QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD (Q)SAR Toolbox v.4.4.1

Tutorial on how to predict ecotoxicological endpoint by automated workflow

Outlook

- **Background**
- Keywords
- Objectives
- Automated workflow
- The exercise
- Report

Background

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

Outlook

- Background
- **Keywords**
- Objectives
- Automated workflow
- The exercise
- Report

Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options

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**
- Automated workflow
- The exercise
- Report

Objectives

This presentation demonstrates how to:

- Fill data gaps for a single chemical or batch of chemicals by automated workflow for ecotoxicological endpoint

Outlook

- Background
- Keywords
- Objectives
- **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.

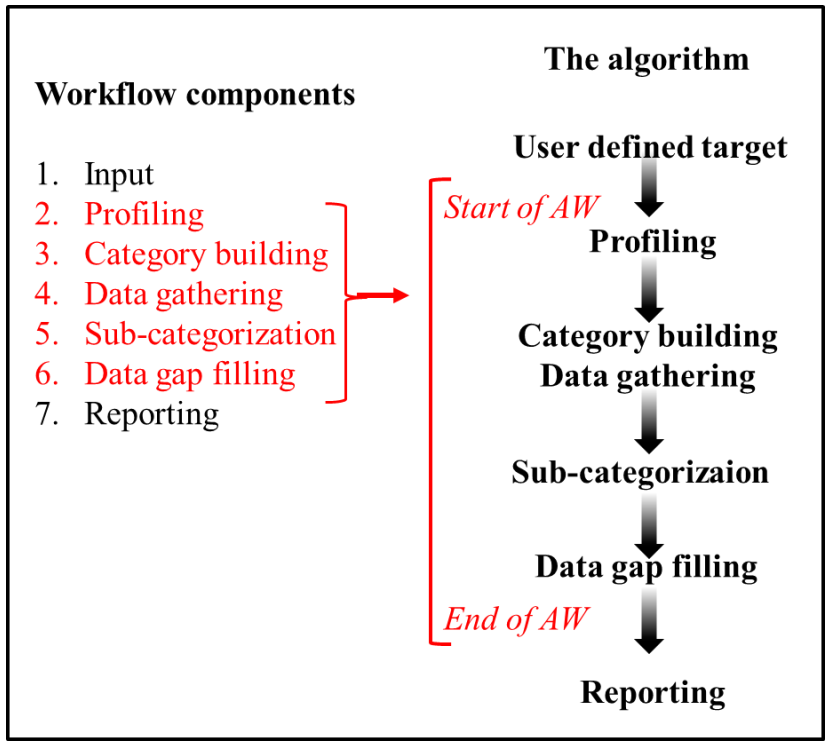
Ecotoxicological endpoint is:

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

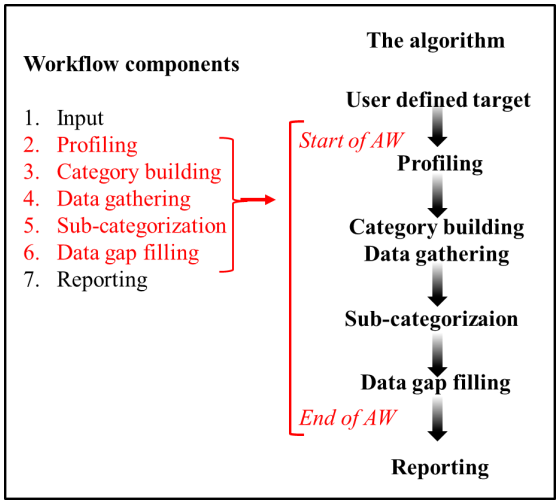
Automated workflow (AW) for ecotoxicity

Workflow components:

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



Automated workflow (AW) for ecotoxicity



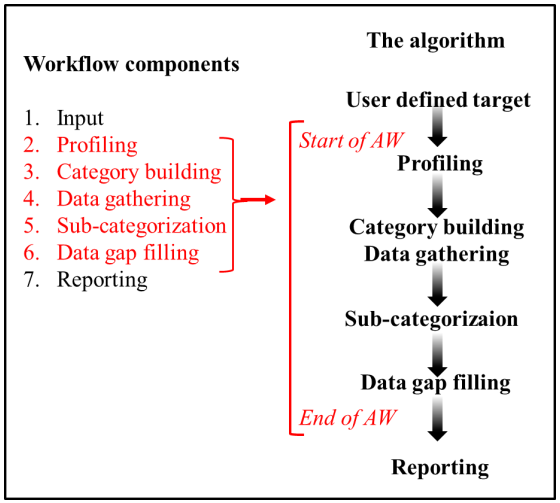
Workflow components: Profiling

- 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%)

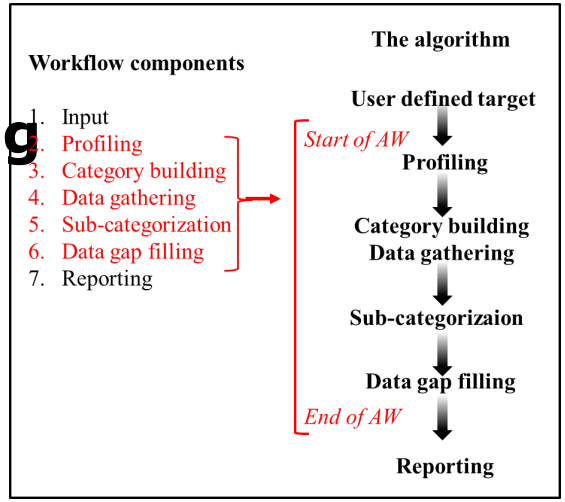
Automated workflow (AW) for ecotoxicity

Workflow components: Data

- Data used in this workflow is for acute aquatic toxicity endpoint (e.g. LC50)
- Databases used in AW are:
  - Aquatic OASIS
  - ECOTOX
  - ECHA REACH
- Endpoint data for LC50 used in AW:
  - *P. promelas* ~ 5400 data for ~ 2400 chemicals
  - *O. mykiss* ~ 6000 data for ~ 2000 chemicals



Automated workflow (AW) for ecotoxicity



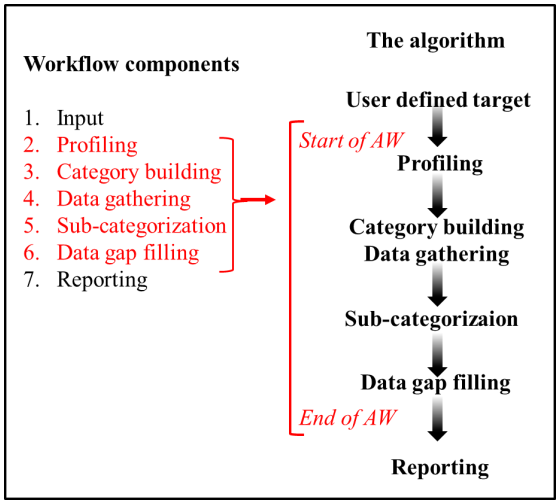
Workflow components: Category building

- 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

- The most populated group with data based on the above mentioned profilers is used for data gap filling

Automated workflow (AW) for ecotoxicity

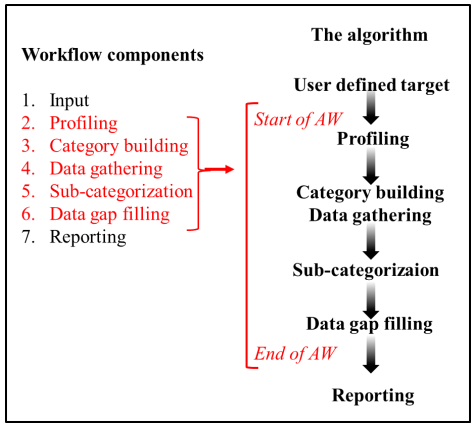
Workflow components: Data Gap Filling



- The aim is to fill the 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

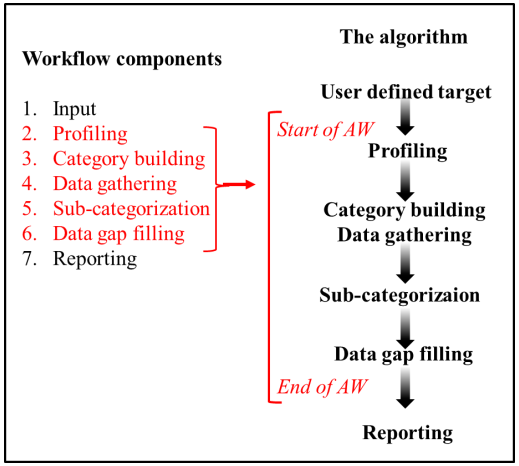
Sequence of subcategorization steps:

Mandatory:

1. *Substance type* – eliminates not discrete chemicals
2. *Water solubility* (WSKOWWIN + WATERNT) – eliminates chemicals with *LC50* > *WS*

Not mandatory:

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



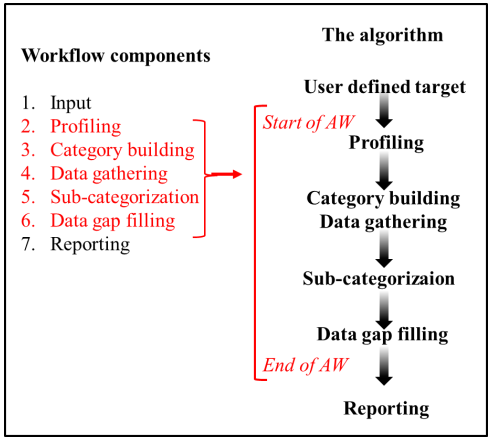
Automated workflow (AW) for ecotoxicity

Workflow components: Subcategorization

Criteria for acceptance of subcategorization step:

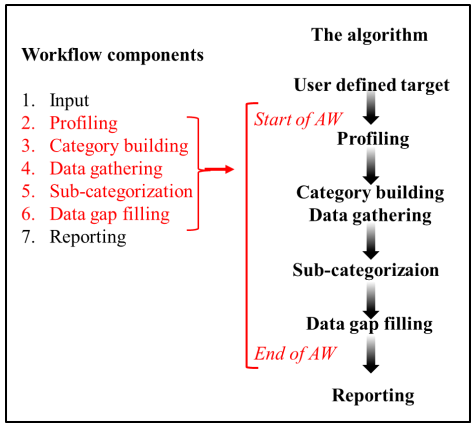
- Depends on the specific statistical and structural criteria (e.g, experimental error, 95% of residuals, log K_{OW}, range of variation of the analogues etc.)

- 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

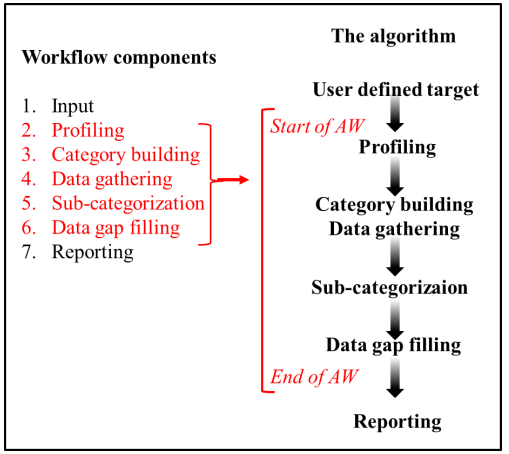
Correlation coefficient R^2 ↑: R^2 increases **OR** R^2 ≥ 0.85

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

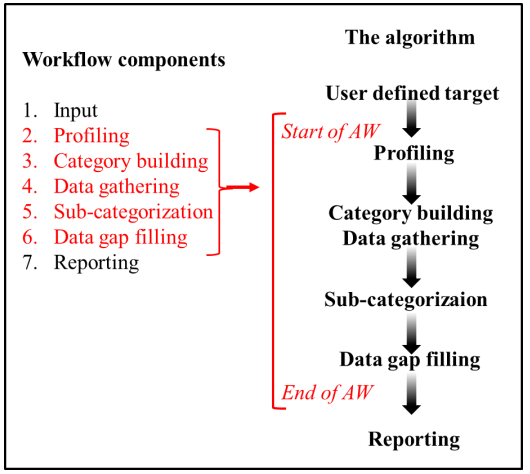
LC50 ↓*:* for the 5 closest analogues the range of variation of *LC50* decreases **OR** range of variation is ≤ 2 log units

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:**

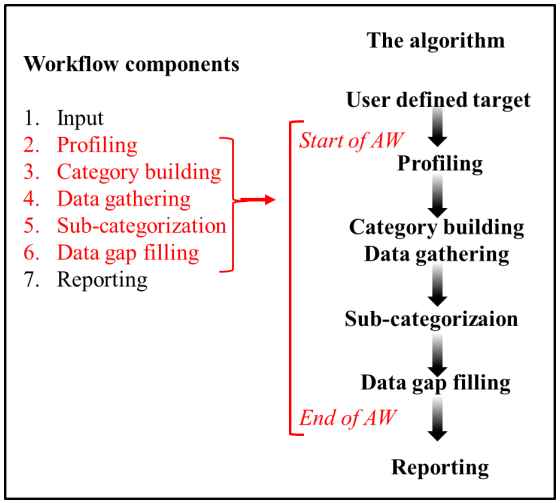
- After the subcategorization step is accepted then the prediction could be accepted if the criteria are met:

After sub-categorization by all PGPs and PS

IF $R^2 \geq 0.7$ **and** $NA \geq 10$
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 the subcategorization step is accepted then the prediction could be accepted if the criteria are met:

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

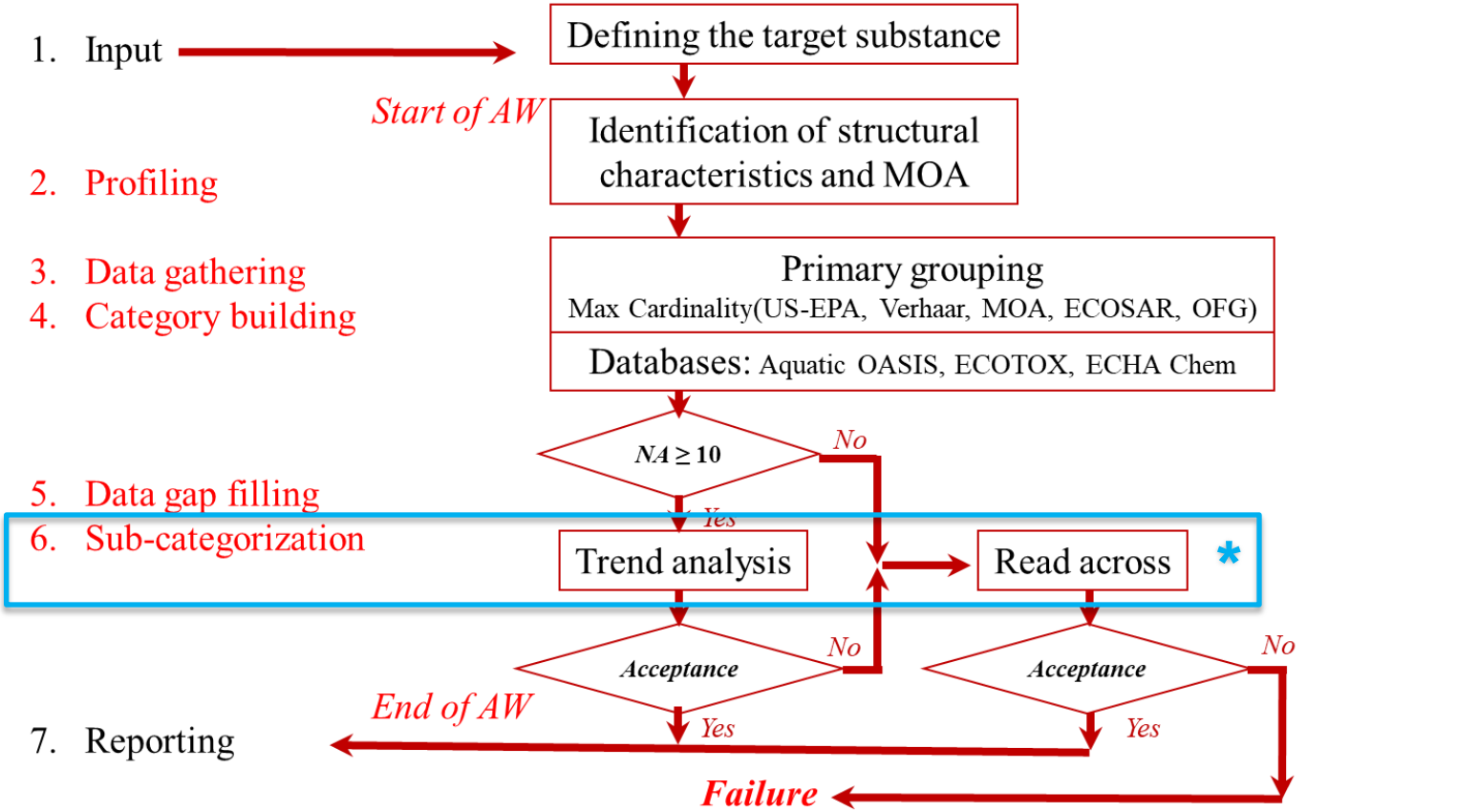
LC50 ≤ 2*:* for the 5 closest analogues the range of variation of *LC50* is ≤ 2 log units

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

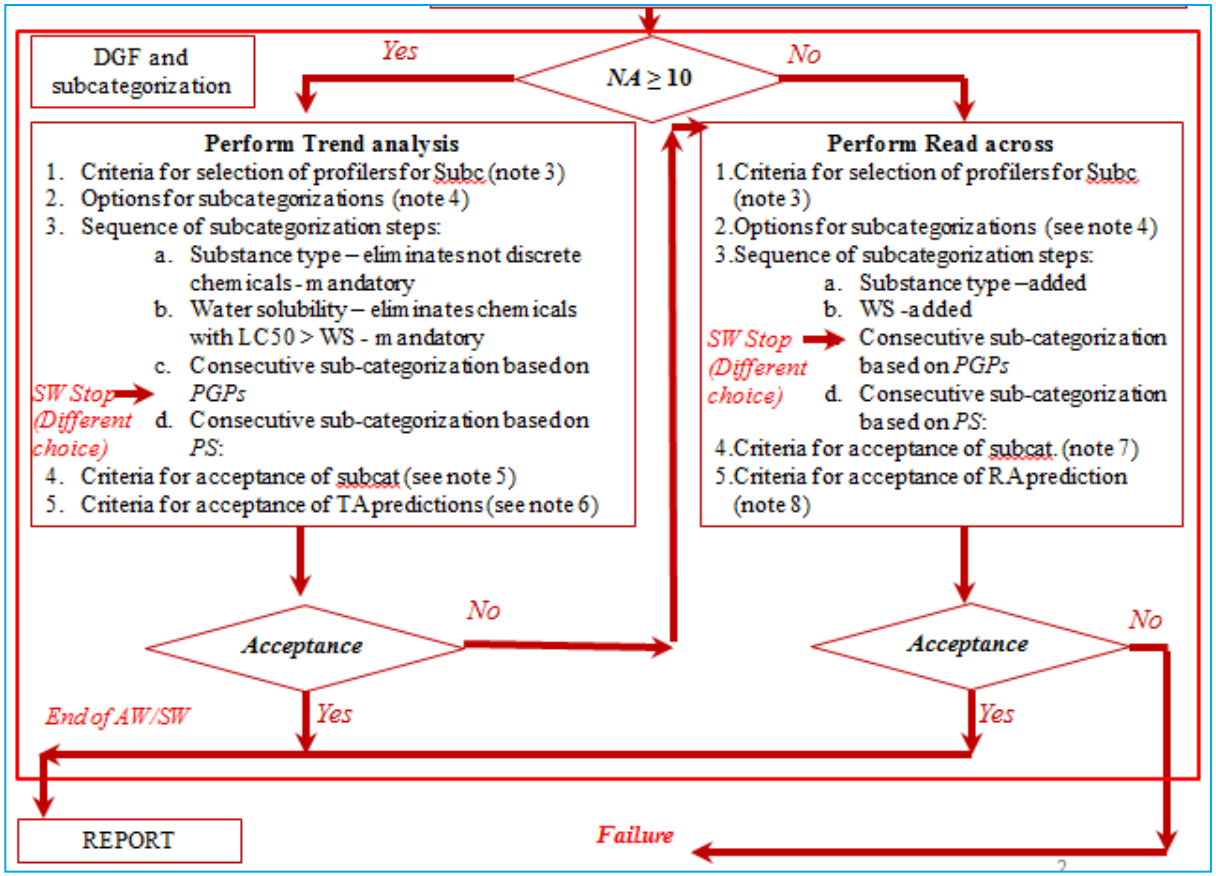
Workflow algorithm - illustration



More details for this section could be seen on next slide

Execution of AW for Ecotoxicity

Algorithm of Ecotoxicological workflow



Automated workflow for skin sensitization
Summary

- The automated workflow for ecotoxicological endpoint is based on LC50 experimental data only;
- The AW follows the implemented logic and finishes with a prediction;
- The automated workflow (AW) is designed to apply data gap filling for discrete chemicals only;
- The AW does not allow interactions during the workflow process;
- Analogous sets of chemicals are selected based on the hypothesis that the toxicological effects of each member of the category will show a common behaviour;
- AW can be executed for one chemical as well as for a batch of chemicals.

Outlook

- Background
- Keywords
- Objectives
- Automated workflow
- **The exercise**
- Report

The Exercise

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

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

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

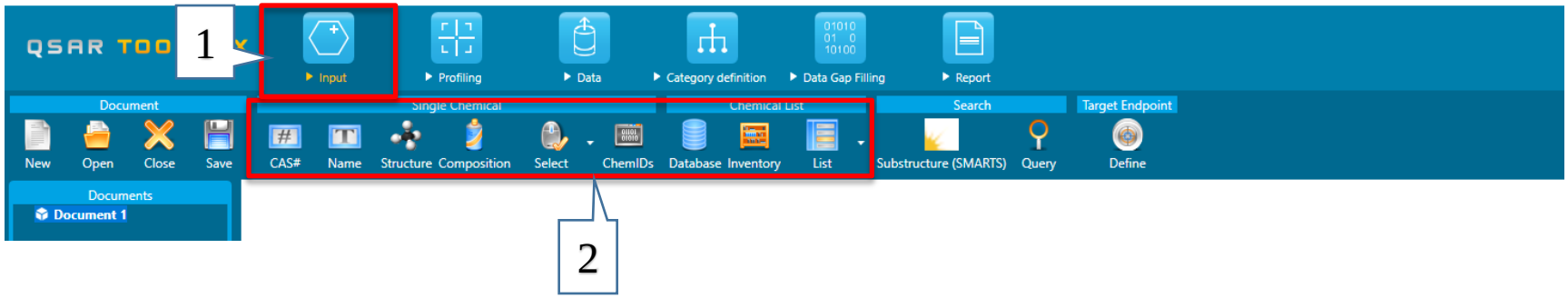
Execution of AW for Ecotoxicity

Chemical Input: Single target chemical

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

Execution of AW for Ecotoxicity

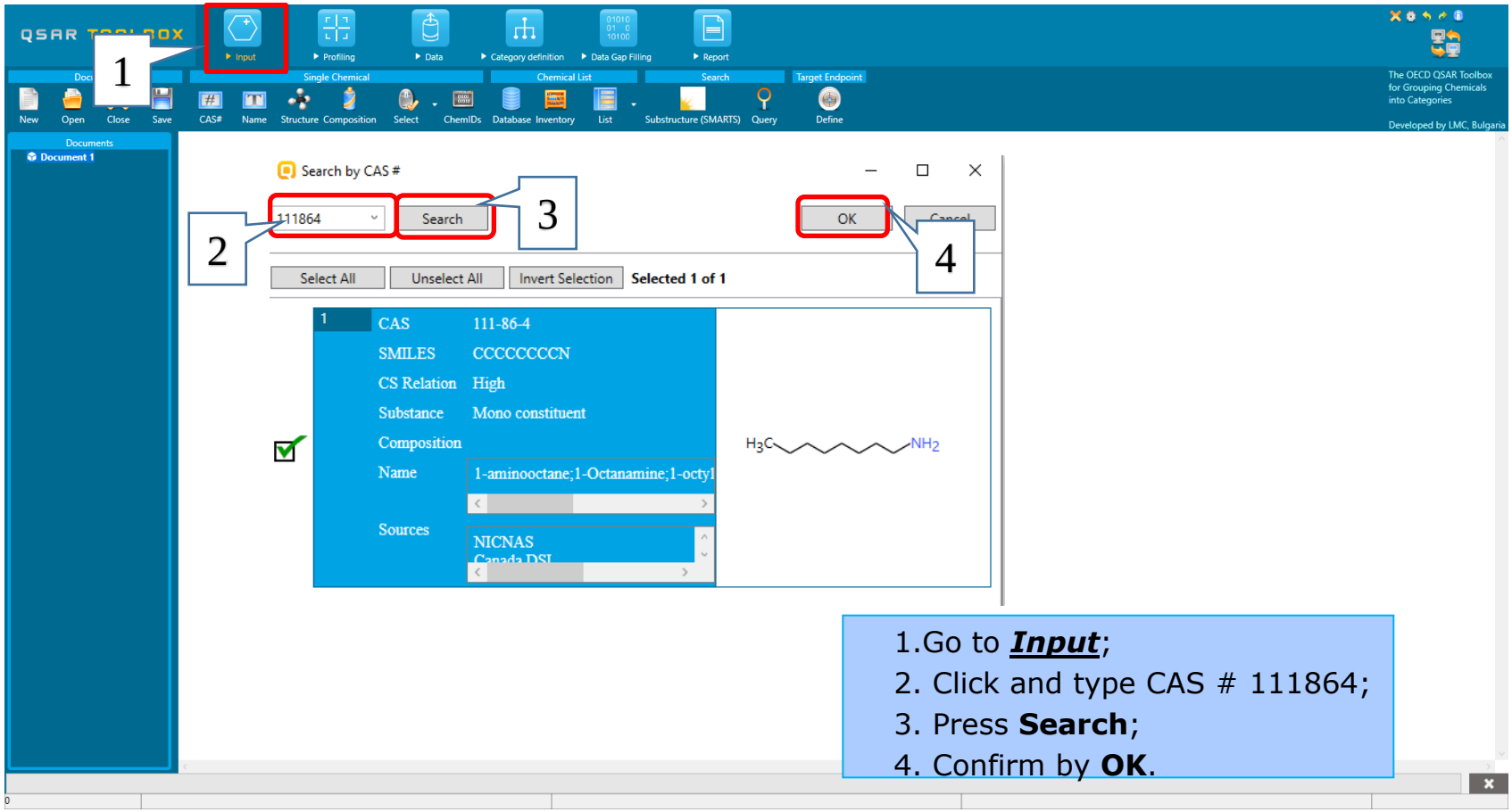
Chemical Input
Single target chemical



1. Click on ***Input***
2. Differ ways of entering the chemical

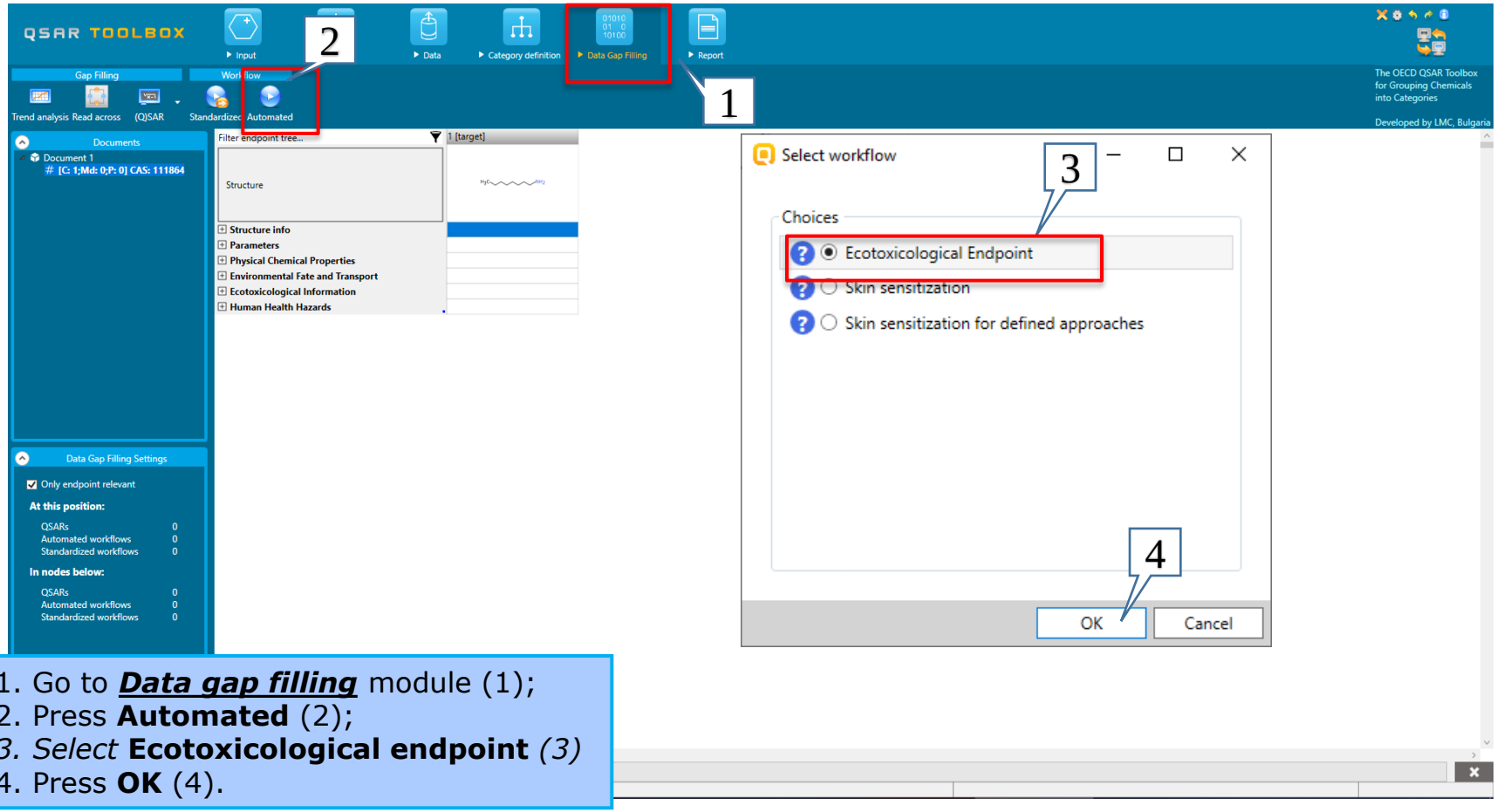
Execution of AW for Ecotoxicity

Input Single target chemical
by CAS RN



Execution of AW for Ecotoxicity

Location of the Automated workflow for Ecotoxicity



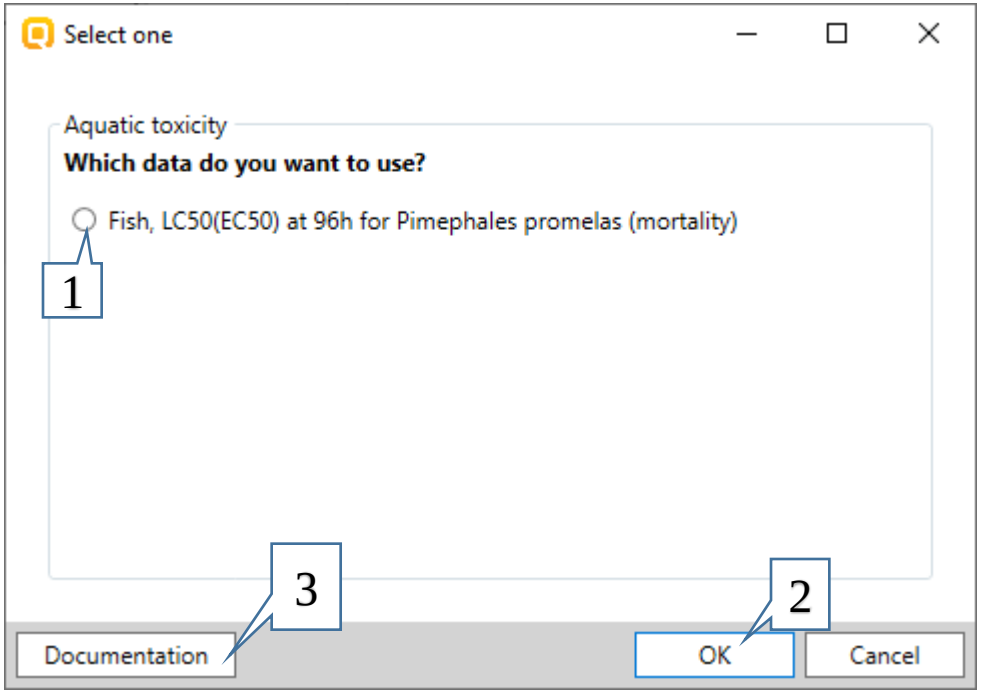
1. Go to ***Data gap filling*** module (1);
2. Press **Automated** (2);
3. *Select* **Ecotoxicological endpoint** *(3)*
4. Press **OK** (4).

Execution of AW for Ecotoxicity

1. A dialogue window gives the user a choice to select the endpoint (1);

2. Then select **OK** (2);

3. A documentation is available (3).



Execution of AW for Ecotoxicity

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

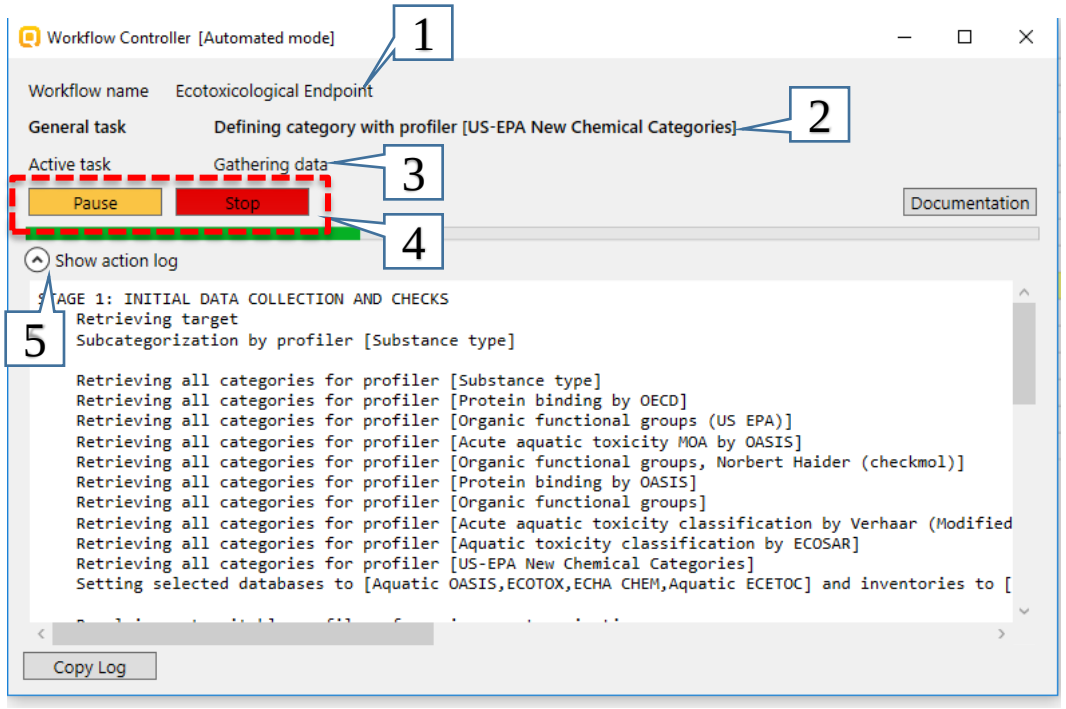
1) Workflow name (1);

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).



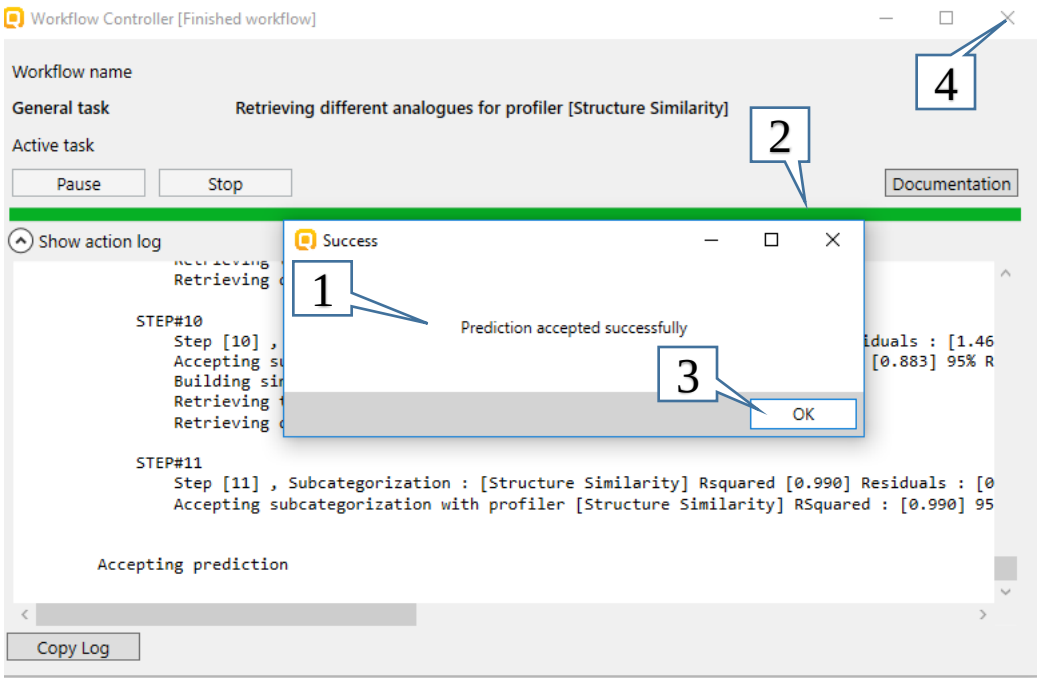
Execution of AW for Ecotoxicity

1) When the workflow finishes a message is displayed that the prediction is accepted (1).

2) Also the progress bar is completely filled (2).

3) The user has to press **OK** (3).

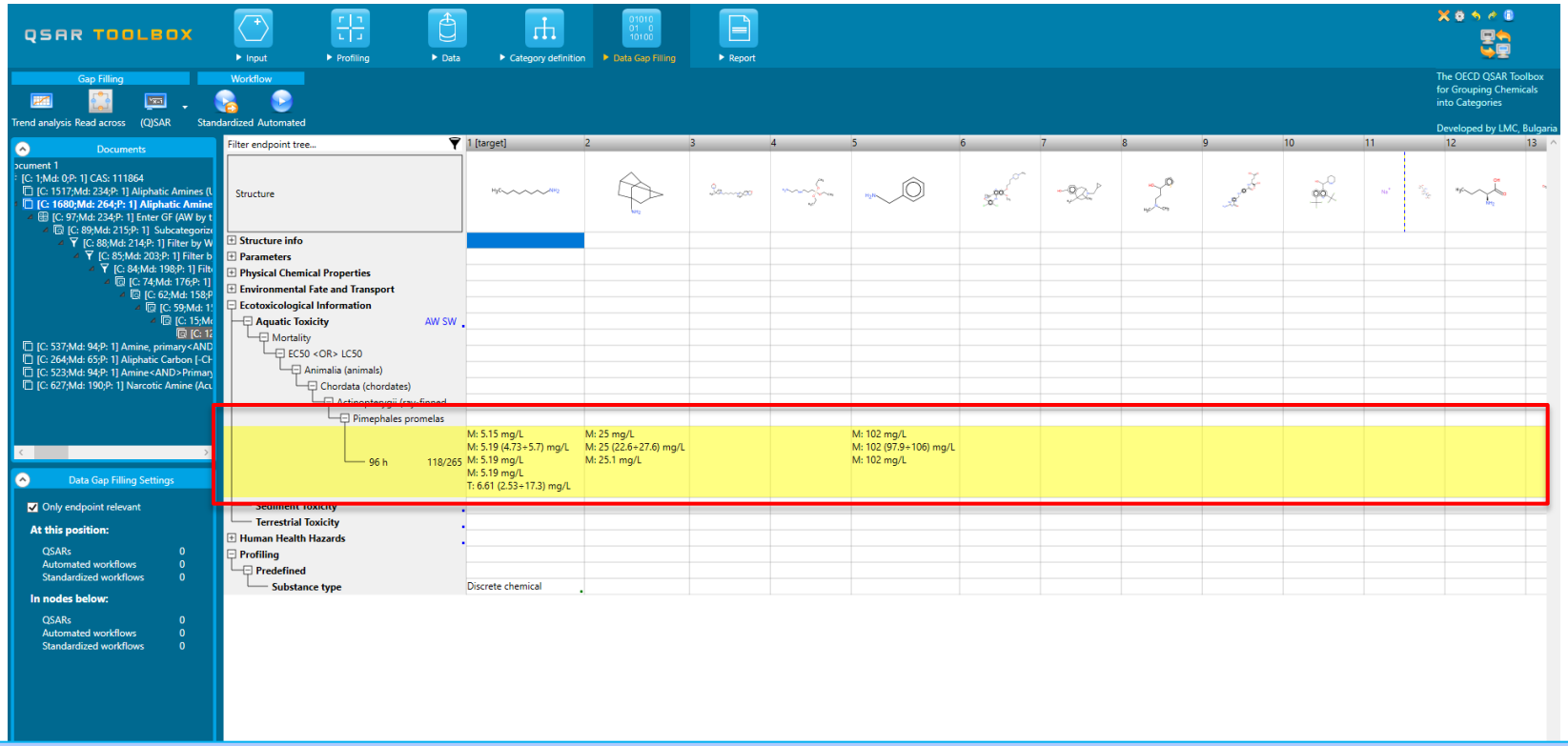
4) Then close the workflow window by pressing **X** button (4).



Execution of AW for Ecotoxicity

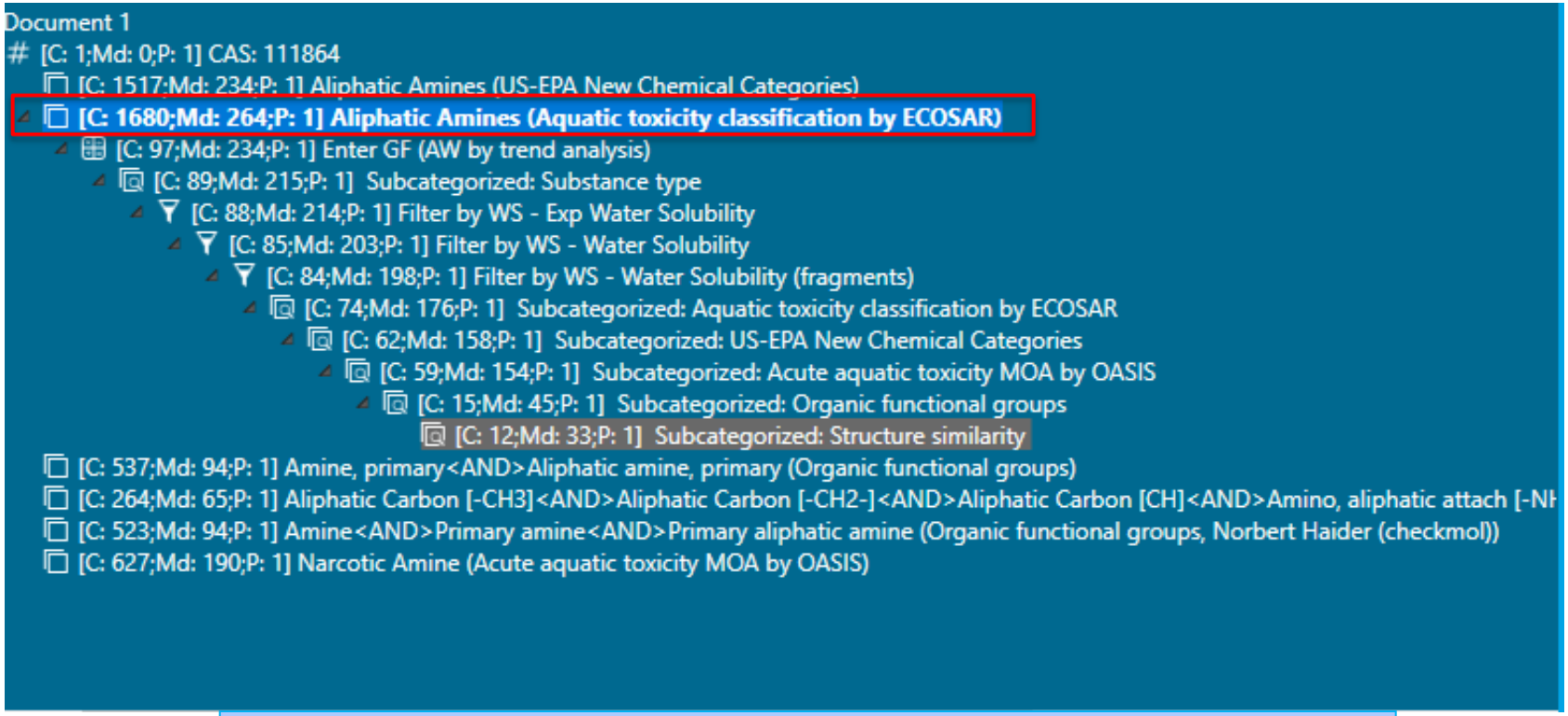
Data Gap Filling

Automated workflow: Single chemical



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

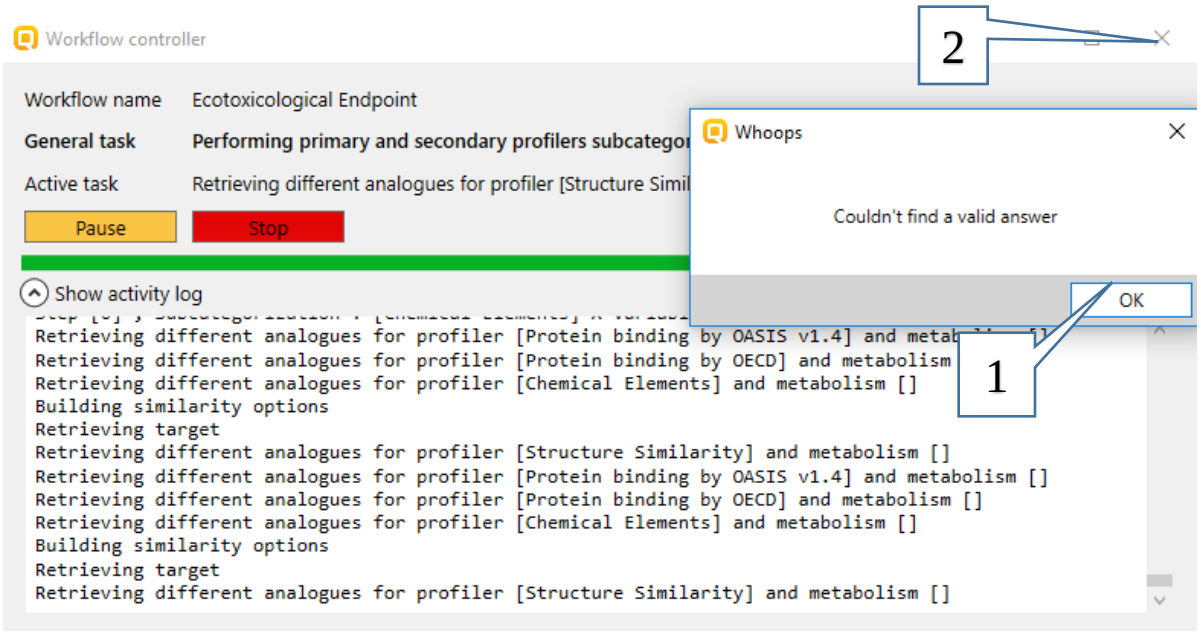
Execution of AW for Ecotoxicity



- The steps executed in the AW are listed in the Documents panel;
- Grey highlighted level indicates that at this level a prediction is accepted
- 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)



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

The Exercise

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

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

Execution of AW for Ecotoxicity

Data Gap Filling

Automated workflow: Batch mode

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

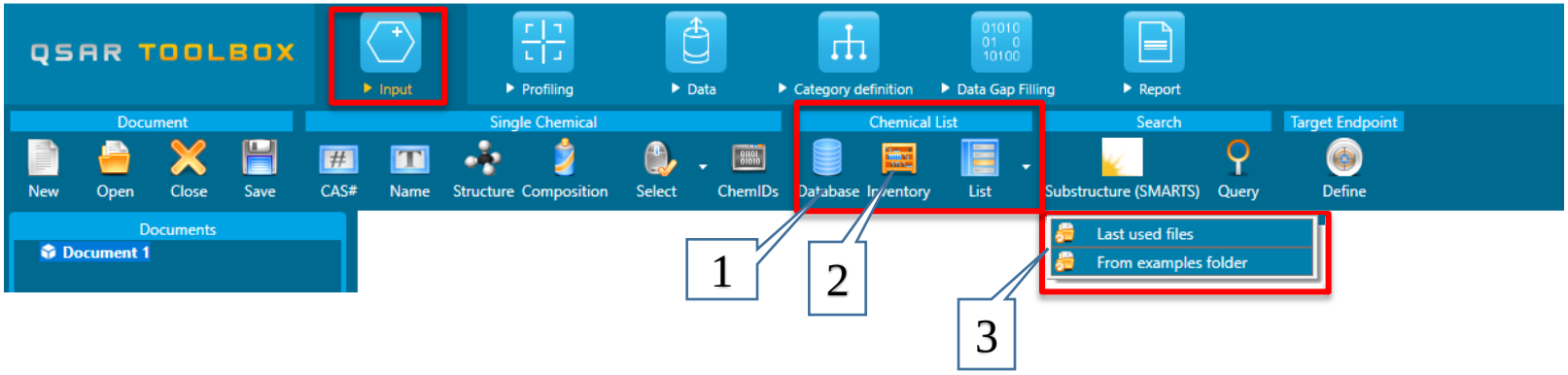
- Selection of chemicals from databases/inventories;
- Loading of chemicals from user's file.

(see next slide)

Execution of AW for Ecotoxicity

List with chemicals (batch work)

Input: Ways of Entering a Chemical List



1) Database
2) Inventory
3) List: Last used files/ From examples folder

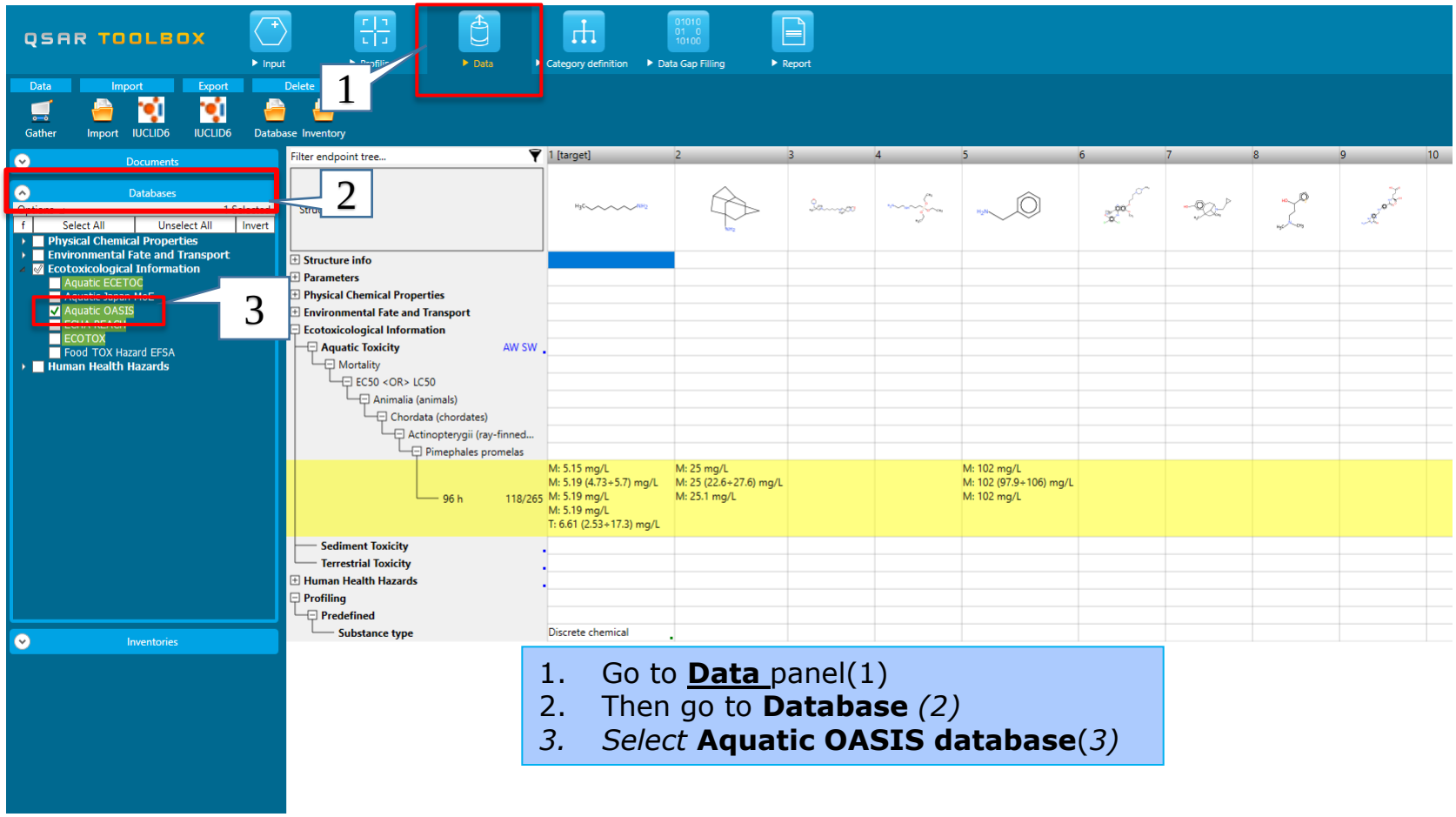
Execution of AW for Ecotoxicity

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

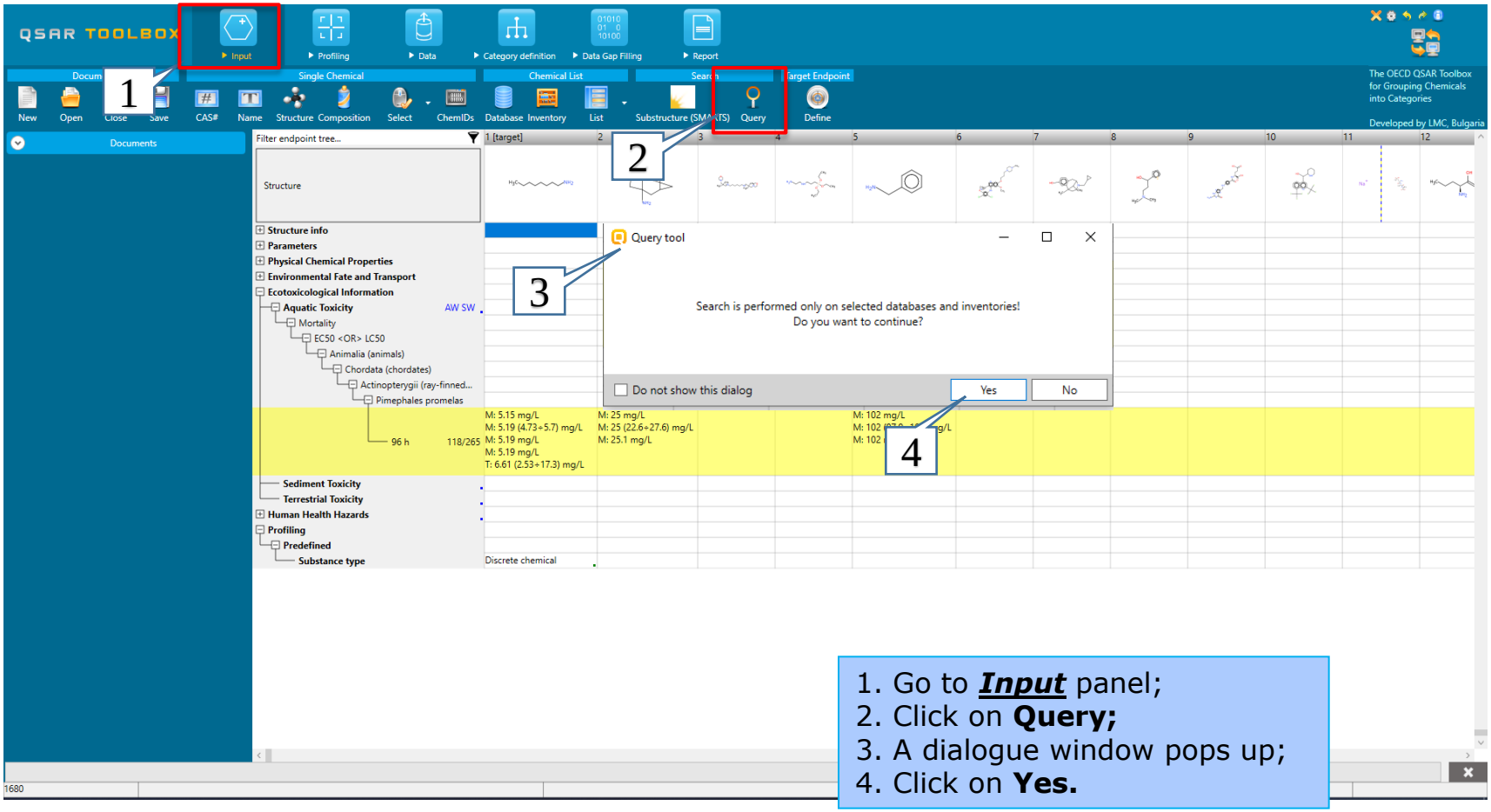


1. Go to **Data** panel(1)
2. Then go to **Database** *(2)*
3. *Select* **Aquatic OASIS database**(*3*)

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

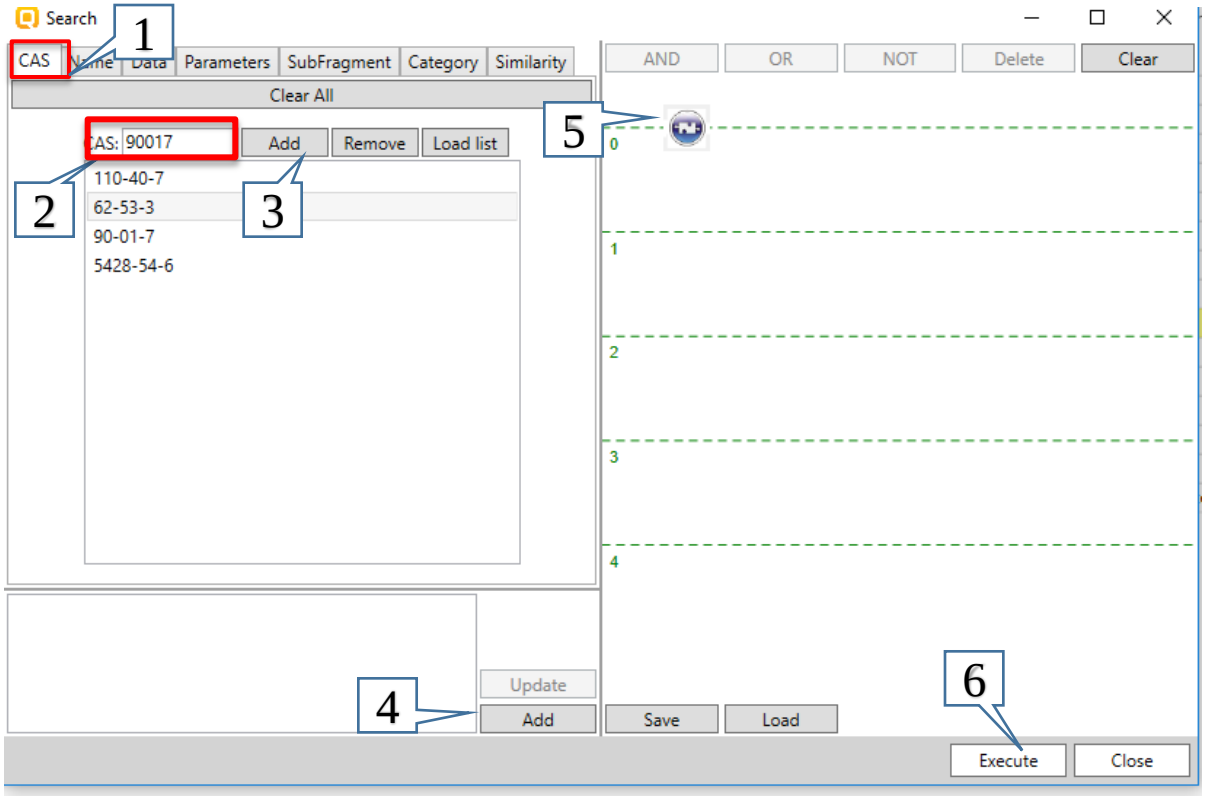
Data Gap Filling
Automated workflow: Batch mode

1) Go to **CAS** tab(1);
2) Type in the empty field the CAS #: 5428546 (2)

3) Click on **Add** *(3)* to add CAS #. Repeat steps 2 and 3 to add the next 3 CAS #: 90017; 62533; 110407;

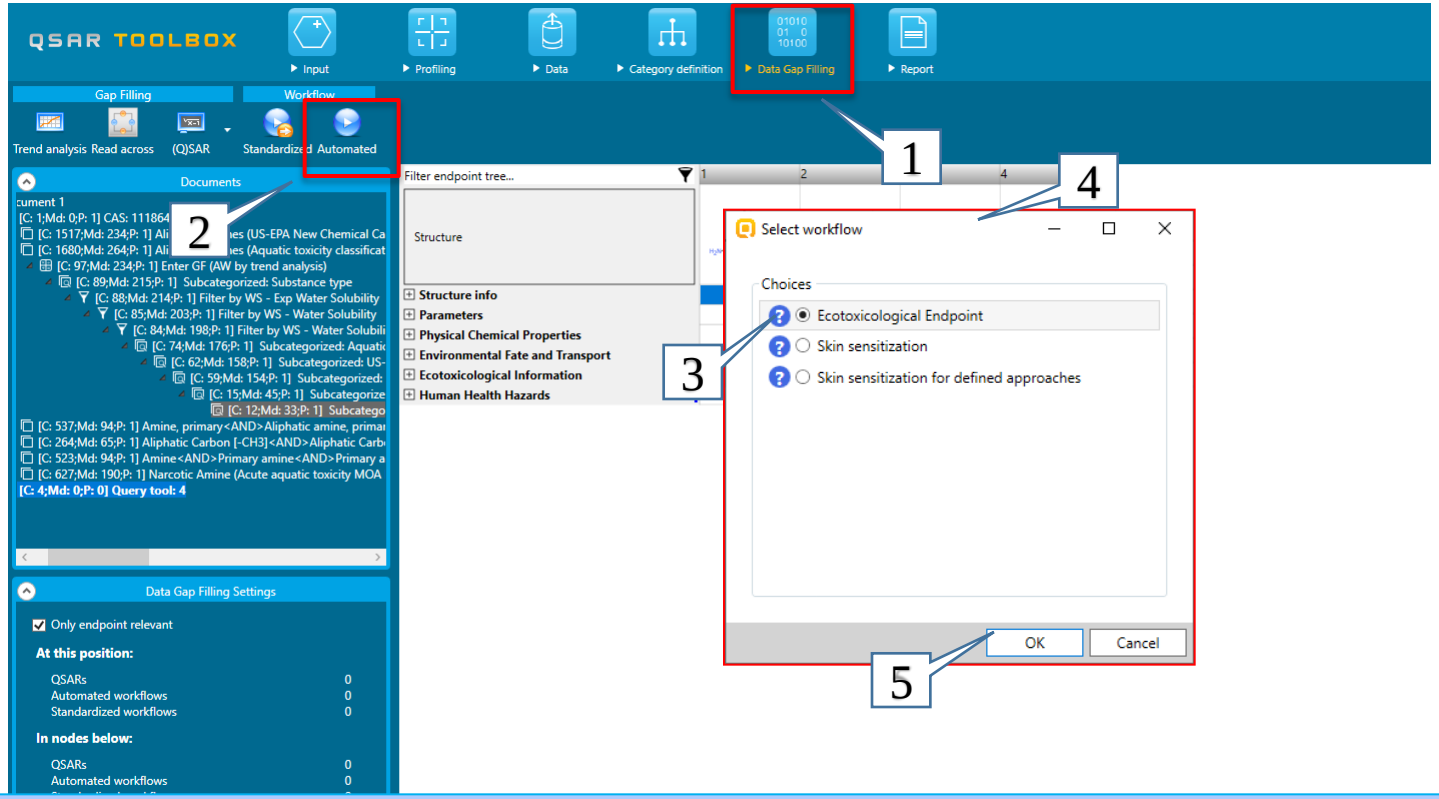
4) Click on **Add** (4)to create the query (5);

5) Click on **Execute** button(6).



Execution of AW for Ecotoxicity

Data Gap Filling
Automated workflow: Batch mode



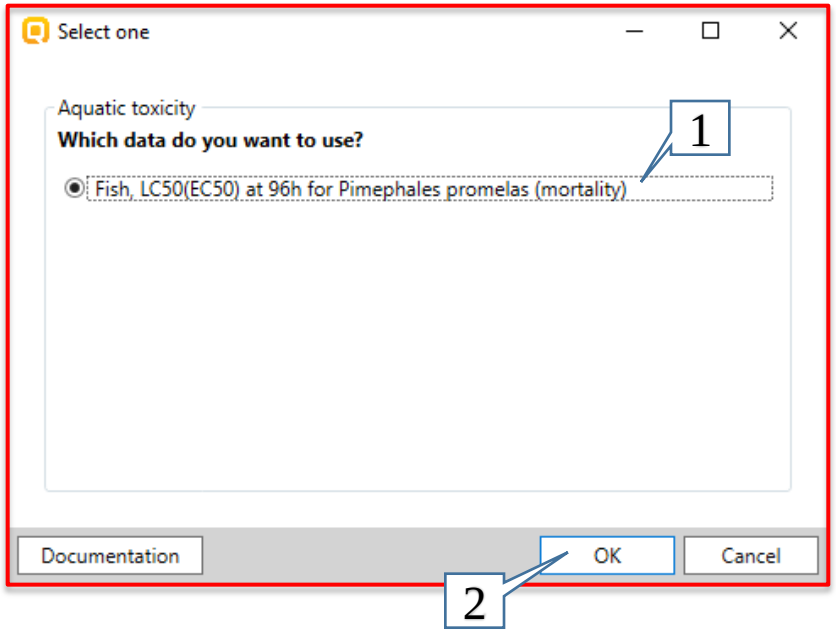
1) Select ***Data gap filling*** tab (1);
2) Click on **Automated** (2);
3) Select **Ecotoxicological endpoint** (3) from the pop-up window (4);
4) Click on **OK** (5).

Execution of AW for Ecotoxicity

Data Gap Filling
Automated workflow: Batch mode

1) A dialogue window gives the user a choice to select the endpoint (1);

2) Then select **OK** (2).



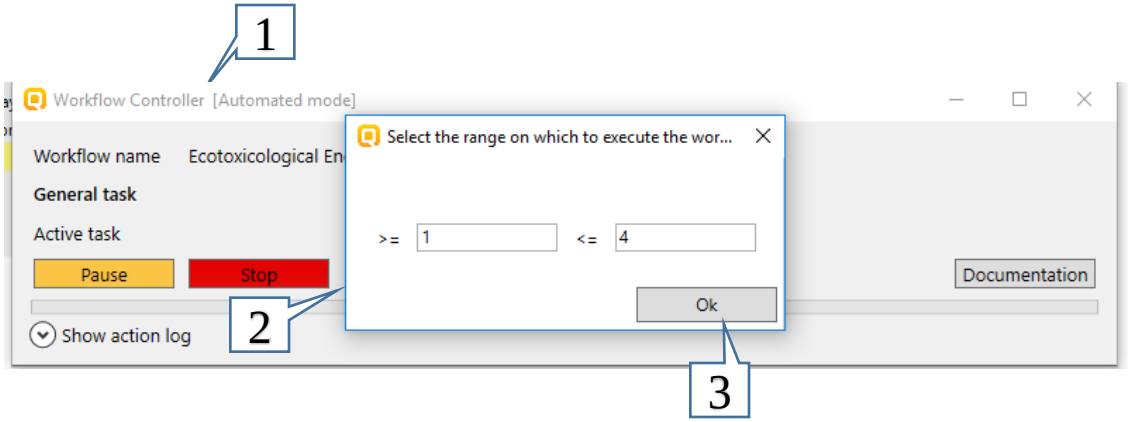
Execution of AW for Ecotoxicity

Data Gap Filling
Automated workflow: Batch mode

1) A *Workflow controller* window appears, which is not active(1);

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;

1) Finally press *OK* (3).



Execution of AW for Ecotoxicity

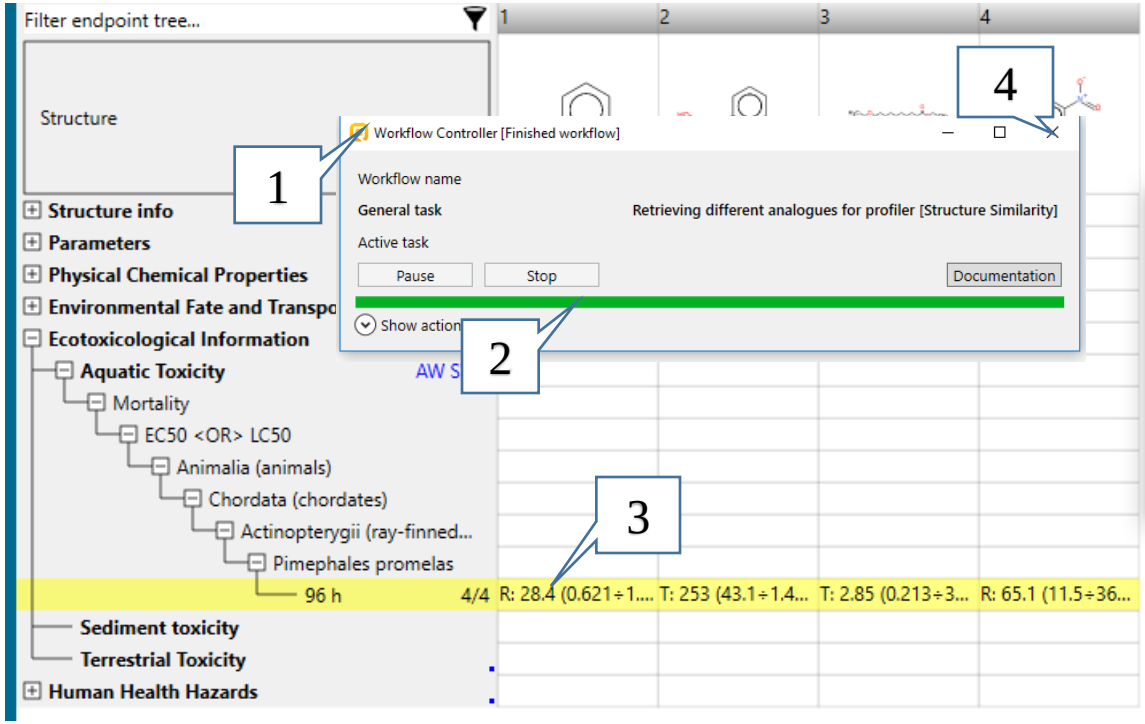
Data Gap Filling
Automated workflow: Batch mode

1) When the workflow finishes, there is an indication in the title of workflow controller(1).

2) Also the progress bar is completely filled (2).

3) The predictions are displayed on the matrix(3). There is also an indication that 4 out of 4 chemicals are predicted

4) Finally close the workflow window by pressing **X** button (4).



Outlook

- Background
- Keywords
- Objectives
- Automated workflow
- The exercise
- **Report**

Execution of AW for Ecotoxicity

List with chemicals (batch work)

Report: Generation of report

An this stage, it is not possible to generate a report for chemicals predicted in batch mode.

The system kindly informs you that *Predictions from batch mode cannot be reported.*

Congratulations!

- You have completed the tutorial on the automated workflow for ecotoxicological endpoint.
- You have been introduced to the automated workflow in a single and in a batch mode.
- You have now been introduced to the consecutive steps of the AW for ecotox and the rationale behind each step.
- Note, proficiency comes with practice!