QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD (Q)SAR Toolbox v.4.4.1

Example for predicting Skin Sensitization of a mixture with known components

Outlook

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

Background

This is a step-by-step presentation designed to take the user of the Toolbox through the workflow for prediction of skin sensitization of a mixture.

Outlook

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

Keywords

TARGET CHEMICAL - chemical of interest, in this case it is a mixture with defined components

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**
- The exercise
- Workflow
- Save the prediction

Objectives

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

Outlook

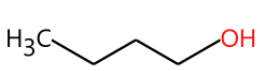
- Background
- Keywords
- Objectives
- **The exercise**
- Workflow
- Save the prediction

Exercise

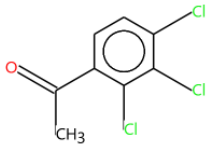
- In this exercise we will:
  - predict skin sensitization of target substance, which represent a mixture with defined constituents
  - Investigate the mode of action for each component of the mixture,
  - Gather available experimental data for target chemical,
  - Investigate skin sensitization of non-tested component,
  - Apply read across for non-tested component, and
  - Predict skin sensitization potential of mixture based on experimental data of tested compounds and predicted data of non-tested one.

- The target substance will consists of three constituents:

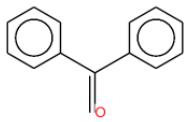
Constituent 1



Constituent 2



Constituent 3



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**

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 mixture

User alternatives for defining mixtures with known compositions:

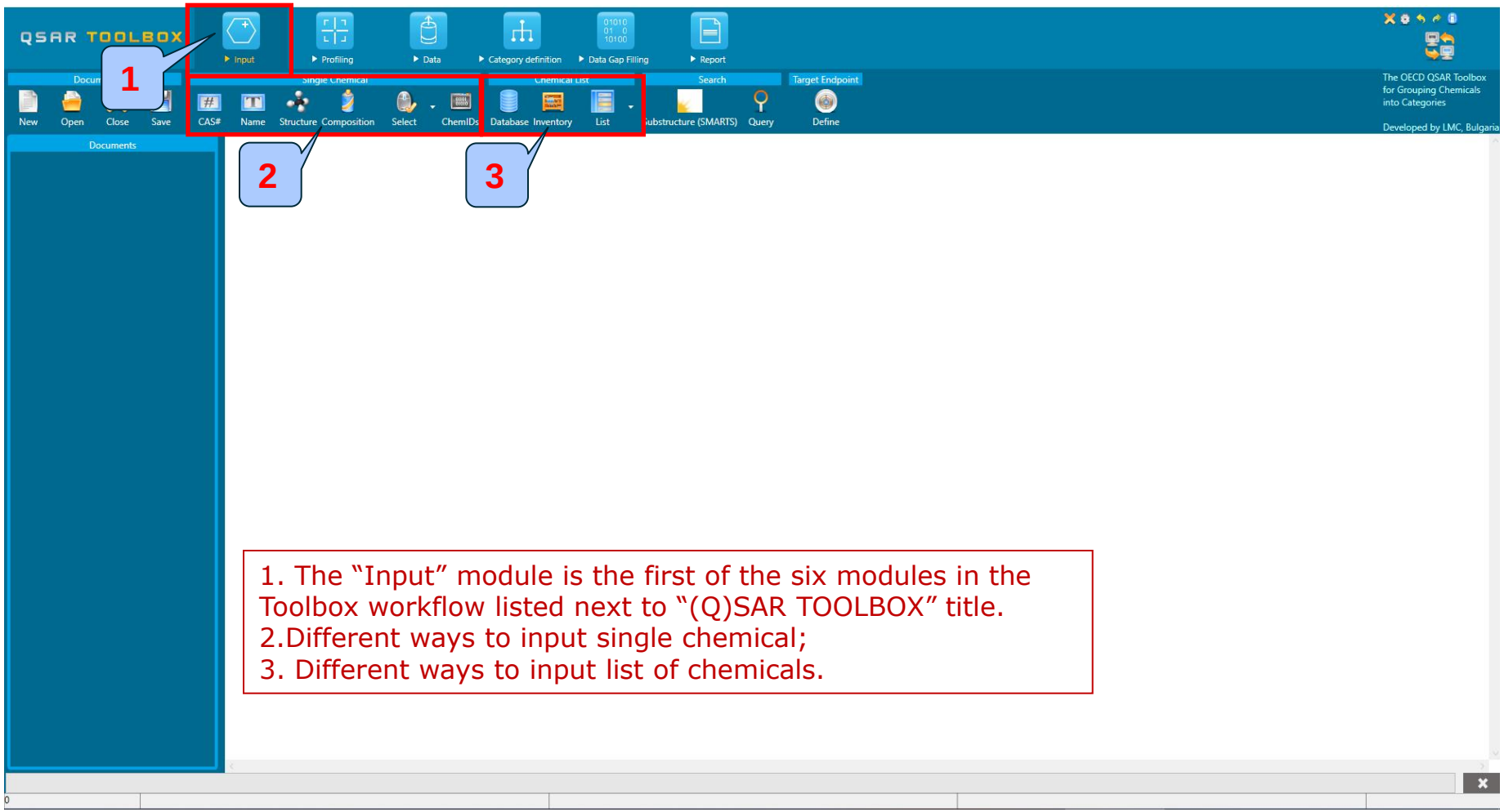
A. Single target substance

- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing mixture constituents and defining their quantities
- Select from User List/Inventory/Databases

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Chemical Input
Input Screen



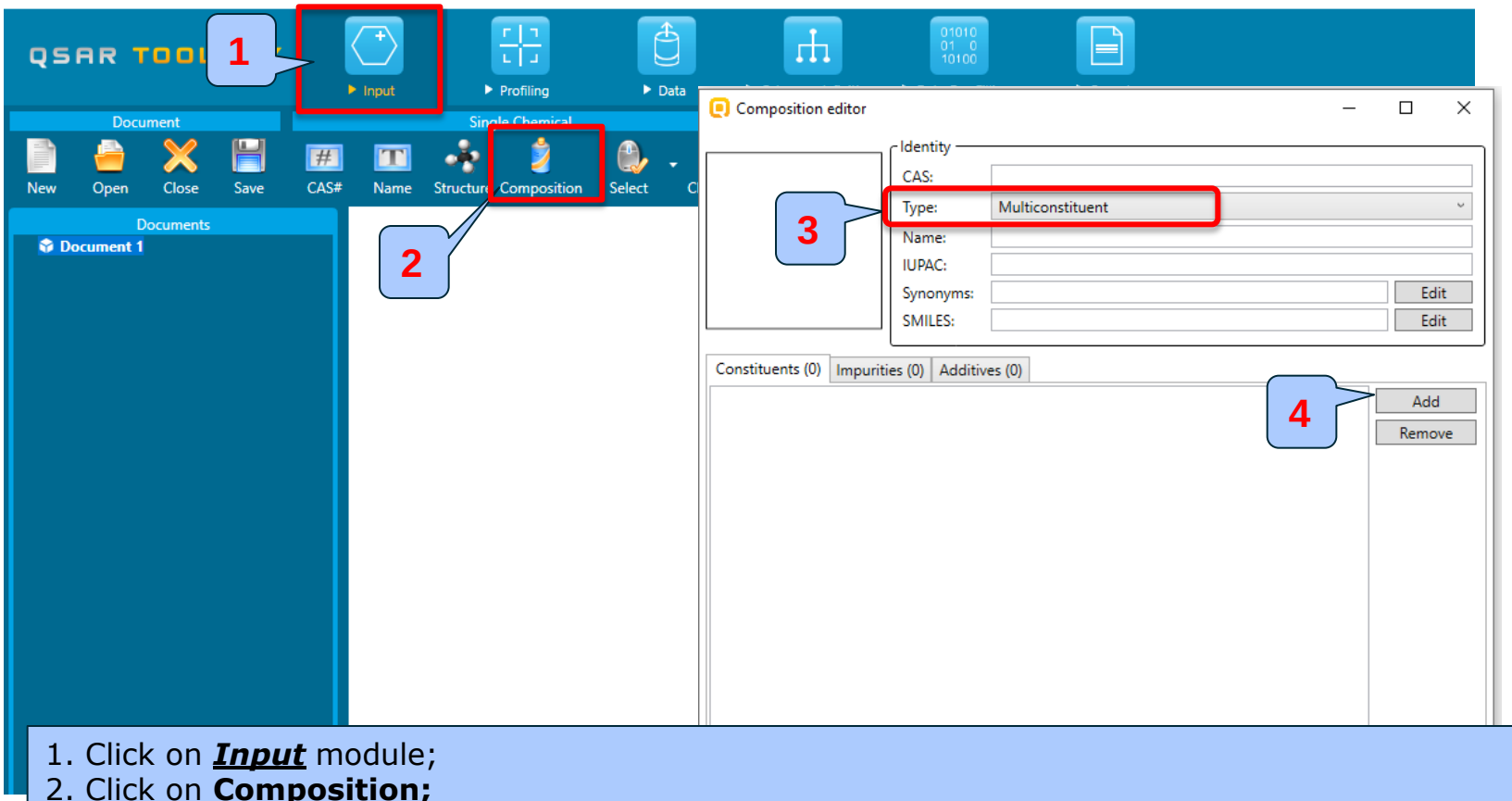
1. The "Input" module is the first of the six modules in the Toolbox workflow listed next to "(Q)SAR TOOLBOX" title.
2.Different ways to input single chemical;
3. Different ways to input list of chemicals.

Input
Input a mixture

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

Chemical input

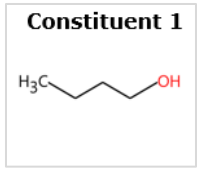
Input target substance by drawing

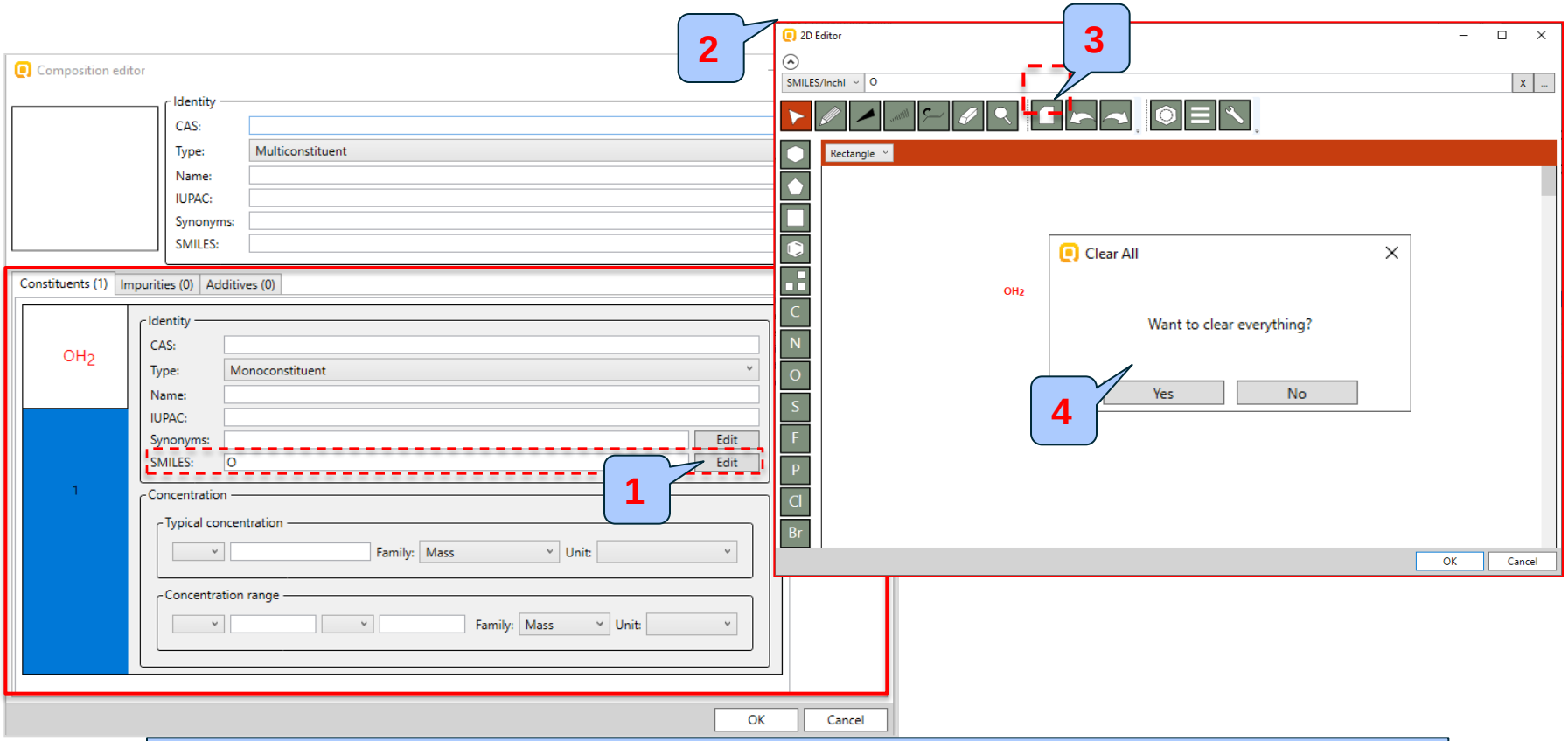


1. Click on ***Input*** module;
2. Click on **Composition;**
3. From composition editor select type: **Multiconstituent;**
4. Click **Add** in order to add constituent. Our target substance consists of three constituents, so click three times on the Add button.

Chemical input

Drawing Constituent 1 of the target mixture

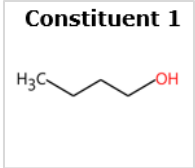


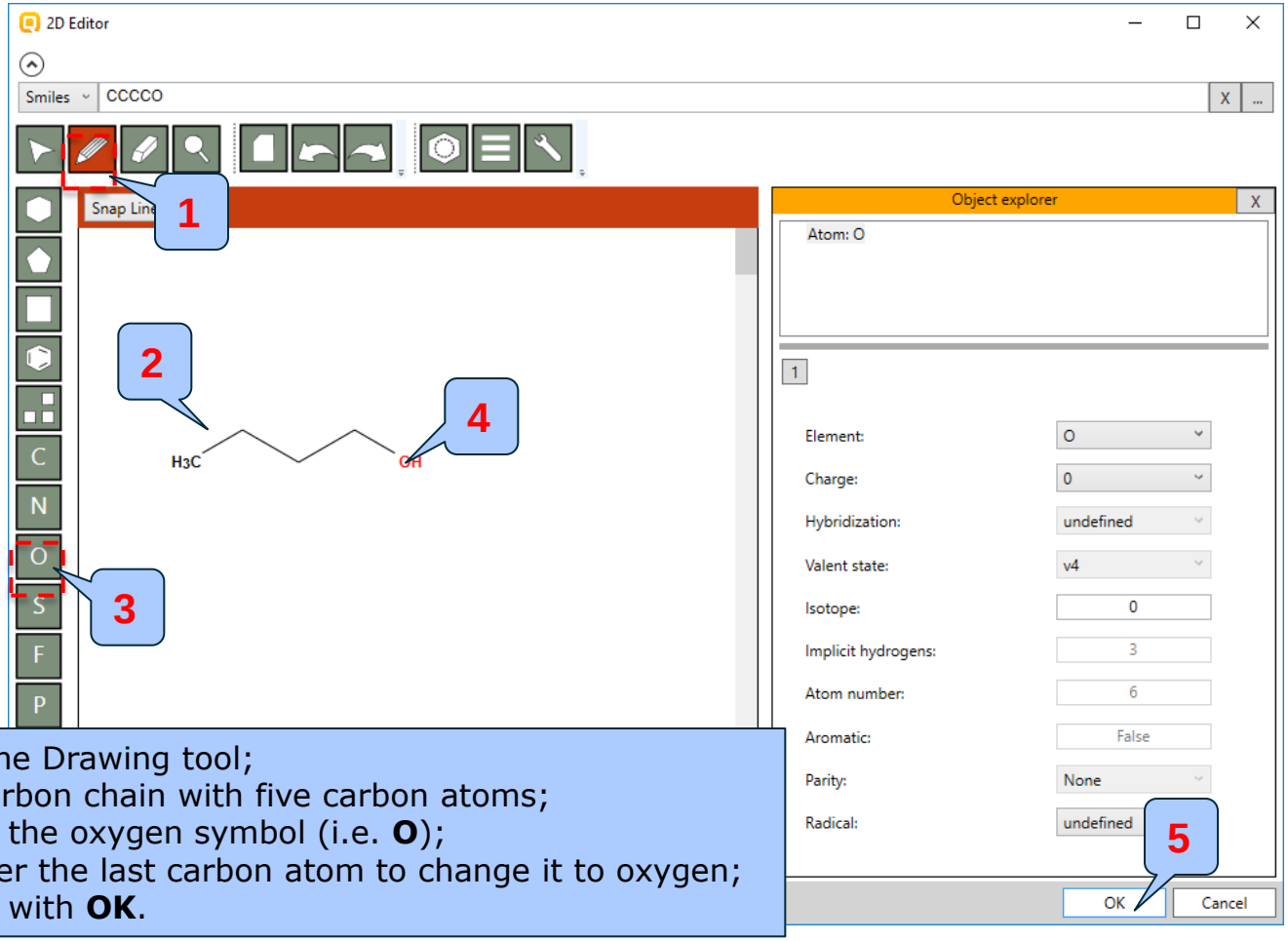


1. Click **Edit** on the SMILES row to define the structure of the first constituent;
2. The 2D editor appears;
3. Click the **Clean button** to clean everything.
4. Confirm with **Yes**.

Chemical input

Drawing Constituent 1 of the target mixture

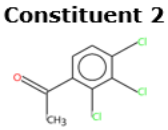


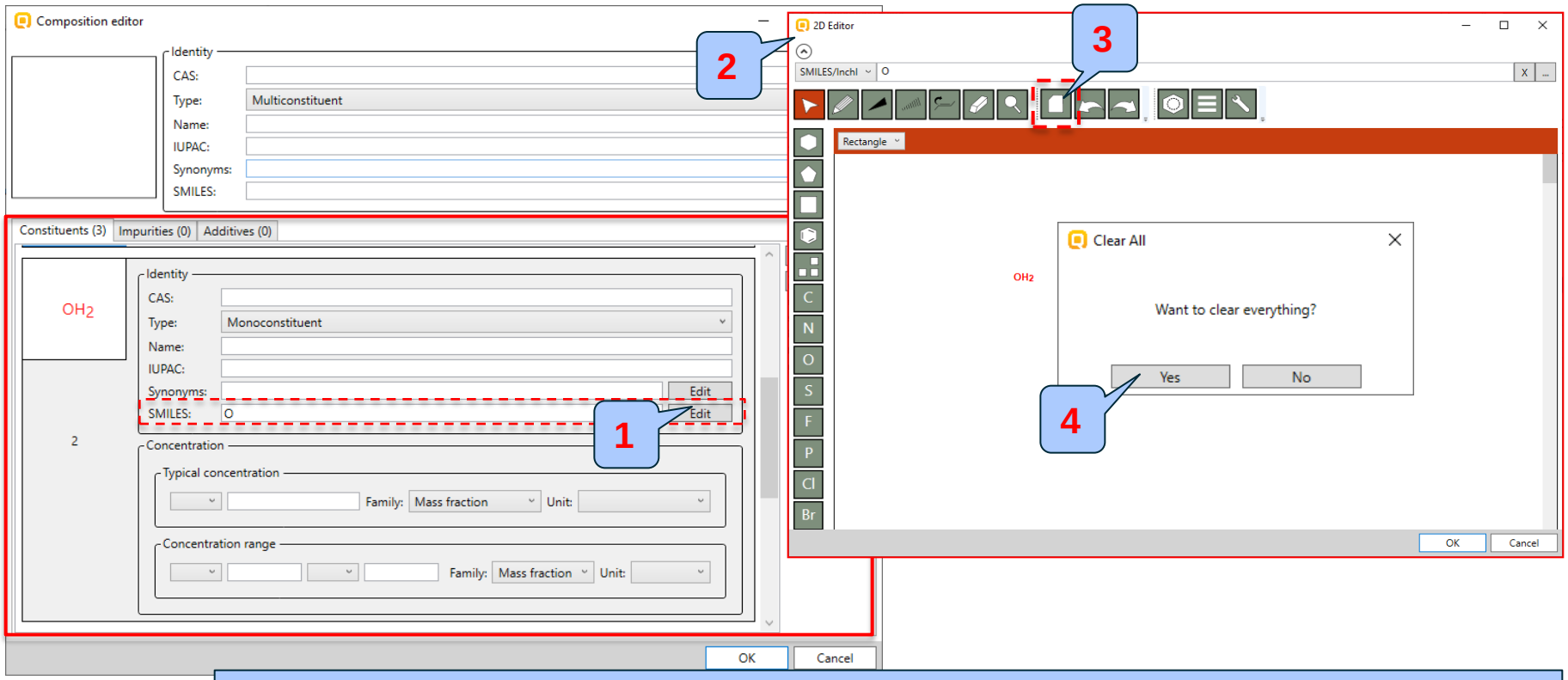


1. Select the Drawing tool;
2. Draw carbon chain with five carbon atoms;
3. Click on the oxygen symbol (i.e. **O**);
4. Click over the last carbon atom to change it to oxygen;
5. Confirm with **OK**.

Chemical input

Drawing Constituent 2 of the target mixture





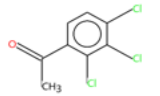
Now we move down to the second constituent and repeat the same steps:

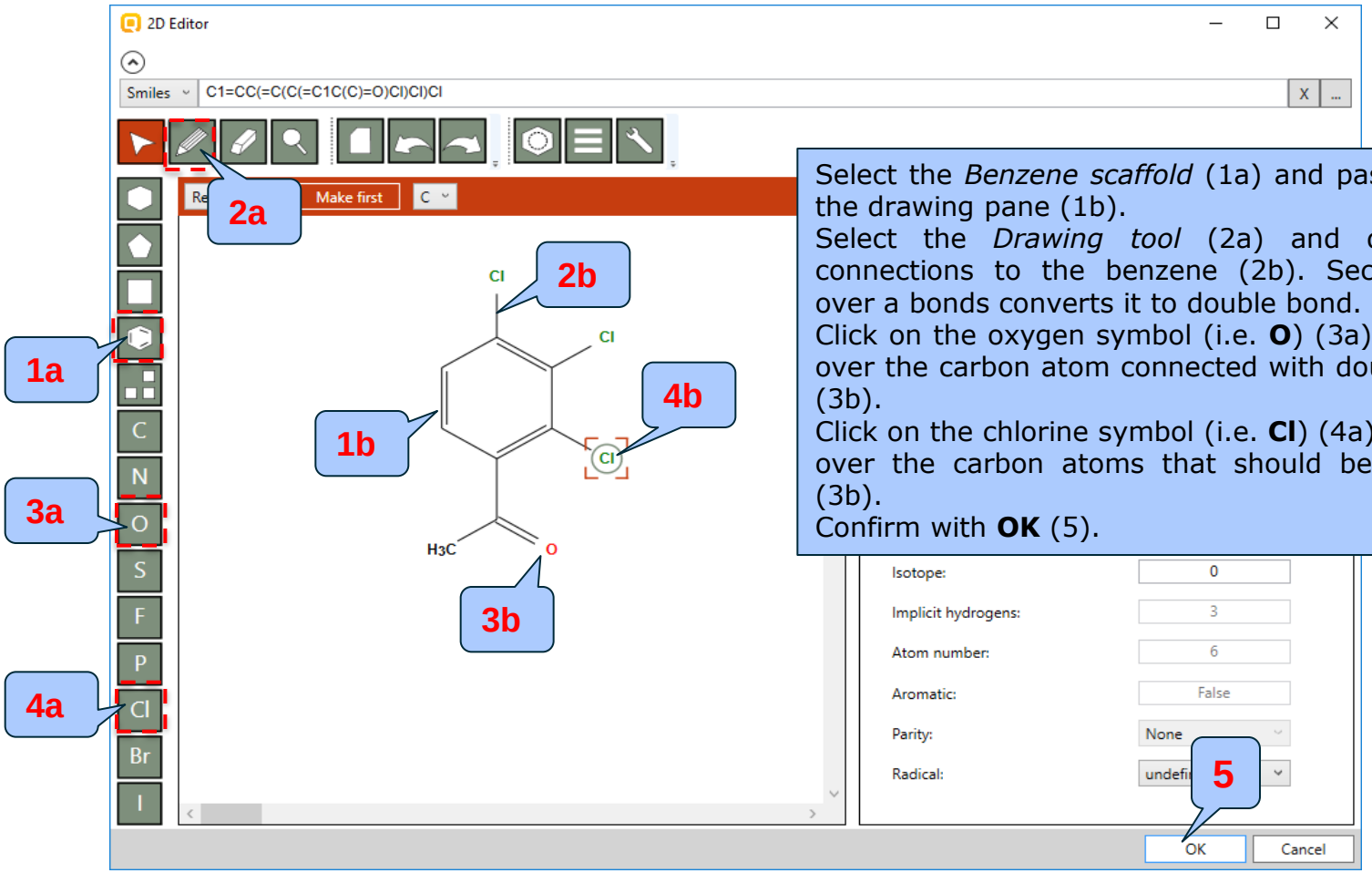
1. Click **Edit** on the SMILES row to define the structure of the second constituent;
2. The 2D editor appears;
3. Click the **Clean button** to clean everything.
4. Confirm with **Yes**.

Chemical input

Drawing Constituent 2 of the target mixture

Constituent 2



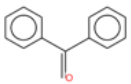


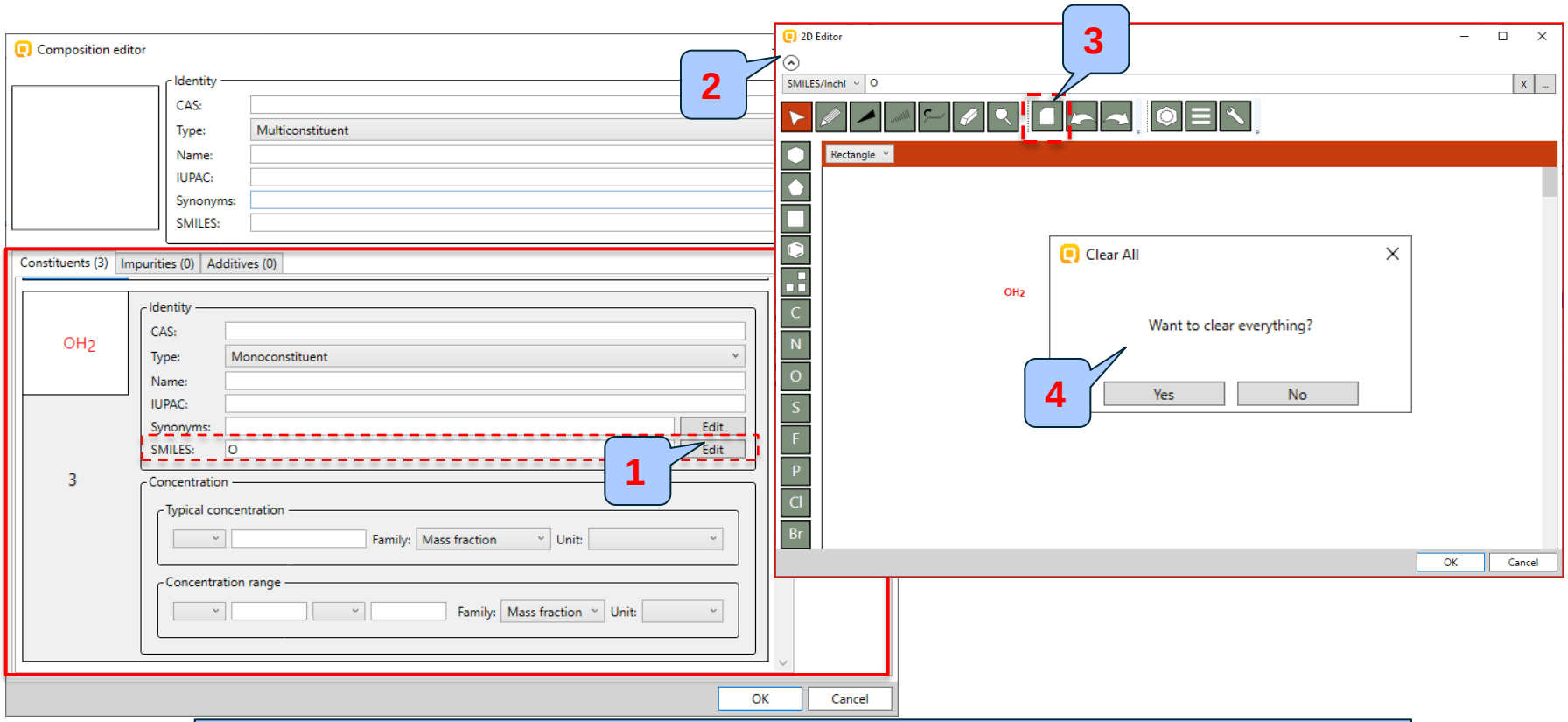
Select the *Benzene scaffold* (1a) and paste it into the drawing pane (1b).
Select the *Drawing tool* (2a) and draw the connections to the benzene (2b). Second click over a bonds converts it to double bond.
Click on the oxygen symbol (i.e. **O**) (3a) and click over the carbon atom connected with double 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).

Chemical input

Drawing Constituent 3 of the target mixture

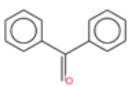
Constituent 3





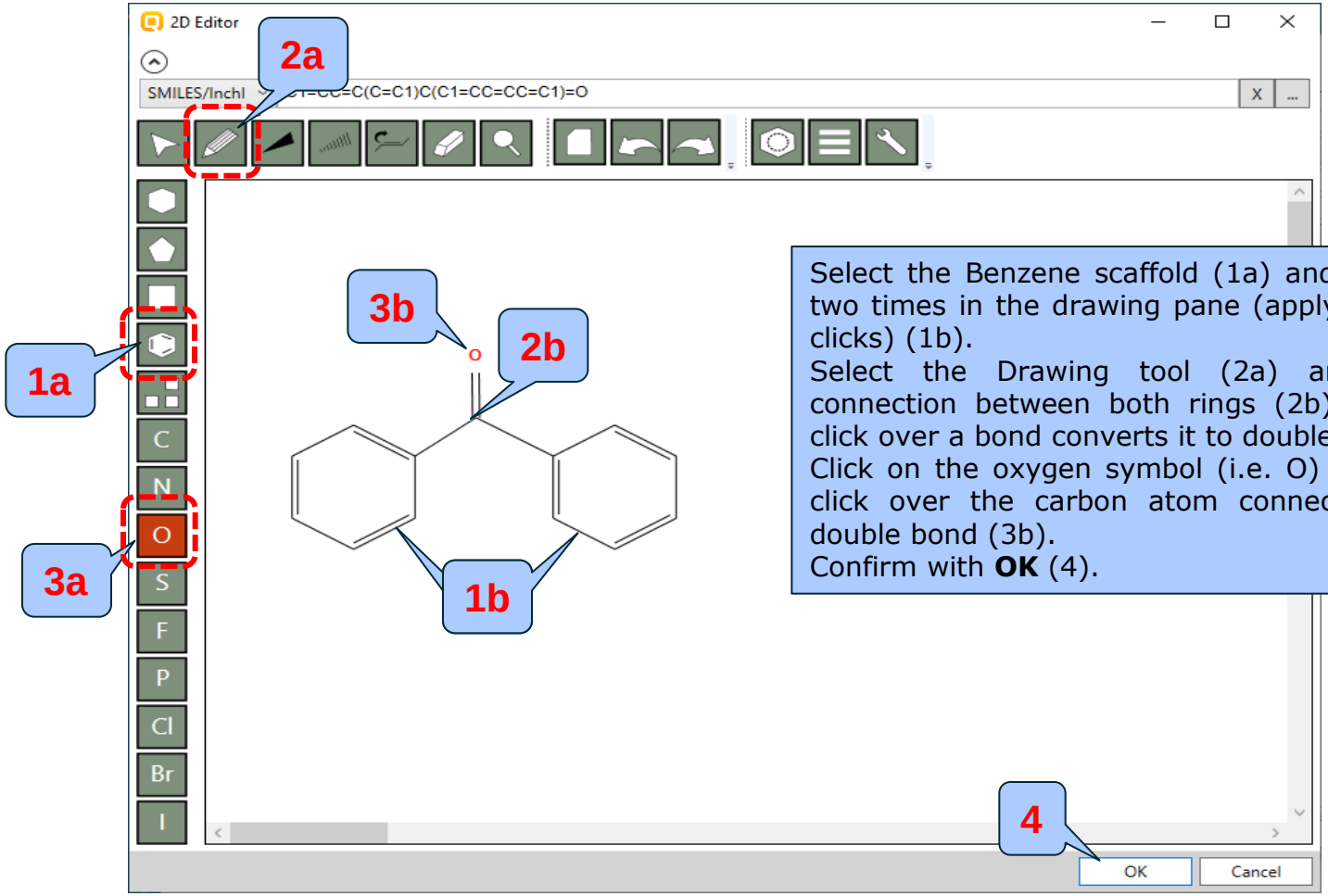
Now we move down to the last (third) constituent and repeat the same steps:
1. Click **Edit** on the SMILES row to define the structure of the second constituent;
2. The 2D editor appears;
3. Click the **Clean button** to clean everything.
4. Confirm with **Yes**.

Constituent 3



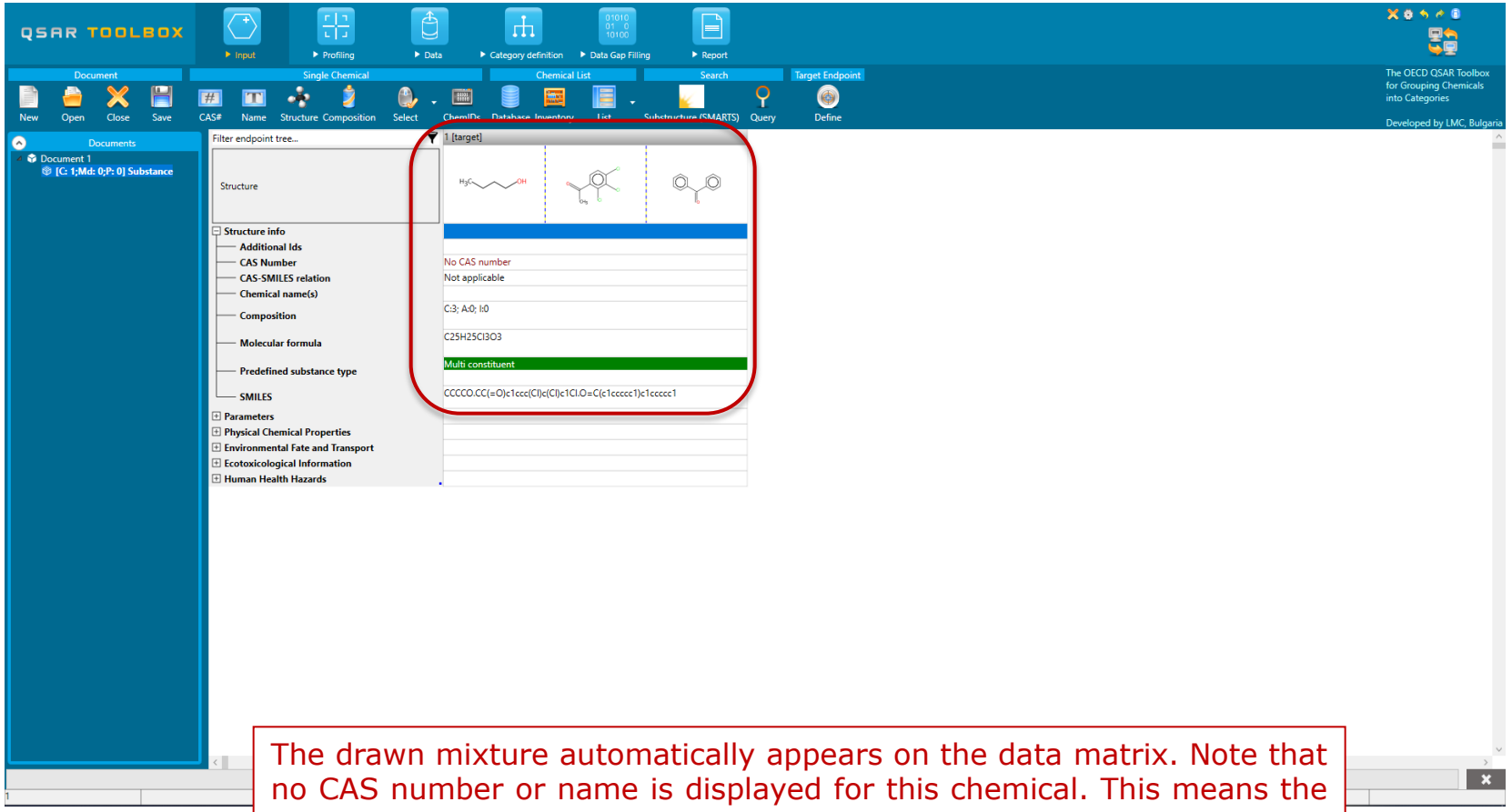
Chemical input

Drawing Constituent 3 of the target mixture



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).

Chemical input
Target substance identity

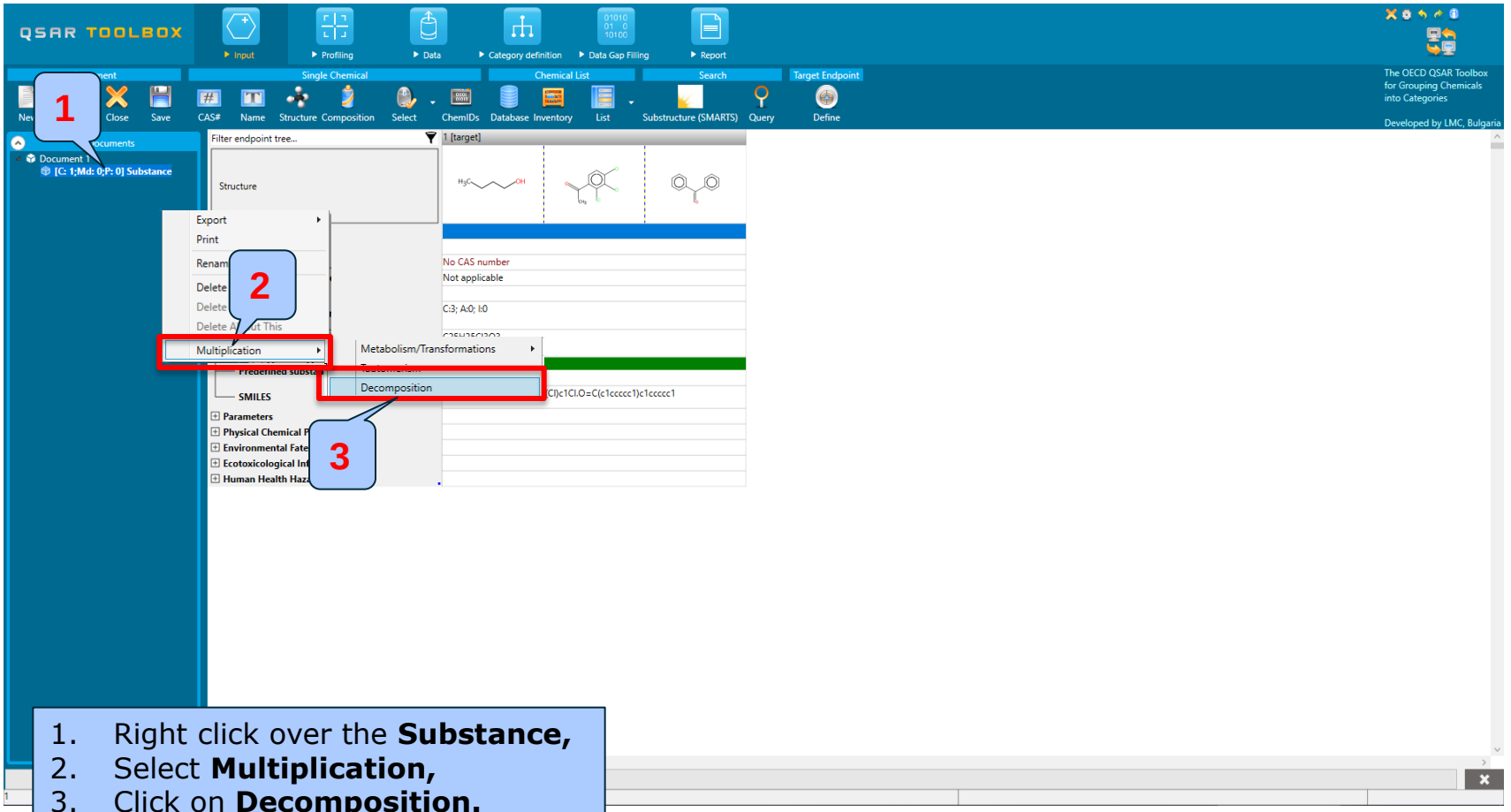


The drawn mixture automatically appears on the data matrix. Note that no CAS number or name is displayed for this chemical. This means the target substance is not listed in the chemical inventories/databases implemented in the Toolbox

Chemical input
Mixture decomposition

- In the current example we will predict the skin sensitization 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.

Chemical input
Input mixture

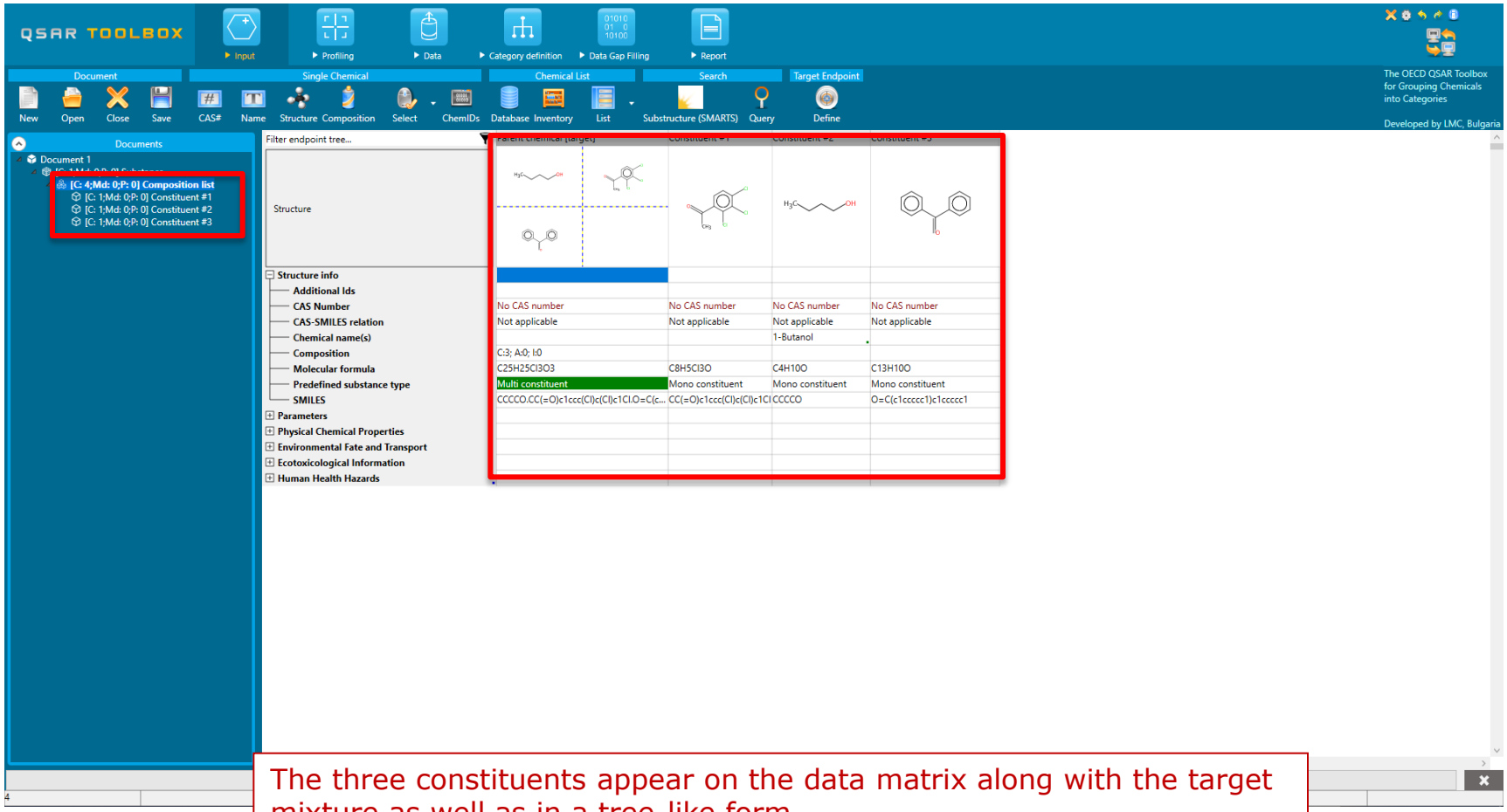


1. Right click over the **Substance,**
2. Select **Multiplication,**
3. Click on **Decomposition.**

Chemical Input
Target chemical identity

- The already drawn target structures automatically appear on the data matrix.
- Note that no CAS number or name is associated with this chemical.
- This means the target chemical is not listed in the chemical inventories/databases available in Toolbox (see next slide).

Chemical Input
Target chemical identity



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

Outlook

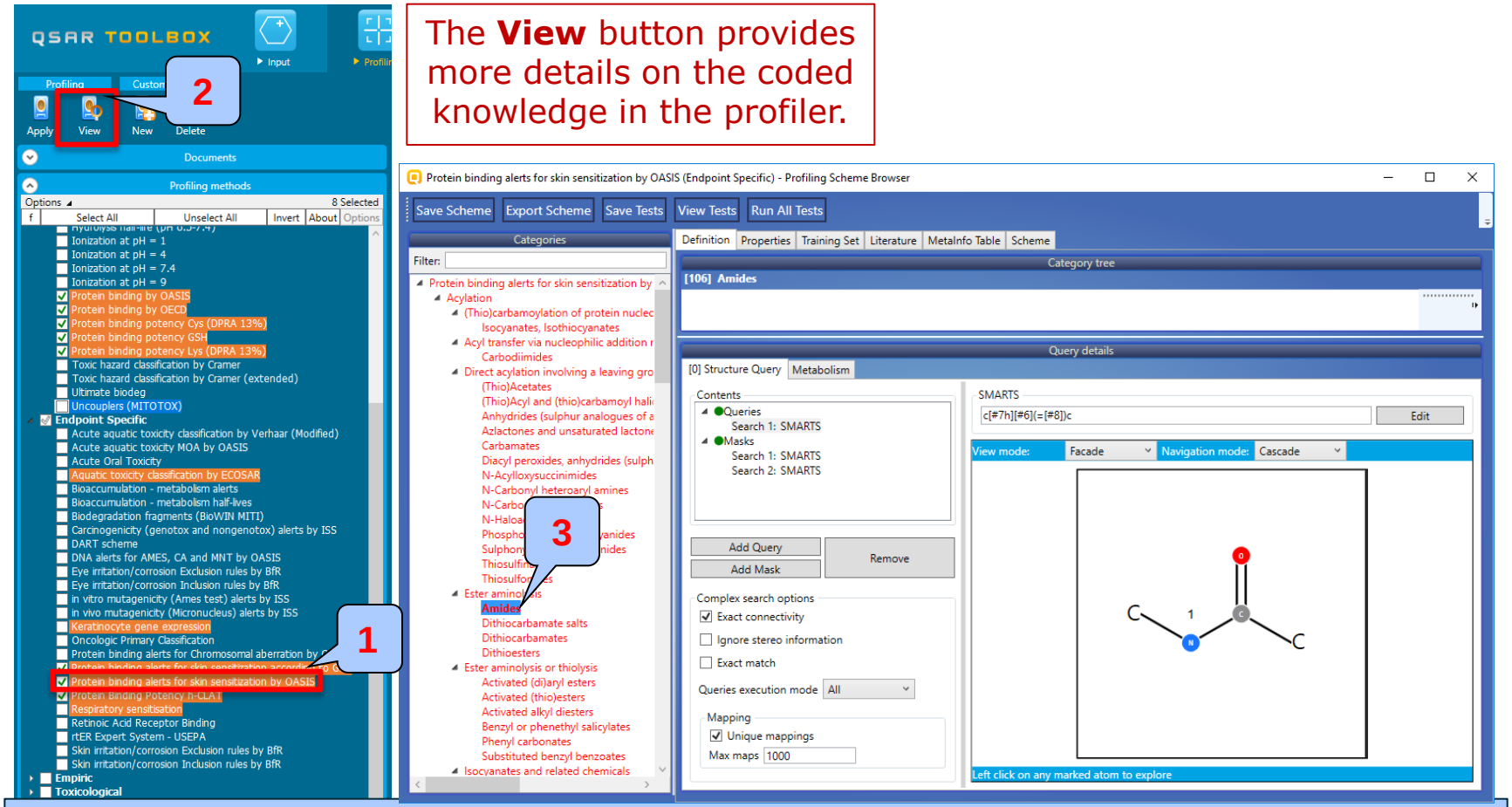
- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - **Profiling**

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.
- 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 highlighting one of the profilers (for example, Protein binding alerts for SS by OASIS and clicking on "About" or "View" (see next screen shots).

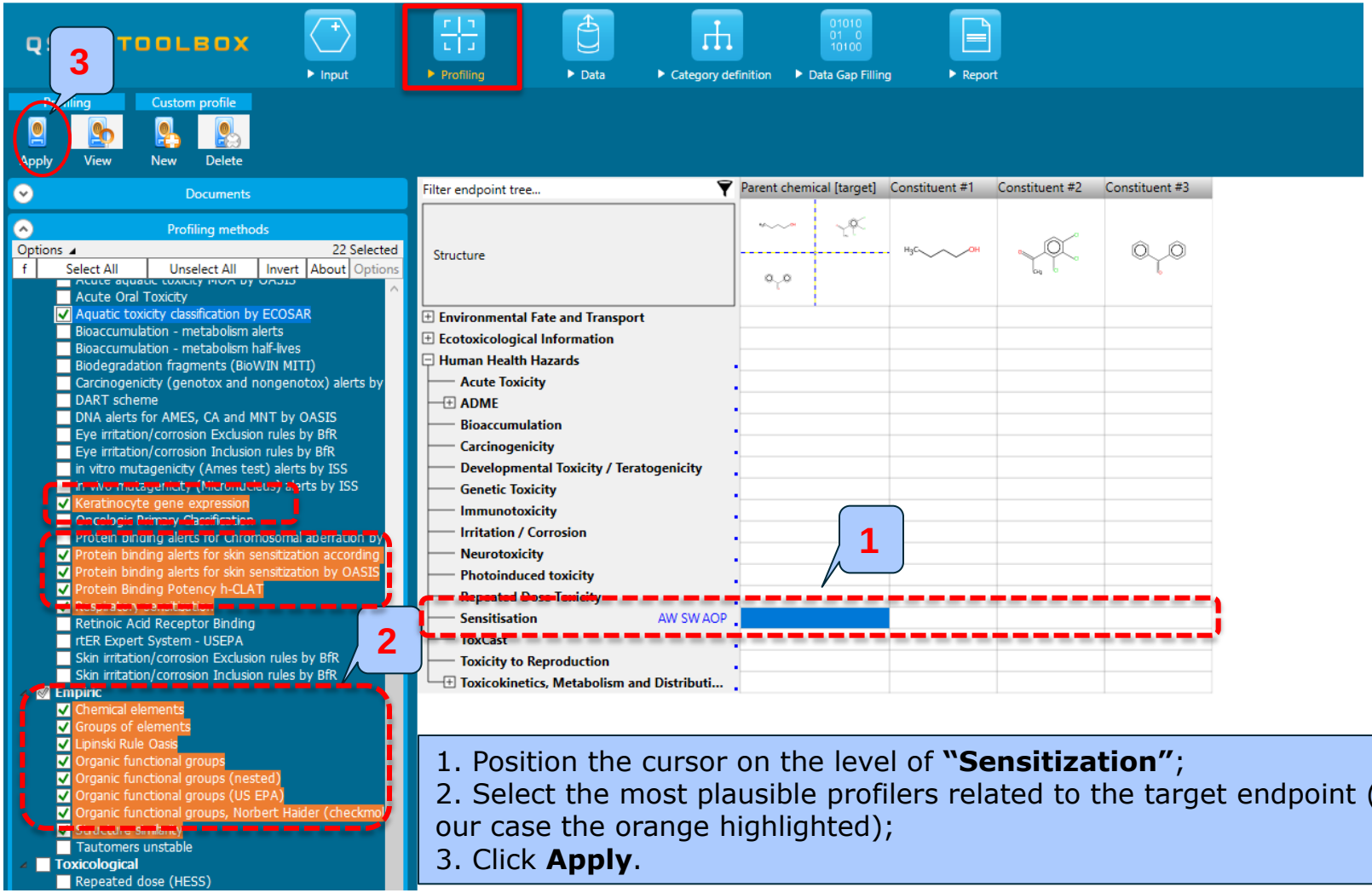
Profiling
Side-Bar to Profiling

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



1. Selected profiler related to the investigated endpoint: **Protein binding alerts for SS by OASIS**;
2. Click on the **"View"** button;
3. Click for example on category **Amides** to see the structural boundaries used to code the knowledge.

Profiling
Side-Bar to Profiling



1. Position the cursor on the level of **"Sensitization"**;
2. Select the most plausible profilers related to the target endpoint (in our case the orange highlighted);
3. Click **Apply**.

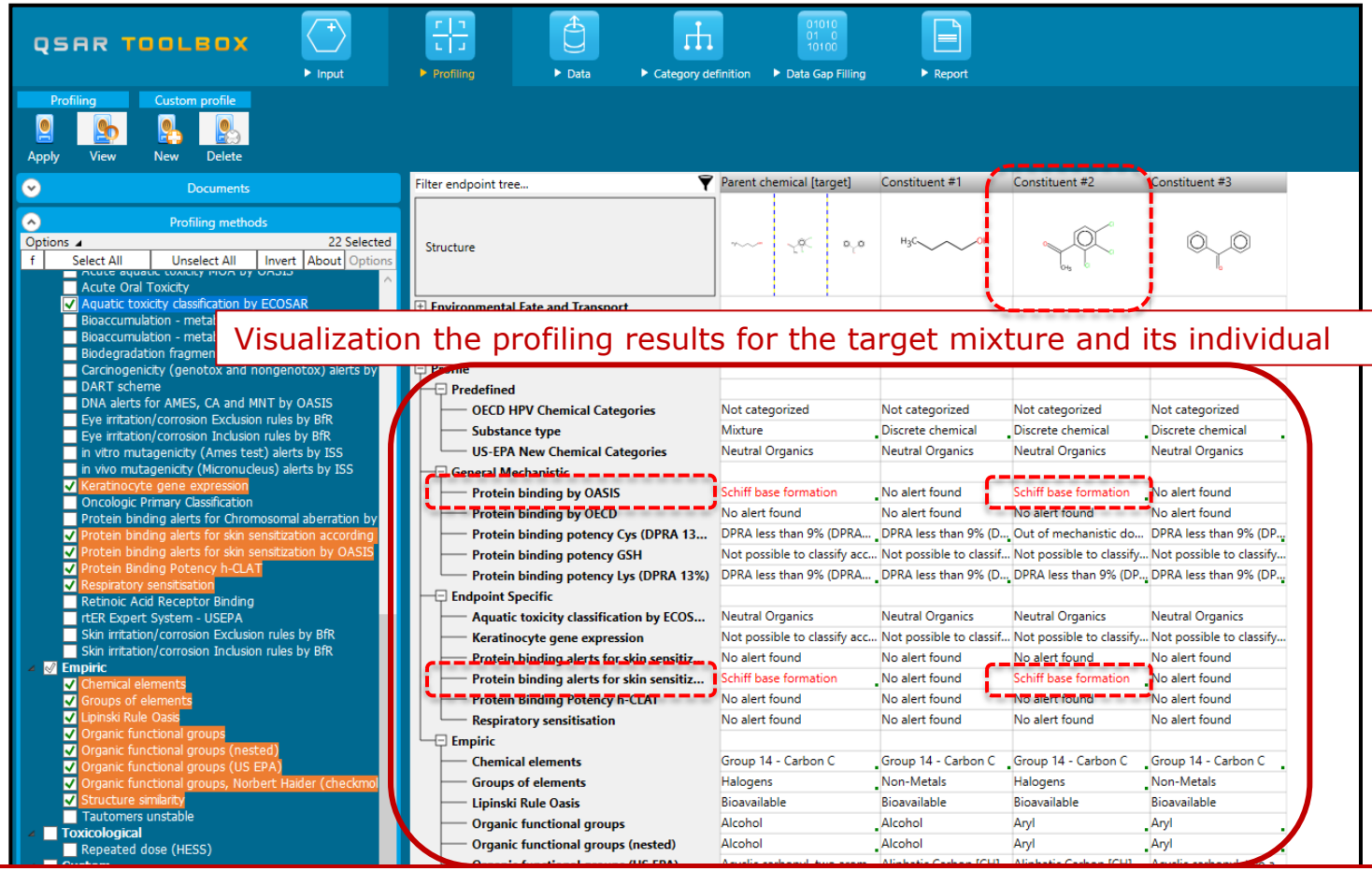
Profiling
Profiling the target substance

- 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 plausible (highlighted in orange) for skin sensitization (only endpoint-specific are listed here):

  - Aquatic toxicity classification by ECOSAR;
  - Keratinocyte gene expression;
  - Protein binding alerts for skin sensitization according to GHS
  - Protein binding alerts for skin sensitization by OASIS
  - Protein binding potency h-CLAT
  - Respiratory sensitization

Profiling

Profiling the target substance



Visualization the profiling results for the target mixture and its individual

The profiling results for all the constituents are consistent with one exception (Constituent #2). The constituent #2 reacts with proteins via Schiff-base formation according to general and endpoint-specific Protein binding alerts for SS profiler.

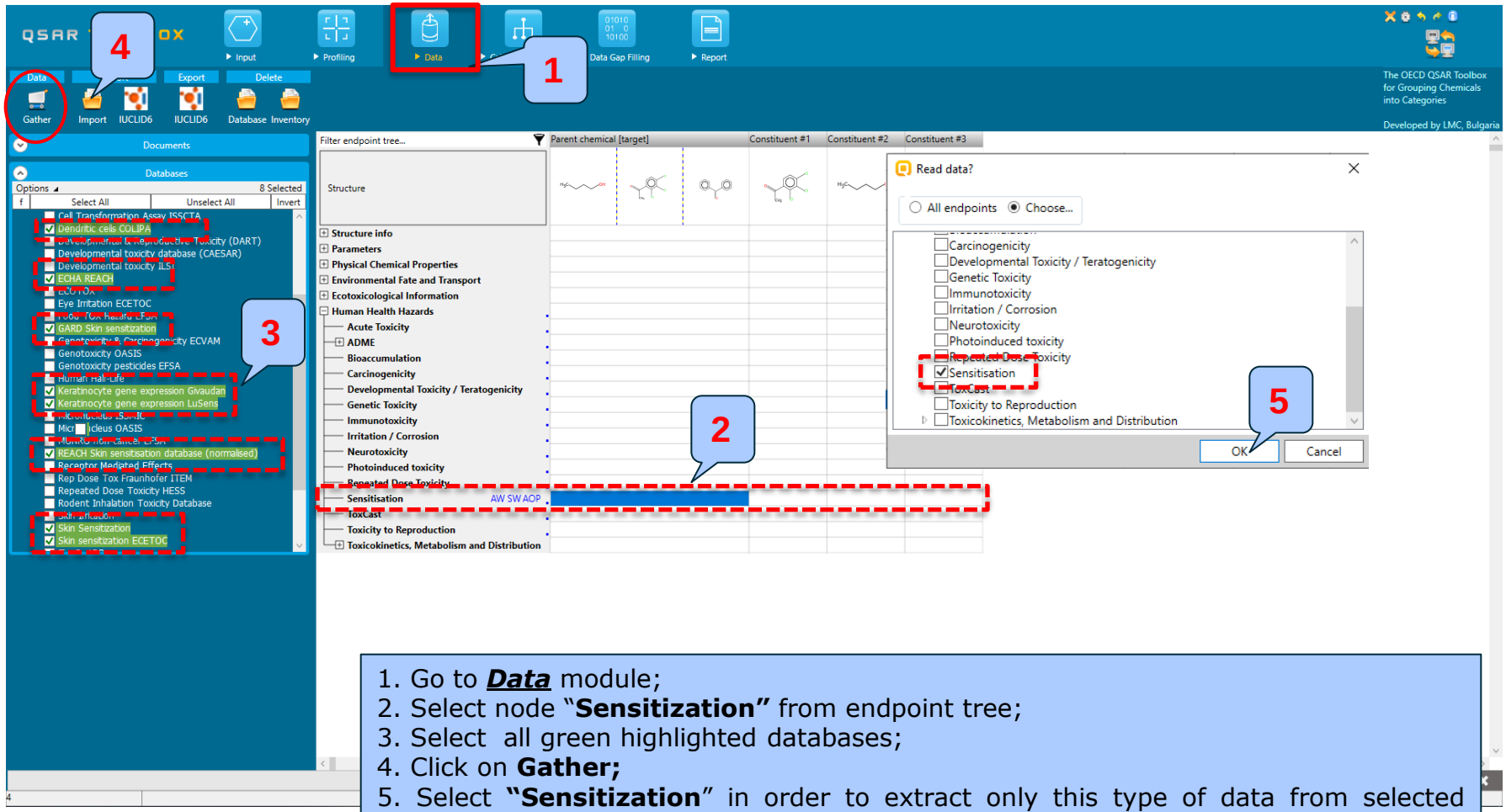
Outlook

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

Data

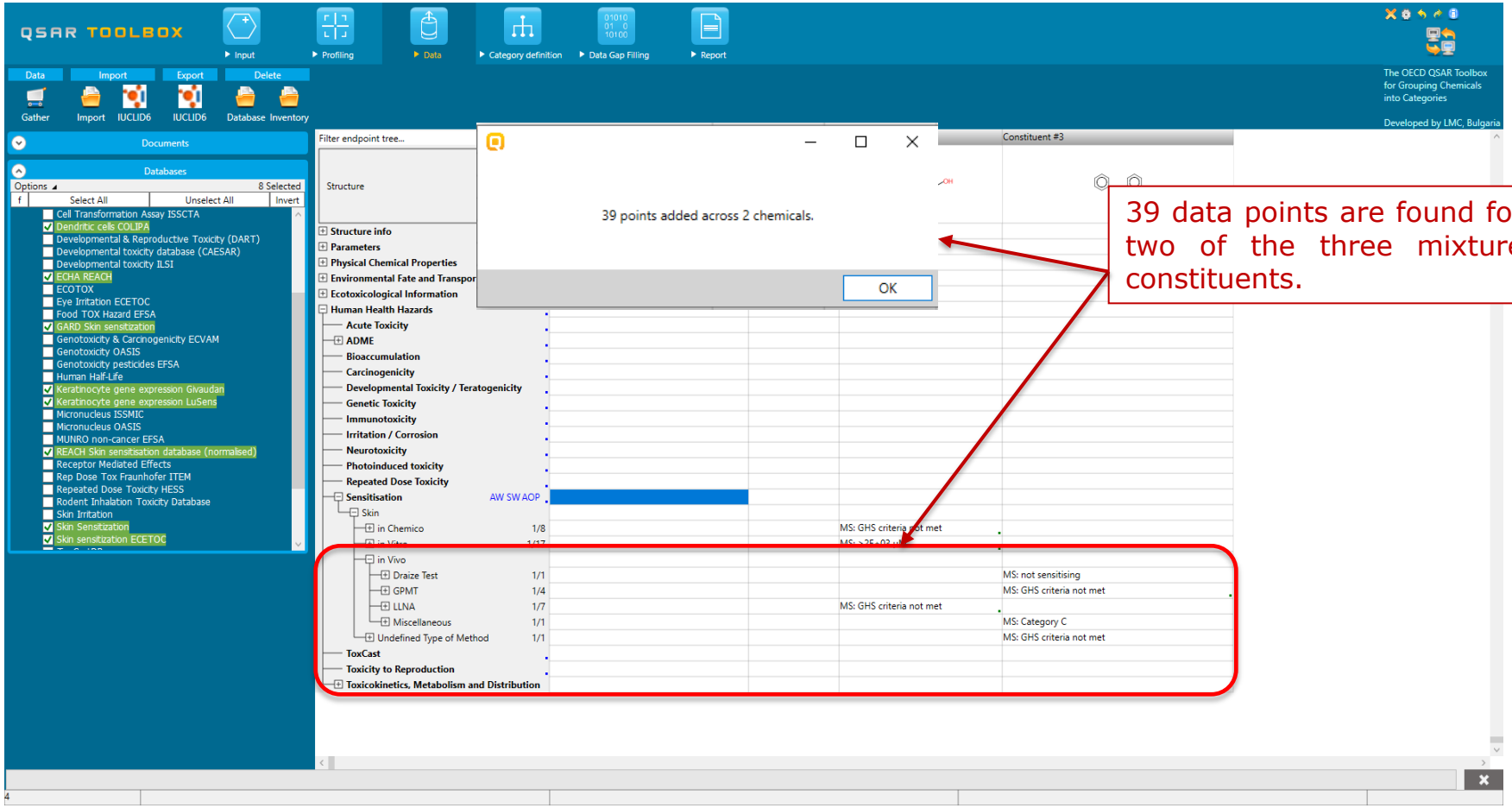
- "Data" refers to the electronic process of retrieving the environmental fate, eco-toxicity and toxicity data that are residing in the Toolbox.
- Data gathering can be executed in a global fashion (i.e. collecting all data of all endpoints) or on a more narrowly defined basis (e.g. collecting data for a single or limited number of endpoints).
- In this example, we limit our data gathering to the common Skin endpoints from databases associated with Skin Sensitization endpoint. The relevant databases are become highlighted in green based on the selected target endpoint

Data



1. Go to ***Data*** module;
2. Select node "**Sensitization"** from endpoint tree;
3. Select all green highlighted databases;
4. Click on **Gather;**
5. Select **"Sensitization"** in order to extract only this type of data from selected databases. Click **OK.**

Data
Process of collecting data



Recap

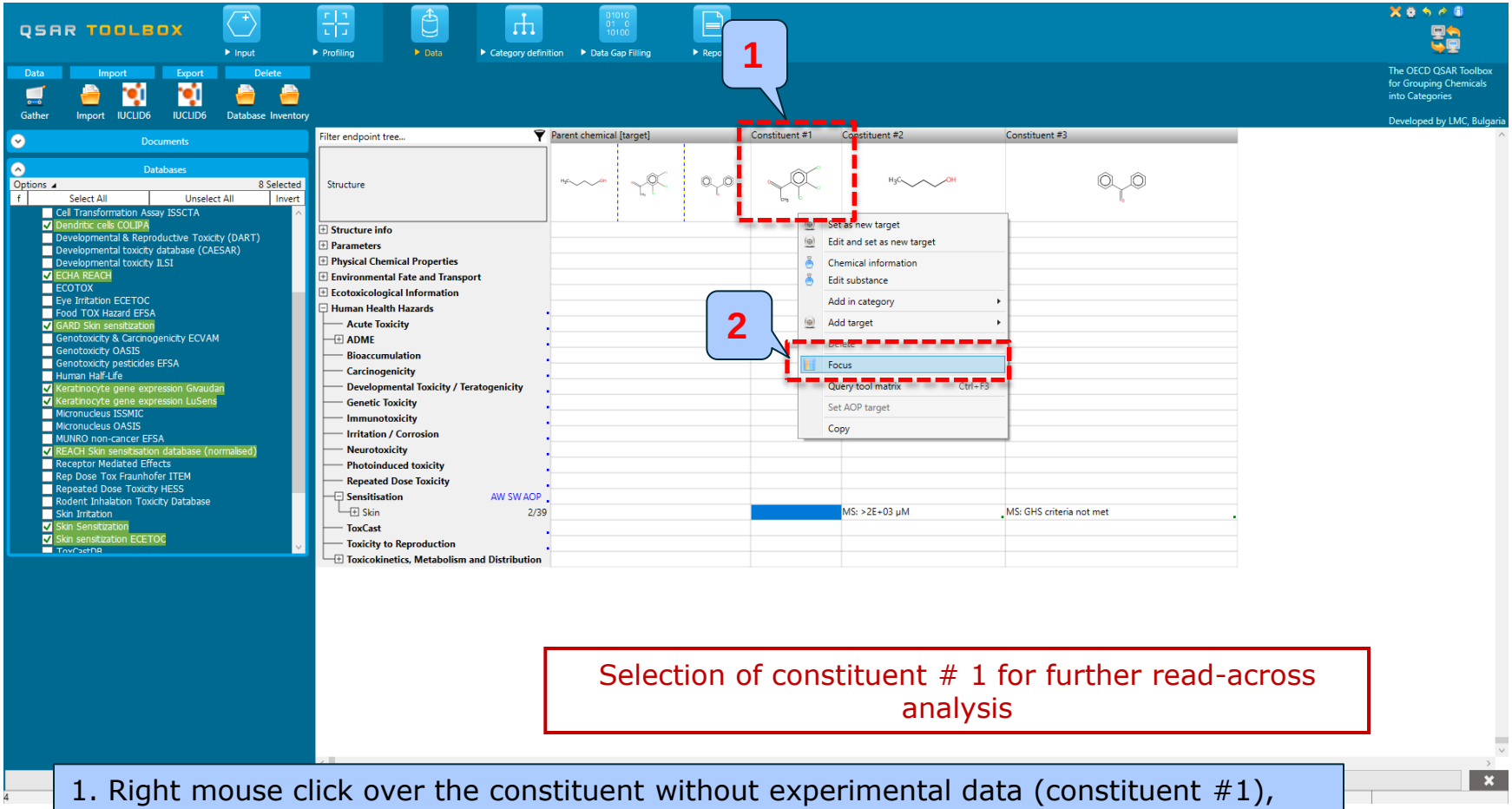
- We have entered the mixture with defined components.
- The profiling results showed no protein binding alerts for two of the mixture constituents (constituents # 1 and #3). The third constituent (constituent #2) has positive protein binding alerts and could elicit skin sensitization effect.
- Negative experimental data has been found for two of the mixture constituents (constituents # 2 and #3). No experimental data has been found for the third constituent (constituent #1).
- The constituent without experimental data and positive protein binding alert (constituent #1) will be used for further read across analysis. Then, all of the available data – experimental and predicted will be used for skin sensitization prediction of the mixture.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - **Read across prediction of constituent without data**
    - **Focus constituent without experimental data**

Read across prediction of constituent without data

Focus constituent

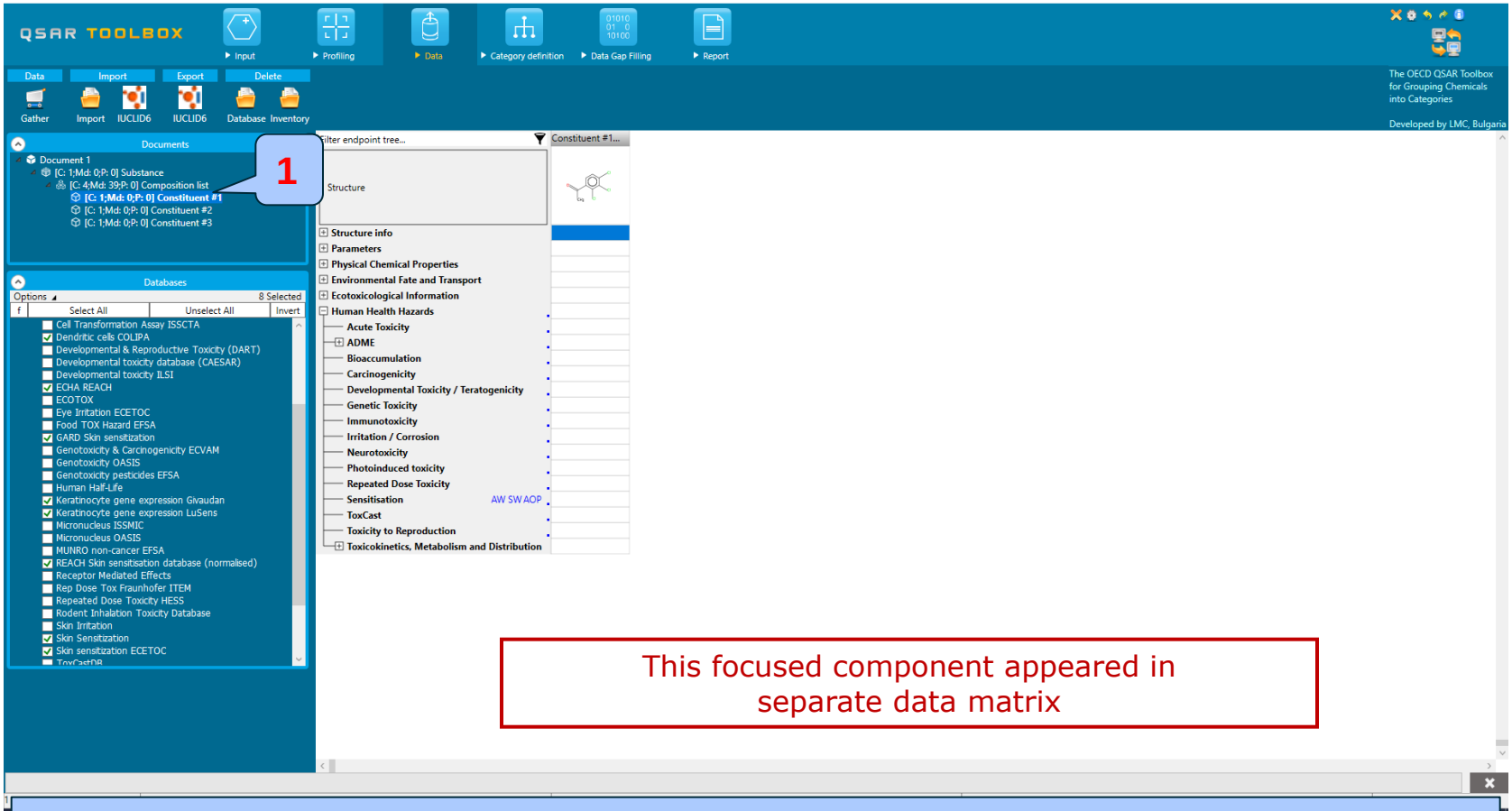


Selection of constituent # 1 for further read-across analysis

1. Right mouse click over the constituent without experimental data (constituent #1),
2. Select **Focus** from the drop down menu.

Read across prediction of constituent without data

Focus constituent



This focused component appeared in separate data matrix

1. A documented tree with focused constituent #1 is automatically selected. The workflow could move further with collecting analogues of the focused constituent #1.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - **Read across prediction of constituent without data**
    - Focus constituent without experimental data
    - **Define category**

Category Definition
Overview

- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- This is the critical step in the workflow.
- Several options are available in the Toolbox to assist the user in refining the category definition.
- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity”.

Basic guidance for category formation and assessment

Suitable categorization phases:

1. Structure-related profilers.
2. Endpoint specific profilers (for sub-cat).
3. Additional structure-related profilers, if needed to eliminate dissimilar chemicals (to increase the consistency of category) (e.g. chemical elements).

Performing categorization:

1. The categorization phases should be applied successively
2. The application order of the phases depend on the specificity of the data gap filling
3. More categories of same Phase could be used in forming categories
4. Some of the phases could be skipped if consistency of category members is reached

Graphical illustration of suitable categorization phases is shown on next slide

Suitable Categorization/Assessment Phases

Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Broad grouping
Endpoint Non-specific

Repeating Phase I due to Multifunctionality of chemicals

Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- Genotoxicity/carcinogenicity
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules

Subcategorization
Endpoint Specific

Metabolism accounted for

Phase III. Eliminating dissimilar chemicals

Apply Phase I – for structural dissimilarity
Filter by test conditions – for Biological dissimilarity

Subcategorization
Endpoint Specific

Read across prediction of constituent without data

Forming category for studied endpoint

Suitable Categorization/Assessment Phases

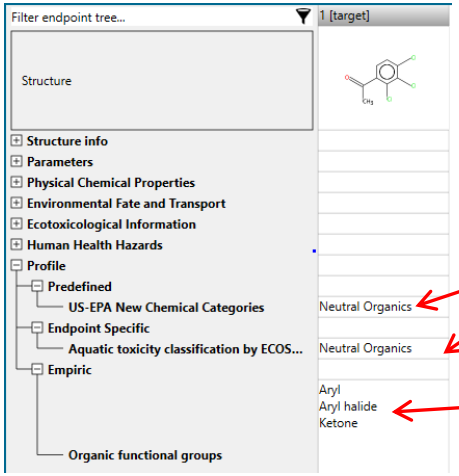
Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Repeating Phase I due to Multifunctionality of chemicals

Broad grouping
Endpoint Non-specific

Phase I categorization in Toolbox



It is not recommended to use "Neutral organic" * as phase I

97 analogues are identified as Aryl halides by OFG

*Neutral organic category include chemicals having different functionalities as alcohols, ketones, ethers etc. In this respect the basic principle that structurally similar chemicals may elicit similar effects would not be preserved, because Neutral organic mixed many different functionalities

Read across prediction of constituent without data

Forming category for studied endpoint

- Based on the above recommendations the OFG is used as initial categorization group
- Refinement of the initial group is based on endpoint-specific protein binding profiler:
    - Protein binding alerts for skin sensitization.

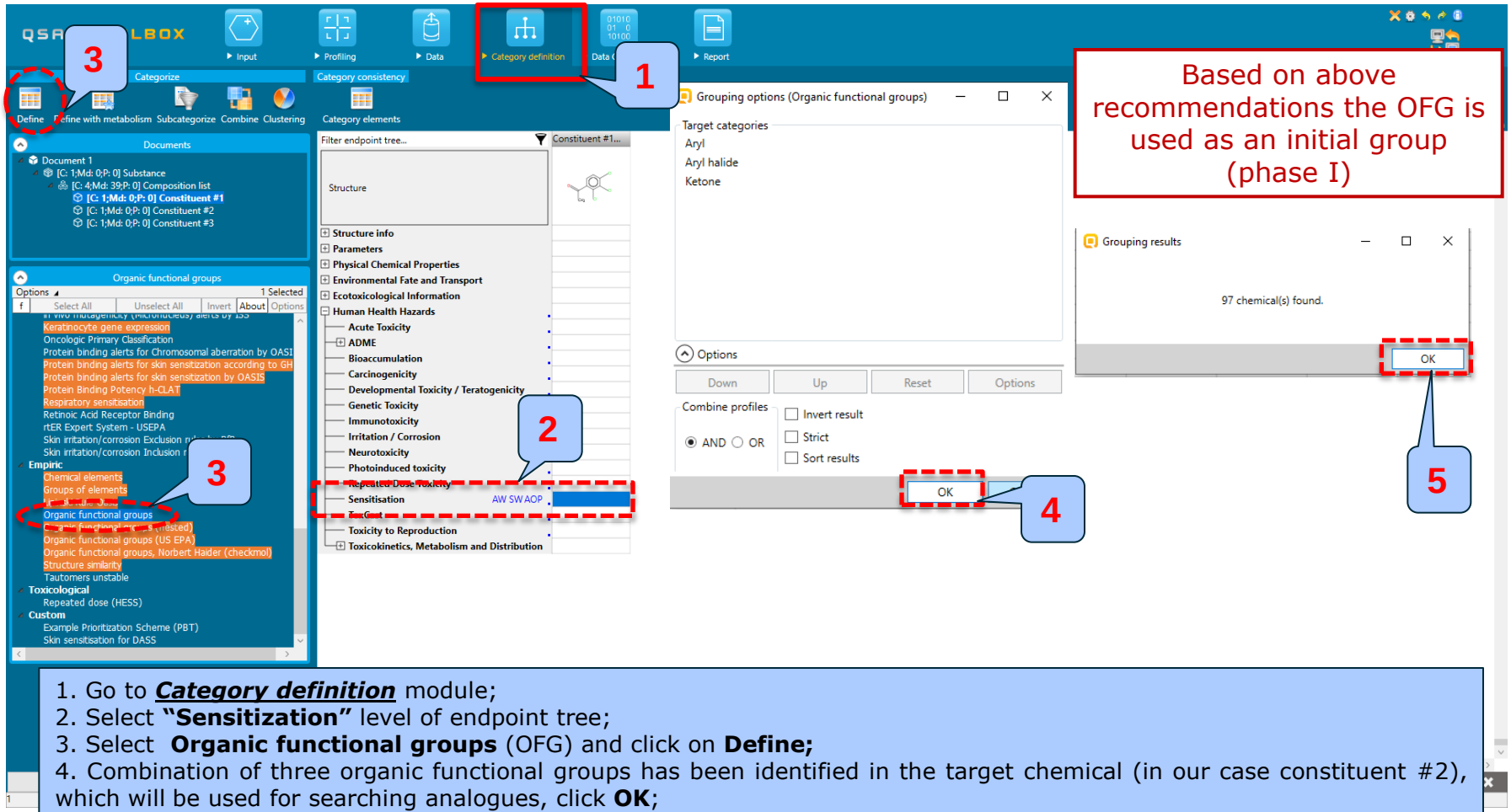
Category definition is a tool for grouping chemicals, which allows to group chemicals based on different measures of "similarity". For more details see tutorials posted on QSAR Toolbox website:

https://qsartoolbox.org/support/

See next slides

Read across prediction of constituent without data

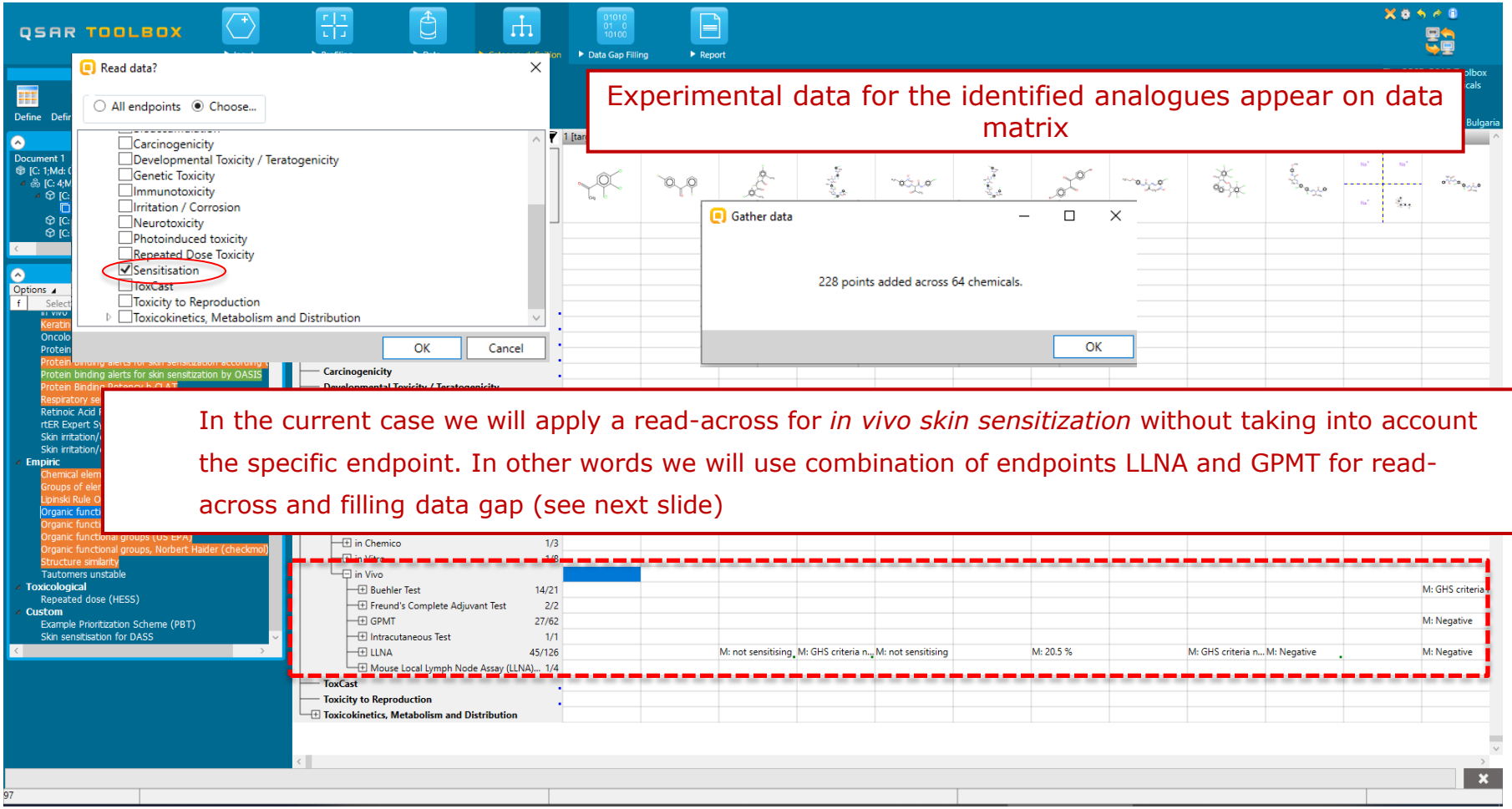
Define category by OFG



1. Go to ***Category definition*** module;
2. Select **"Sensitization"** level of endpoint tree;
3. Select **Organic functional groups** (OFG) and click on **Define;**
4. Combination of three organic functional groups has been identified in the target chemical (in our case constituent #2), which will be used for searching analogues, click **OK**;
5. a list of 97 chemicals has been found having all the three categories identified in the target chemical; gather data for the analogues (see next slide)

Read across prediction of constituent without data

Gather data for analogues chemicals

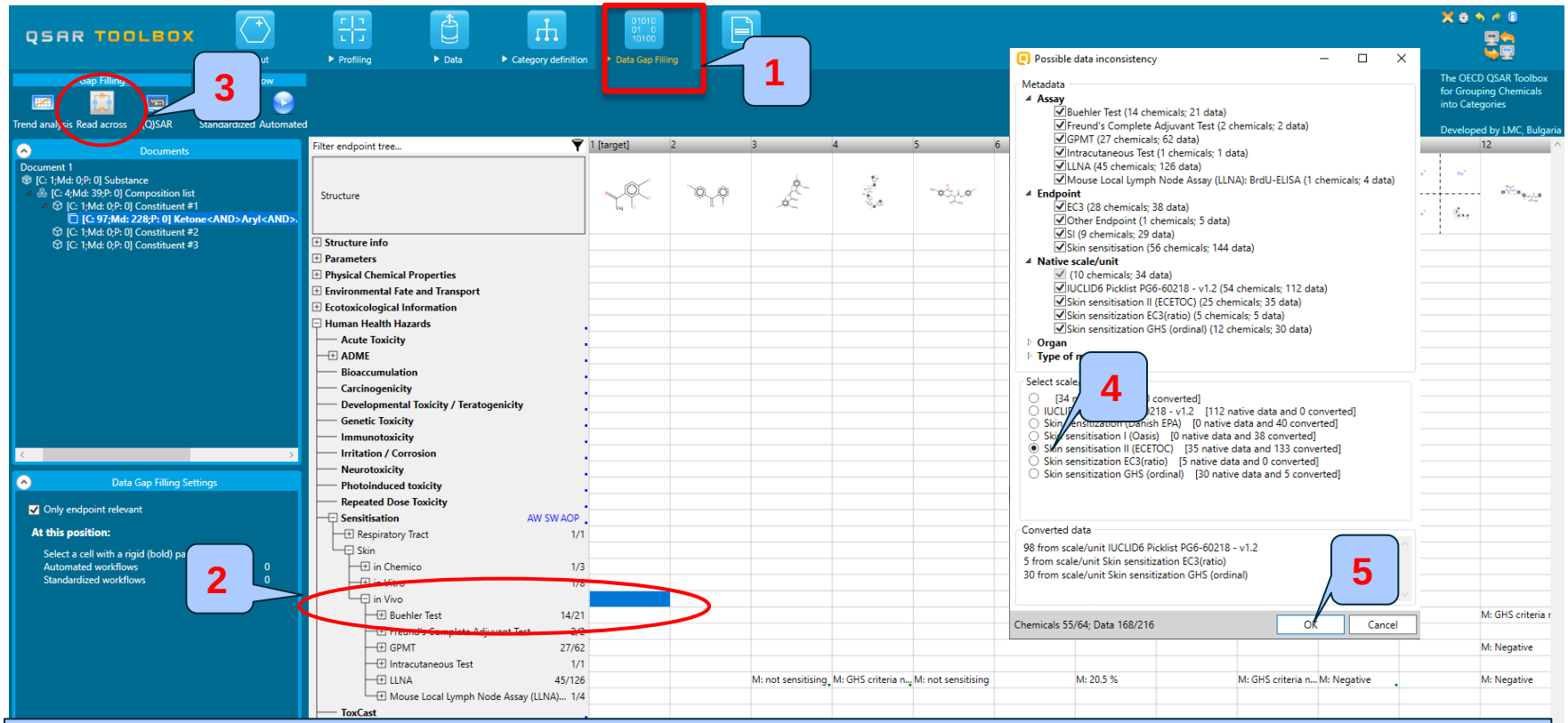


Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - **Read across prediction of constituent without data**
    - **Focus constituent without experimental data**
    - **Define category**
    - **Apply read across**

Read across prediction of constituent without data

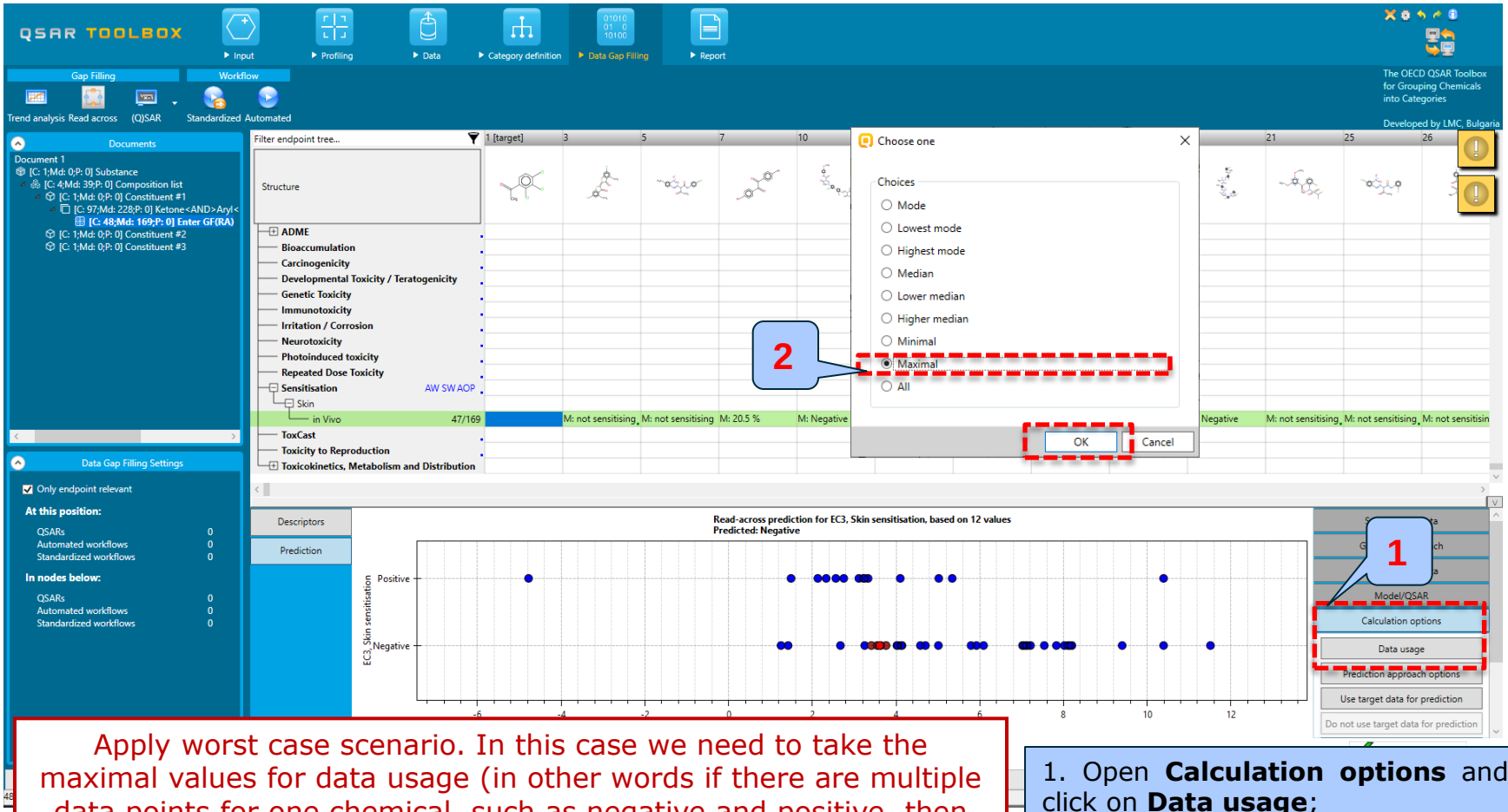
Apply read across



1. Go to ***Data Gap filling*** module;
2. Click on the cell corresponding to Skin Sensitization in Vivo (i.e. in this case we will combine all the data stored under "In vivo" level);
3. Click on **Read-across** button;
4. Select scale/unite **Skin sensitization II(ECETOC)**;
5. Click **OK** (in this case we mix all endpoints and assays).

Read across prediction of constituent without data

Apply read across



Apply worst case scenario. In this case we need to take the maximal values for data usage (in other words if there are multiple data points for one chemical, such as negative and positive, then the positive data will be taken for read-across prediction)

1. Open **Calculation options** and click on **Data usage**;
2. Select Maximal and click **OK**

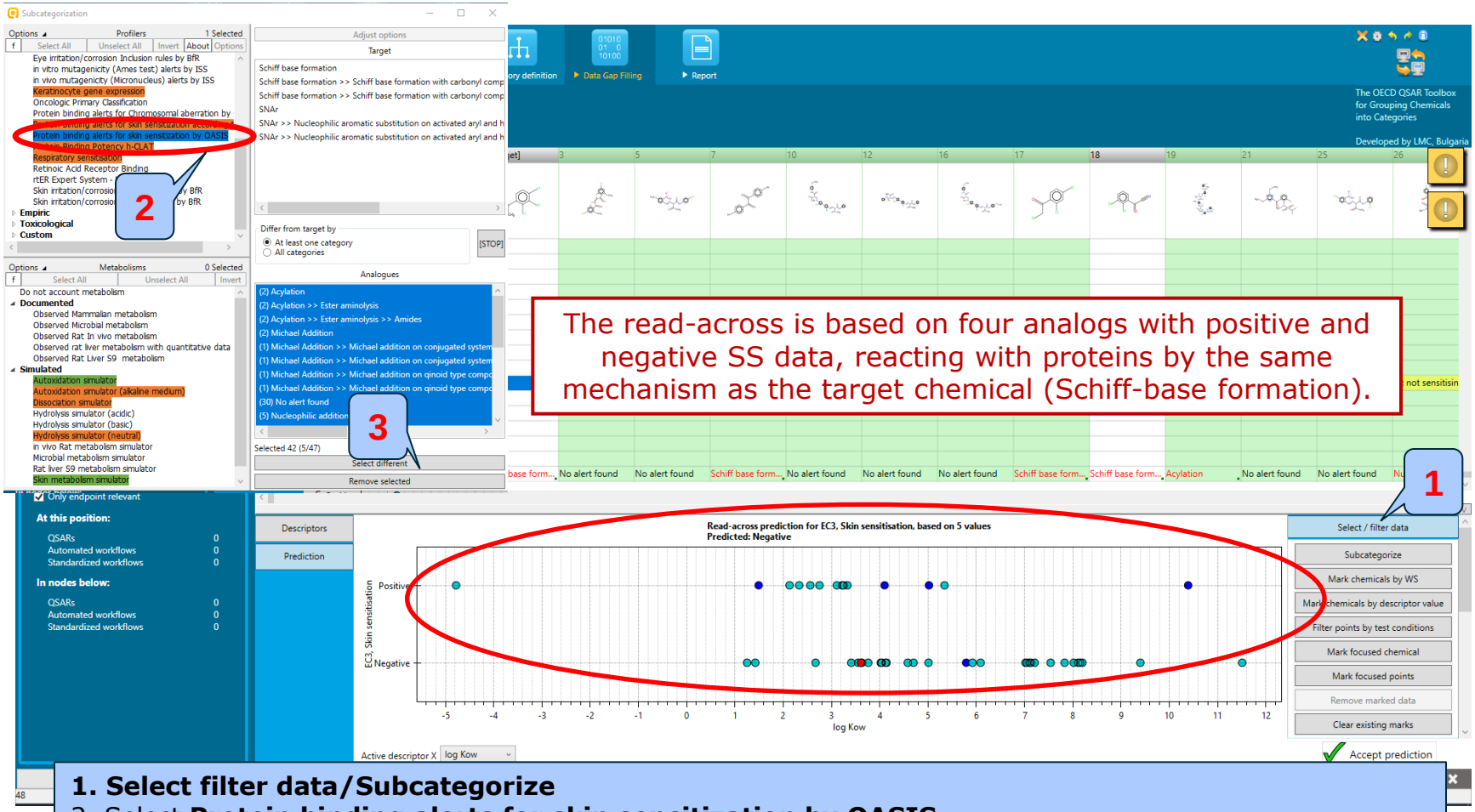
Read across prediction of constituent without data
Subcategorization

- The initial category could be refined by subcategorizing the analogues according to the "Protein binding alerts for skin sensitization by OASIS" and Structural similarity profilers.

- These steps are summarized in the next screen shots.

Read across prediction of constituent without data

Subcategorization by Protein binding alert for SS



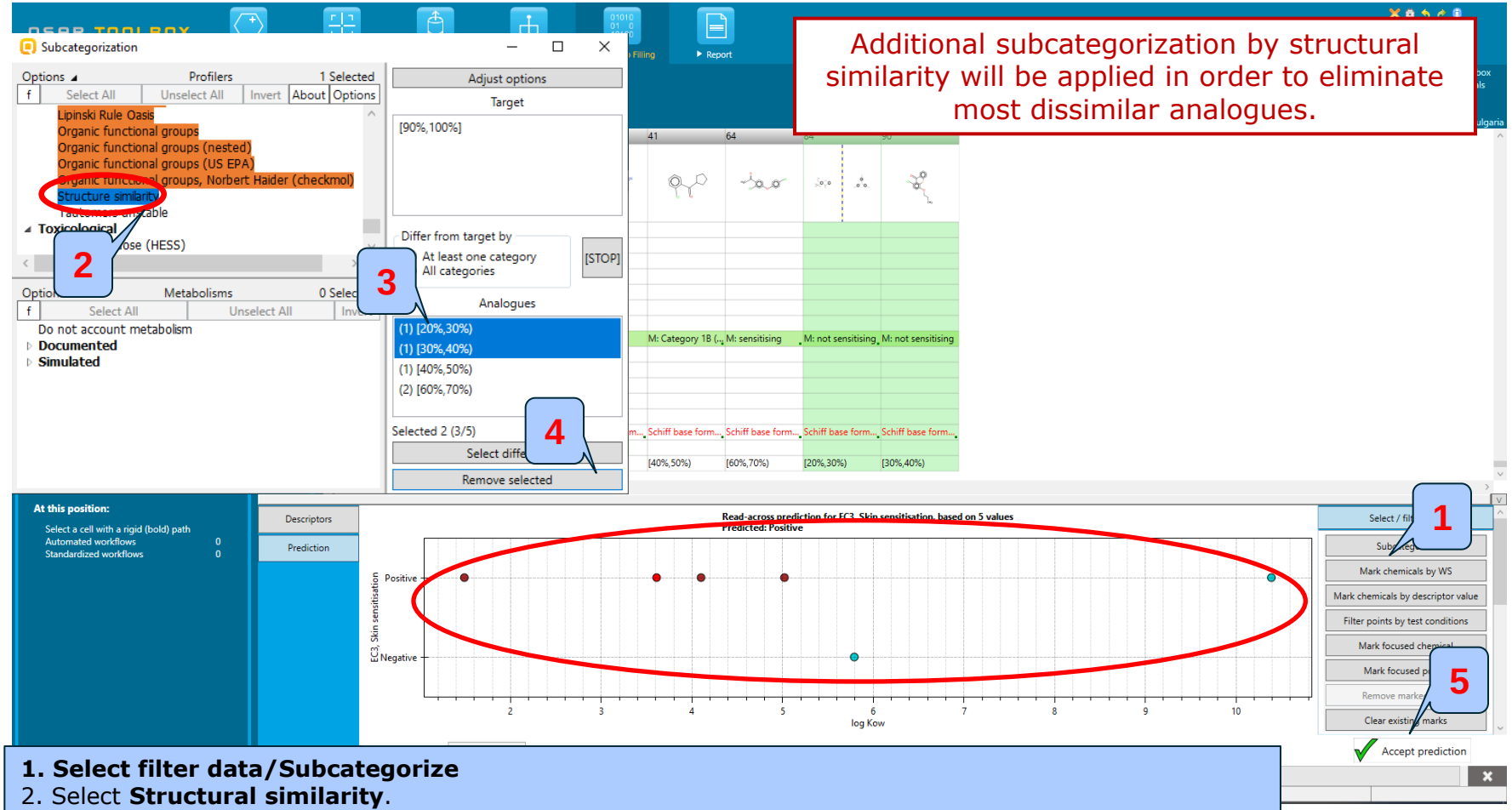
1. Select filter data/Subcategorize

2. Select **Protein binding alerts for skin sensitization by OASIS**.

3. Remove selected to eliminate dissimilar chemicals, reacting by different protein binding mechanisms;

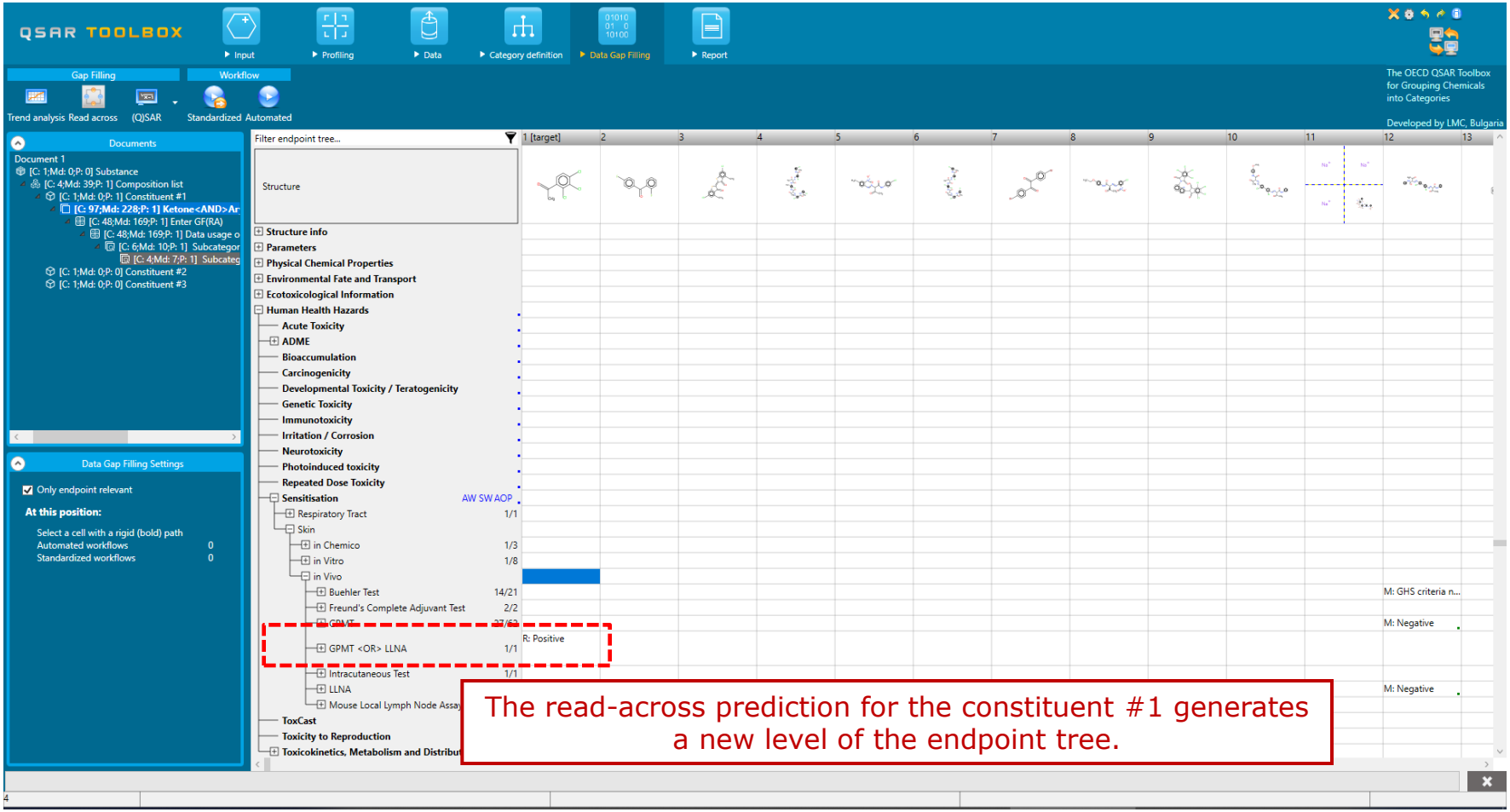
Read across prediction of constituent without data

Subcategorization by Structural similarity



1. Select filter data/Subcategorize
2. Select **Structural similarity**.
3. Select first two bins with most dissimilar analogues ([20%,30%] and [30%,40%])
4. Remove selected to eliminate dissimilar chemicals;
5. Click **Accept prediction**

Read across prediction of constituent without data



Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Read across prediction of constituent without data
  - **Filling data gap for skin sensitization of mixture**

Data Gap Filling
Overview

- "Data Gap Filling" module gives 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 MOA could be found in F1 help
- In this particular case all components of the current mixture are with dissimilar mode of action. In this respect Independent 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

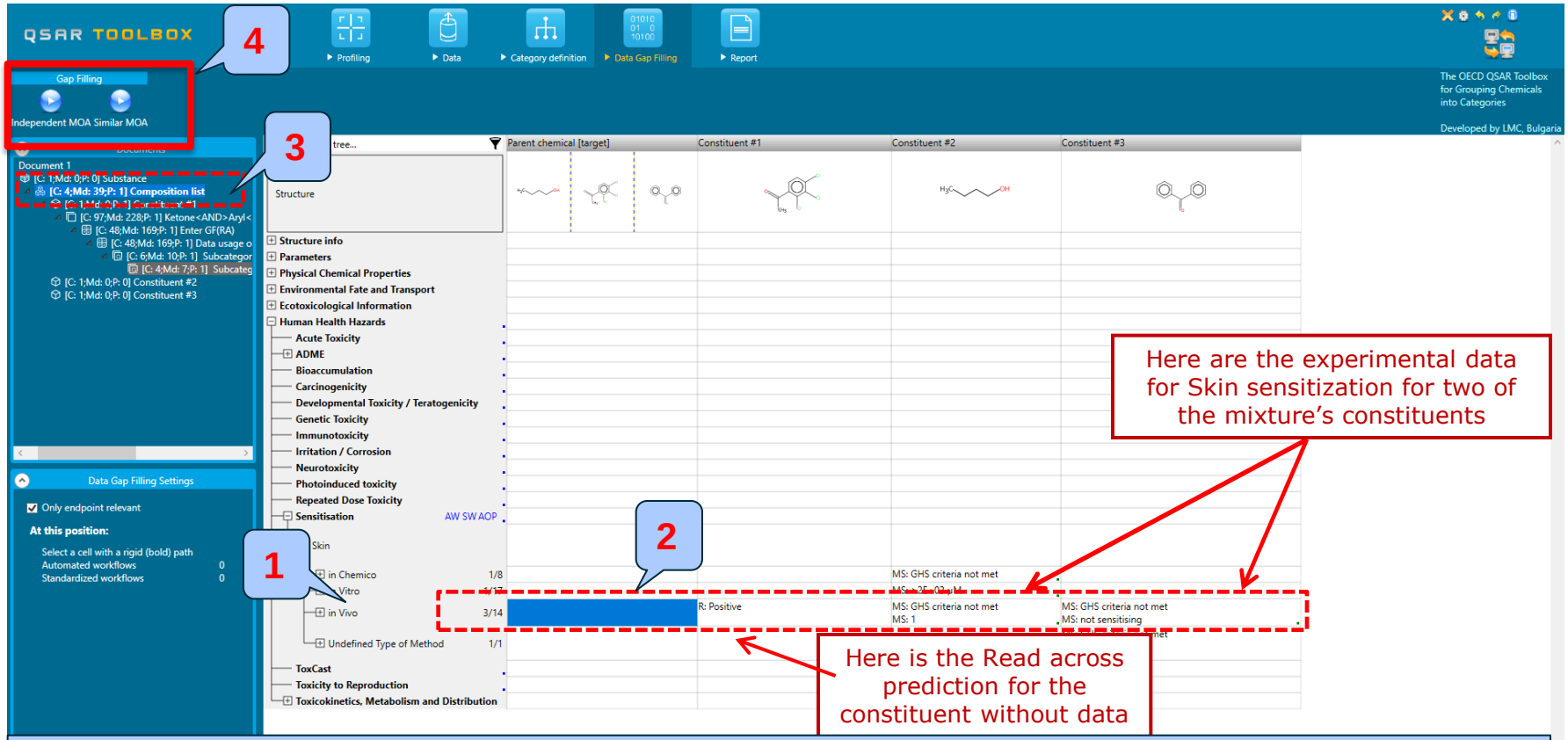
Data Gap Filling
Case study

- In this particular case components of the current mixture have different modes of action (constituent #1 and #3 have same mode, they do not interact with proteins (see slide 32), however constituent #2 interacts with proteins via Schiff-base mechanism). In this respect Independent MOA is applied;

- Application of Independent MOA for this case study is illustrated on the next slides

Filling data gap for skin sensitization of mixture

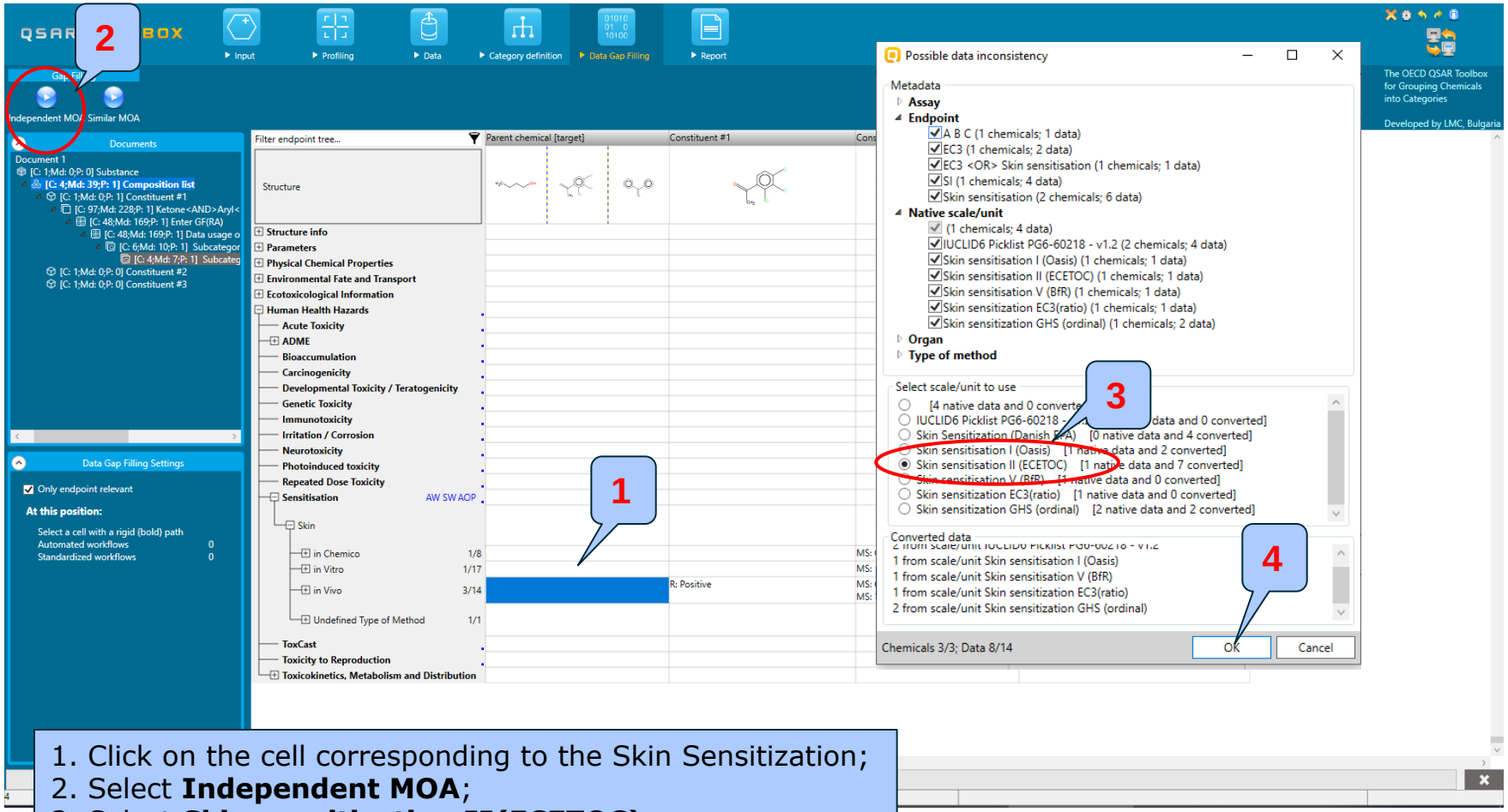
Applying Independent Mode of Action



1. Once you in the Data gap filling module, collapse the endpoint tree to the level of "in Vivo" as shown on the snapshot above;
2. Click on the cell corresponding to this mixture;
3. Click on **Composition list** (highlighted) from documented three;
4. The latter action automatically activate two mode of actions: **Independent MOA** and **Similar MOA.**

Filling data gap for skin sensitization of mixture

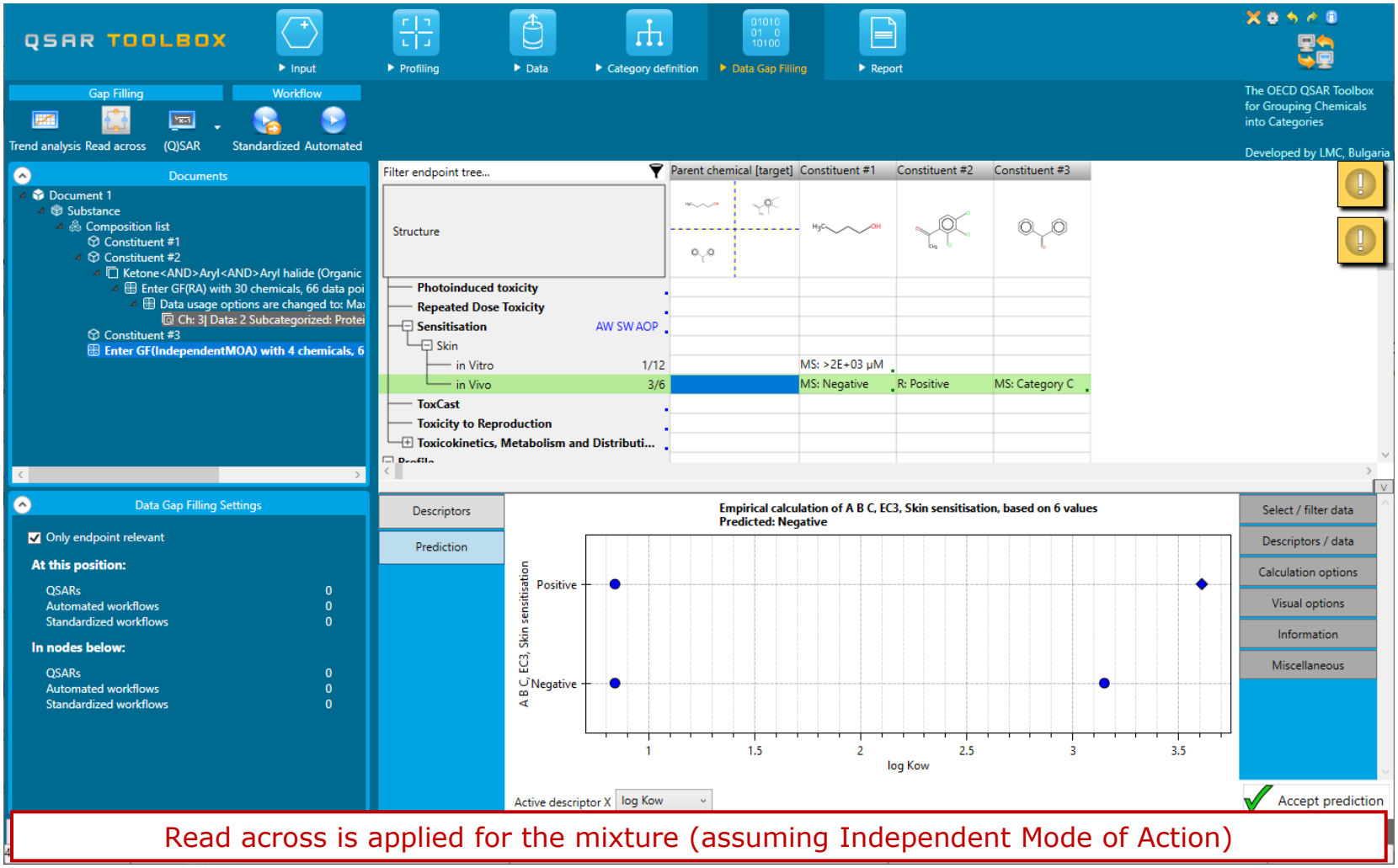
Applying Independent Mode of Action



1. Click on the cell corresponding to the Skin Sensitization;
2. Select **Independent MOA**;
3. Select **Skin sensitization II(ECETOC);**
4. Click **OK.**

Filling data gap for skin sensitization of mixture

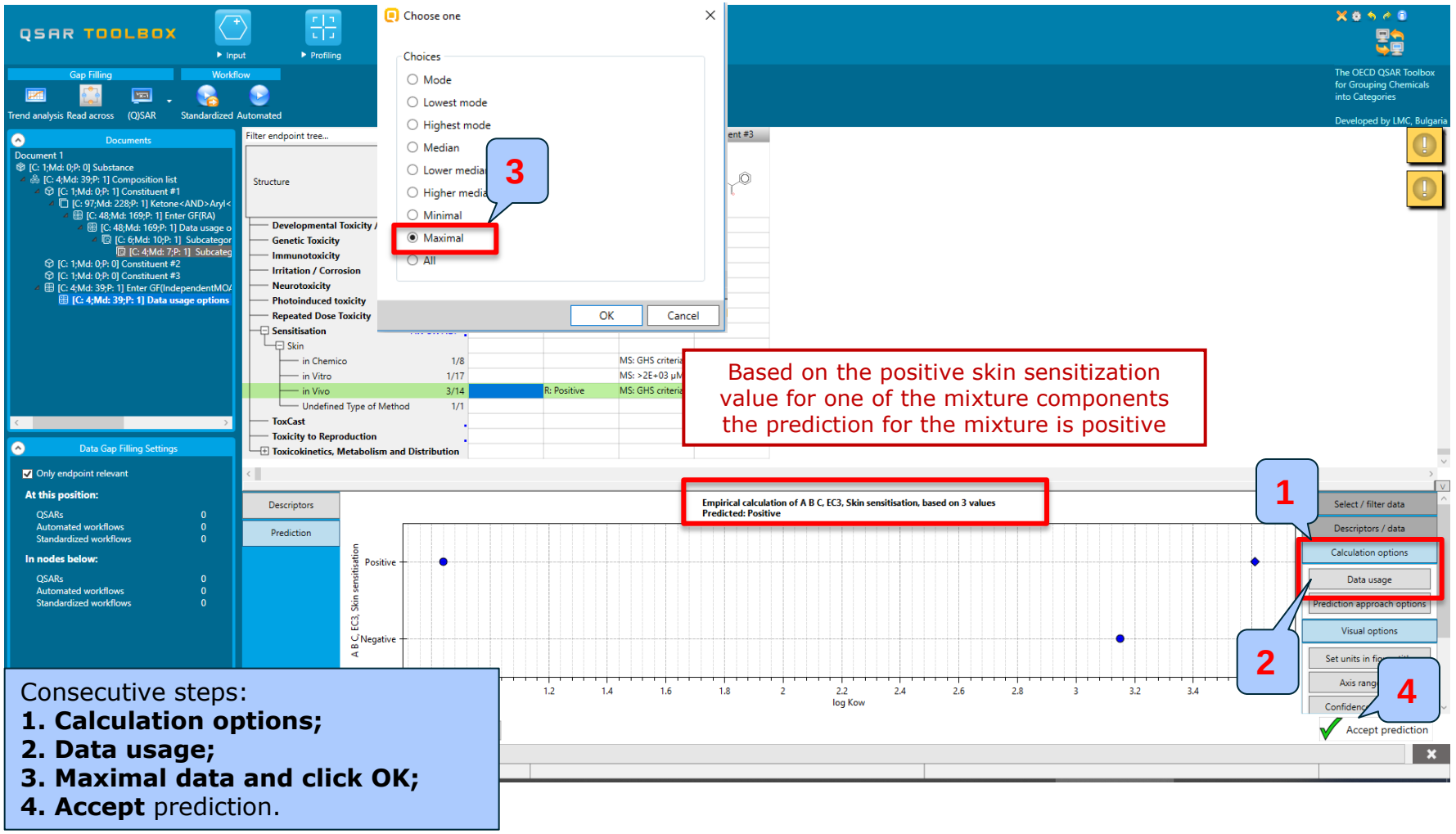
Applying Independent Mode of Action



Read across is applied for the mixture (assuming Independent Mode of Action)

Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action

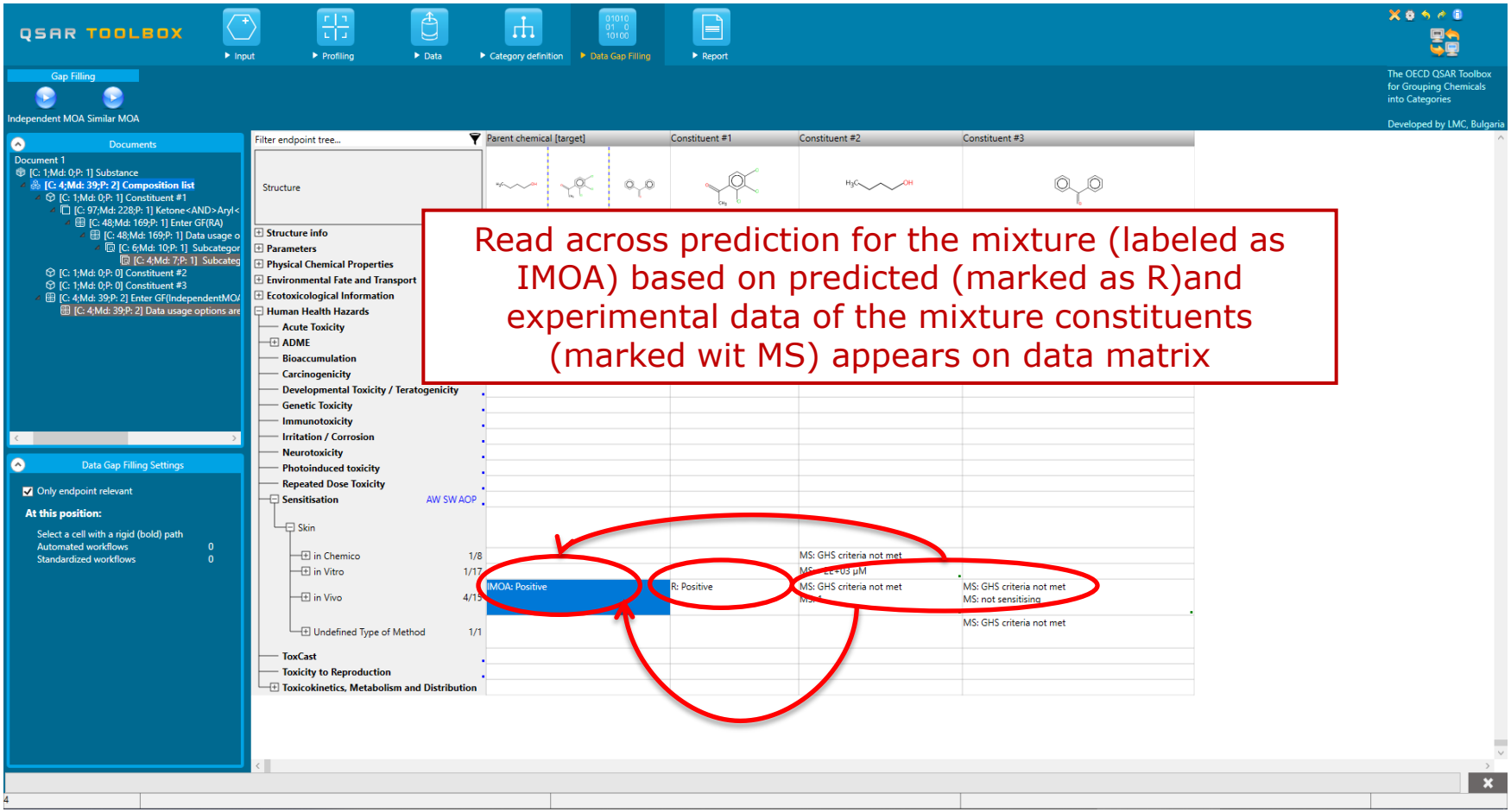


Consecutive steps:

1. **Calculation options;**
2. **Data usage;**
3. **Maximal data and click OK;**
4. **Accept** prediction.

Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action



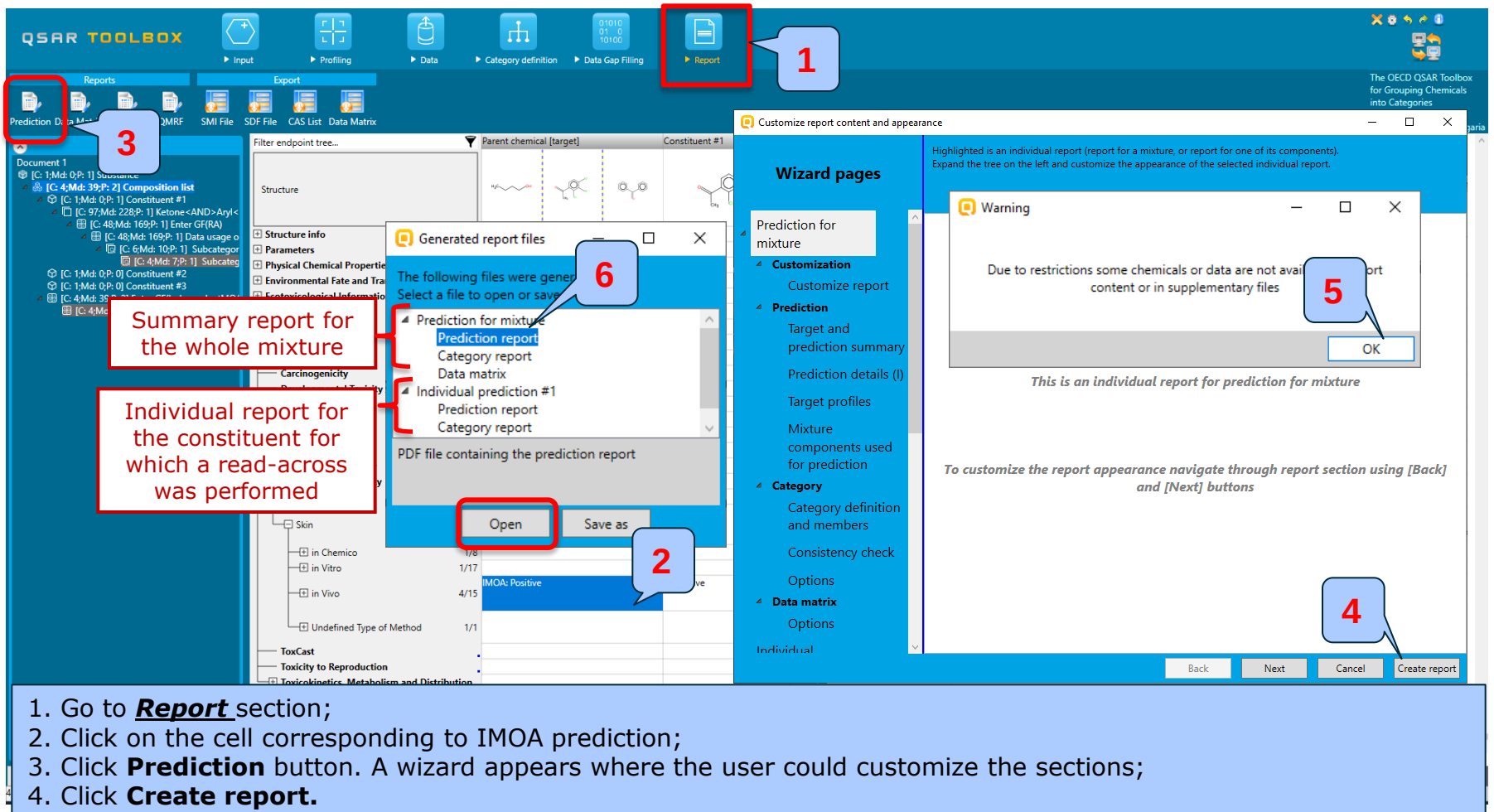
Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Read across prediction of constituent without data
  - Filling data gap for skin sensitization of mixture
  - **Generating report for mixture**

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 IMOA prediction;
3. Click **Prediction** button. A wizard appears where the user could customize the sections;
4. Click **Create report.**
5. Click **OK** on the appeared message; Two reports are generated: one summary report for the mixture and one for the read-across prediction for the constituent #1.
6. Select **prediction of mixture** and click **Open** button.

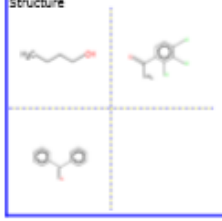
Report

QSAR Toolbox prediction for multicomponent substance

Based on observed and predicted data for mixture components

Date: 13 Apr 2020
Author(s):
Contact details:

Target information

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

Prediction summary

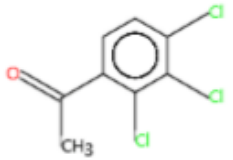
Predicted endpoint: A B C, EC3, Skin sensitisation; No effect specified; No species specified; No duration specified; No guideline specified
Predicted value: Positive
Unit/scale: Skin Sensitization (Danish EPA)
Data gap filling method: Independent mode of action
Summary: *manually editable field*
Not provided by the user

	Quantity, mol %	in Vivo, Skin Sensitization (Danish EPA)
Target mixture	-	

Mixture component #1	33.3	*Positive*
Mixture component #2	33.3	Positive
Mixture component #3	33.3	Category C

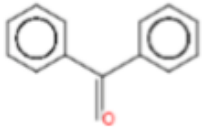
NOTE: Predicted values are written in italic

Mixture component #1
CC(=O)c1ccc(Cl)c(Cl)c1Cl



Mixture component #2
CCCCO

Mixture component #3
O=C(c1ccccc1)c1ccccc1



Information of the mixture's constituents

Prediction summary

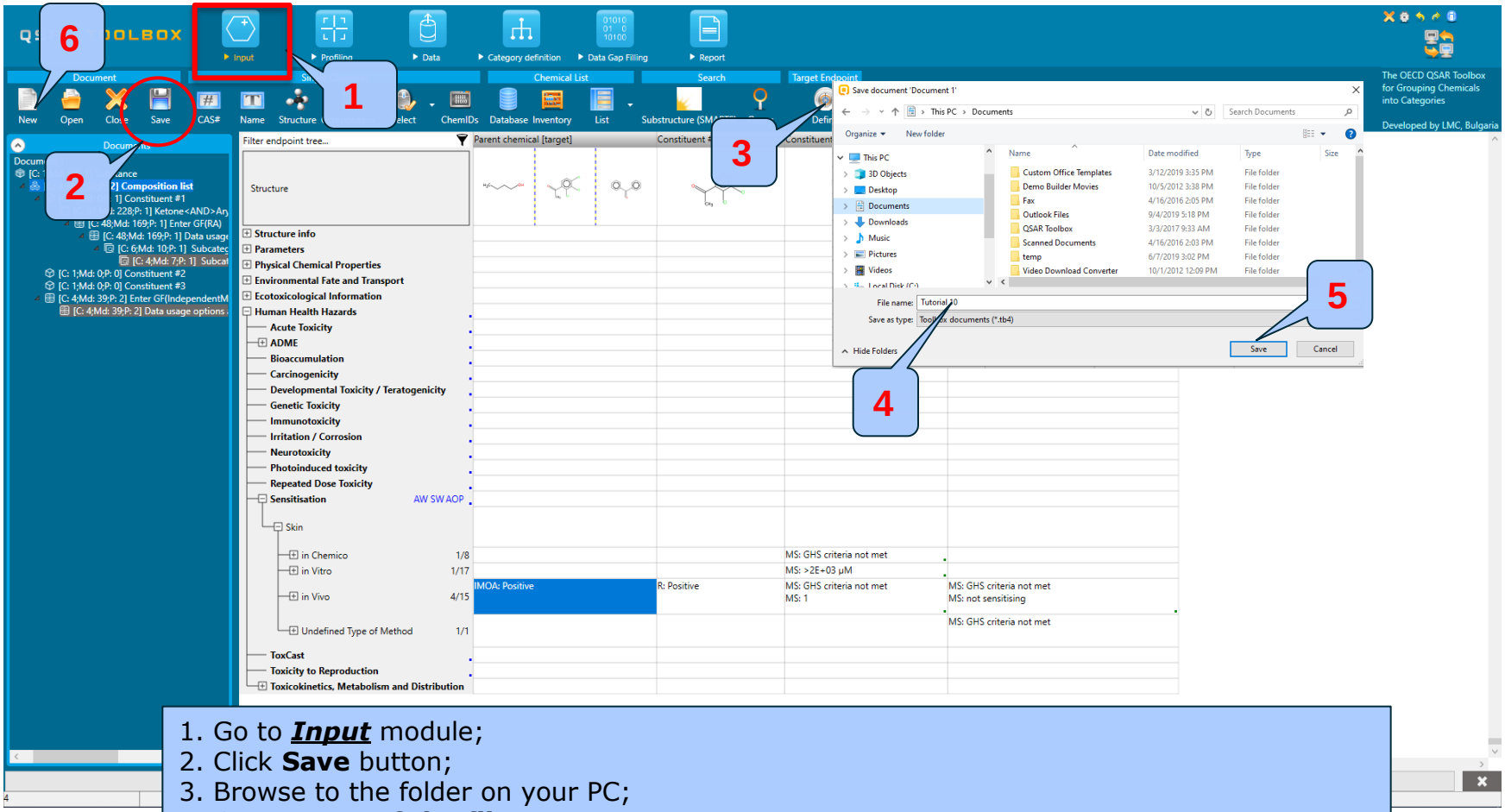
Outlook

- Background
- Keywords
- Objectives
- The exercise
- Workflow
- **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 next screenshot.

Saving the prediction result



1. Go to ***Input*** module;
2. Click **Save** button;
3. Browse to the folder on your PC;
4. Gave **name of the file**;
5. Click **Save** button. The file is saved and could be opened later by using **Open** button (6).