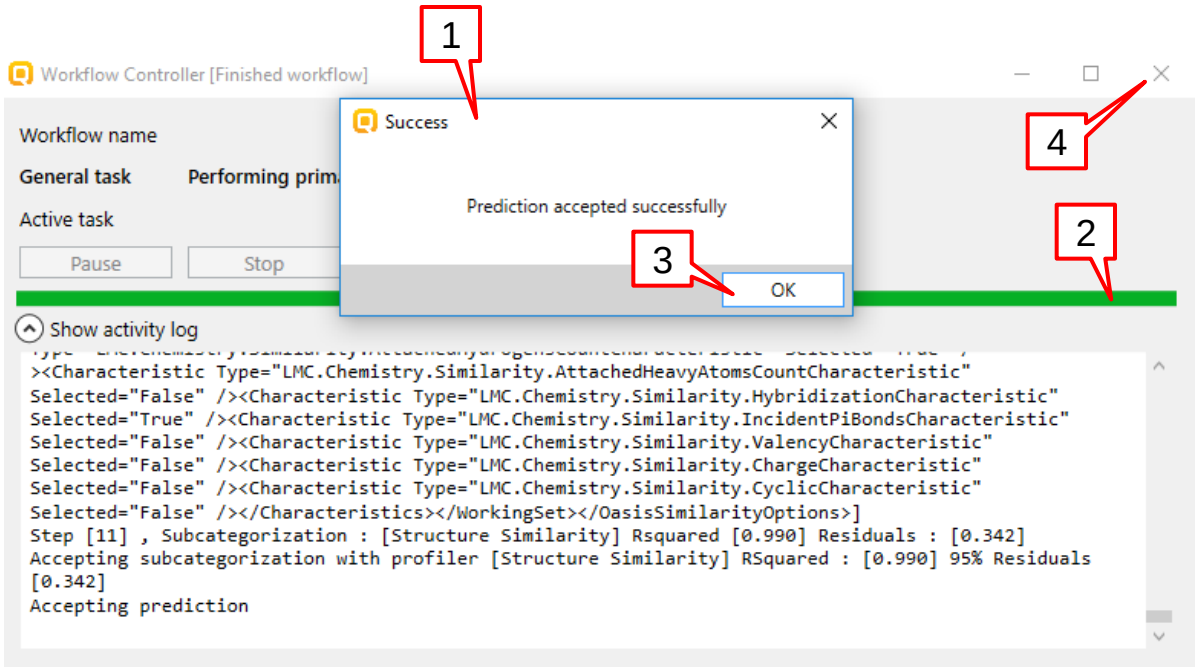
A workflow controller window is displayed throughout the automated workflow procedure. It includes:

- 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).



Execution of AW for Ecotoxicity

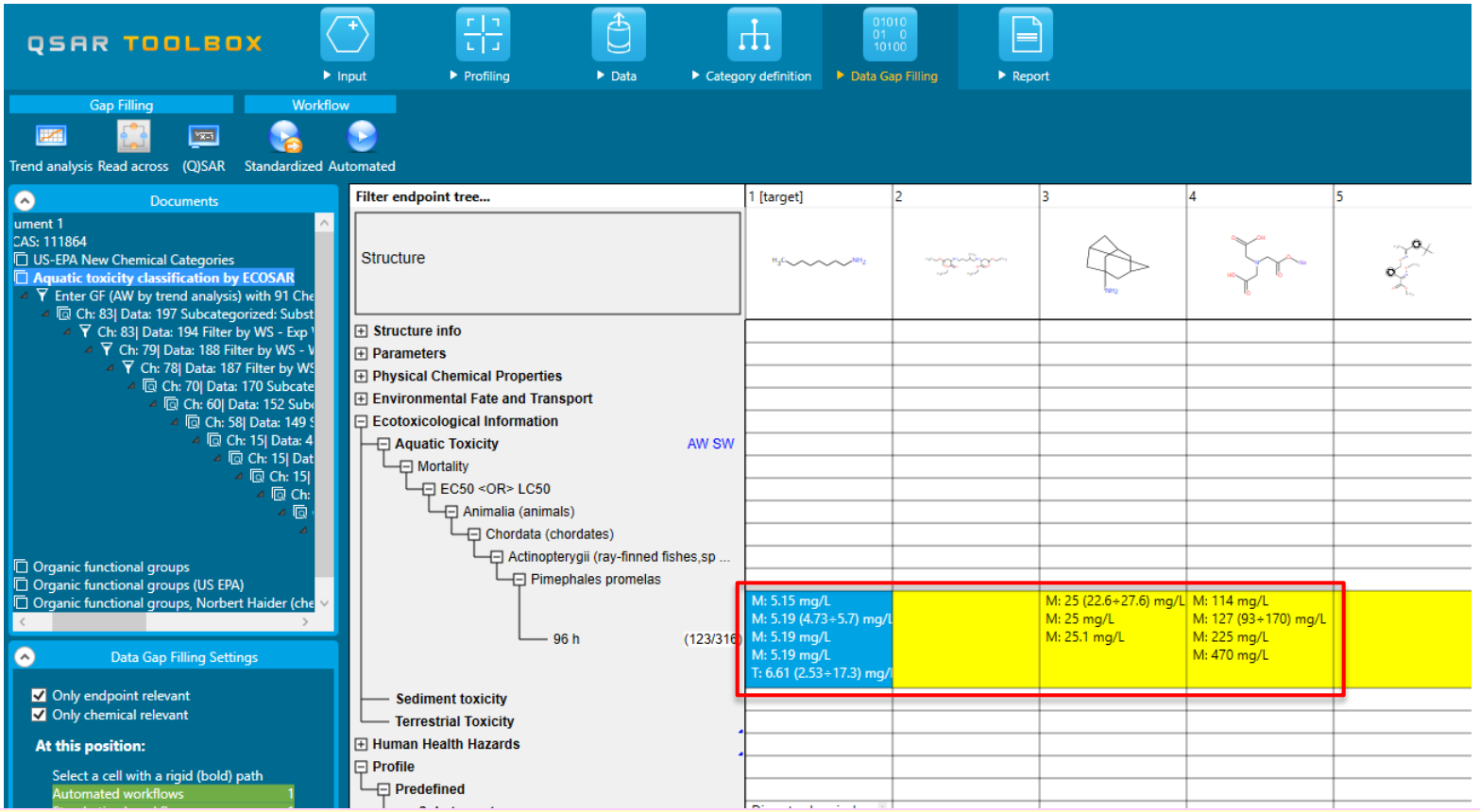
- When the workflow finishes a message is displayed that the prediction is accepted (1).
- Also the progress bar is completely filled (2).
- The user has to press *OK* (3).
- Then close the workflow window by pressing **X** button (4).



Execution of AW for Ecotoxicity

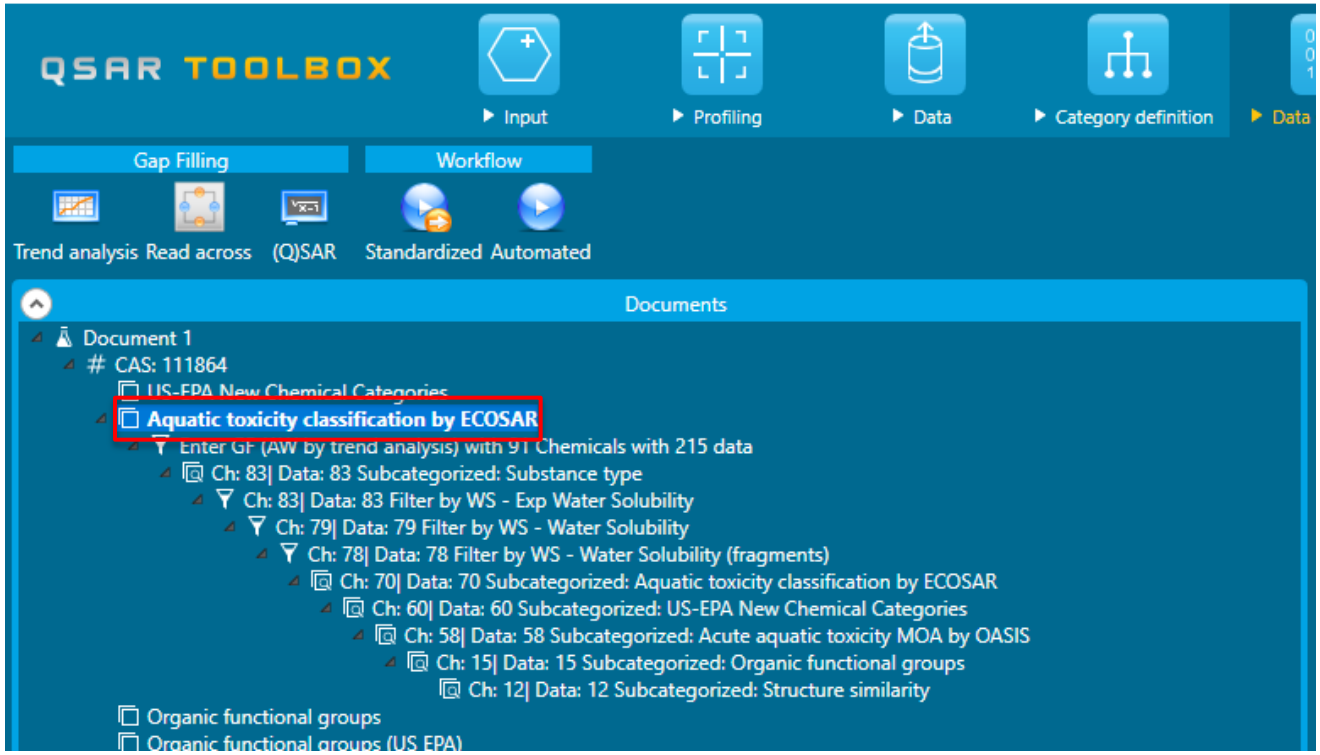
Data Gap Filling

Automated workflow: Single chemical



- The prediction is displayed on the matrix labeled with "T", which stands for trend analysis;
- "M" stands for measured data.

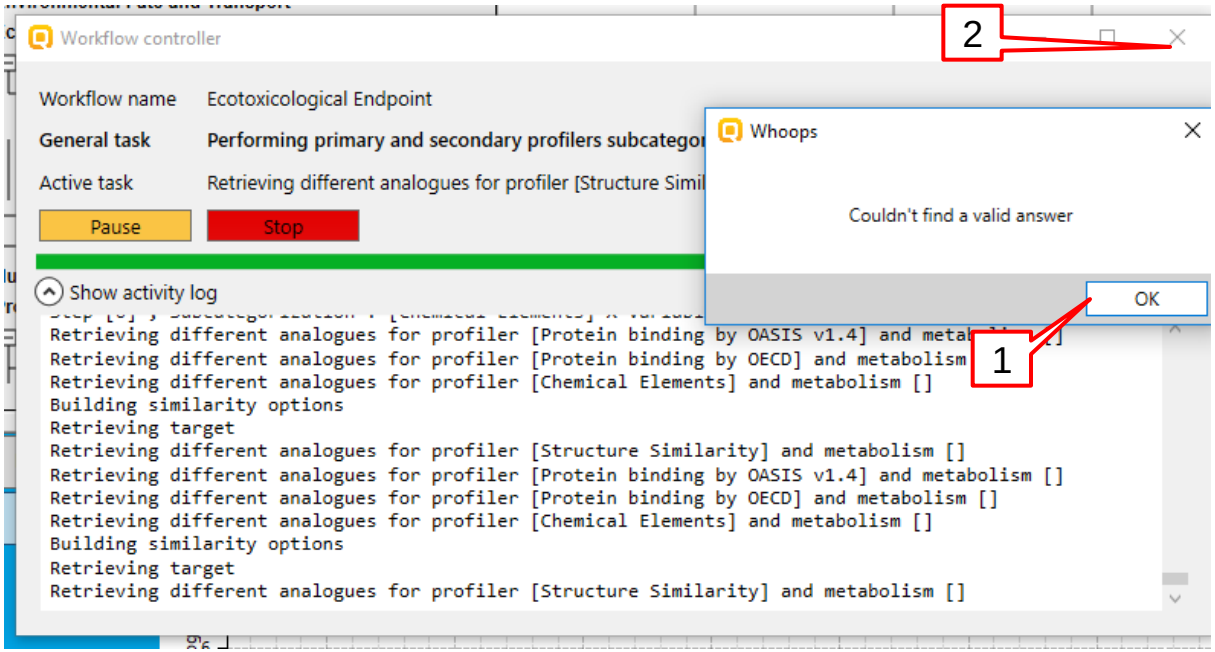
Execution of AW for Ecotoxicity



- The steps executed in the AW are listed in the Documents panel;
- AW always finishes at the level of primary grouping.

Execution of AW for Ecotoxicity

In case the prediction does not answer the criteria for acceptance of the prediction (1) or not enough data is collected for primary grouping then the corresponding messages appears, such as: "No enough data to build primary group" or Couldn't find a valid answer"(1)



- Click *OK* (1);
- Then close the workflow controller window (2)

The Exercise

- In this exercise we will predict the skin sensitization effect for:
  - Single chemical – CAS# 111-86-4;
  - **Batch of chemicals**

- This prediction will be accomplished by using of the automated workflow for skin sensitization.

Execution of AW for Ecotoxicity

Data Gap Filling

Automated workflow: Batch mode

There are several ways to load a batch of chemicals amongst which:

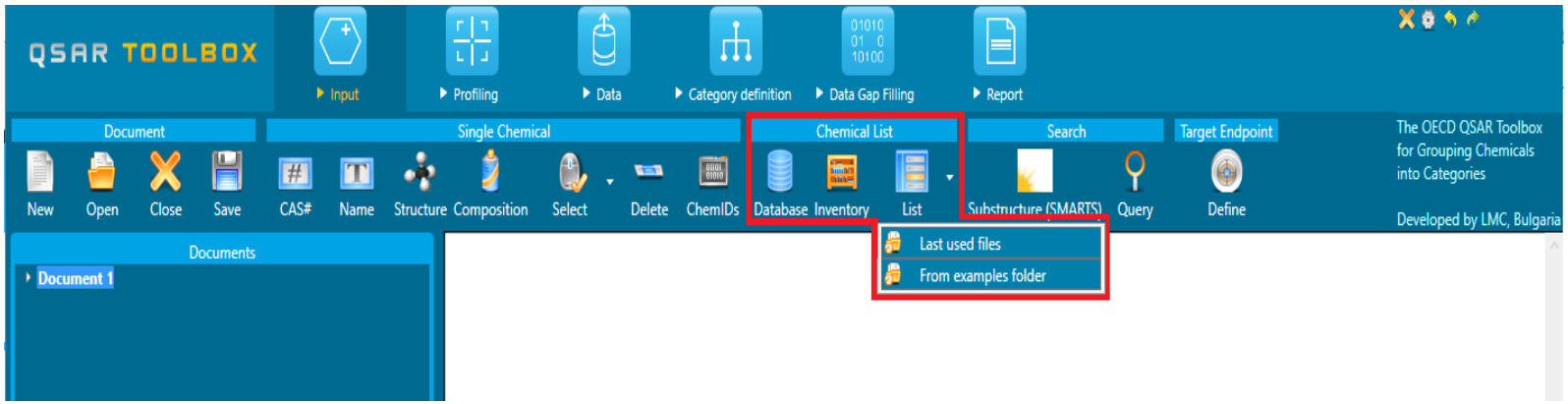
- Selection of chemicals from databases/inventories;
- Loading of chemicals from user's file.

(see next slide)

Execution of AW for Ecotoxicity

List with chemicals (batch work)

Input: Ways of Entering a Chemical List



- Database
- Inventory
- List: Last used files/ From examples folder

Execution of AW for Ecotoxicity

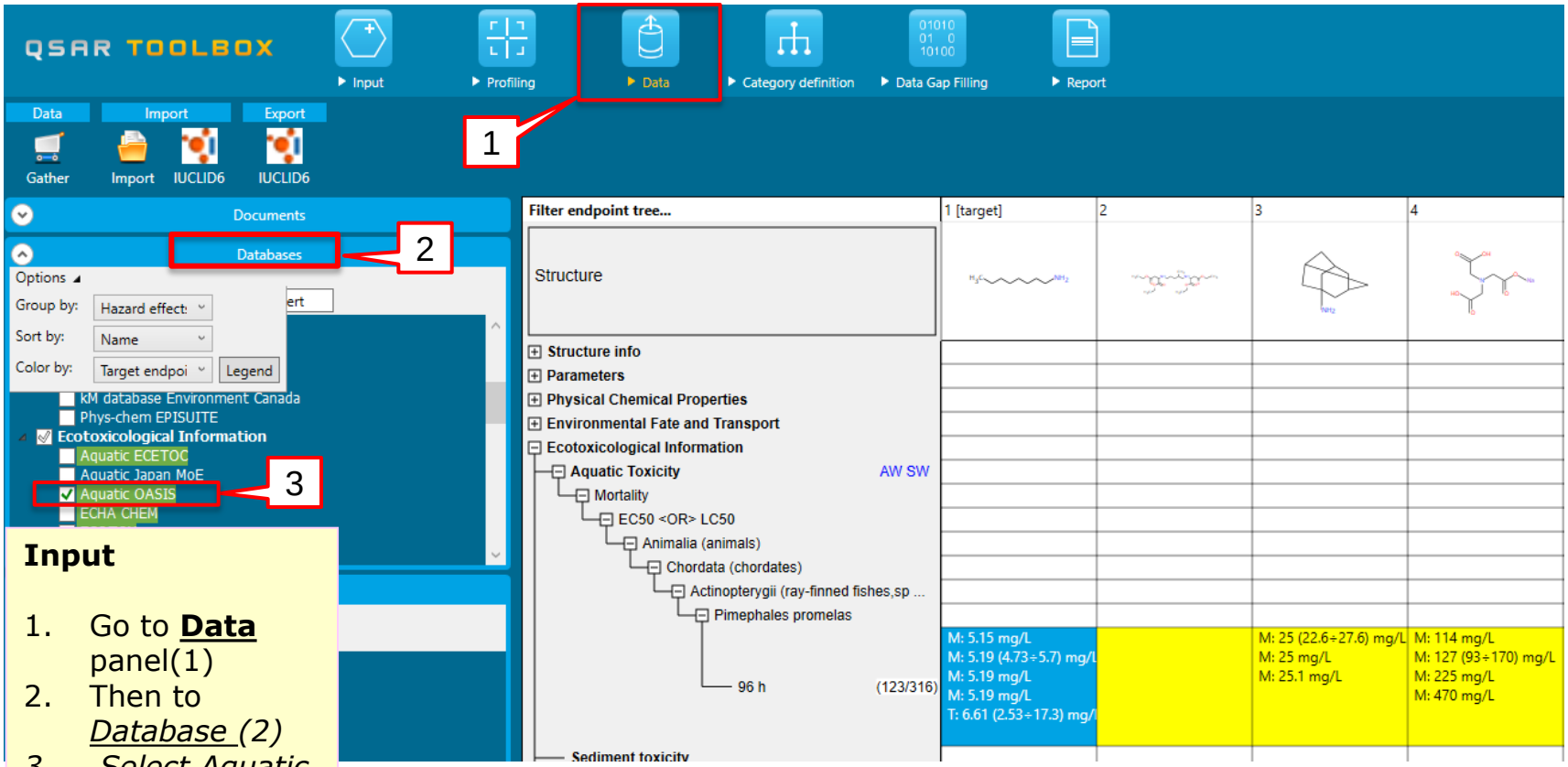
Data Gap Filling
Automated workflow: Batch mode

In this tutorial, *Query tool* functionality is used to load chemicals with known CAS RNs.

Execution of AW for Ecotoxicity

Data Gap Filling

Automated workflow: Batch mode



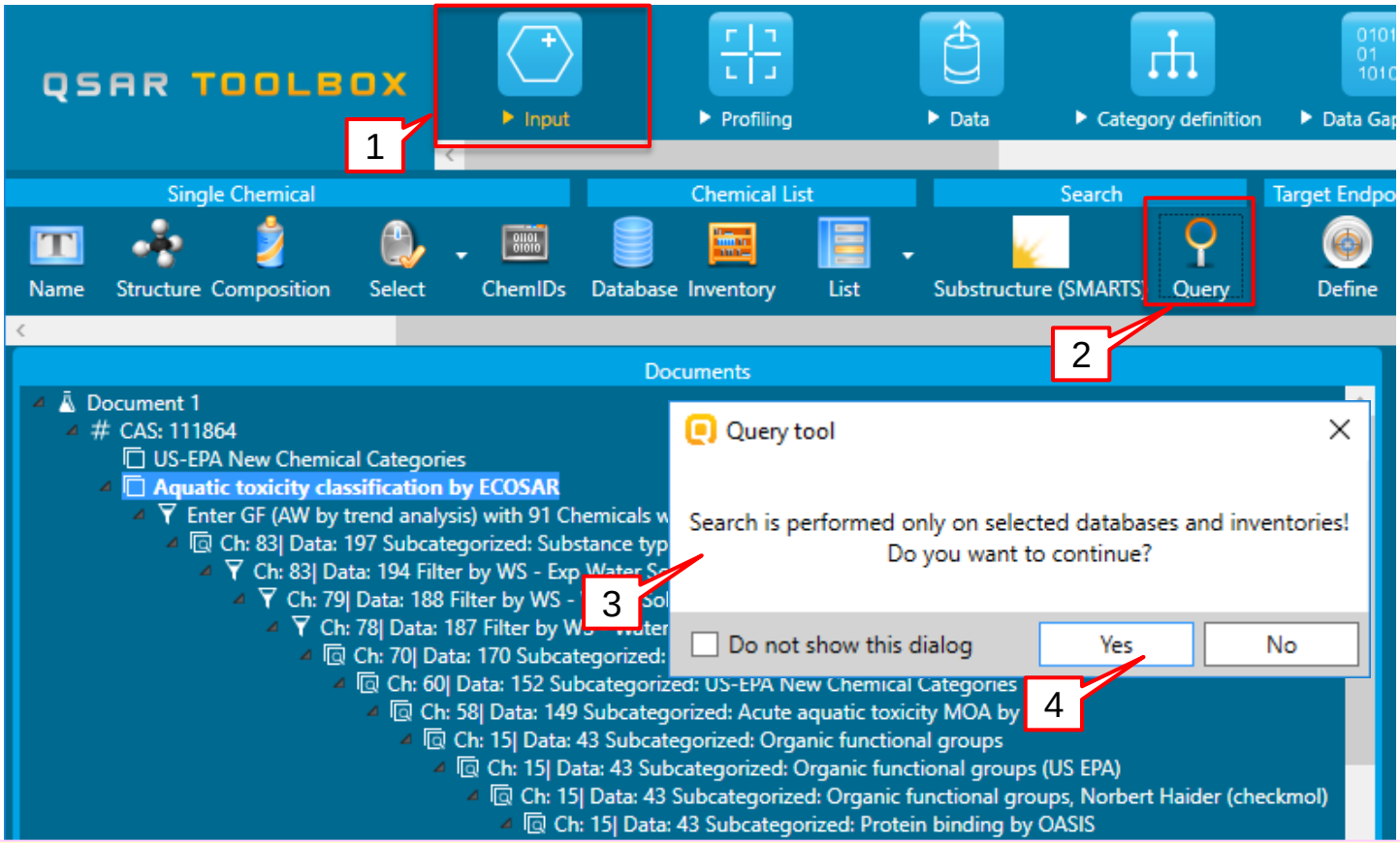
Input

1. Go to **Data** panel(1)
2. Then to *Database (2)*
3. *Select Aquatic OASIS database(3)*

Execution of AW for Ecotoxicity

Data Gap Filling

Automated workflow: Batch mode

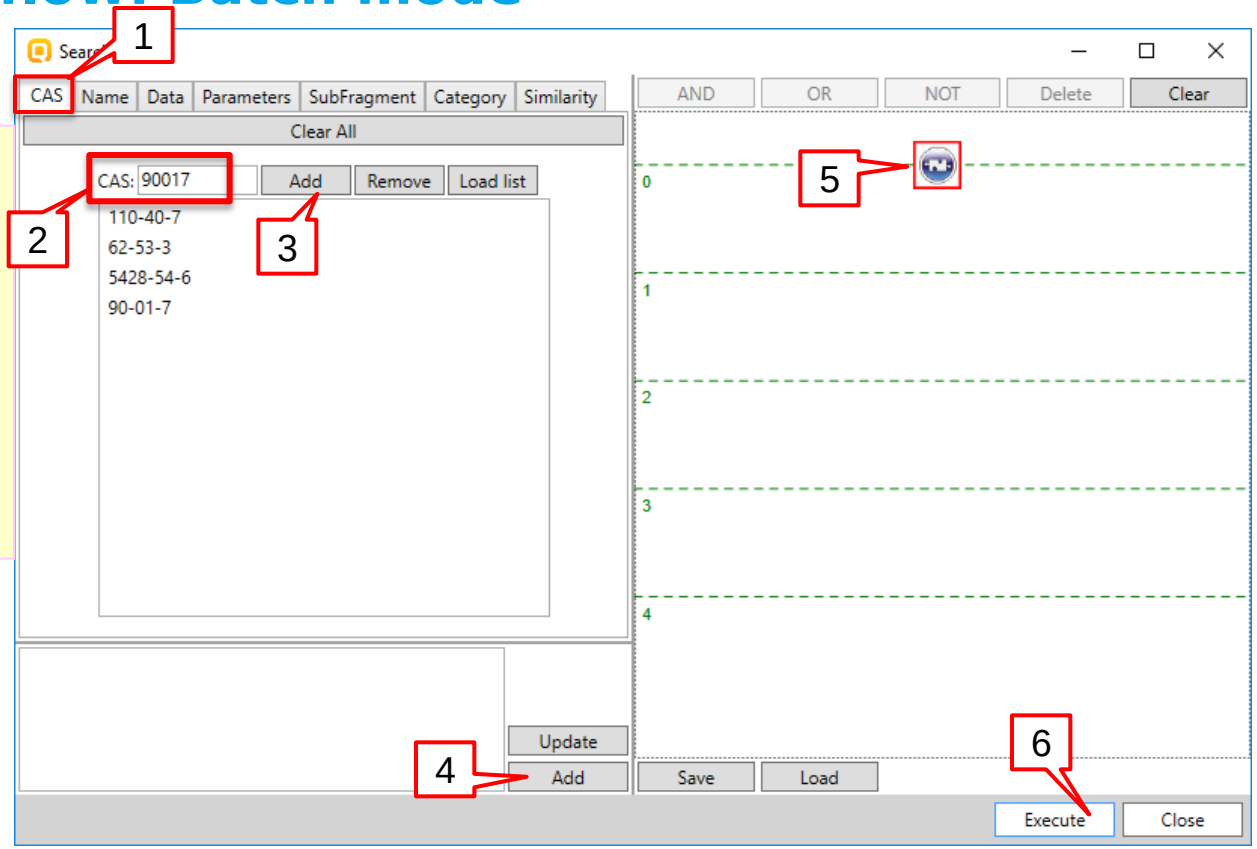


1.Go to *Input* panel; 2.Click on *Query;* 3. A dialogue window pops up; 4. Click on *Yes.*

Execution of AW for Ecotoxicity

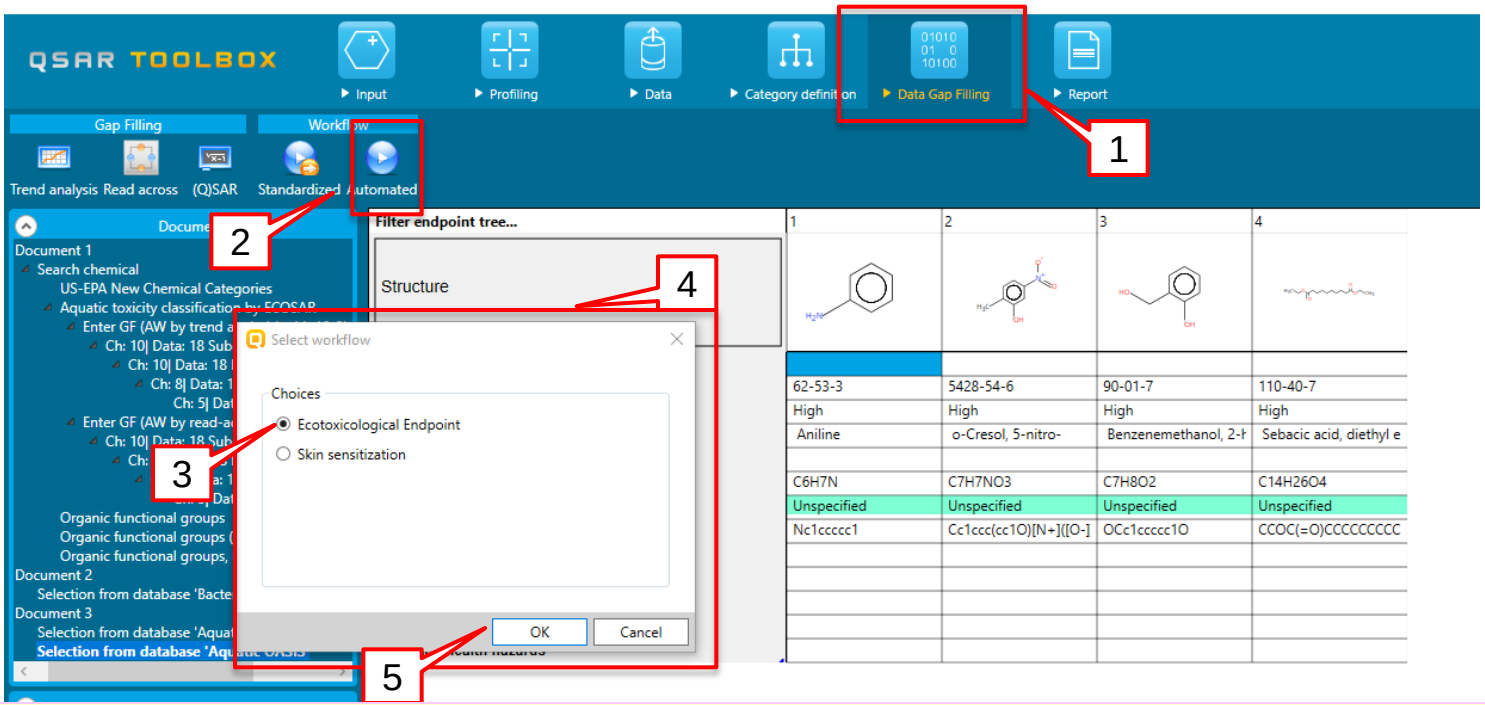
Data Gap Filling
Automated workflow: Batch mode

1. Go to *CAS* tab(1);
2. Type in the CAS RN in the cell (2)
3. Click on *Add (3)* to add all 4 CAS RNs.
4. Click on *Add* (4)to create the query (5) ;
5. Click on *Execute* (6).



Execution of AW for Ecotoxicity

Data Gap Filling
Automated workflow: Batch mode

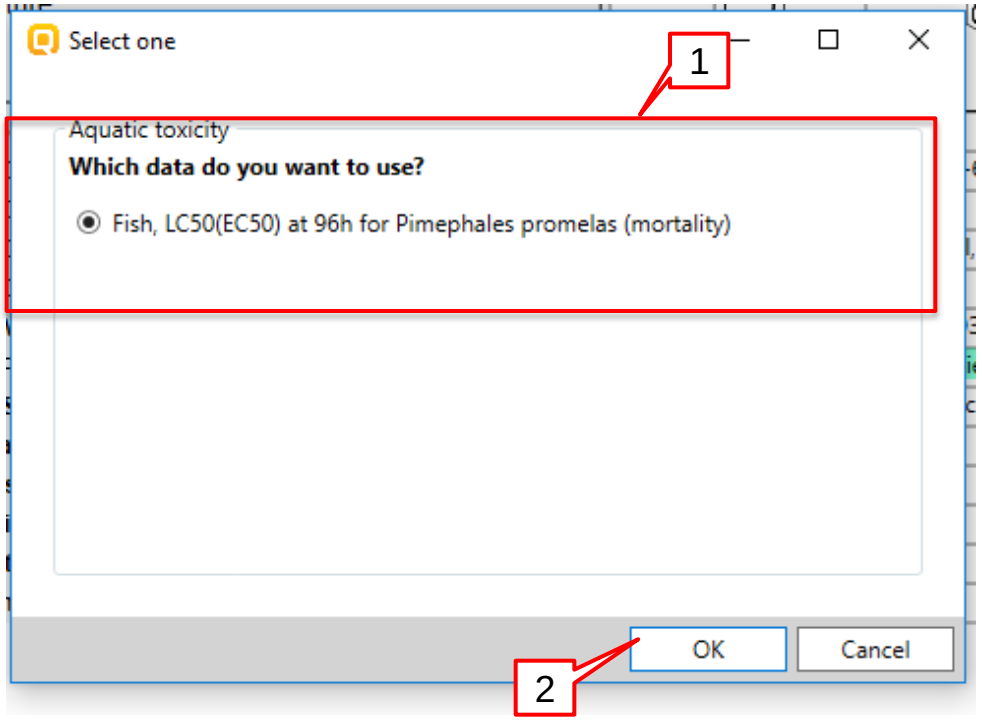


1. Select *Data gap filling* tab (1);
2. Click on *Automated* (2);
3. Select *Ecotoxicological endpoint* (3) from the pop-up window (4);
4. Click on *OK* (5).

Execution of AW for Ecotoxicity

Data Gap Filling
Automated workflow: Batch mode

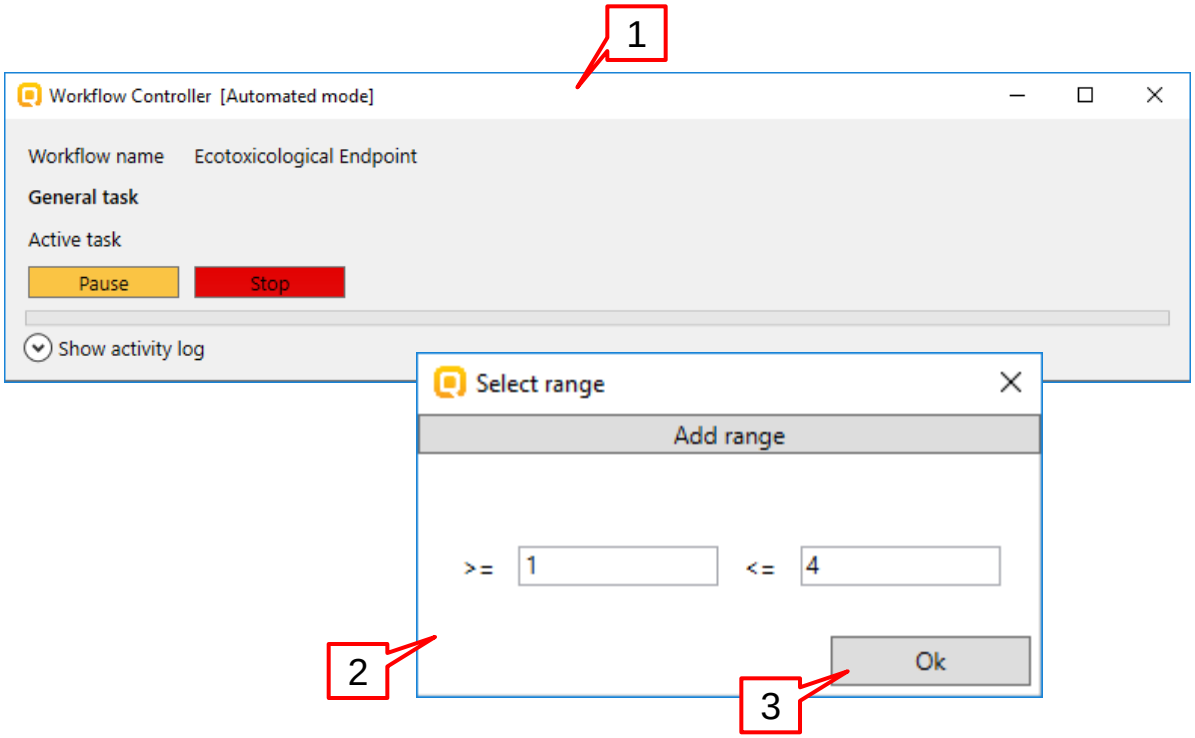
- A dialogue window gives the user a choice to select the end-point (1);
- Then select *OK* (2).



Execution of AW for Ecotoxicity

Data Gap Filling
Automated workflow: Batch mode

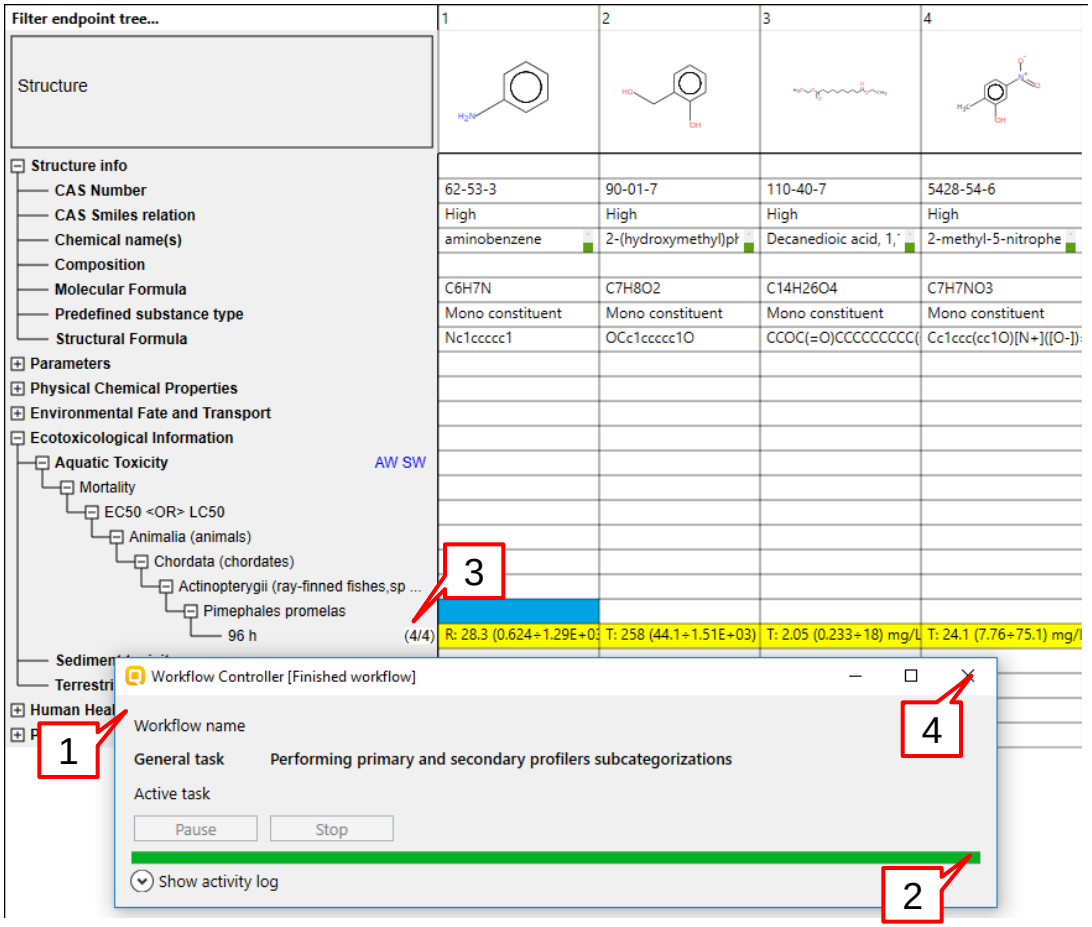
- A *Workflow controller* window appears, which is not active(1);
- The pop-window *Select range (2)* is displayed where the user has to select the range of chemicals from the set, which has to be predicted;
- Finally press *OK* (3).



Execution of AW for Ecotoxicity

Data Gap Filling
Automated workflow: Batch mode

- When the workflow finishes, there is an indication in the workflow controller(1).
- Also the progress bar is completely filled (2).
- The predictions are displayed on the matrix(3). There is also an indication that 4 out of 4 chemicals are predicted
- Finally close the workflow window by pressing **X** button (4).



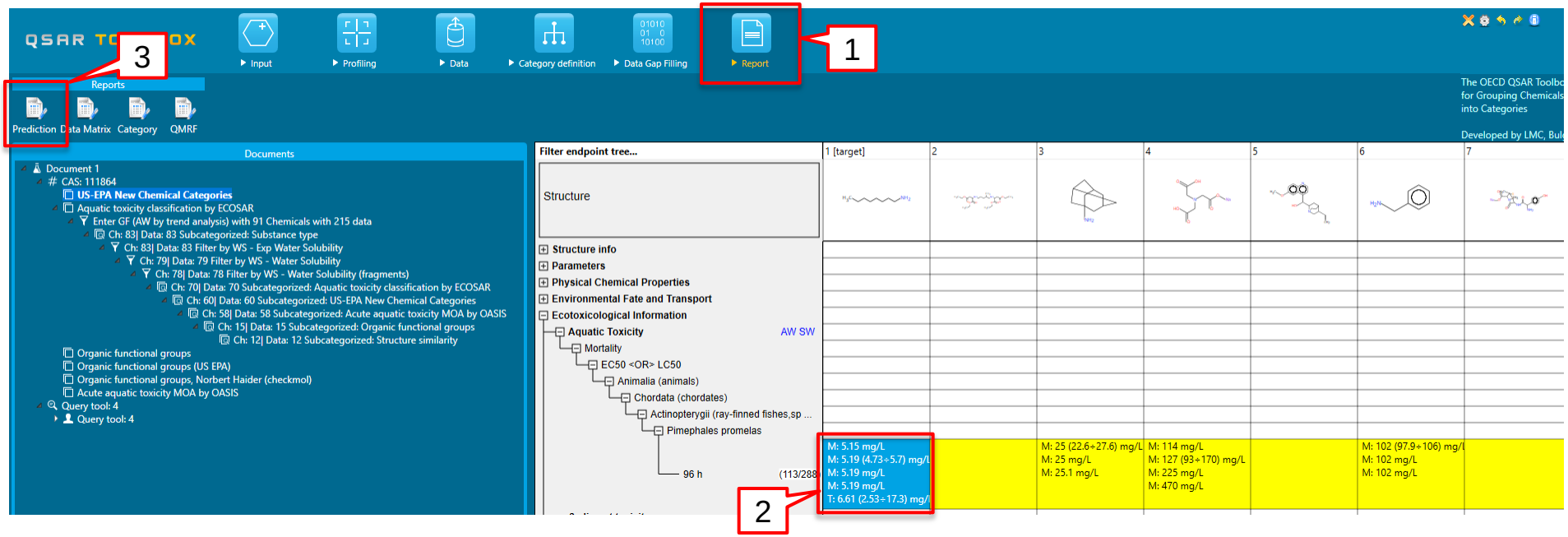
Outlook

- Aim
- Automated workflow
- The exercise
- **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

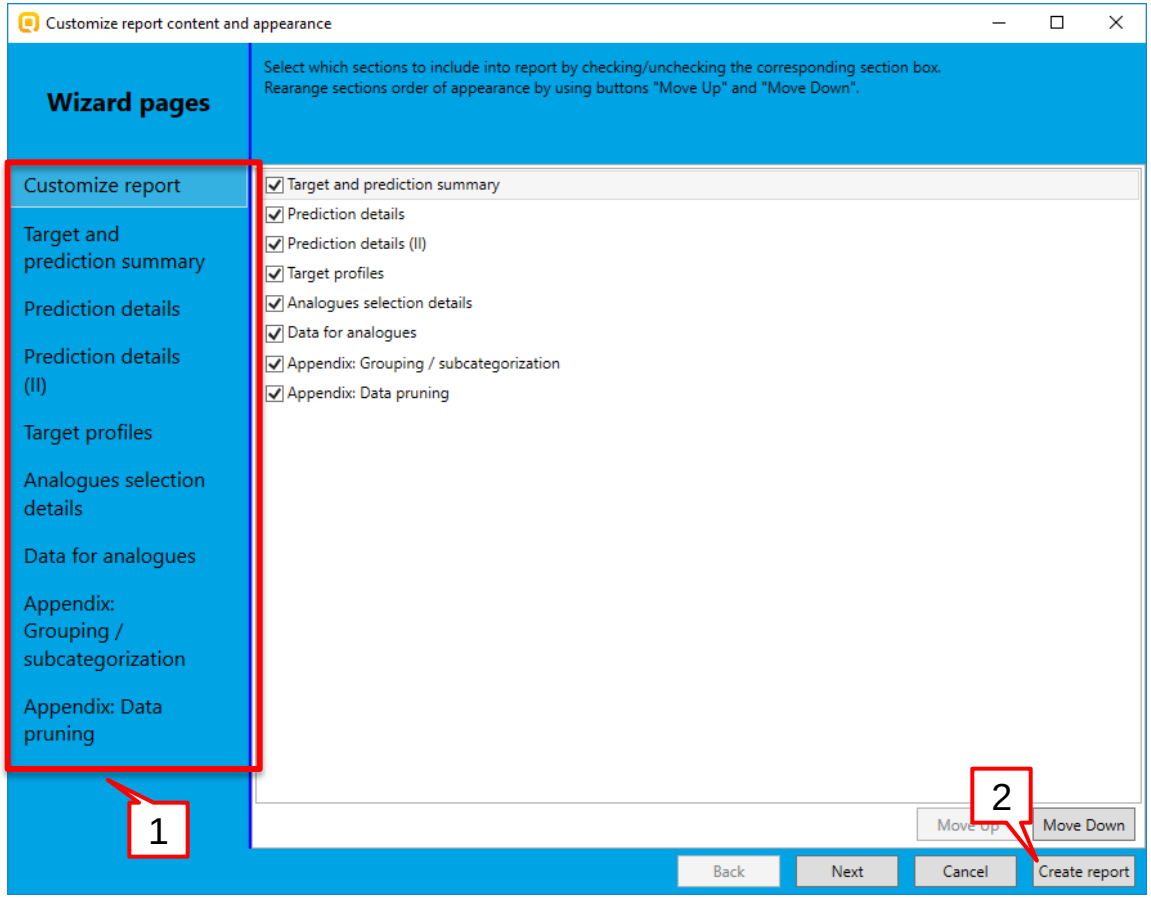


(1) Go to *Report* section;(2) Select the cell with the prediction; (3)Click on *Prediction.*

Report
Generation report

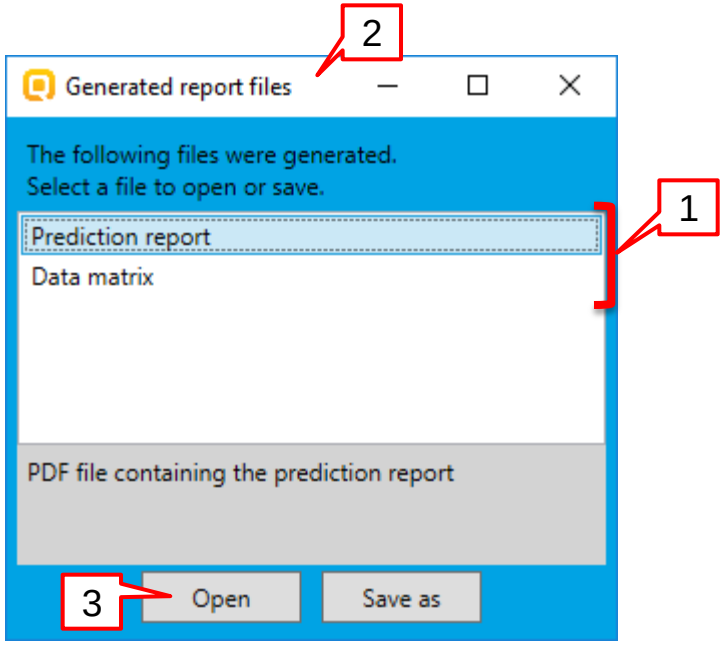
- Select different levels to customize the information that is going to be shown in the report (1)

- Select *Create report* (2) to display the report



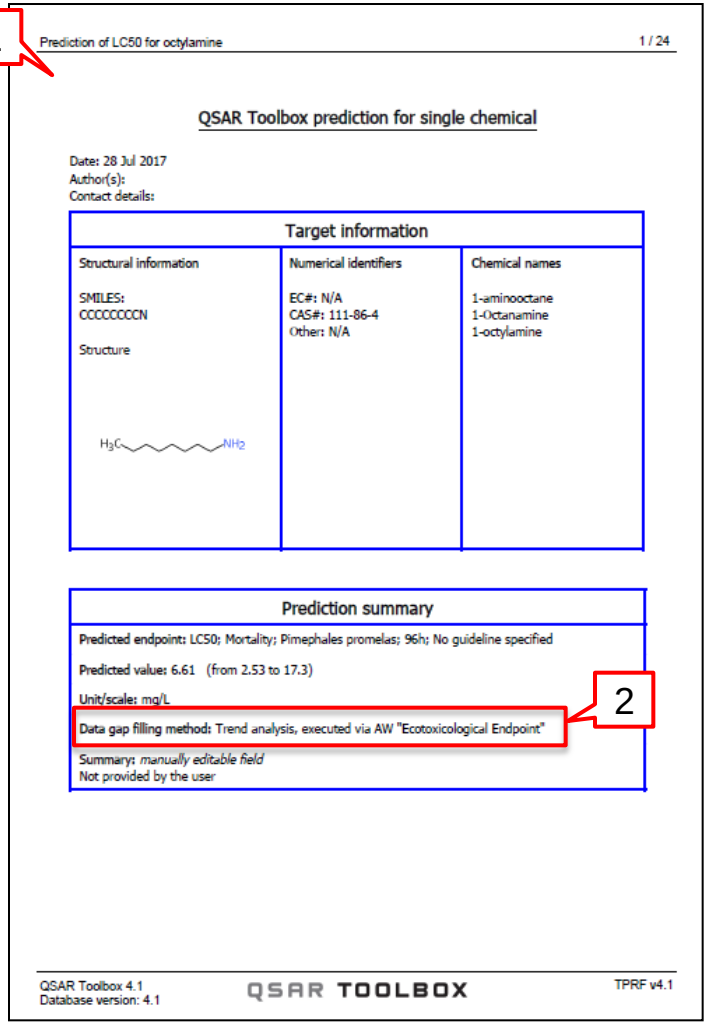
Report
Generation report

Two files (1) are generated, which can be selected from the Generated report files window (2) by clicking *Open (3);*



Report
Overview

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


1

Prediction of LC50 for octylamine 1 / 24

QSAR Toolbox prediction for single chemical

Date: 28 Jul 2017
Author(s):
Contact details:

Target information

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

Prediction summary

Predicted endpoint: LC50; Mortality; Pimephales promelas; 96h; No guideline specified

Predicted value: 6.61 (from 2.53 to 17.3)

Unit/scale: mg/L

Data gap filling method: Trend analysis, executed via AW "Ecotoxicological Endpoint"

2

Summary: *manually editable field*
Not provided by the user

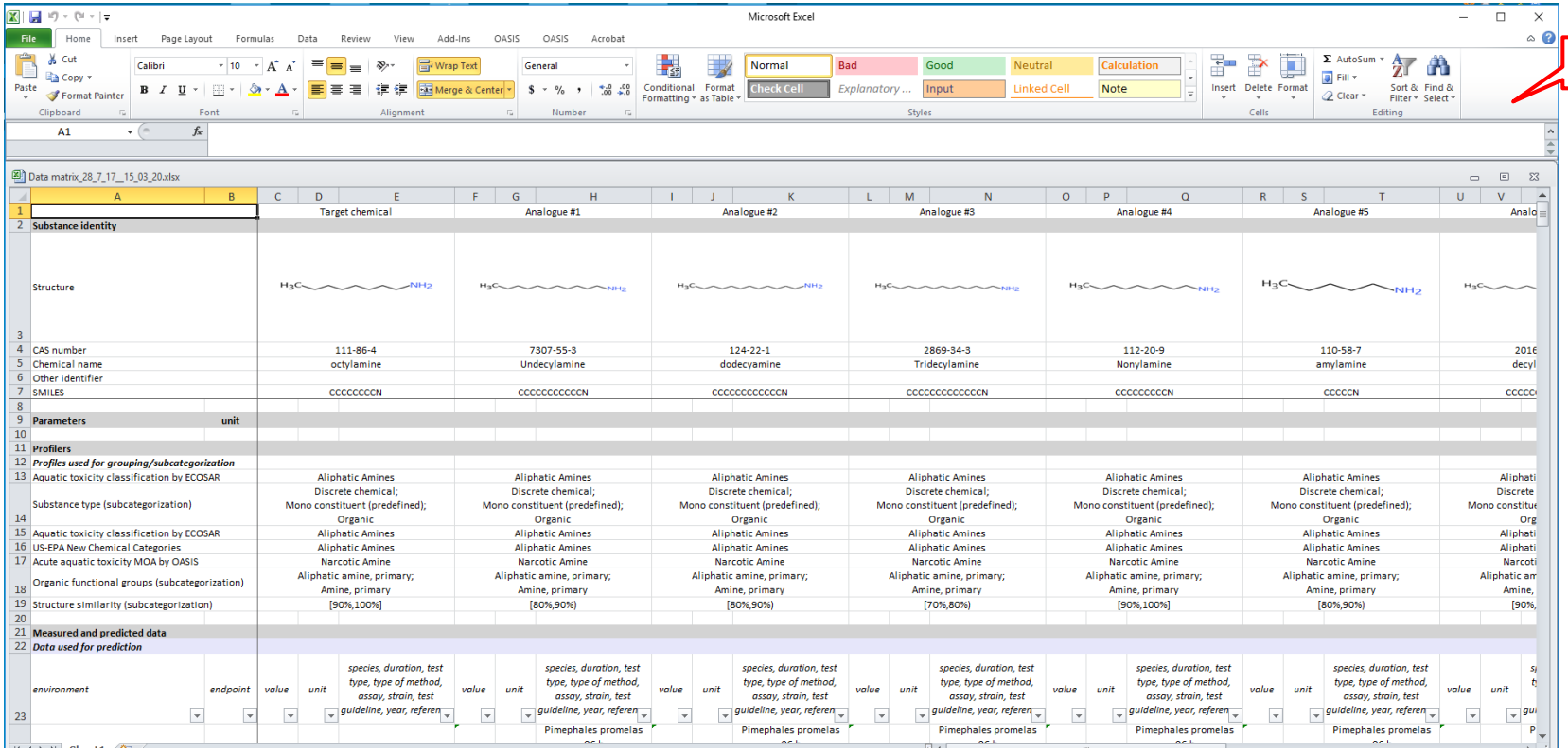
QSAR Toolbox 4.1
Database version: 4.1

QSAR TOOLBOX

TPRF v4.1

Report

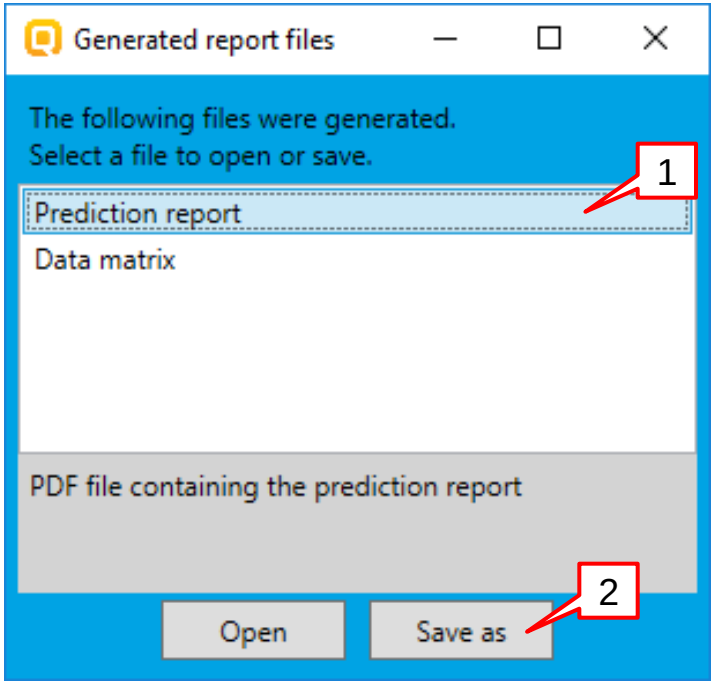
Overview



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

Report
Saving the prediction

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



Report
Saving the prediction

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