QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.4.1

Example for predicting acute aquatic toxicity to fish of a mixture with known components

Outlook

- **Background**
- Keywords
- Objectives
- The exercise
- Workflow
- Save/Load

Background

- This is a step-by-step presentation designed to take the user of the Toolbox through the workflow of predicting acute aquatic toxicity to fish of a mixture with known components.

Outlook

- Background
- **Keywords**
- Objectives
- The exercise
- Workflow
- Save/Load

Keywords

TARGET CHEMICAL - chemical of interest, in this case this is mixture with define components

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

MOAD of ACTION(MOA) – profiling pattern of the chemicals available on data matrix with respect to the applied profilers

Outlook

- Background
- Keywords
- **Objectives**
- The exercise
- Workflow
- Save/Load

Objectives

- **This presentation reviews a number of functionalities of the Toolbox:**
  - The 2D editor for defining Mixture components
  - Filling data gaps by Similar mode approach

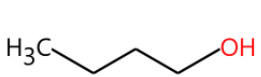
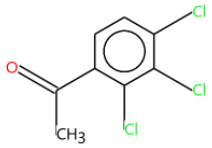
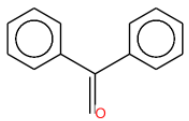
Outlook

- Background
- Keywords
- Objectives
- **The exercise**
- Workflow
- Save the prediction

Exercise

- In this exercise we will:
  - predict the aquatic toxicity to fish (LC50, mortality, 96h, *Pimephales promelas*) of a target substance which represents a mixture with defined constituents;
  - investigate the mode of action of constituents of the mixture;
  - gather available experimental data for target and its constituents;
  - Predict acute aquatic toxicity using Similar mode approach.

- The target substance will consists of three constituents with quantities as follows:

Constituent 1	Constituent 2	Constituent 3
		
Weight = 90%	Weight = 1%	Weight = 9%

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
- Save the prediction

Workflow

- **The Toolbox has six modules which are used in a sequential workflow:**
  - Input
  - Profiling
  - Data
  - Category Definition
  - Data Gap Filling
  - Report

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - **Input**

Input
Overview

- This module provides the user with several means of entering the chemical of interest (i.e. the target substance).
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target substance is the correct one.

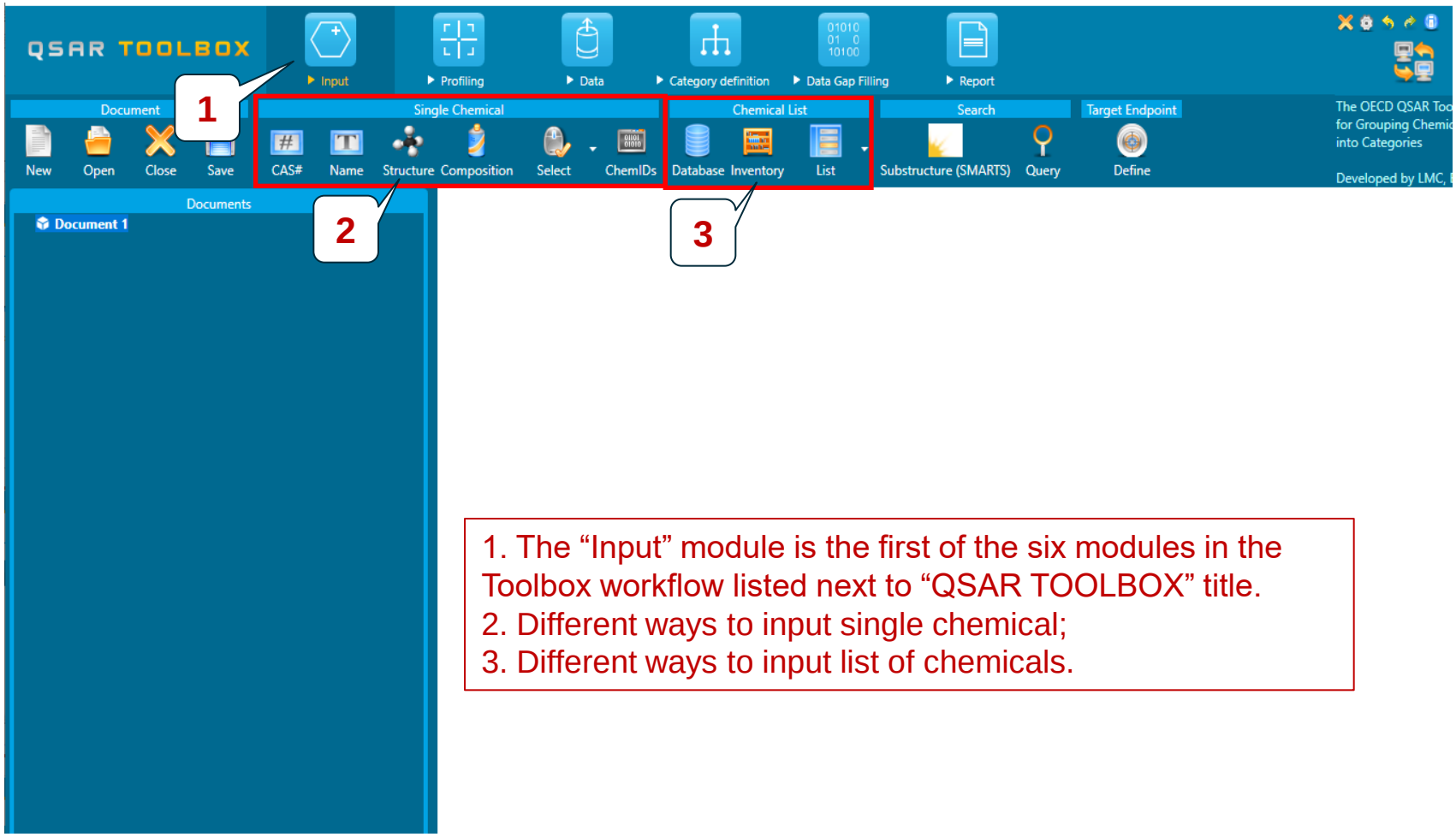
Input
Input chemical(s)

Alternative ways to input chemical(s):

A. Single target substance
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases

B. Group of chemicals
- User List/Inventory
- Specialized Databases

Input
Input Screen



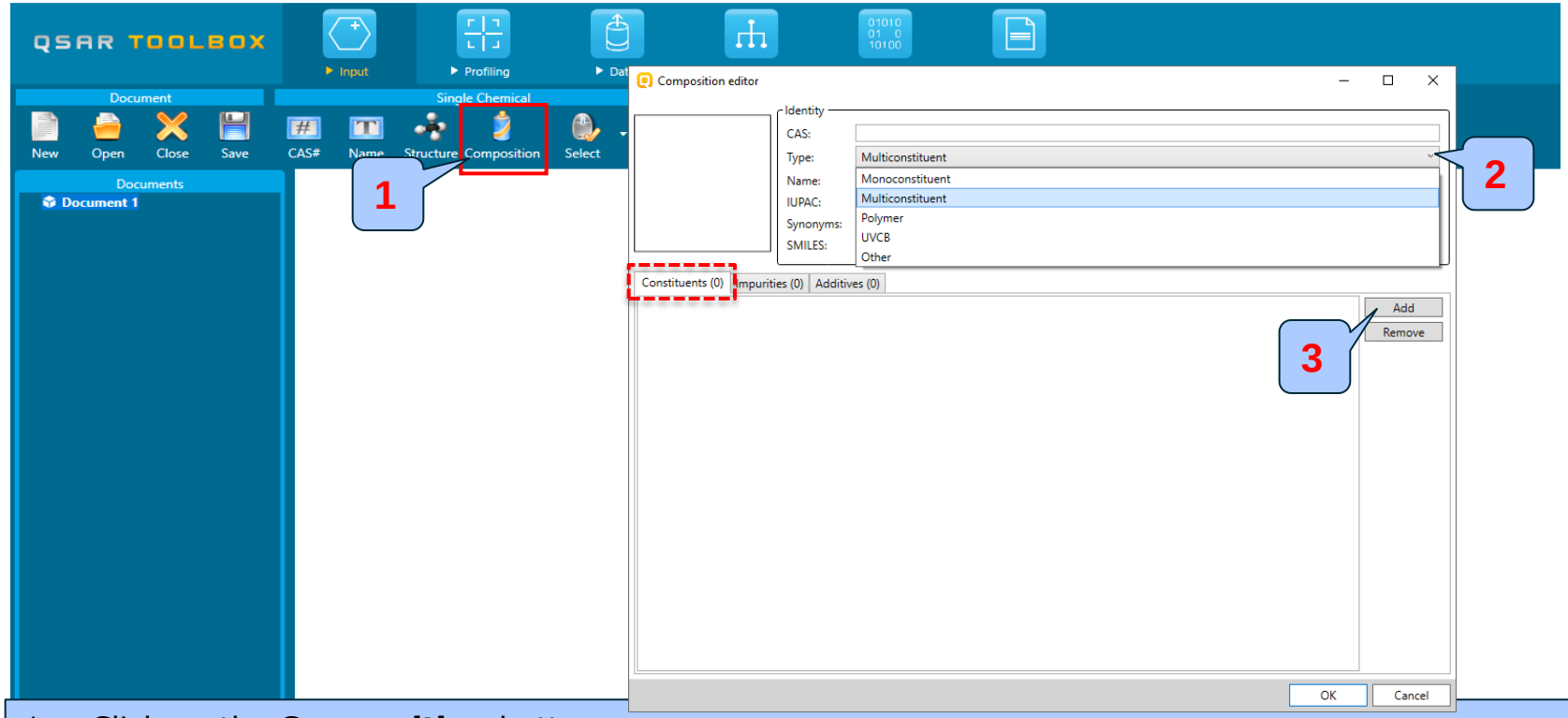
1. The “Input” module is the first of the six modules in the Toolbox workflow listed next to “QSAR TOOLBOX” title.
2. Different ways to input single chemical;
3. Different ways to input list of chemicals.

Input
Input a mixture

- In the current example our target substance will be a mixture.
- We will draw its components within the “Composition” tool.
- Consecutively, the quantities of the mixture components will be defined.

Input

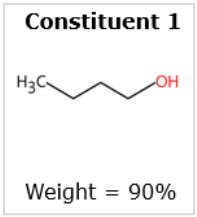
Input target substance by drawing

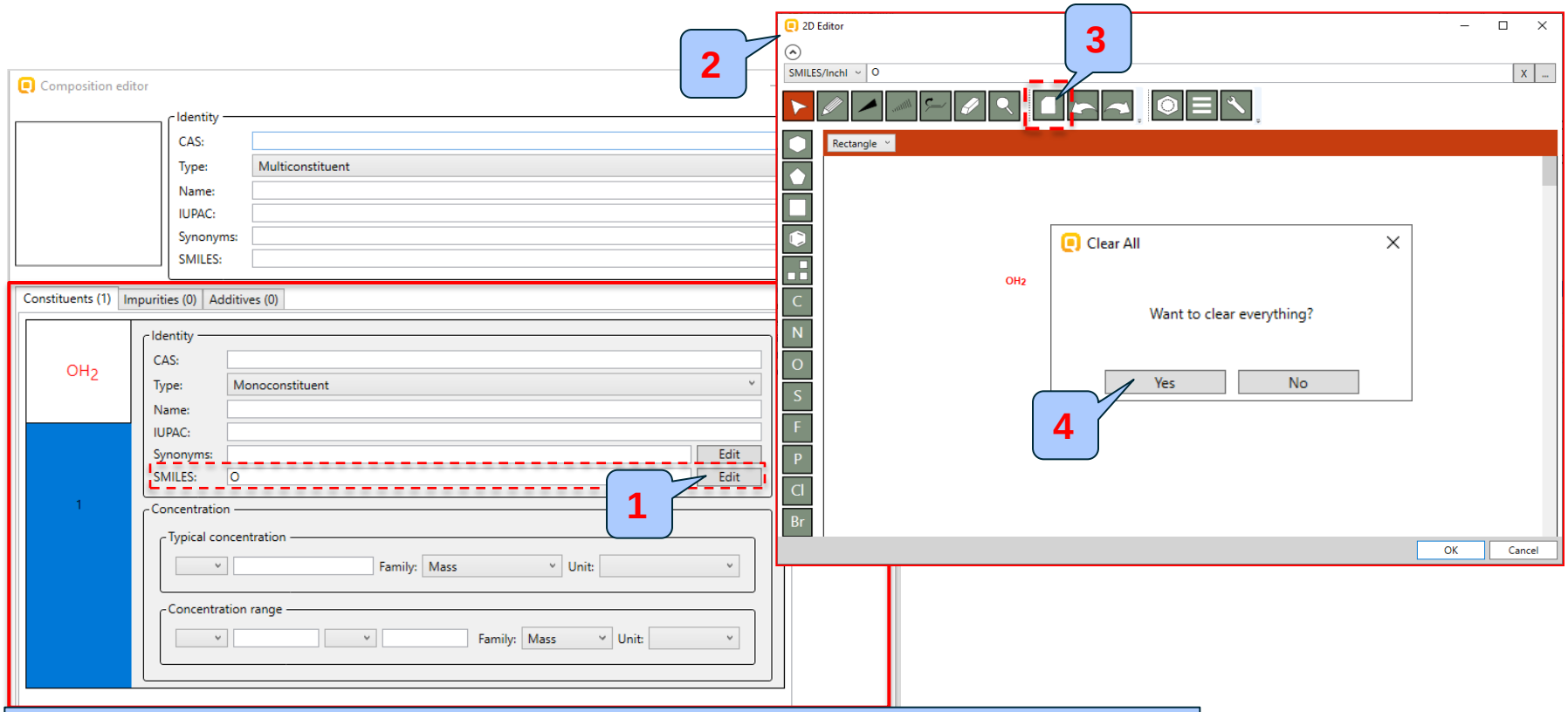


1. Click on the **Composition** button;
2. Select "Multiconstituent" from the drop-down menu for the type of the substance;
3. Click **Add** in order to add constituent. Our target substance consists of three constituents, so click three times on the Add button.

Input

Drawing Constituent 1 of the target mixture

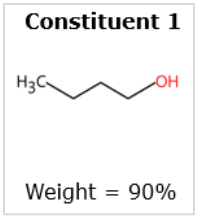


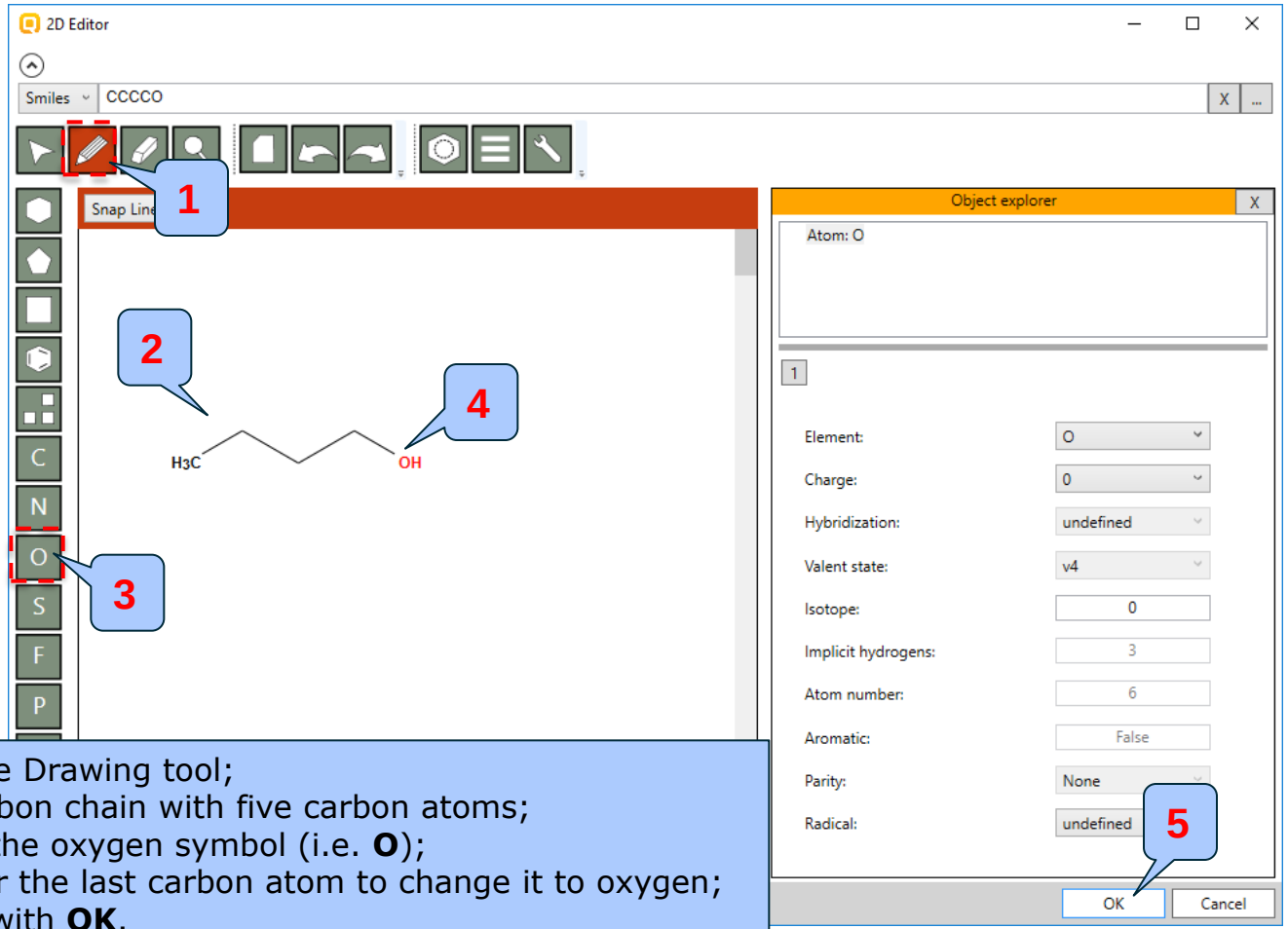


1. Click **Edit** on the SMILES row to define the structure of the first constituent;
2. The 2D editor appears;
3. Click the **Clean button** to clean everything;
4. Confirm with **Yes**.

Input

Drawing Constituent 1 of the target mixture

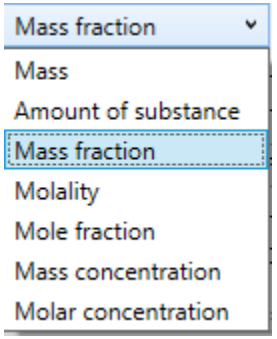




1. Select the Drawing tool;
2. Draw carbon chain with five carbon atoms;
3. Click on the oxygen symbol (i.e. **O**);
4. Click over the last carbon atom to change it to oxygen;
5. Confirm with **OK**.

Input
Define quantities of mixture`s constituents

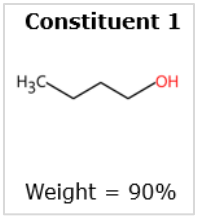
- Quantities of the constituents should be added manually.
- There are several ways to add mixture quantity:
  - Mass fraction
  - Mass
  - Amount of substance
  - Molality
  - Mole fraction
  - Mass concentration
  - Molar concentration



- In the current example we will select “Mass fraction %” in “Weight %”

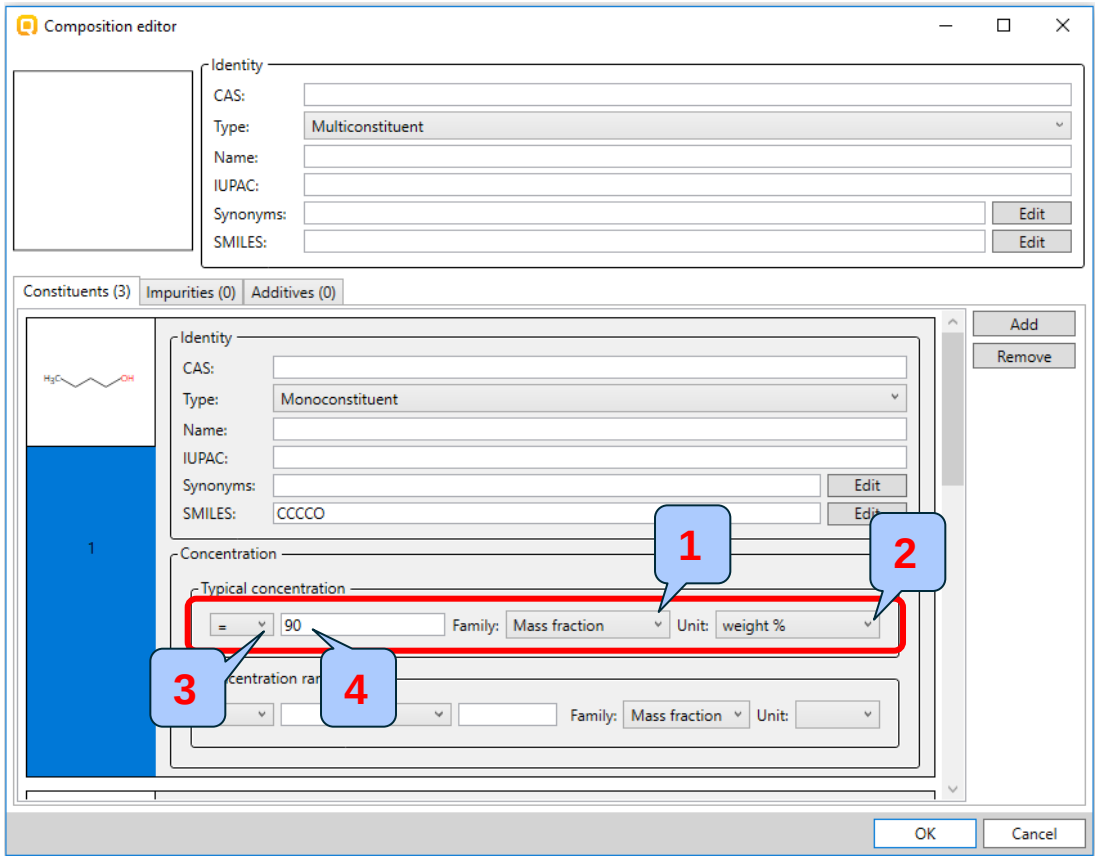
Input

Define the quantity of Constituent 1



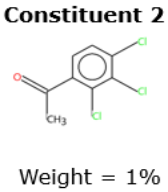
For Constituent 1 of the target mixture define:

1) Family: **Mass fraction**
2) Unit: **weight %**
3) qualifier: Equal to (i.e. **=**)
4) quantity: **90**



Input

Drawing Constituent 2 of the target mixture



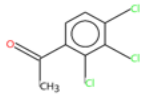
Now we move down to the second constituent and repeat the same steps:

1. Click **Edit** on the SMILES row to define the structure of the second constituent;
2. The 2D editor appears;
3. Click the **Clean button** to clean everything;
4. Confirm with **Yes**.

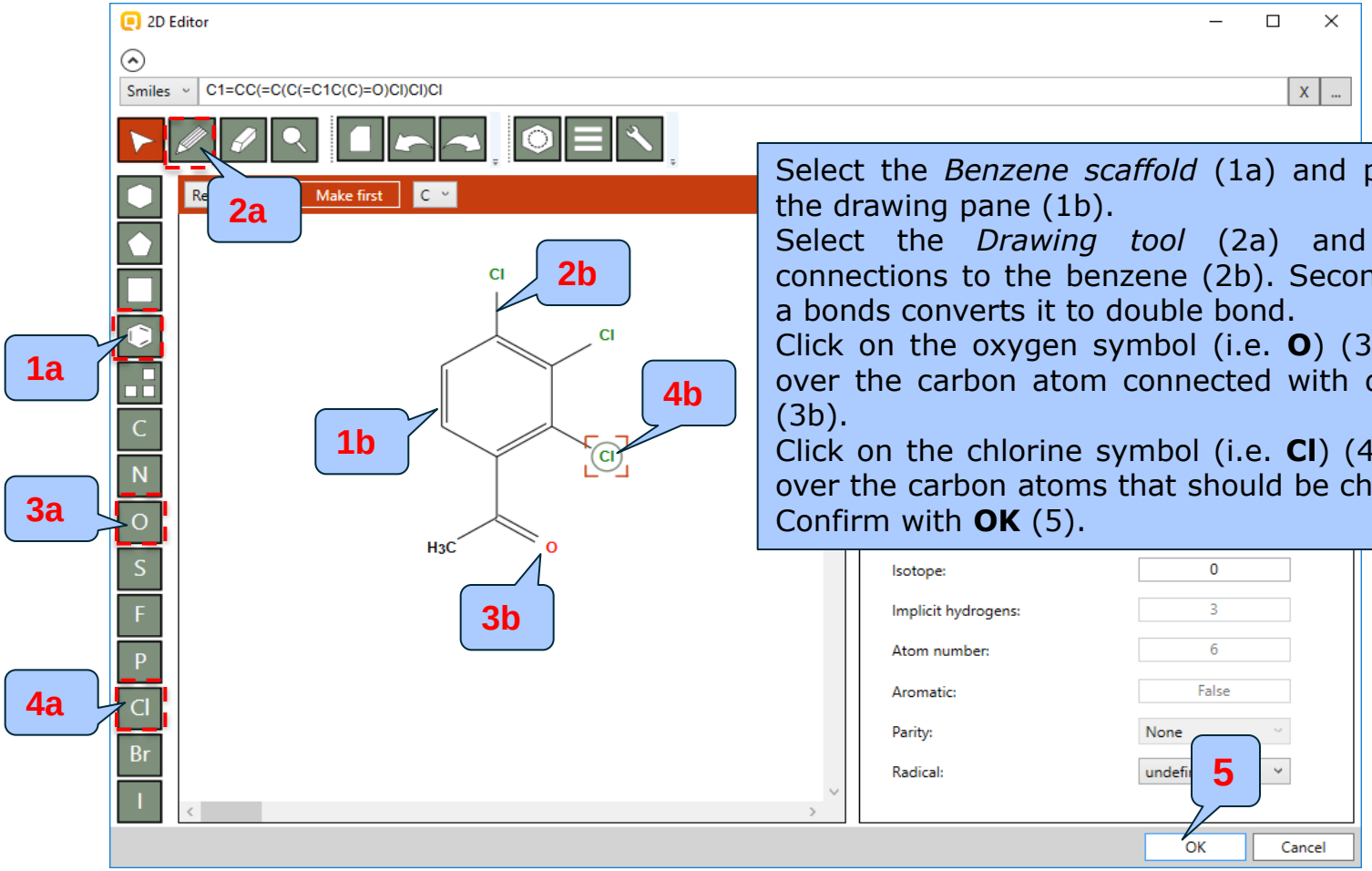
Input

Drawing Constituent 2 of the target mixture

Constituent 2



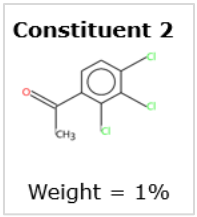
Weight = 1%



Select the *Benzene scaffold* (1a) and paste it into the drawing pane (1b).
Select the *Drawing tool* (2a) and draw the connections to the benzene (2b). Second click over a bonds converts it to double bond.
Click on the oxygen symbol (i.e. **O**) (3a) and click over the carbon atom connected with double bond (3b).
Click on the chlorine symbol (i.e. **Cl**) (4a) and click over the carbon atoms that should be changed (3b).
Confirm with **OK** (5).

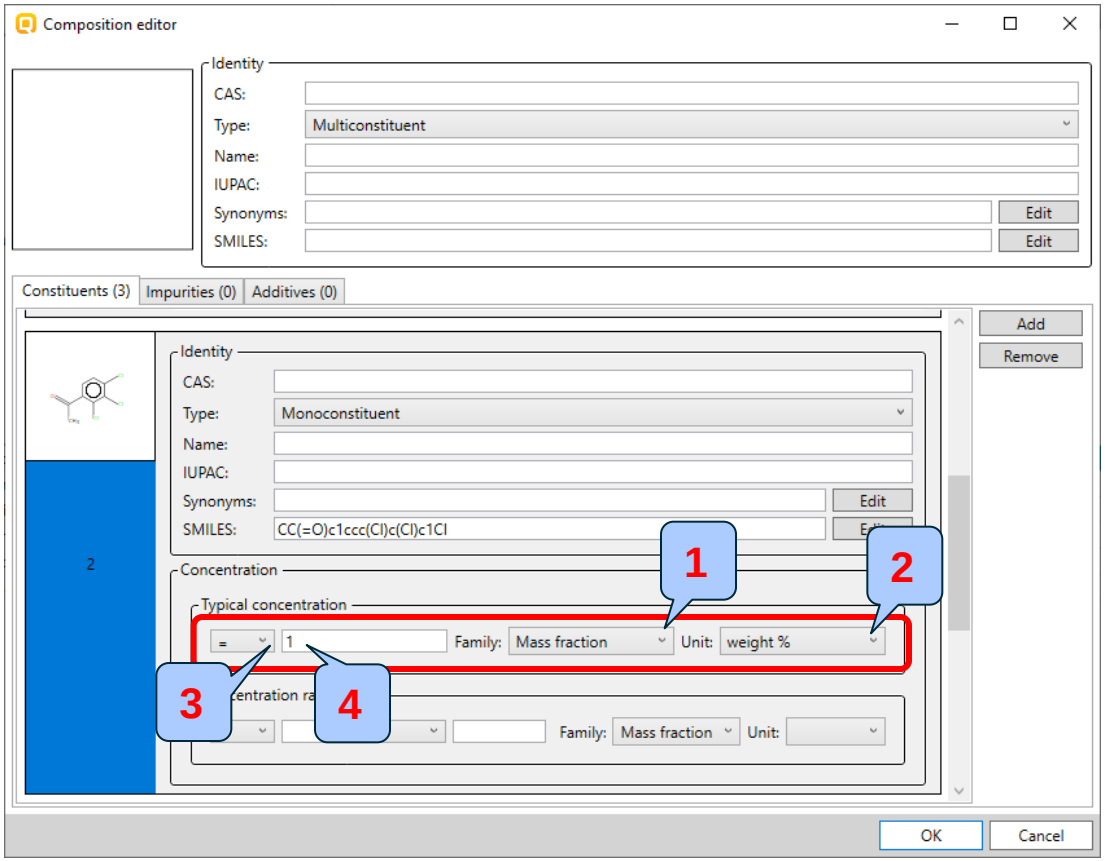
Input

Define the quantity of Constituent 2



For Constituent 2 of the target mixture define:

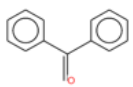
1) *Family*: **Mass fraction**
2) *Unit:* **weight %**
3) *qualifier:* Equal to (i.e. **=**)
4) *quantity:* **1**



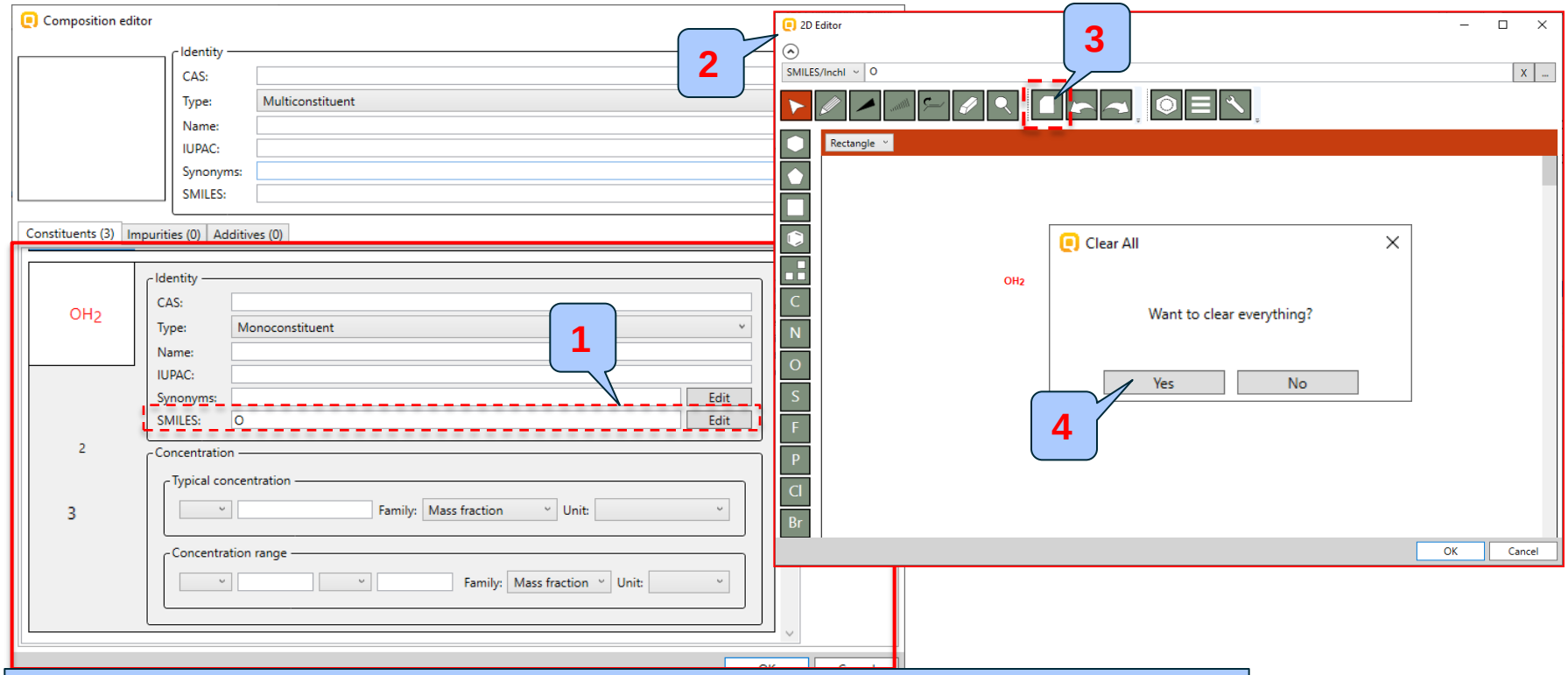
Input

Drawing Constituent 3 of the target mixture

Constituent 3



Weight = 9%



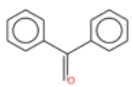
Now we move down to the last (third) constituent and repeat the same steps:

1. Click **Edit** on the SMILES row to define the structure of the second constituent;
2. The 2D editor appears;
3. Click the **Clean button** to clean everything;
4. Confirm with **Yes.**

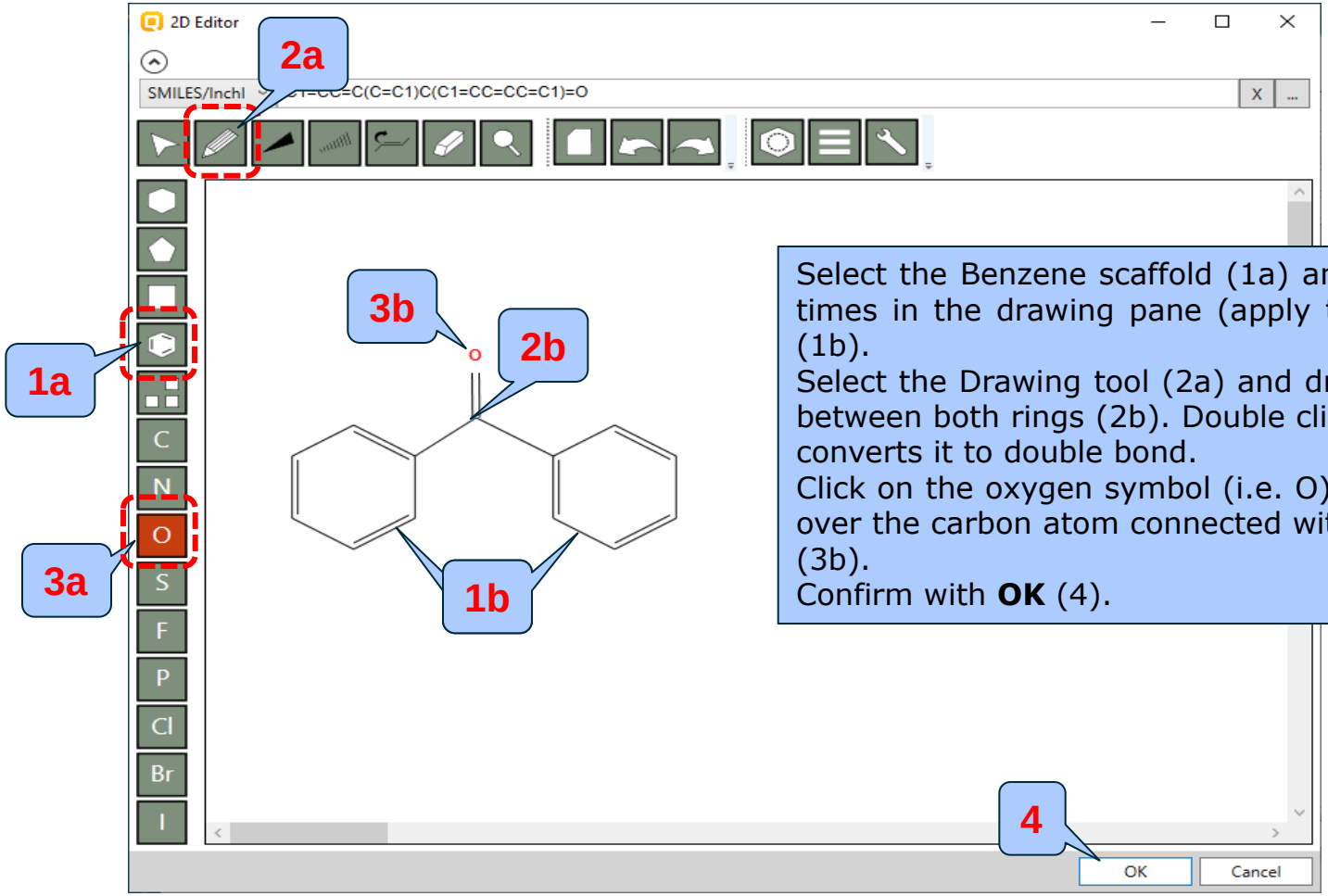
Input

Drawing Constituent 3 of the target mixture

Constituent 3



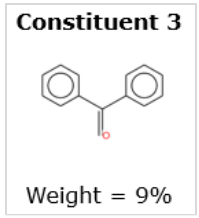
Weight = 9%



Select the Benzene scaffold (1a) and paste it two times in the drawing pane (apply two left clicks) (1b).
Select the Drawing tool (2a) and draw connection between both rings (2b). Double click over a bond converts it to double bond.
Click on the oxygen symbol (i.e. O) (3a) and click over the carbon atom connected with double bond (3b).
Confirm with **OK** (4).

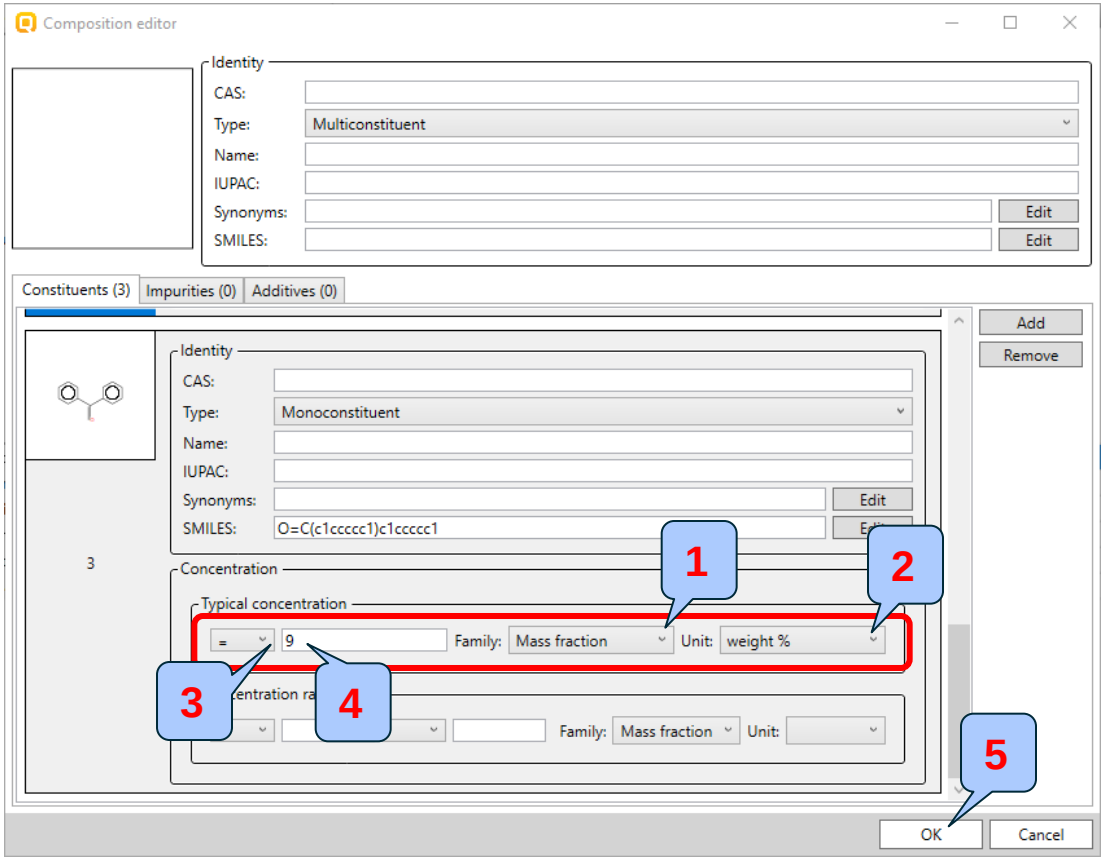
Input

Define the quantity of Constituent 3



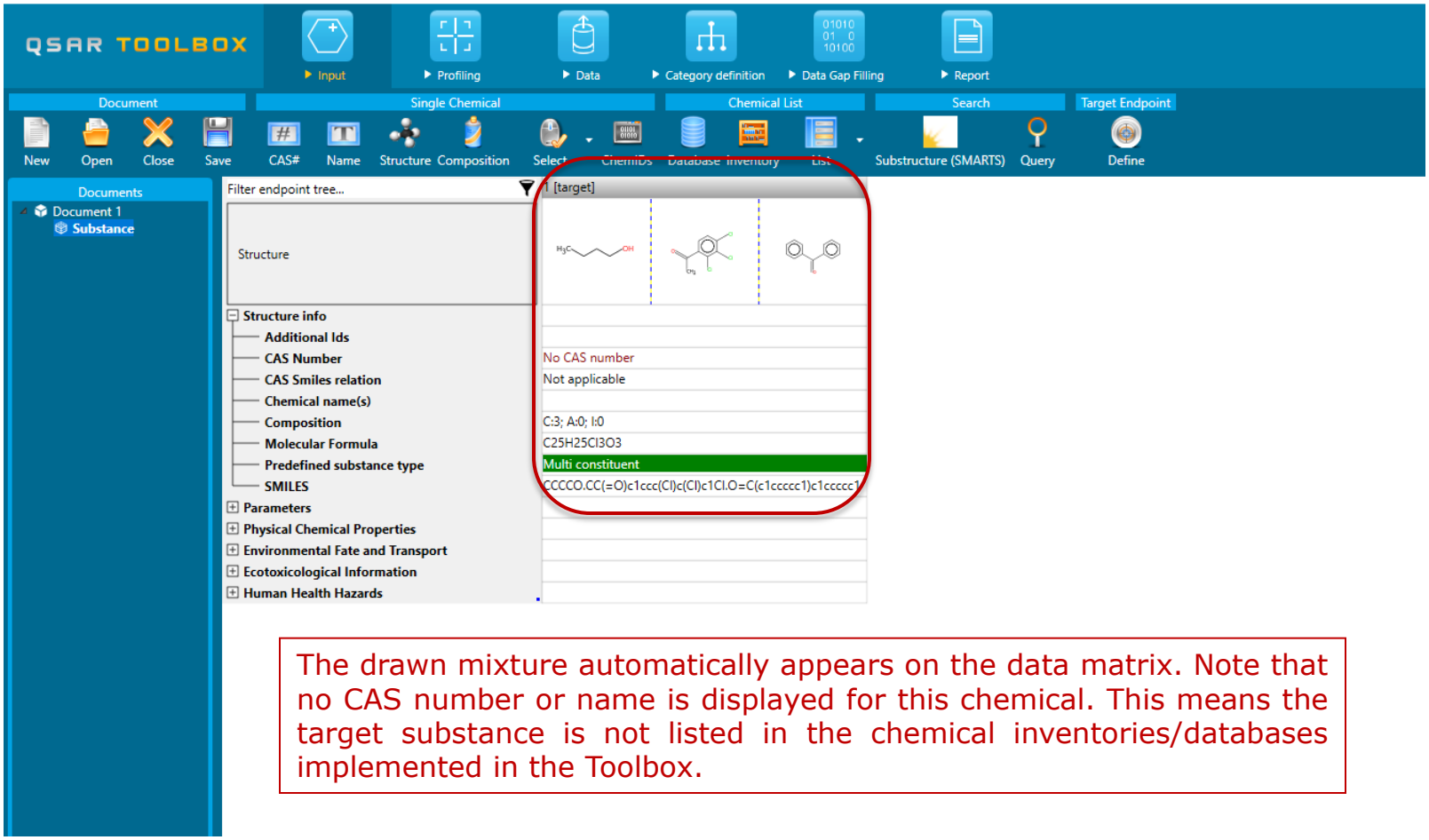
For Constituent 3 of the target mixture define:

1) Family: **Mass fraction**
2) Unit: **weight %**
3) qualifier: Equal to (i.e. **=**)
4) quantity: **9**



Now all three constituents of the target mixture are defined. Click **OK** (5).

Input
Target substance identity

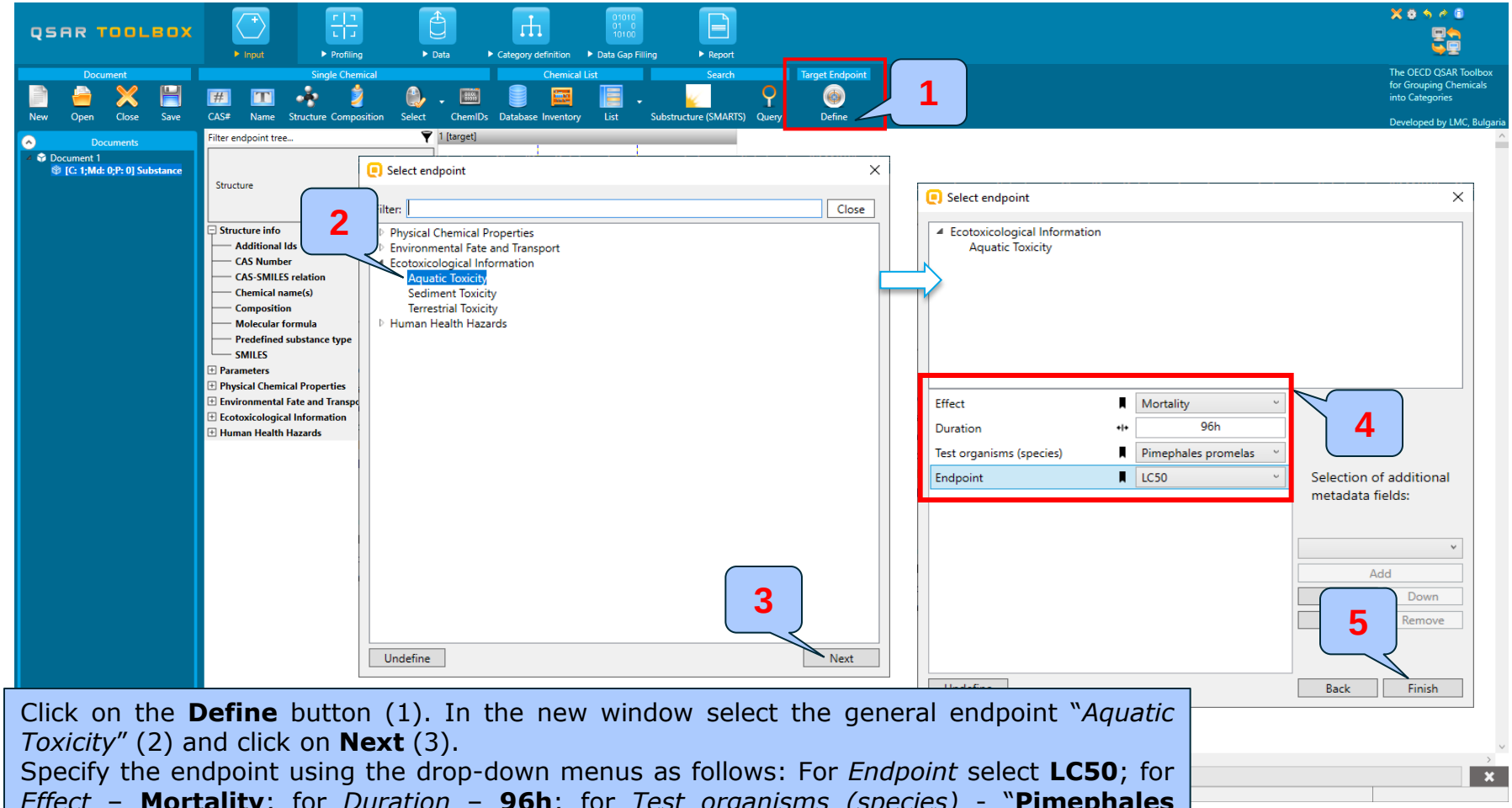


The drawn mixture automatically appears on the data matrix. Note that no CAS number or name is displayed for this chemical. This means the target substance is not listed in the chemical inventories/databases implemented in the Toolbox.

Input
Define target endpoint - overview

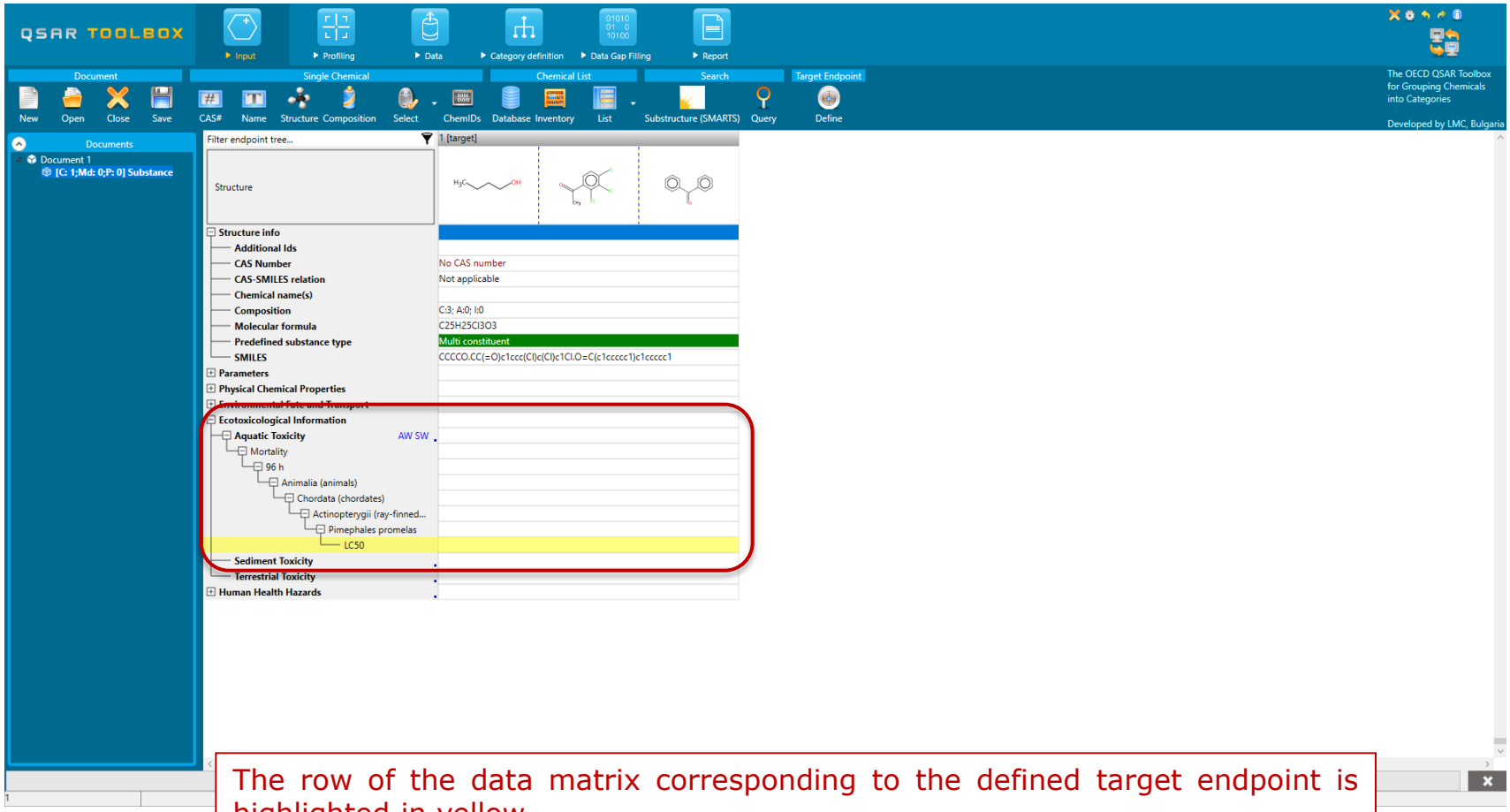
- The Define target endpoint functionality allows entering the endpoint of interest e.g., EC3, LC50, gene mutation etc.
- The relevant profiles and databases become highlighted in color once the targeted endpoint is preliminary defined by this functionality;
- Calculation of alert performance (AP) is only possible if the target endpoint is preliminary defined;
- There are different ways for defining the target endpoint (via the button from the Input module or by right click from the endpoint tree). For more details press F1 button in order to see the online help.

Input
Define target endpoint



Click on the **Define** button (1). In the new window select the general endpoint "*Aquatic Toxicity*" (2) and click on **Next** (3).
Specify the endpoint using the drop-down menus as follows: For *Endpoint* select **LC50**; for *Effect* – **Mortality**; for *Duration* – **96h**; for *Test organisms (species)* - "**Pimephales promelas**" (4). Finally click on **Finish** (5).

Input
Define target endpoint

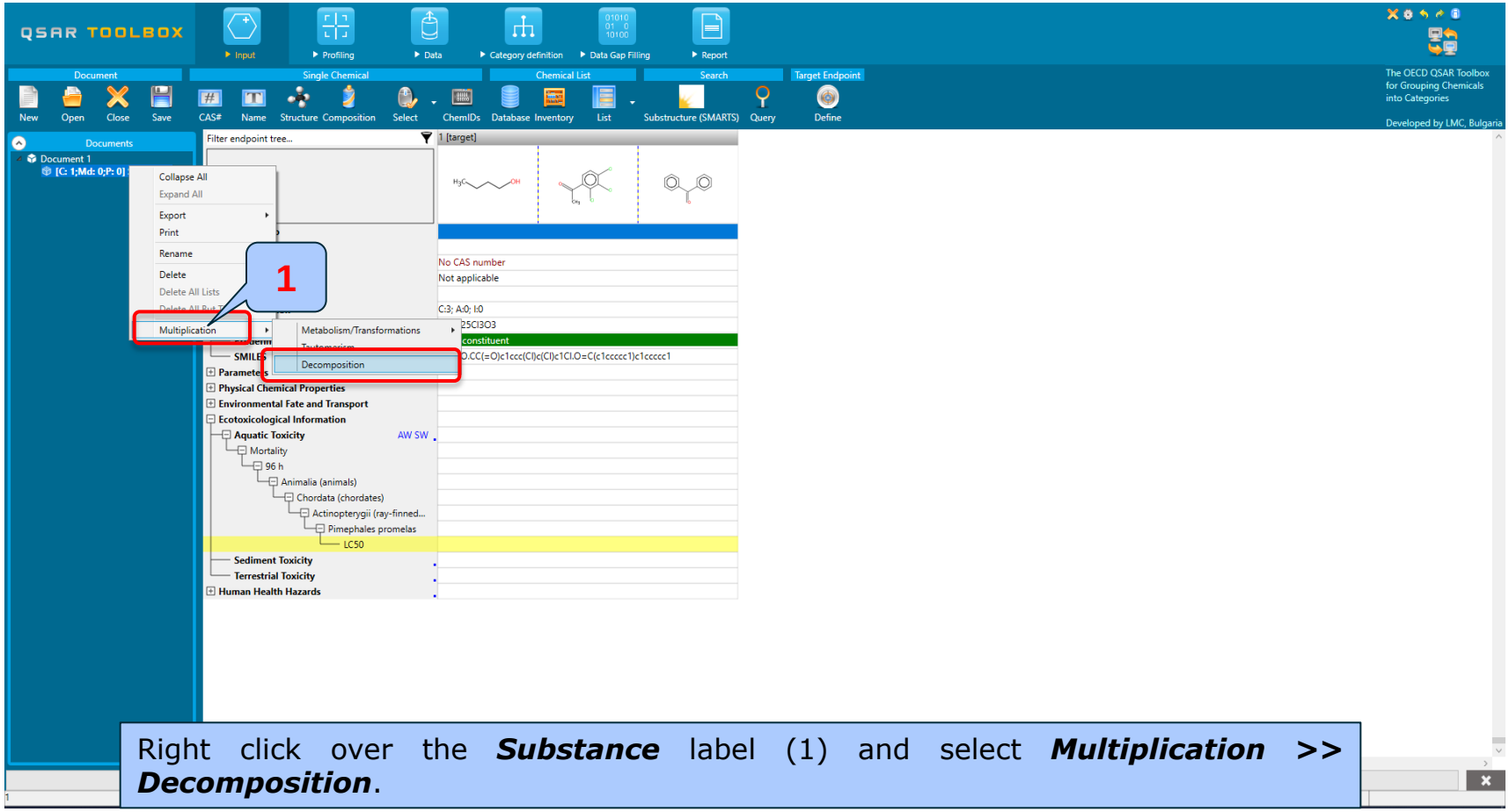


The row of the data matrix corresponding to the defined target endpoint is highlighted in yellow.

Input
Mixture decomposition

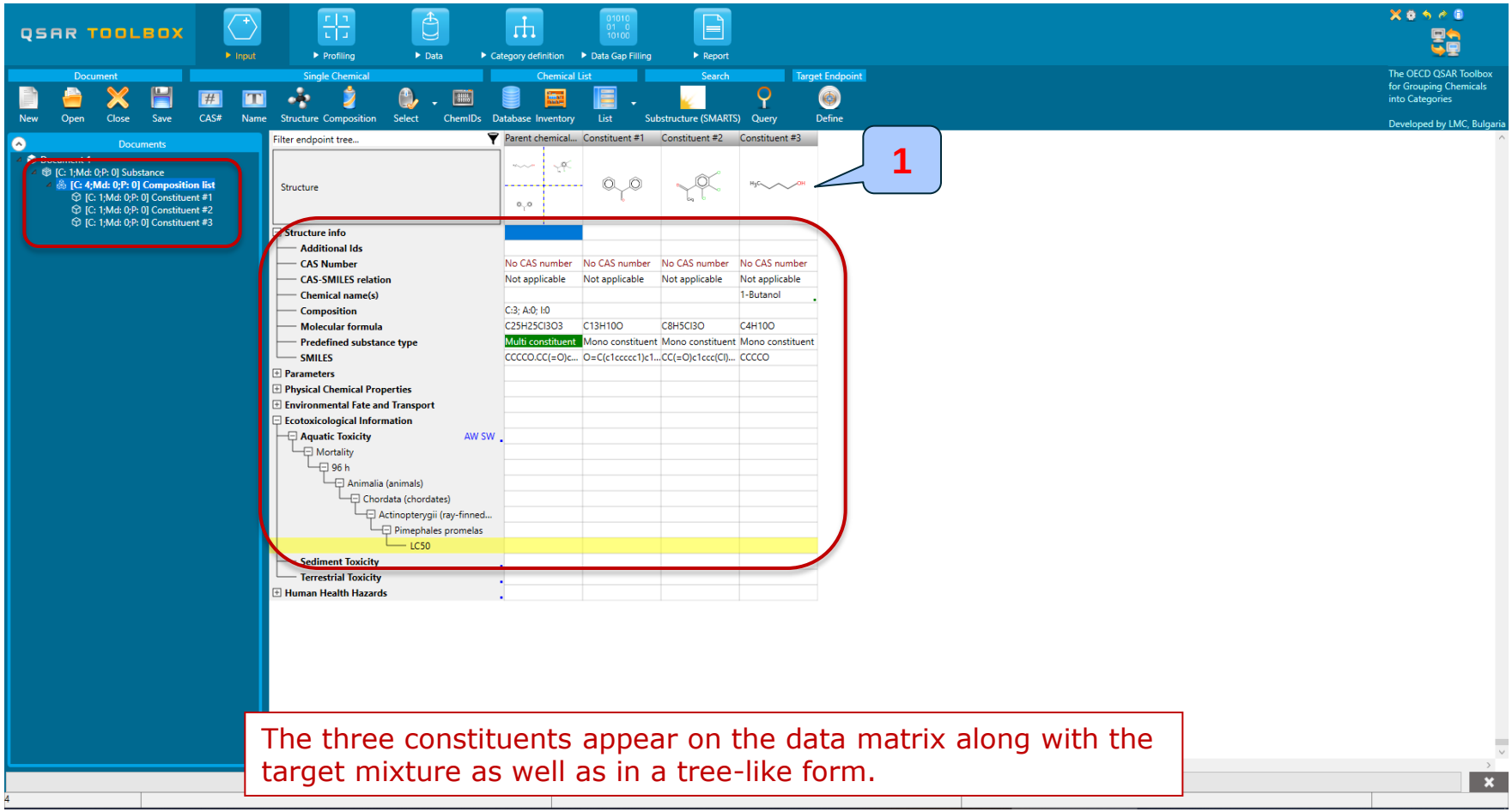
- In the current example we will predict the aquatic toxicity of a mixture based on its constituents.
- A specific option “Decomposition” allows all constituents of a mixture as well as available additives/impurities to be shown in the data matrix.
- Once the constituents are on the data matrix, the user can handle them as individual substances and further, to use them for predicting the whole mixture.

Input
Mixture decomposition



Right click over the ***Substance*** label (1) and select ***Multiplication >> Decomposition***.

Input
Mixture decomposition



The three constituents appear on the data matrix along with the target mixture as well as in a tree-like form.

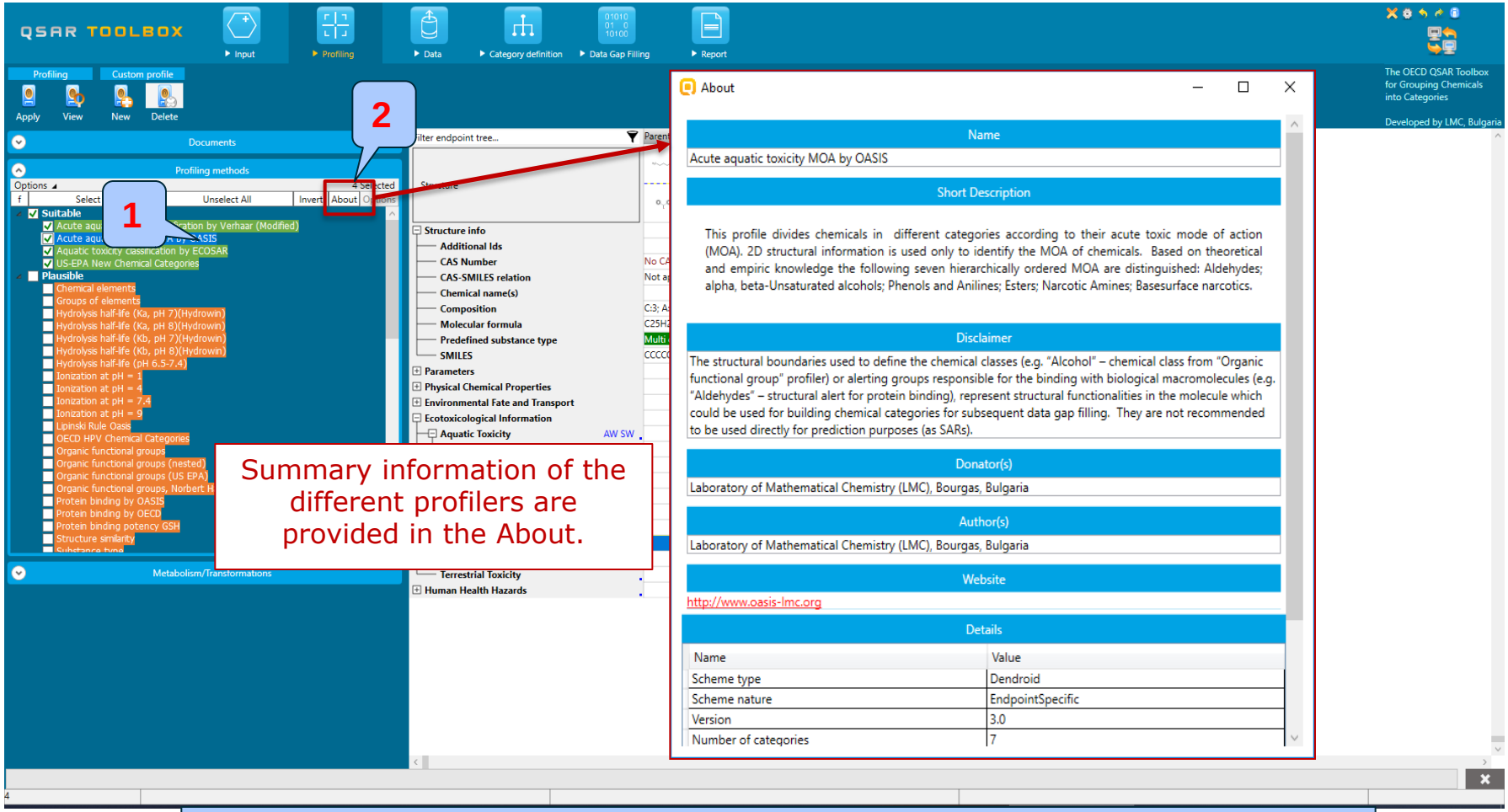
Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - **Profiling**

Profiling
Overview

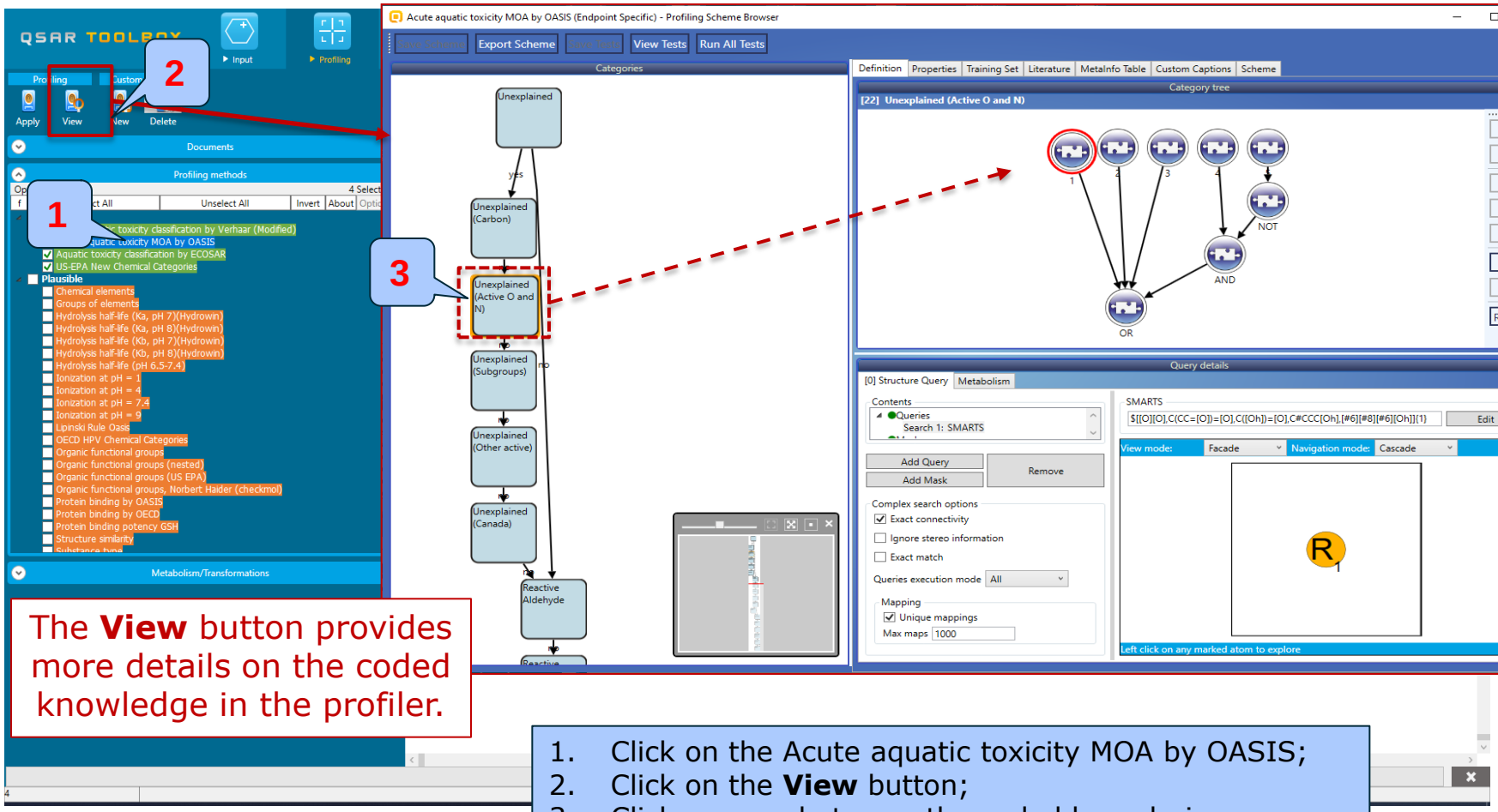
- "Profiling" refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- The available information includes likely mechanism(s) of action, as well as observed or simulated metabolites.
- For most of the profilers, background information can be retrieved by selection of a profilers (for example, Acute aquatic toxicity MOA y OASIS) and then click on "About" or "View" (see the next slide).

Profiling
Side-Bar to Profiling



1. Click on the Acute aquatic toxicity MOA by OASIS; 2. Click on the **About** button.

Profiling
Side-Bar to Profiling

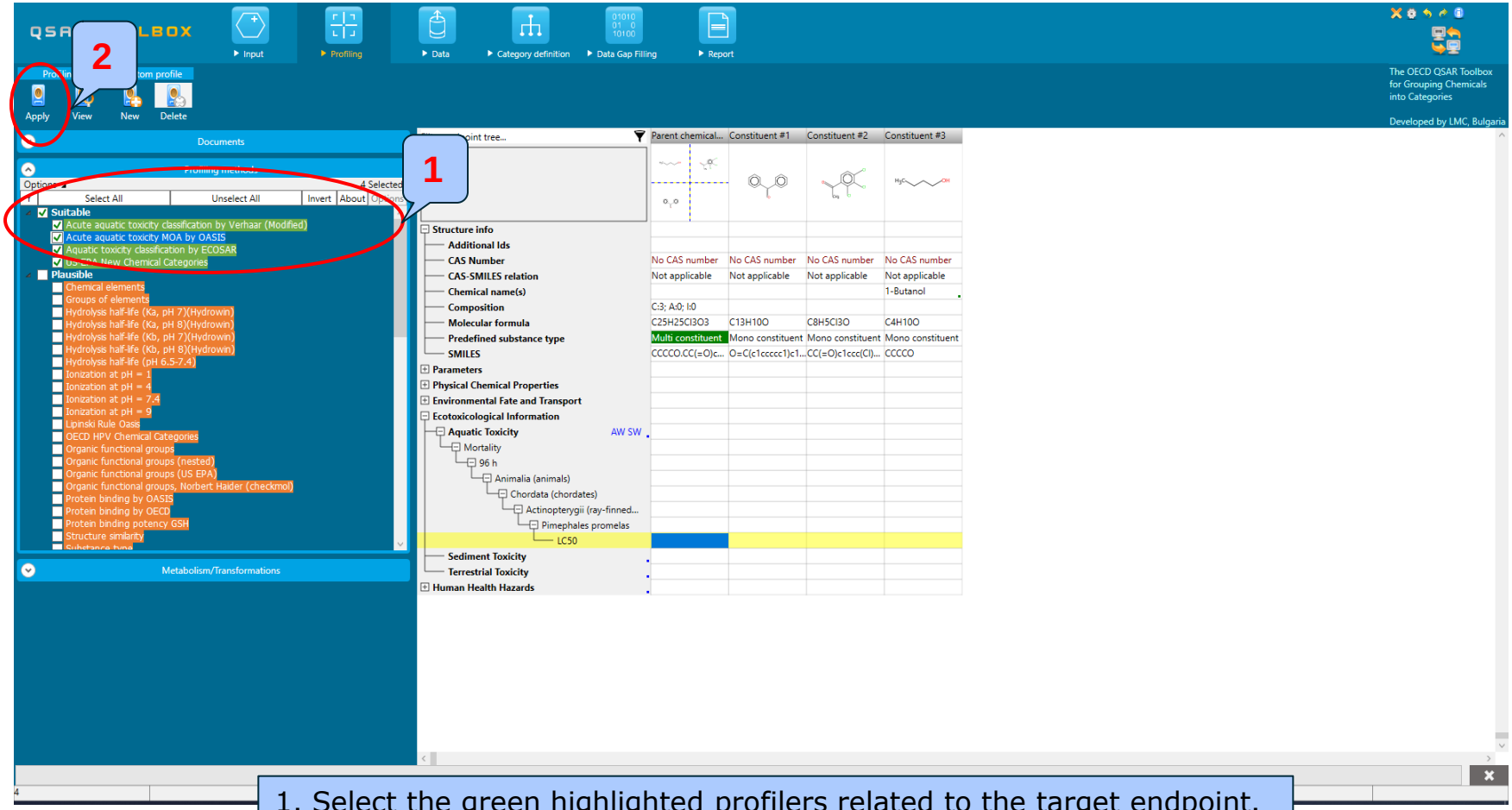


The **View** button provides more details on the coded knowledge in the profiler.

1. Click on the Acute aquatic toxicity MOA by OASIS;
2. Click on the **View** button;
3. Click on a node to see the coded boundaries.

Profiling

Profiling the target substance



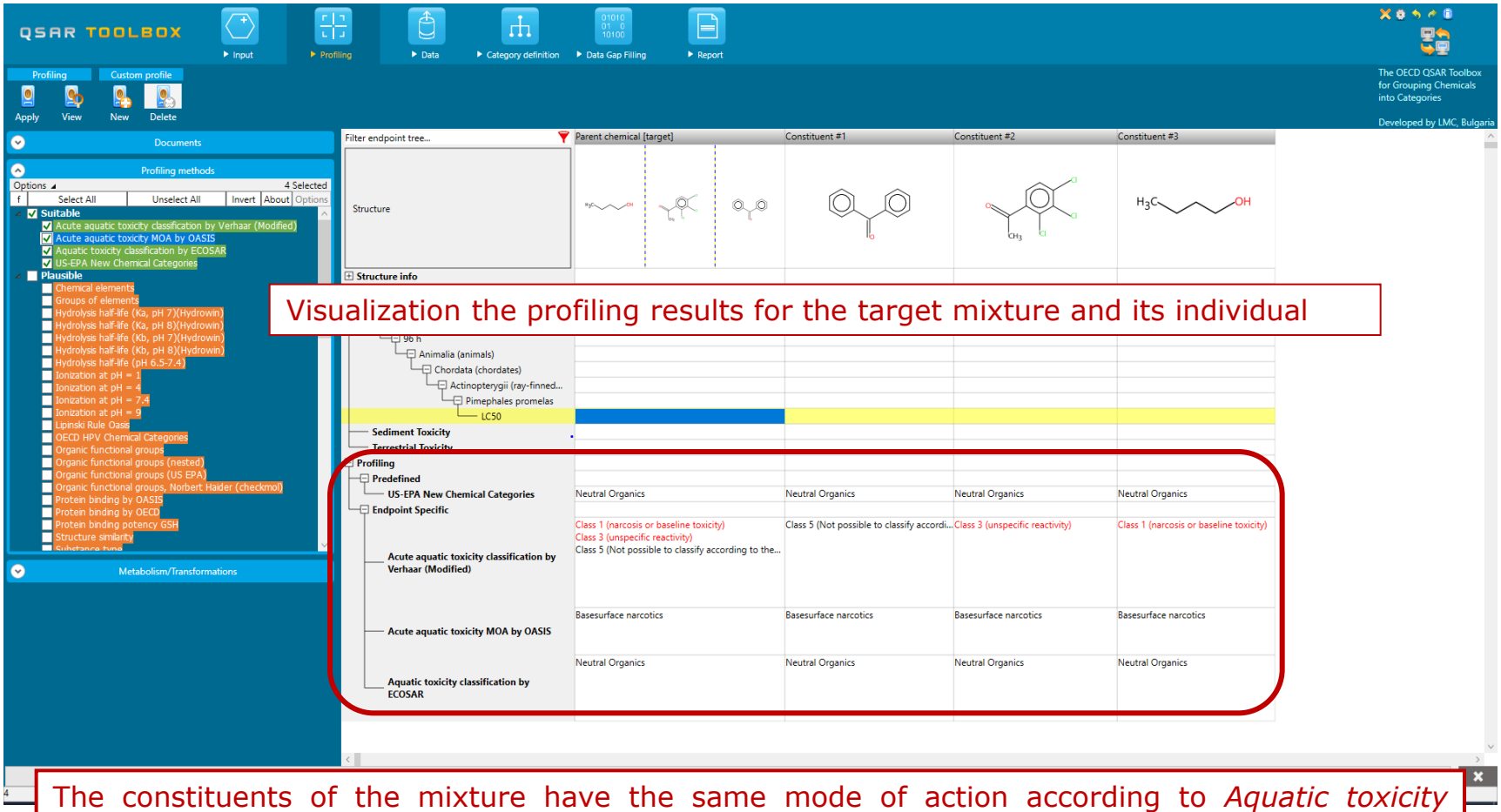
1. Select the green highlighted profilers related to the target endpoint.
2. Click **Apply**

Profiling
Profiling the target substance

- The actual profiling will take several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appear as a dropdown box under the target substance.
- In this example the target mixture and its constituents are profiled by all profilers defined as suitable (highlighted in green) for aquatic toxicity:
    - Acute aquatic toxicity classification by Verhaar (Modified);
    - Aquatic toxicity classification by ECOSAR;
    - US-EPA New Chemical Categories;
    - Acute aquatic toxicity classification MOA by OASIS.

Profiling

Profiling the target substance



Visualization the profiling results for the target mixture and its individual

The constituents of the mixture have the same mode of action according to *Aquatic toxicity classification by ECOSAR*; *US-EPA New Chemical Categories* and *Acute aquatic toxicity classification MOA by OASIS* profilers

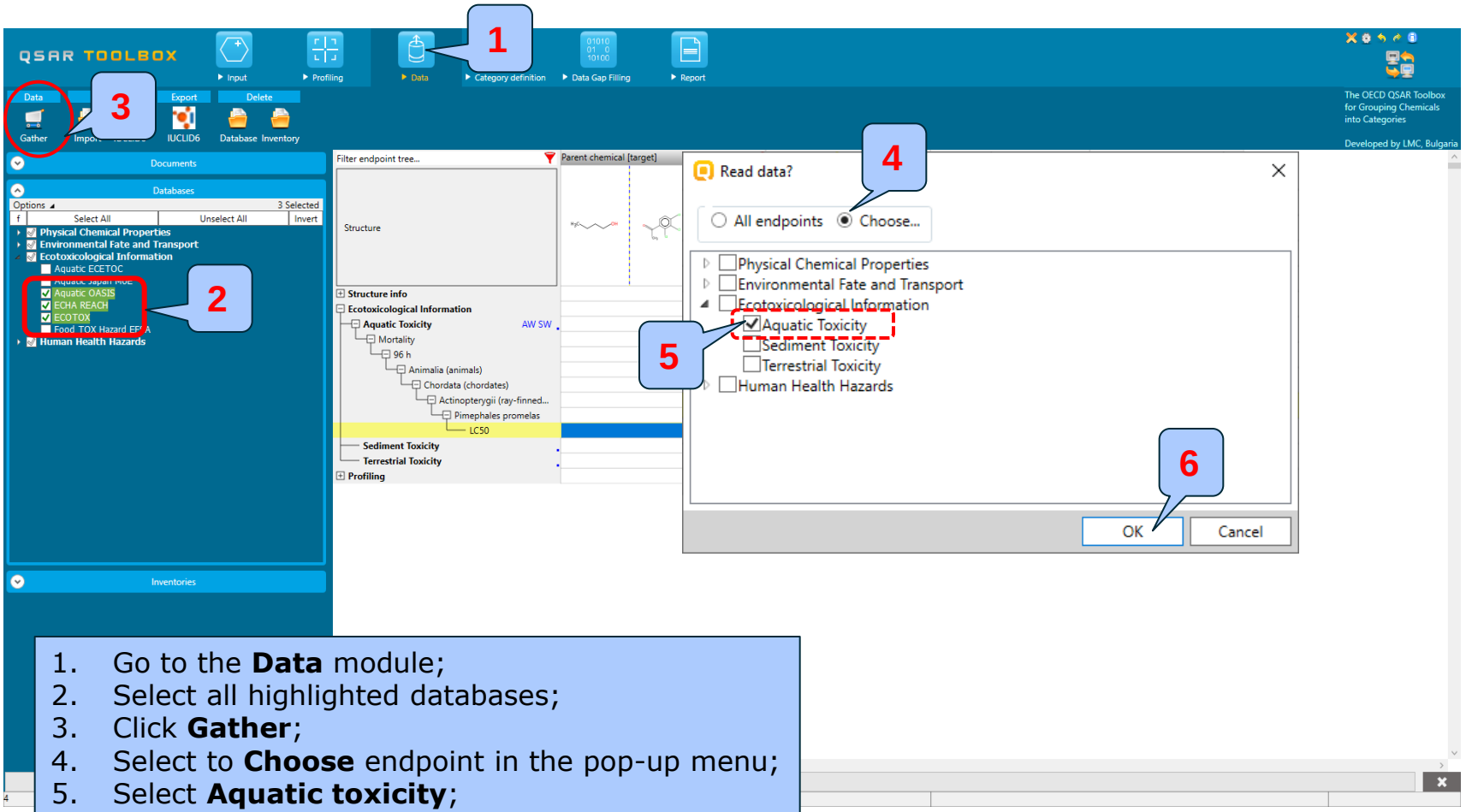
Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - **Data**

Data

- "Data" refer to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox database.
- Data gathering can be executed in a global fashion (i.e., collecting all data of all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).
- In case of defined endpoint, the databases containing data for this endpoint, will be highlighted in green. The user could select all or just some of them.
- In this example, we will use all highlighted databases.

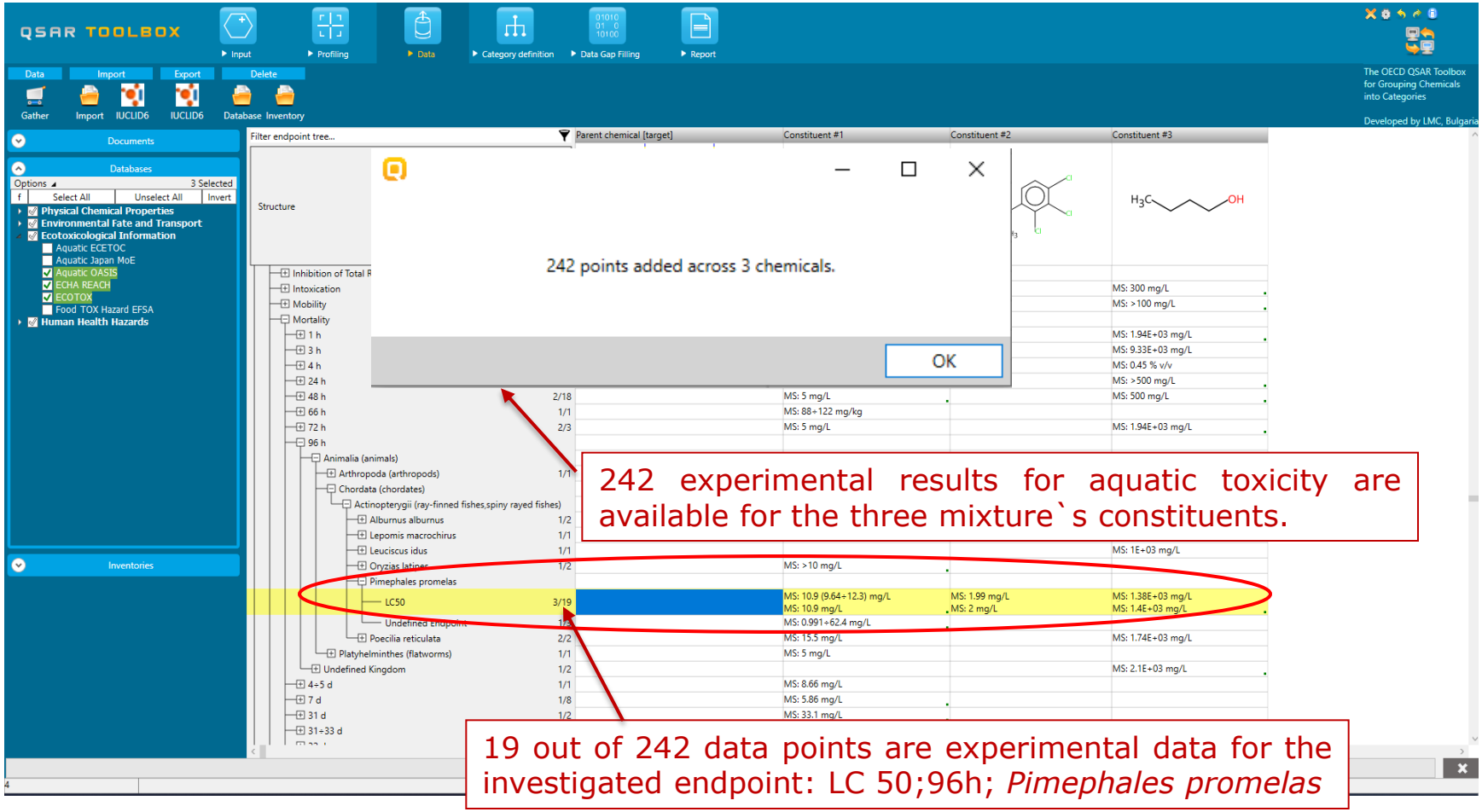
Data
Gather data



1. Go to the **Data** module;
2. Select all highlighted databases;
3. Click **Gather**;
4. Select to **Choose** endpoint in the pop-up menu;
5. Select **Aquatic toxicity**;
6. Confirm by **OK**.

Data

Available experimental data



Recap

- We have entered a mixture with three constituents. The quantity of each constituent has been defined.
- The profiling results showed the same mode of action for the three constituents of the mixture.
- We have collected the available aquatic toxicity experimental data for the constituents of the target mixture. 19 experimental results for the defined target endpoint have been found.
- Now we are ready to continue with next step of the workflow - Data Gap Filling.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - **Data Gap filling**

Data Gap Filling
Overview

- "Data Gap Filling" module give access to two different data gap filling tools:
  - **Independent MOA-** all components are with different mode of action
  - **Similar MOA-** all components are with similar mode of action
- More details about different MOAs could be found in F1 (also more details are given on next six slides)
- In this particular case all components of the current mixture are with similar mode of action. In this respect Similar MOA is applied

Data Gap Filling
Independent MOA

Assumption – combined effect can be calculated from the effects caused by the individual mixture components by following the statistical concept of independent random events

Mixture response: $$E(C_{Mix}) = 1 - \prod_{i=1}^{N} [1 - E(C_i)]$$

$E(C_{Mix})$ - the effect provoked by the total mixture

$E(C_i)$ - the effects that the individual components would cause if applied singly at that concentration at which they are present in the mixture

Problem - dose-response relationships are practically unknown

Data Gap Filling
Similar MOA

Assumption – components in a mixture contribute to the joint effect, in proportion to their prevalence and individual potency

- **Components act at the same target site**
- **Components act by the same mechanism**
- **Components have similar effect (rather than mechanism)**

Method for calculation of toxic effects of a mixture with components acting by same mechanisms is given on next slide

Data Gap Filling
Similar MOA

Toxicity Equivalence Factor

$$TEF_j^{(i)} = \frac{ED_{resp}^{(i)}}{ED_{resp}^{(j)}}$$

i – index (reference) chemical
ED_{resp} – dose (concentration) of a chemical that cause a specified response (fraction of animals that respond, fractional change in a measured physiological value, etc.)

Toxicity Equivalent Concentration

$$TEC_j^{(i)} = TEF_j^{(i)} d_j$$

Dose (concentration) of the reference chemical *i* that will cause the same effect as chemical *j* at dose (concentration) d_j

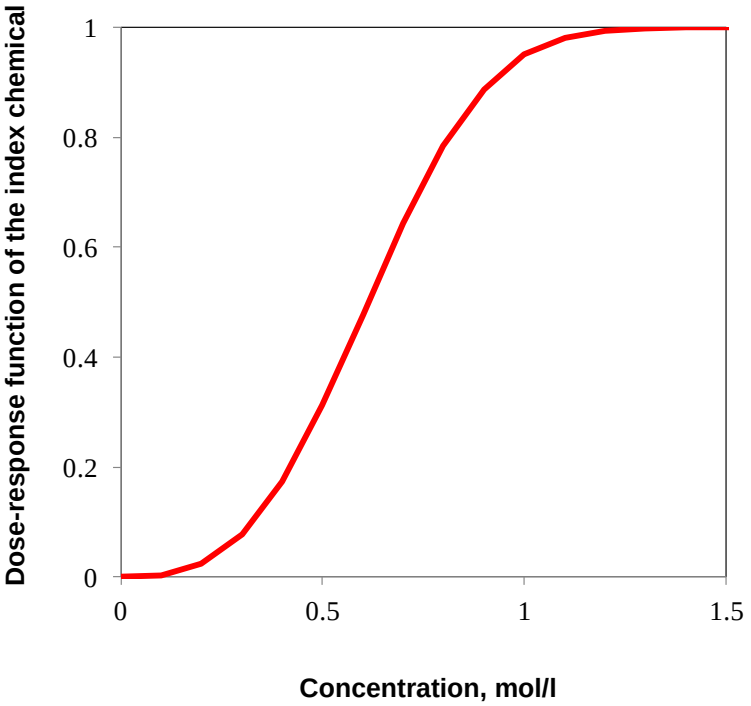
Index Toxicity Equivalent Concentration

$$ITEC = \sum_{j=1}^{J} TEC_j^{(i)} = \sum_{j=1}^{J} TEF_j^{(i)} d_j$$

Equivalent dose (concentration) of the reference chemical *i* that will cause the same effect as the mixture

Data Gap Filling
Similar MOA

Toxic effect of mixture - response (fraction of animals that respond, fractional change in a measured physiological value, etc.) as a result of exposure to mixture



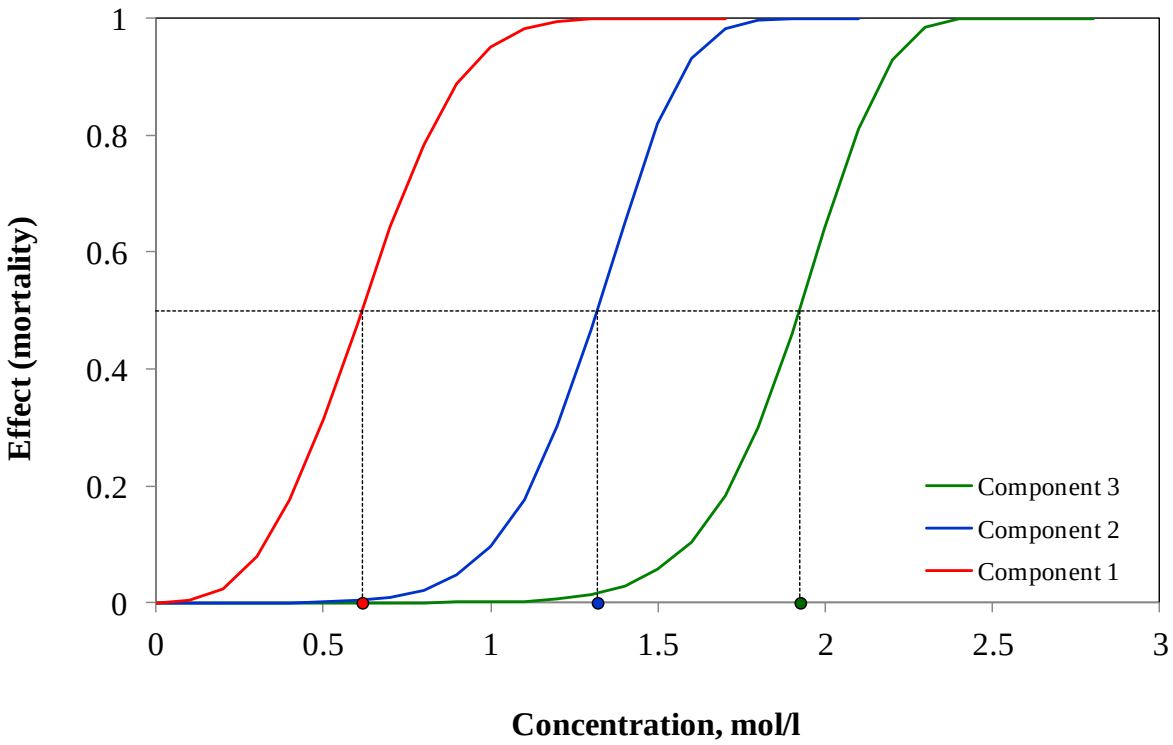
$$Effect^{Mixture} = f_i(ITEC)$$

f_i - dose-response function of the index chemical

Illustration of calculating effect of mixture is given on next two slides

Data Gap Filling
Similar MOA (Illustration)

Reference chemical: Component 1 (*i* = 1)



Toxicity equivalent factors

$$TEF_j^{(1)} = \frac{LC_{50}^{(1)}}{LC_{50}^{(j)}}$$

Toxicity Equivalent concentrations

$$TEC_j^{(1)} = TEF_j^{(1)} C_j$$

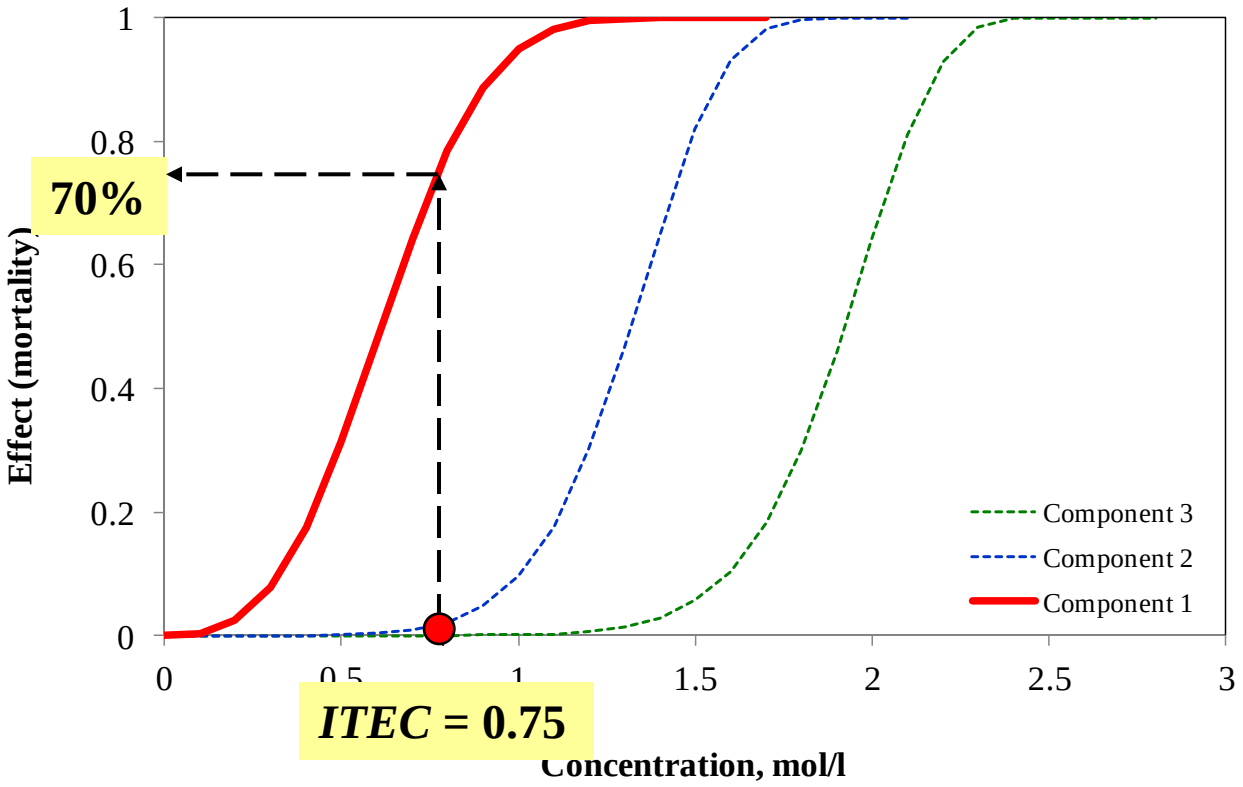
Index Toxicity Equivalent Concentration

$$ITEC = \sum_{j=1}^{J} TEC_j^{(1)}$$

$$ITEC \rightarrow Effect^{Mixture} = f_i(ITEC)$$

Data Gap Filling
Similar MOA (Illustration)

Reference chemical: Component 1 (*i* = 1)



Toxicity equivalent factors

$$TEF_j^{(1)} = \frac{LC_{50}^{(1)}}{LC_{50}^{(j)}}$$

Toxicity Equivalent concentrations

$$TEC_j^{(1)} = TEF_j^{(1)} C_j$$

Index Toxicity Equivalent Concentration

$$ITEC = \sum_{j=1}^{J} TEC_j^{(1)}$$

$$ITEC = 0.75 \Rightarrow Effect^{Mixture} = f_i(ITEC) \approx 70\% \, mortality$$

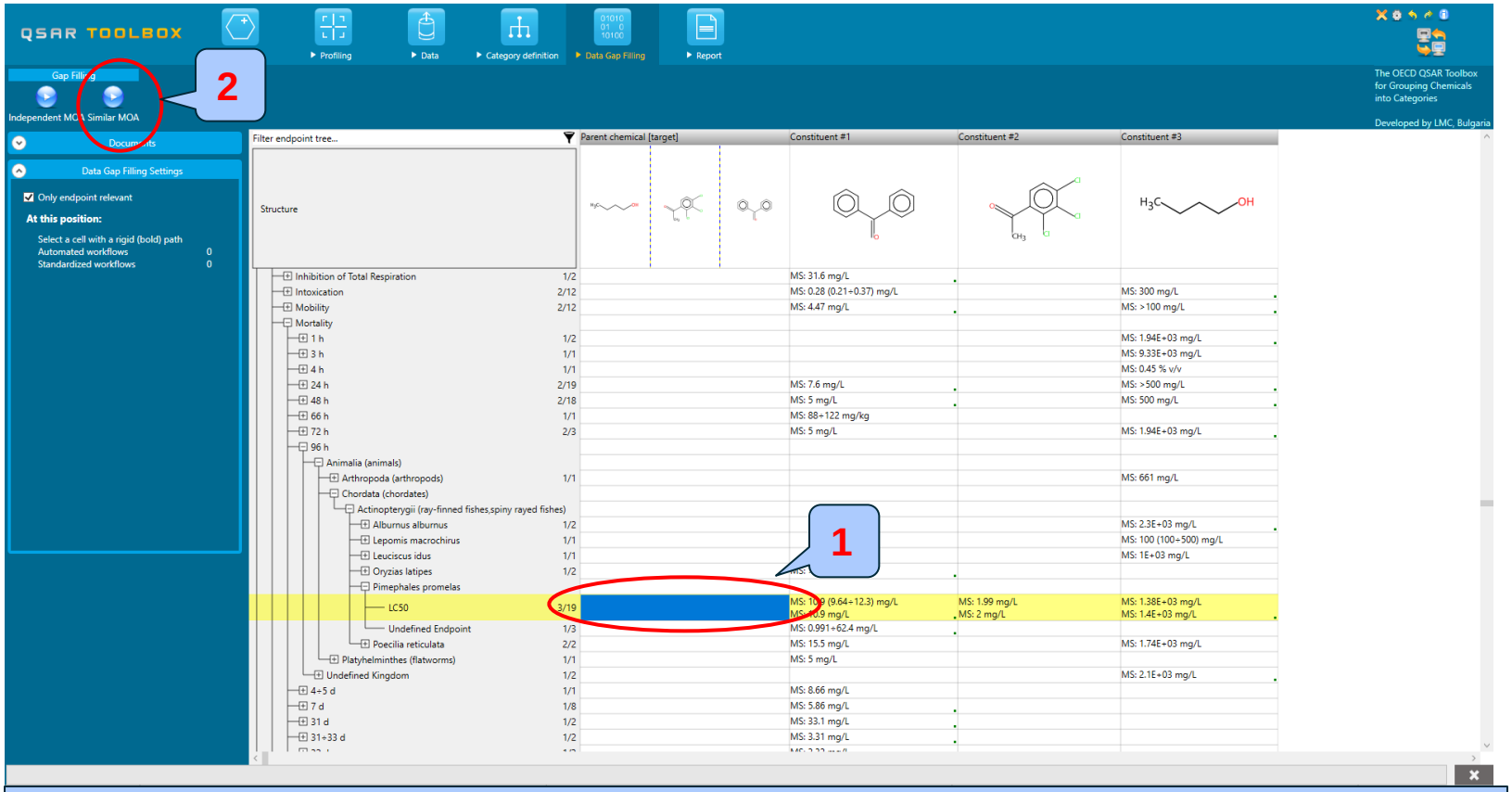
Data Gap Filling
Case study

- In this particular case all components of the current mixture are with similar mode of action. In this respect Similar MOA is applied

- Application of Similar MOA for our case study is illustrated on next slides

Data Gap Filling

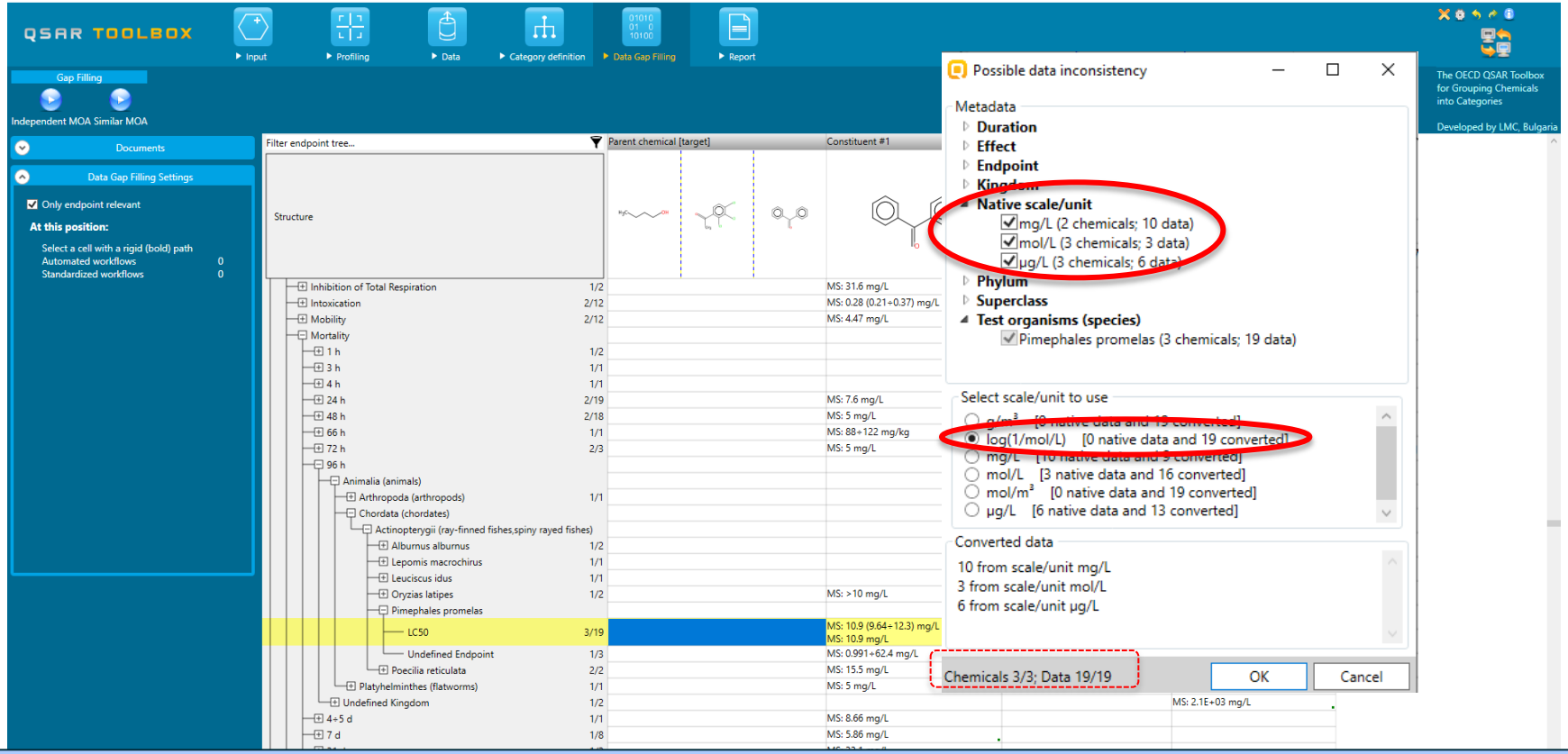
Apply Similar MOA



1. Highlight the data box corresponding to ***Pimephales promelas*/LC50/96h** under the target substance;
2. Click on **Similar MOA** button.

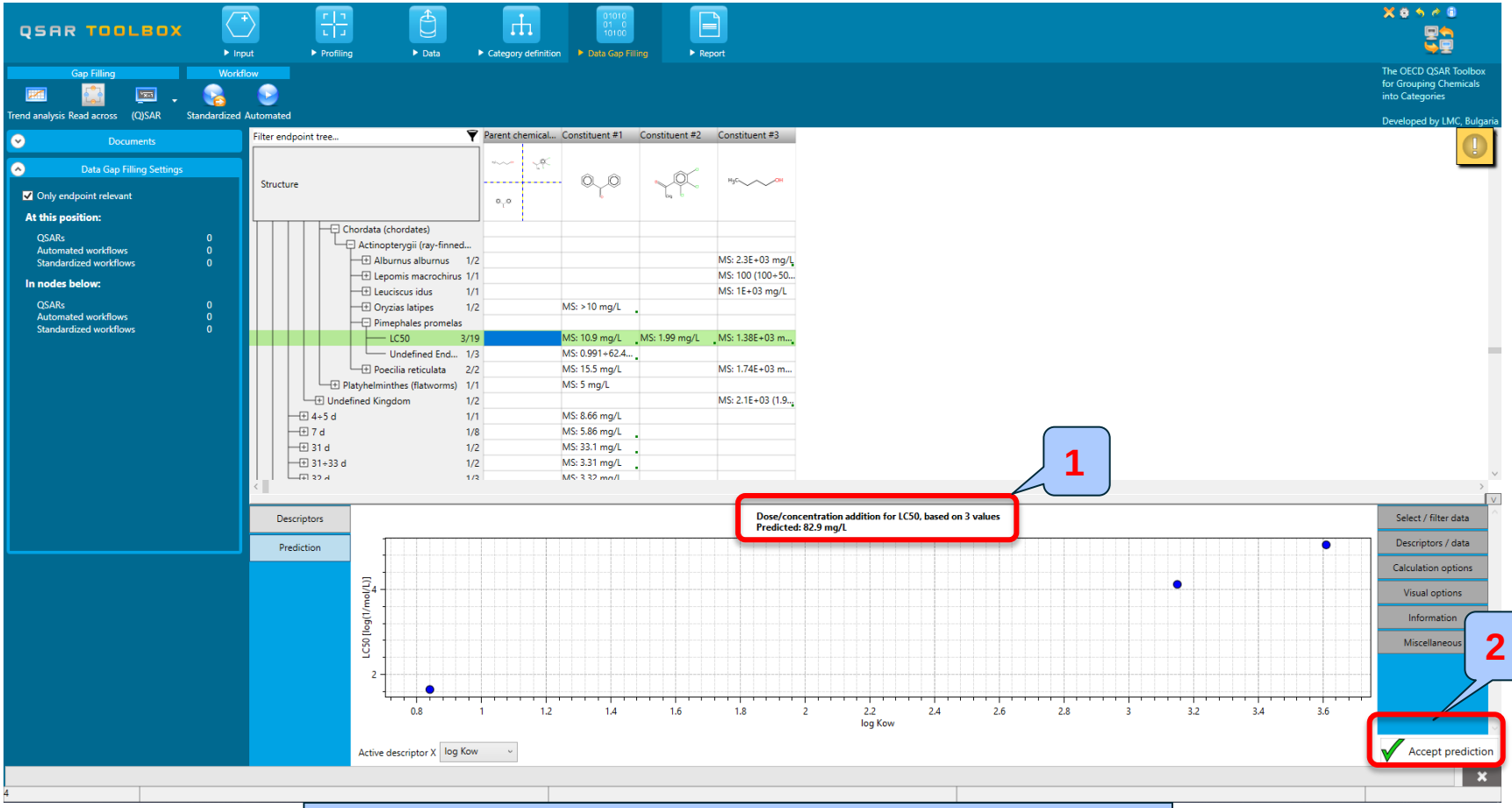
Data Gap Filling

Apply Similar MOA



The user will be informed if there is different experimental data in the appeared window. In our case there are 3 chemicals with 19 data belonging to 3 different units: mg/l; mol/l and µg/l. The data could be converted each other. By default for the gap filling is used log (1/mol/L) unit, which is a expected mathematical expression for linear relationships. Click **OK**

Data Gap Filling
Results of Similar MOA

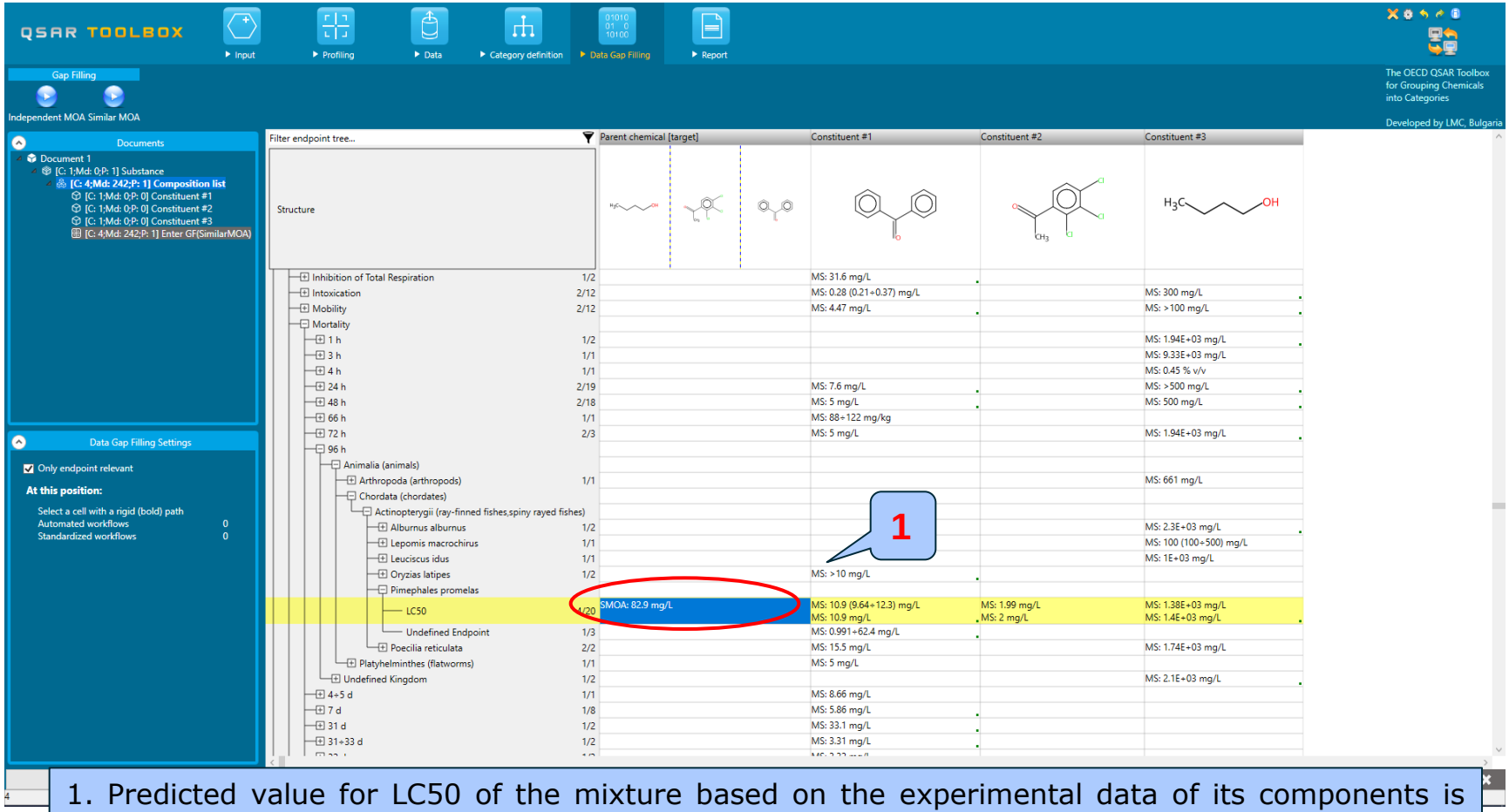


1. Predicted result is 82.9 mg/l; 2. Click Accept prediction

Data Gap Filling
Results

- The components of the mixture have the same mode of action.
- By accepting the prediction the data gap is filled (see next screen shot).

Data Gap Filling
Predicted value for LC50



1. Predicted value for LC50 of the mixture based on the experimental data of its components is **82.9 mg/l**.

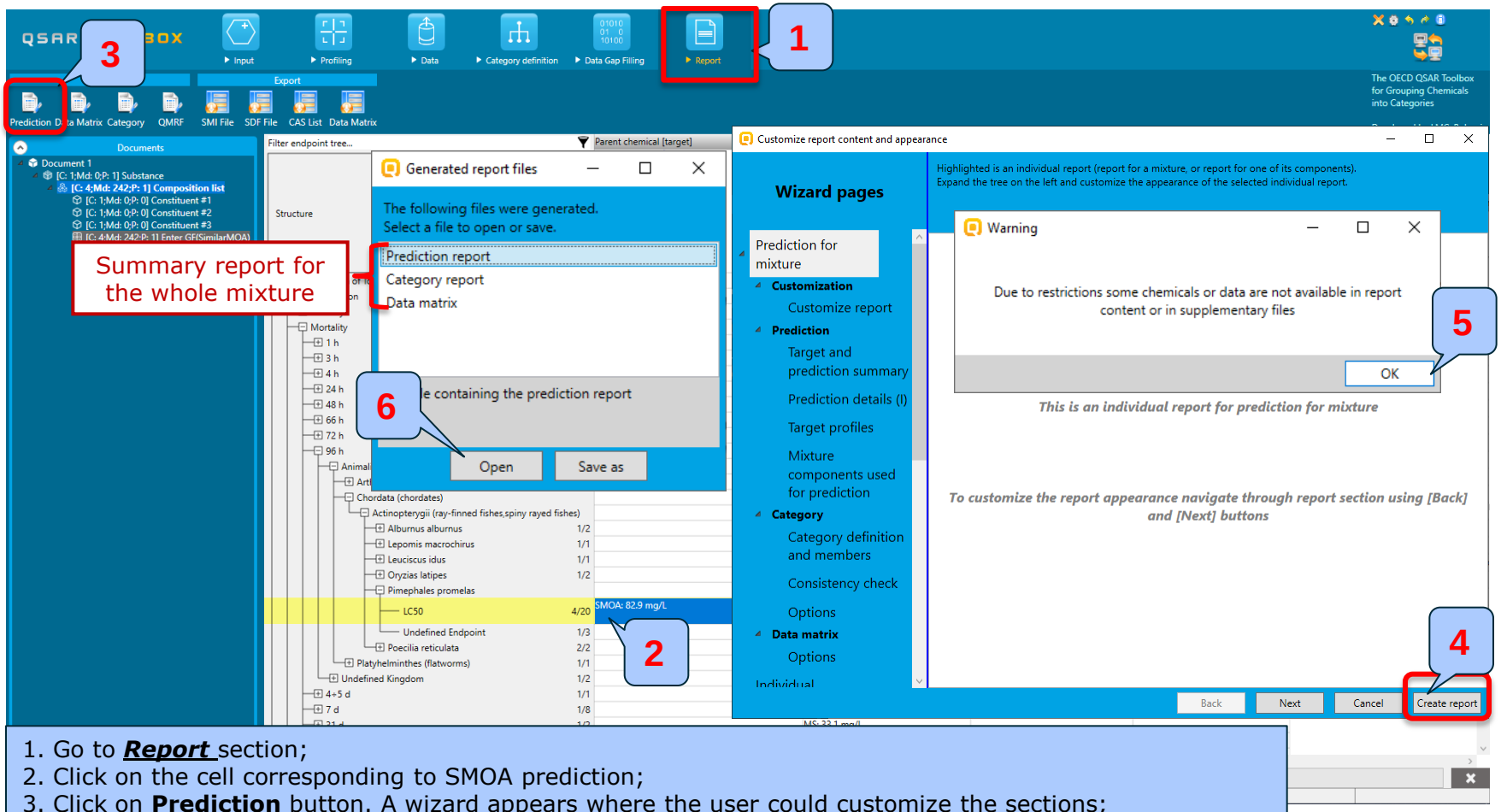
Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Data Gap filling
  - **Report**

Report

- Remember the report module allows the user to generate a report on the predictions performed with the Toolbox.
- The report can be printed or saved in different formats.
- Generating the report is shown on next screenshots.

Report



1. Go to ***Report*** section;
2. Click on the cell corresponding to SMOA prediction;
3. Click on **Prediction** button. A wizard appears where the user could customize the sections;
4. Click **Create report.**
5. Click **OK** on the appeared message; Three report files for the mixture prediction could be generated;
6. Select **one of the reports** and click **Open** button.

Report

QSAR Toolbox prediction for multicomponent substance

Based on observed and predicted data for mixture components

Date: 7 Apr 2020
Author(s):
Contact details:

Target information		
Structural information SMILES: CCCCO.CC(=O)c1ccc(O)c(C)c1Cl.O=C(c1ccccc1)c1ccccc1 Structure	Numerical identifiers CAS#: No CAS number Other: N/A	Chemical names

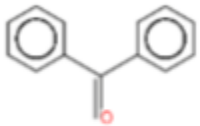
Prediction summary
Predicted endpoint: LC50; Mortality; Pimephales promelas; 96 h; No guideline specified
Predicted value: 82.9
Unit/scale: mg/L
Data gap filling method: Similar mode of action
Summary: *manually editable field*
Not provided by the user

	Quantity, mol %	LC50, mg/L
Target mixture	-	

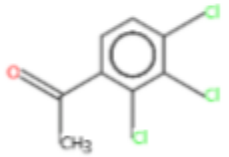
Prediction summary

Mixture component #1	3.90	13.8
Mixture component #2	0.353	2.00
Mixture component #3	95.8	1.68E3

Mixture component #1
O=C(c1ccccc1)c1ccccc1



Mixture component #2
CC(=O)c1ccc(Cl)c(C)c1Cl



Mixture component #3
CCCCO

Information of the mixture's constituents

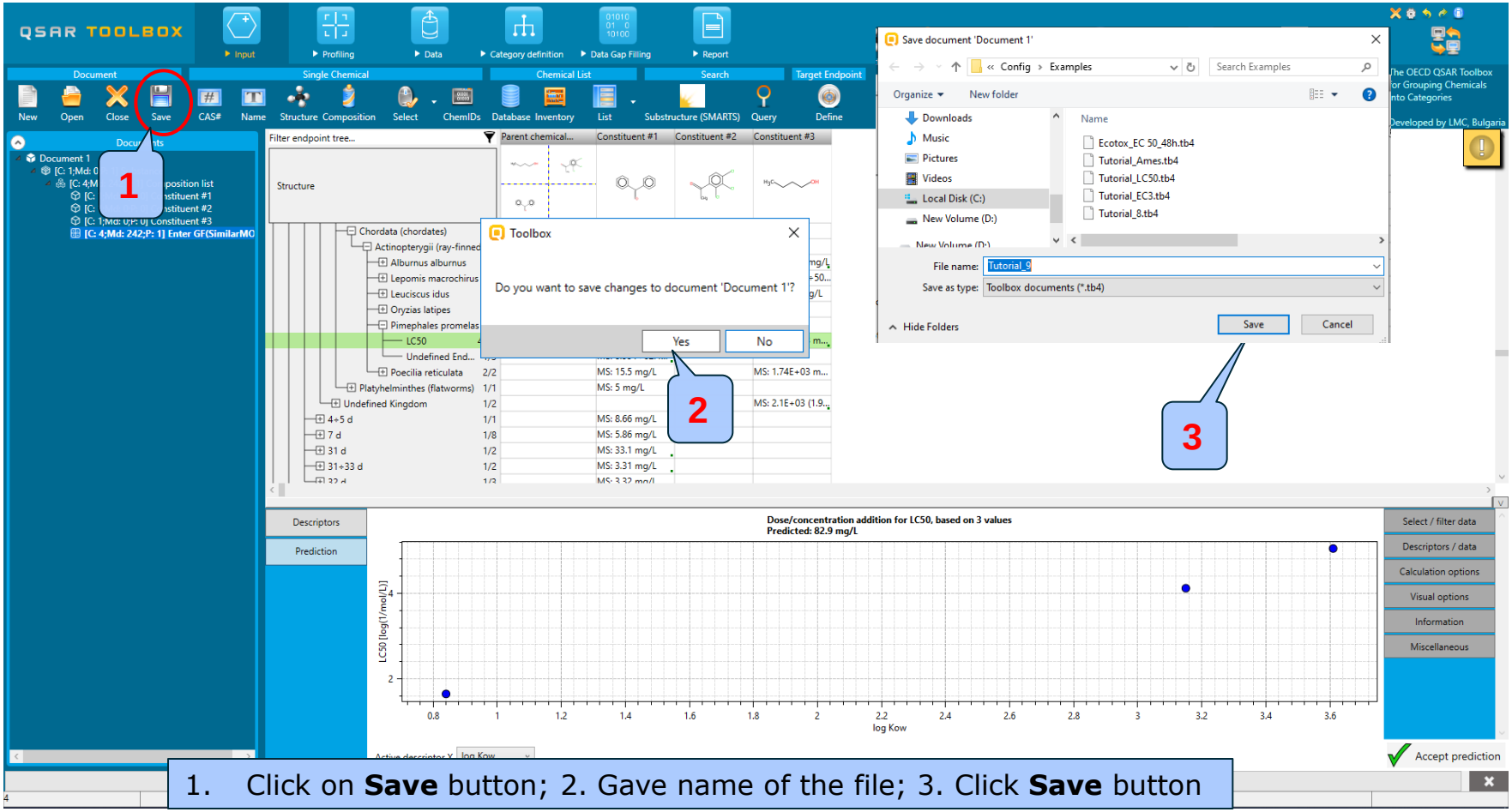
Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Data Gap filling
  - Report
- **Save the prediction result**

Saving the prediction result

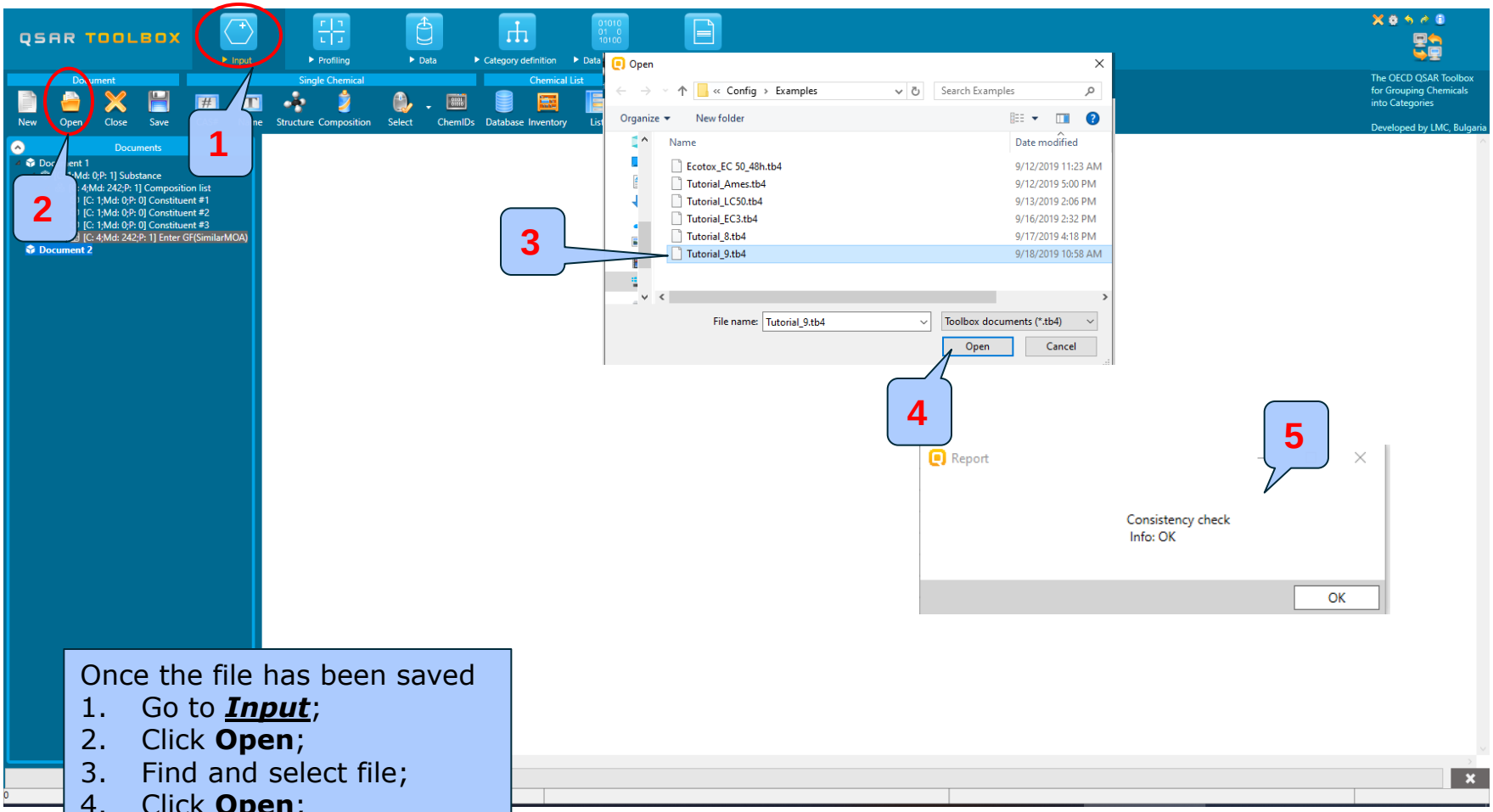
- Saving functionality allows storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions, etc.
- This functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document and later executing these actions in the same sequence in order to get the same result(s).
- Saving the file with TB prediction is illustrated on the next screenshots.

Saving the prediction result



1. Click on **Save** button; 2. Gave name of the file; 3. Click **Save** button

Open saved file



Once the file has been saved

1. Go to ***Input***;
2. Click **Open**;
3. Find and select file;
4. Click **Open**;
5. Click **OK**