QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Example for predicting Skin Sensitization of mixture with known components

Outlook

- **Background**
- 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 skin sensitization of mixture with known components

Outlook

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

Exercise

- In this exercise we will predict the skin sensitization of mixture, which is the "target" chemical.
- Investigate the mode of action for each components of the mixture
- Gather available experimental data for target chemical
- Investigate skin sensitization of non-tested component
- Applying read across for non-tested component
- Predict skin sensitization potential of mixture based on experimental data of tested compounds and predicted data of non-tested one.

Outlook

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

Workflow

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

Outlook

- Background
- 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

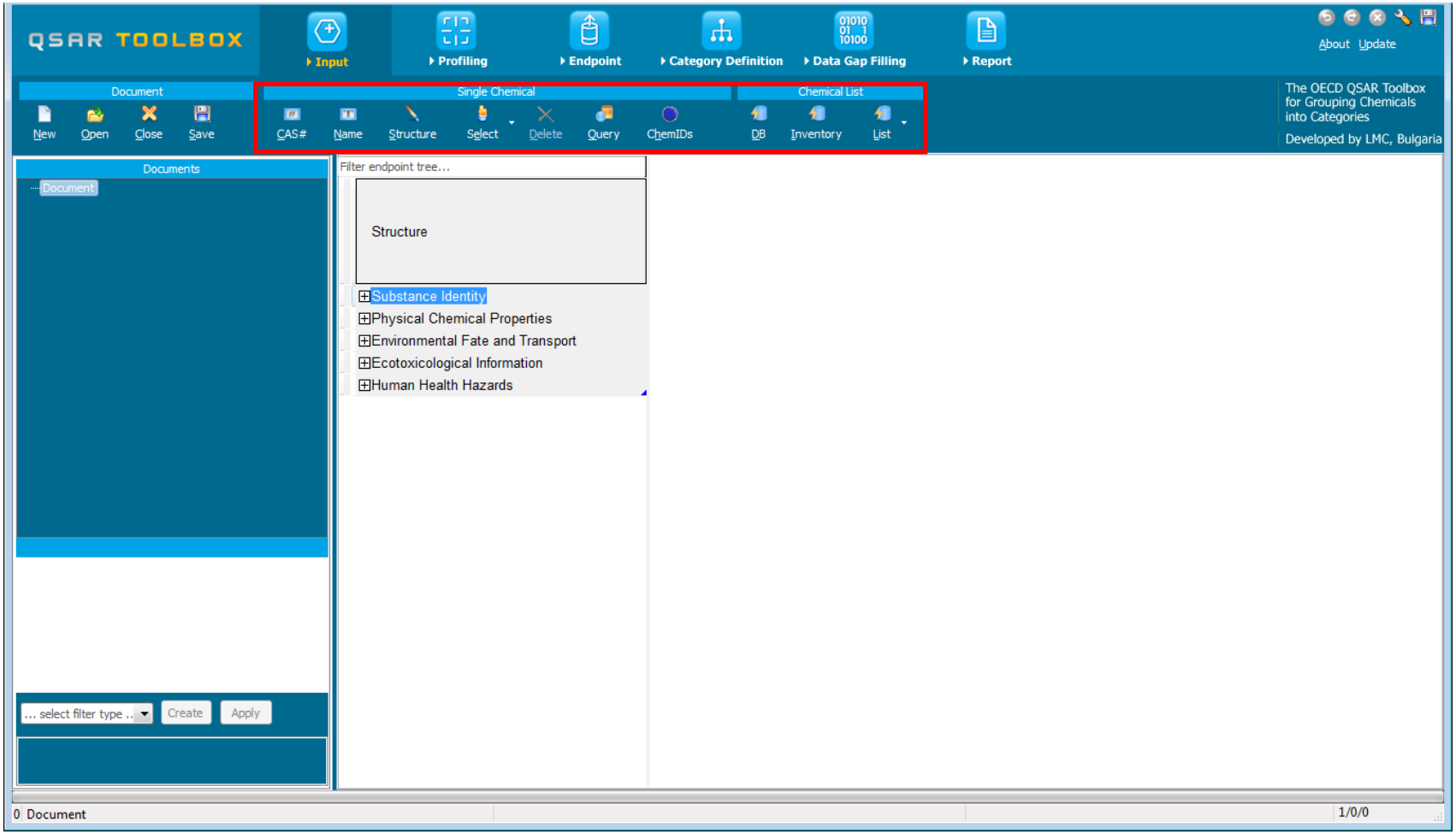
User alternatives for defining mixtures with known compositions:

- 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
- Chemical IDs such as EC number, Einecs number
- Load file with mixture

Getting Started

- Open the Toolbox.
- The six modules in the workflow are seen listed next to "QSAR TOOLBOX".
- Click on "Input" (see next screen shot)

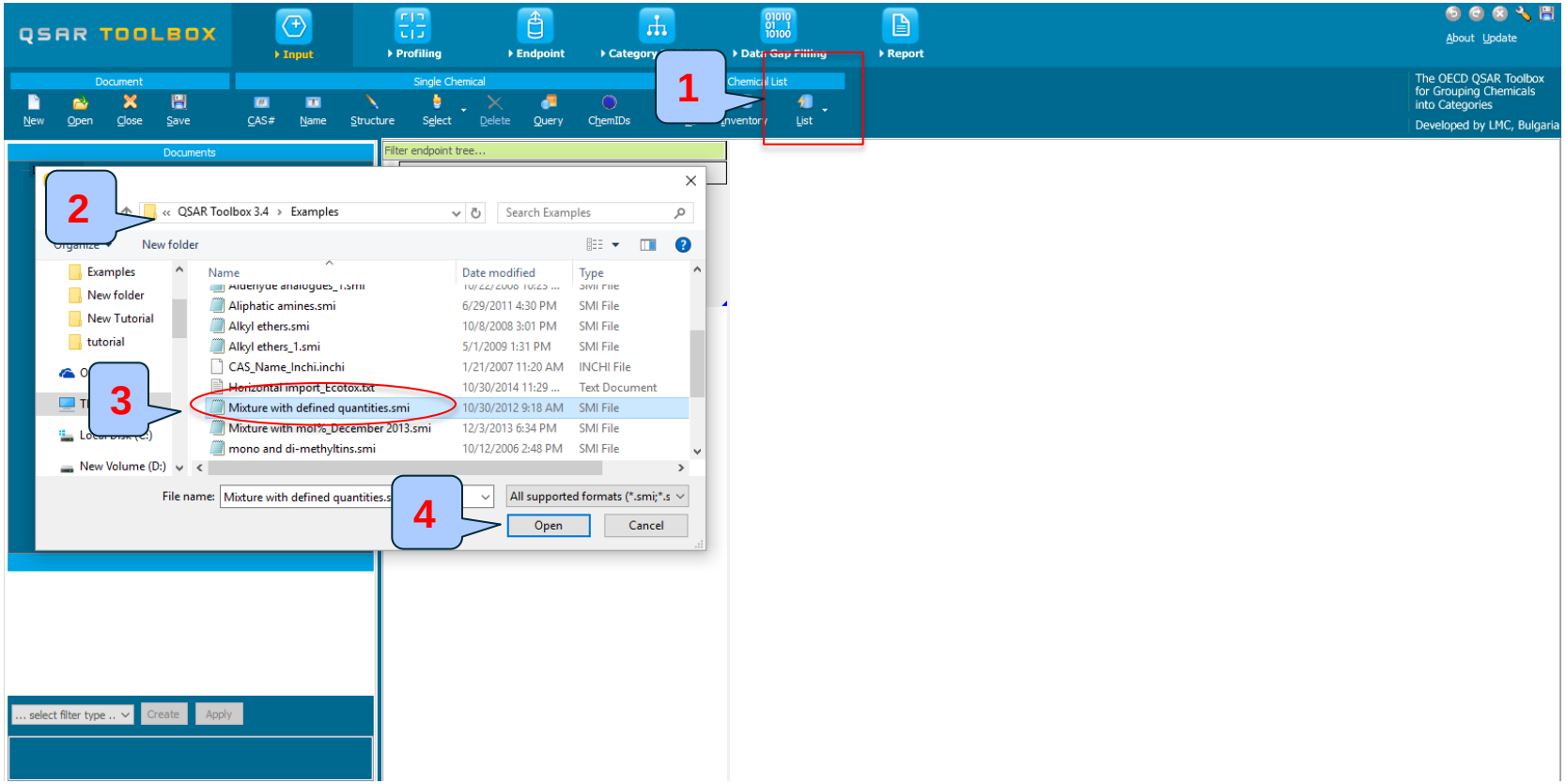
Chemical Input
Input Screen



Chemical input
Load list with chemical mixture

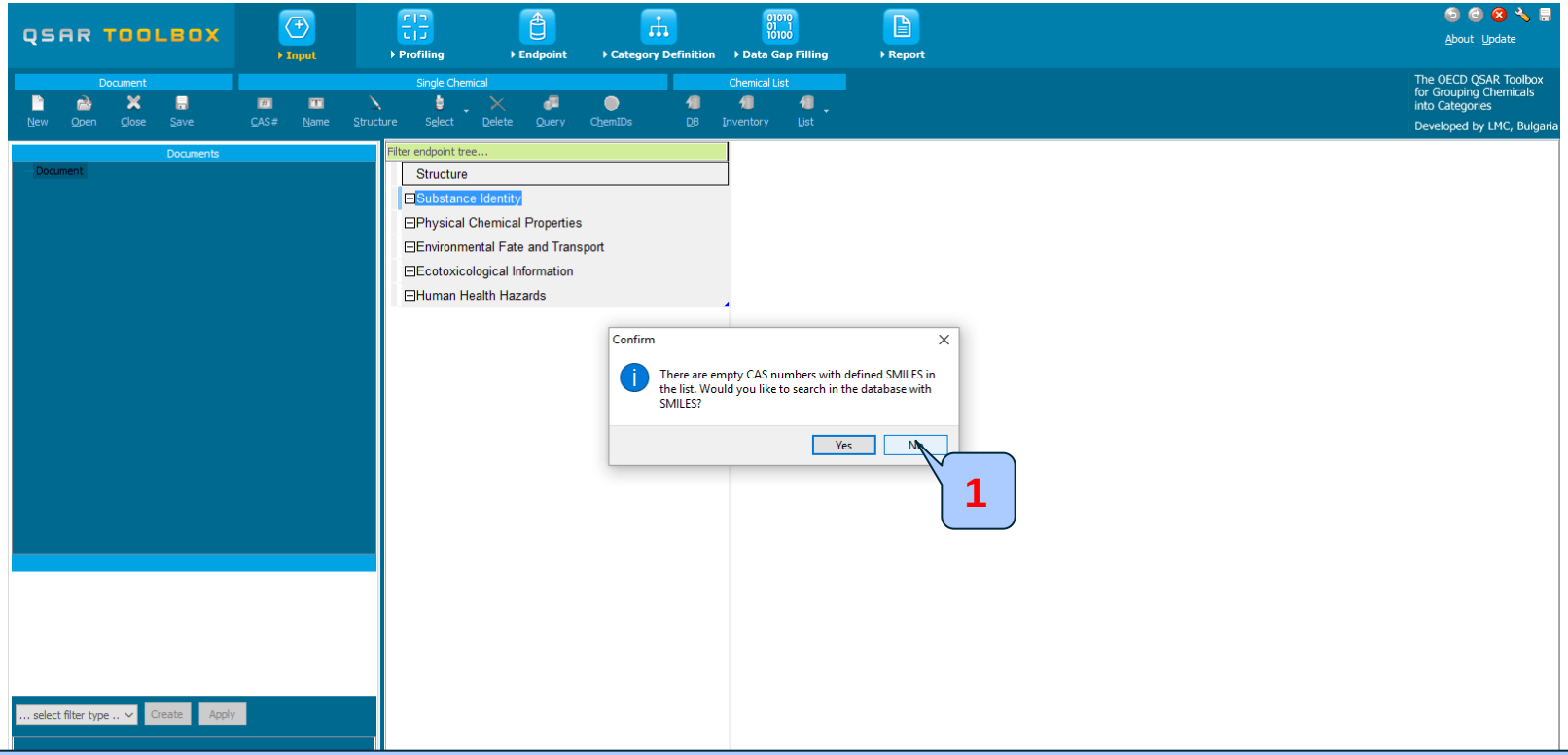
- Toolbox allows to enter target chemicals through tab delimited file
- This requires mixture with defined components to be previously defined in a tab delimited file
- The subsequent series of screen shots will take you through the process of entering the target chemical via tab delimited file
- In this particular case, the example file with mixture is available in the Example directory of Toolbox installation (C:\Program Files (x86)\QSAR Toolbox\QSAR Toolbox 3\Examples)

Chemical input
Load list with chemical mixture



1. **Click** on Chemical list
2. **Browse** and find the file with mixture located at Examples directory
3. **Select** the file
4. **Open** the file "Mixture with defined quantities.smi"

Chemical input
Load list with chemical mixture



The notification message appears, informing the user that there are structures without CAS numbers. If you want the software to search databases for their CAS numbers, click Yes, otherwise click No.

1. **Select** No

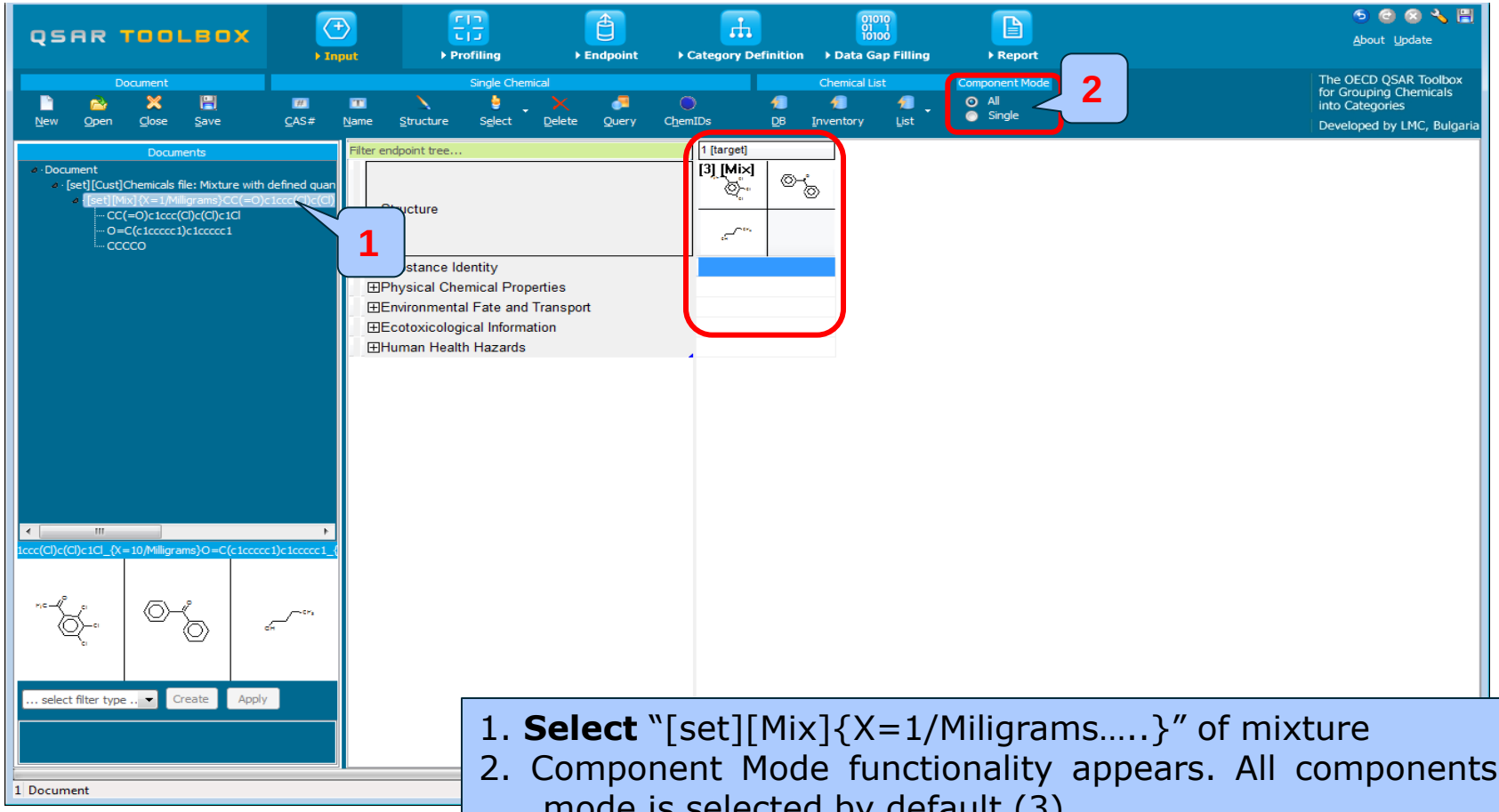
Chemical Input
Target chemical identity

- The already drawn target structure automatically appears on the data matrix

- Note that no CAS number or name is displayed for this chemical. This means the target chemical is not listed in the chemical inventories/databases available in Toolbox(see next slide).

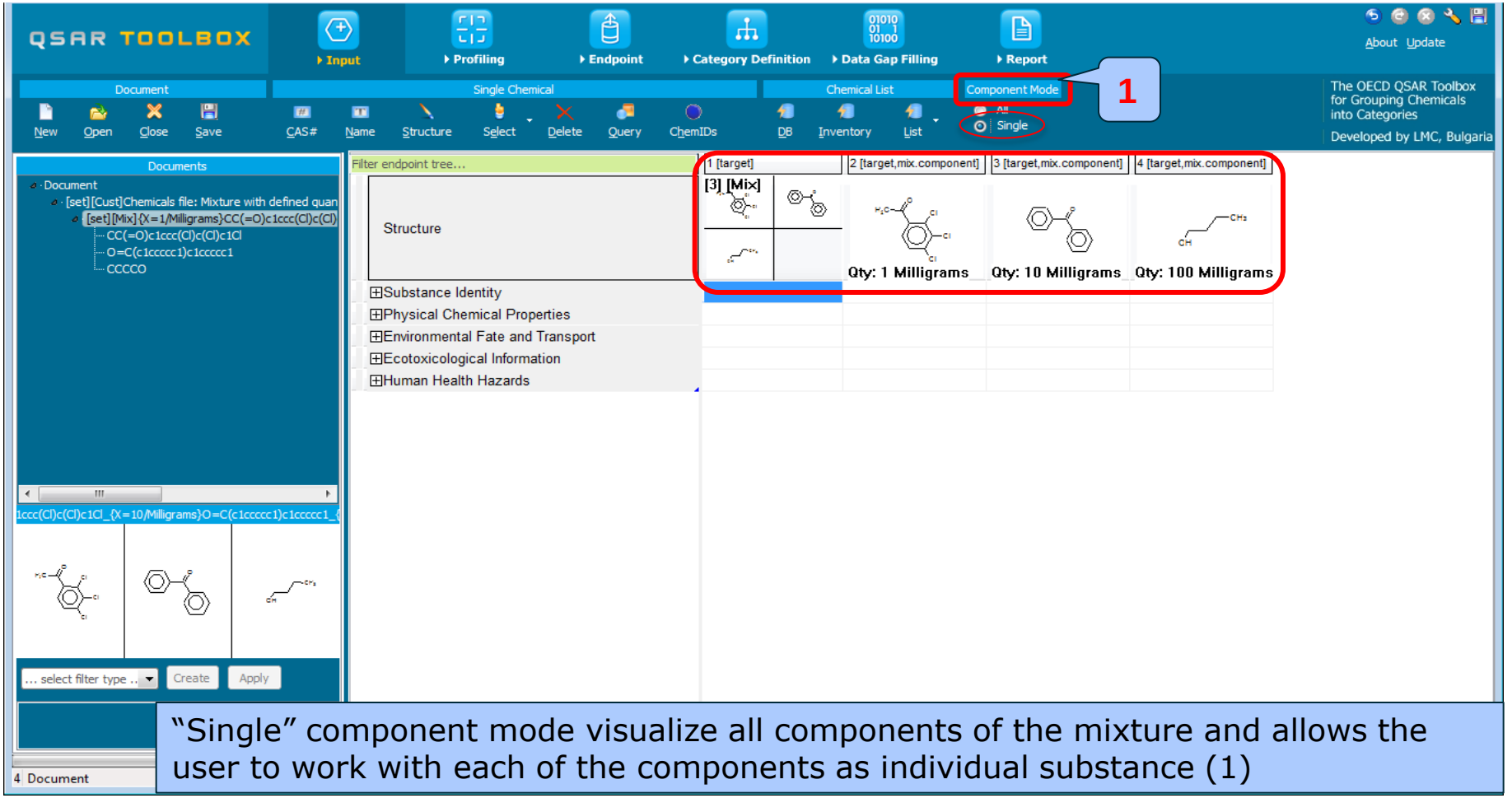
- Visualization of components of the mixture is possible when user selects Single Component Mode

Chemical Input
Target chemical identity



1. **Select** "[set][Mix]{X=1/Miligrams…..}" of mixture
2. Component Mode functionality appears. All components mode is selected by default (3)

Chemical Input
Target chemical identity



"Single" component mode visualize all components of the mixture and allows the user to work with each of the components as individual substance (1)

Outlook

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

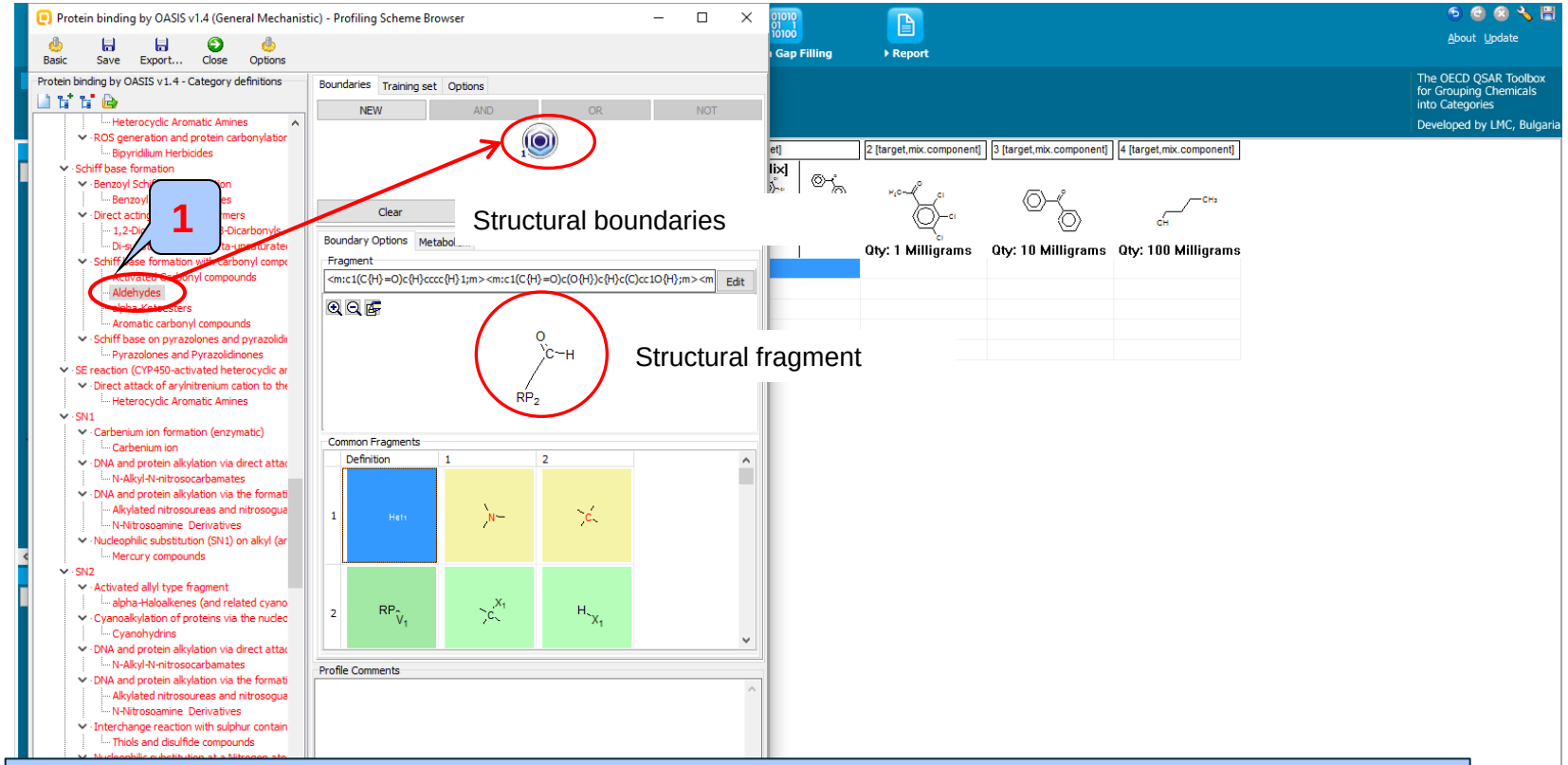
Profiling
Overview

- "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.

Profiling
Side-Bar to Profiling

- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, Protein binding by OASIS v1.4 and clicking on “View” (see next screen shot).

Profiling
Side-Bar to Profiling



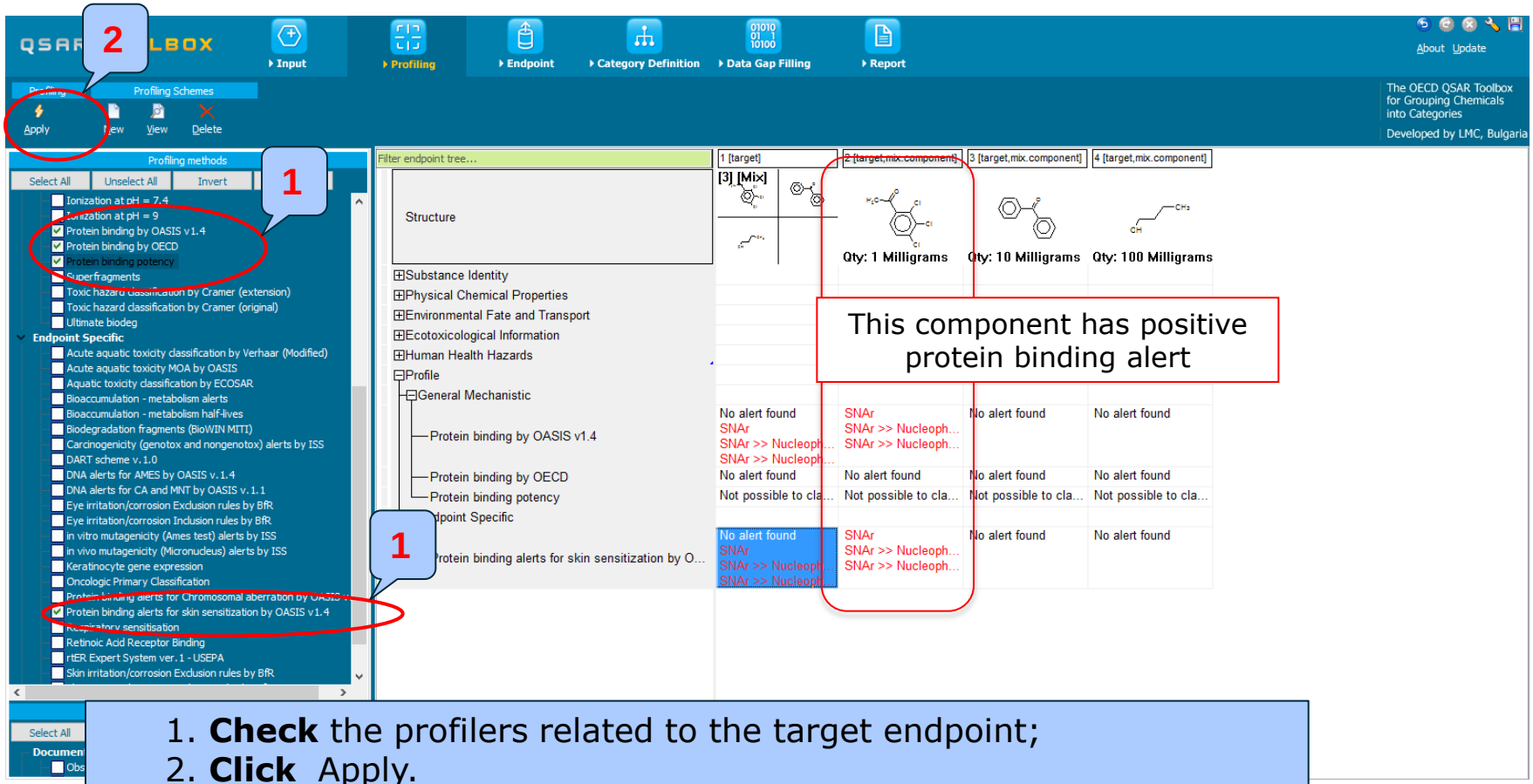
1. Select "Aldehydes"

Profiling
Profiling the target chemical

- Select the "Profiling methods" related to the target endpoint by clicking on the box next to the profilers name.
- This selects (a green check mark appears) or deselects (green check disappears) profilers.
- For this example, select the following profilers relevant to the Skin sensitization (see next screenshot):

  - Protein binding by OASIS v1.4 – general mechanistic
  - Protein binding by OECD – general mechanistic
  - Protein binding potency – general mechanistic
  - Protein binding alerts for skin sensitization by OASIS v1.4 – endpoint specific

Profiling
Side-Bar to Profiling



Outlook

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

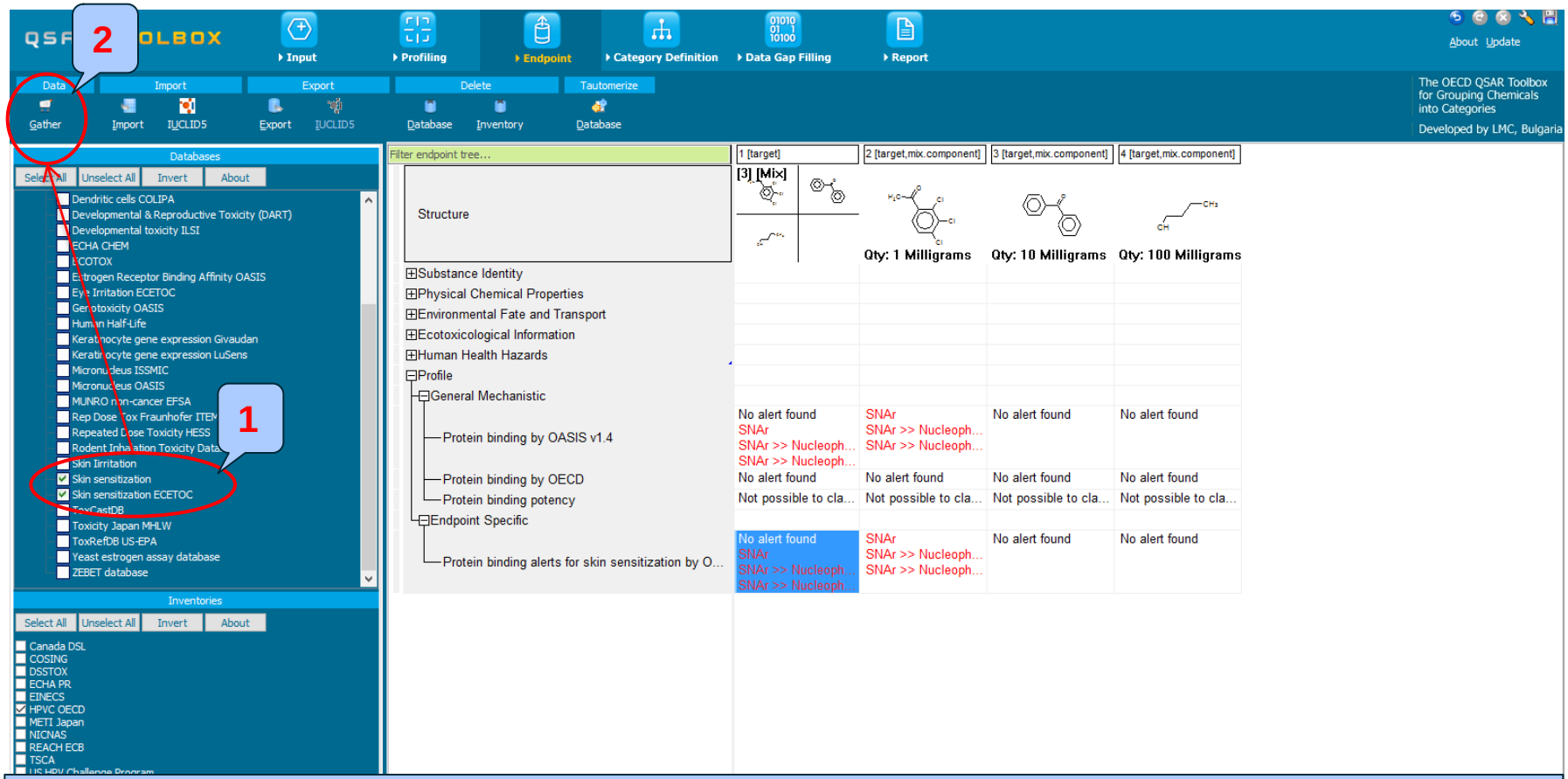
Endpoint

- "Endpoint" 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).

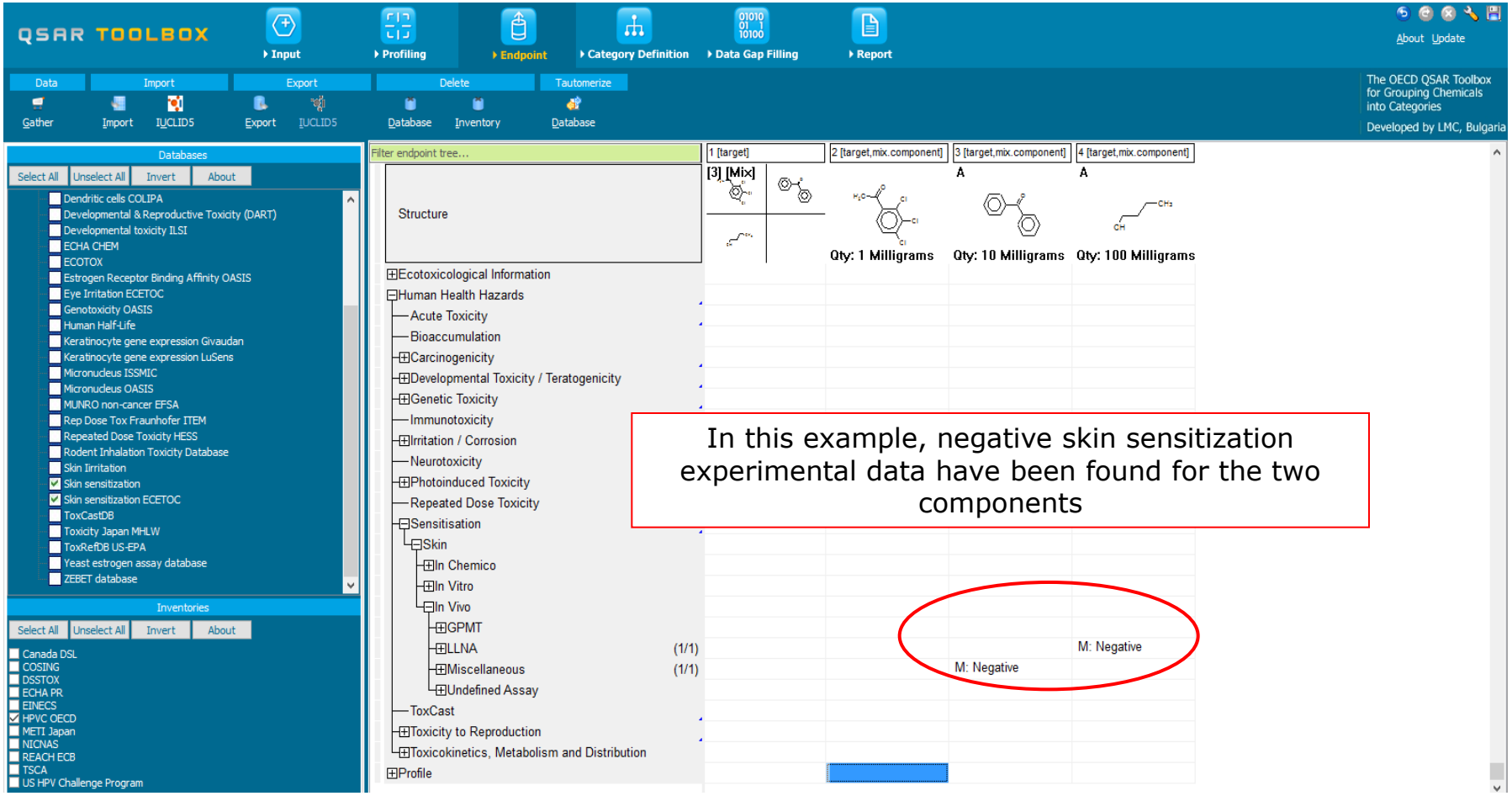
- In this example, we limit our data gathering to the common Skin endpoints from databases containing Skin Sensitization data

Endpoint



1. **Select** databases related to the target endpoint by adding a green check in the box before the database name.
2. **Click** Gather

Endpoint
Process of collecting data



Recap

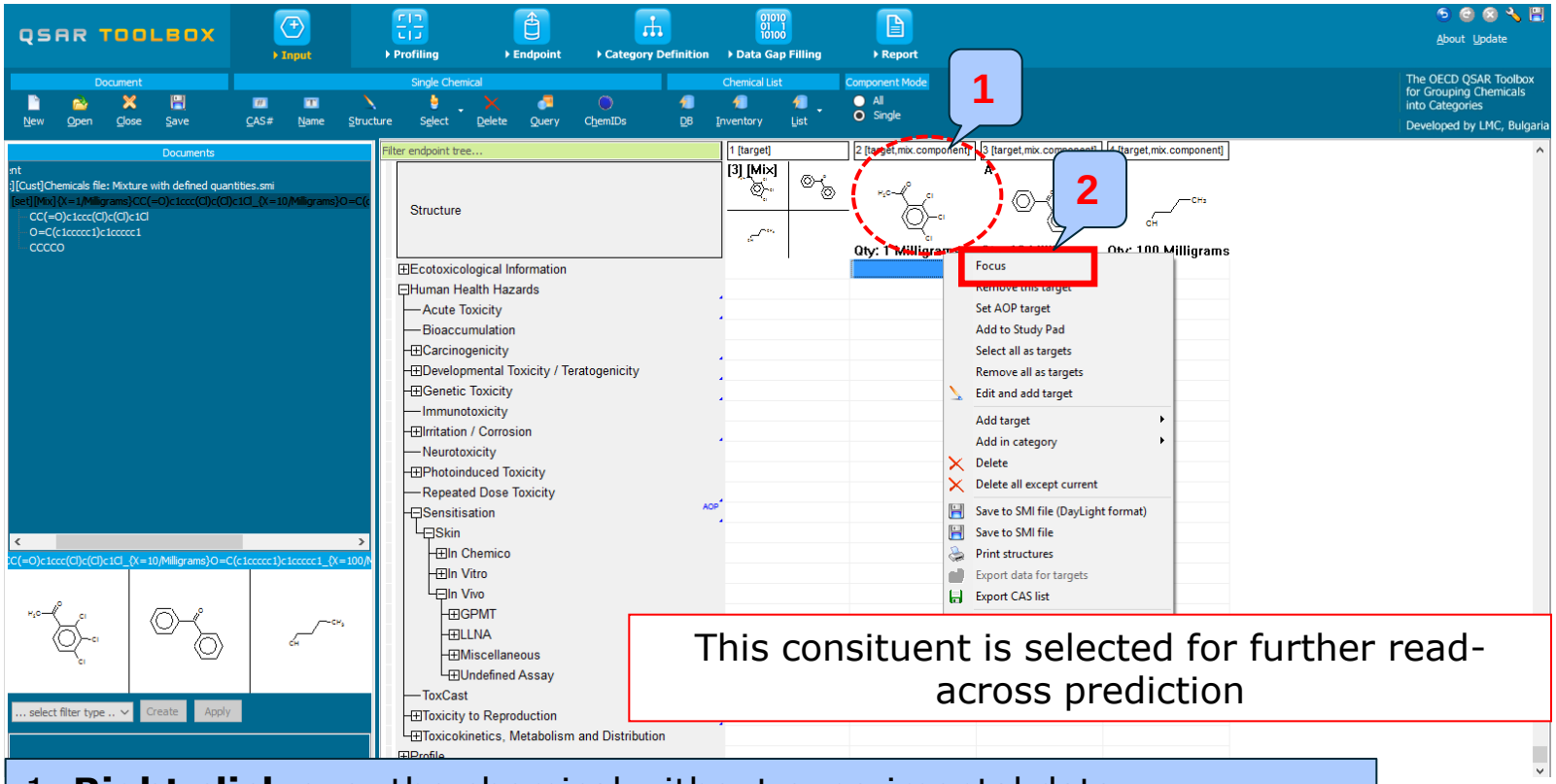
- You have entered the mixture with defined components
- You have profiled the target chemical mixture and found no protein binding alerts for two of the mixture constituents. The third constituent has positive protein binding alerts and could elicit skin sensitization effect
- Negative experimental data has been found for two of mixture components. No experimental data has been found for the third constituent
- The constituent without experimental data and positive protein binding alert has been used for further read across analysis. Then, all of the available data – experimental and predicted will be used for SS prediction of the mixture.
- Now you are ready to continue with "Read across prediction of constituent without data".

Outlook

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

Read across prediction of constituent without data

Focus constituent

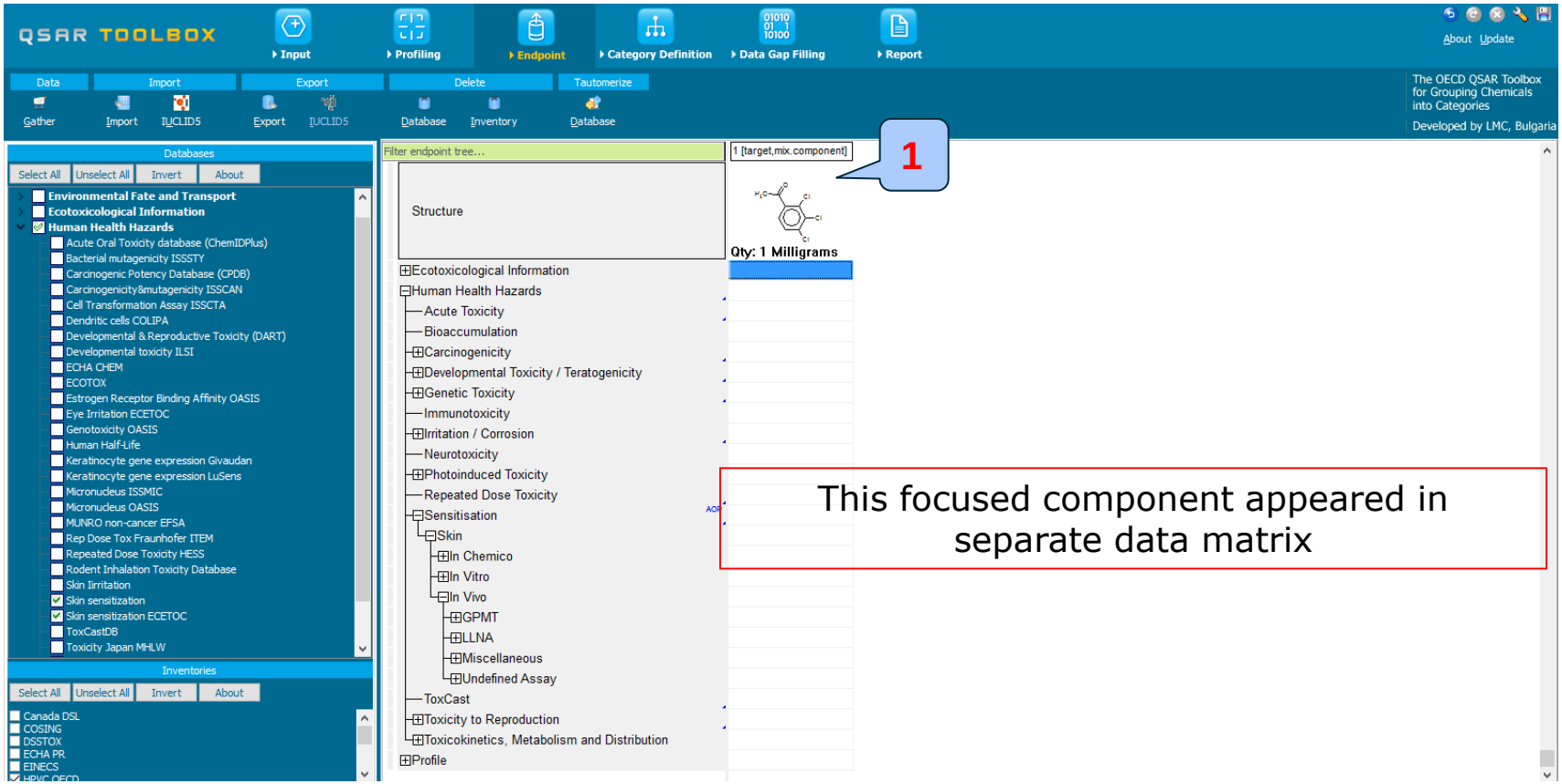


This consituent is selected for further read-across prediction

1. **Right click** over the chemical without experimental data
2. **Select** Focus

Read across prediction of constituent without data

Focus constituent



Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **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.

Category Definition
Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of "similarity".

- Detailed information about grouping chemical (Chapter 4) could be found in document "Manual for Getting started" published on OECD website:

  http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm

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

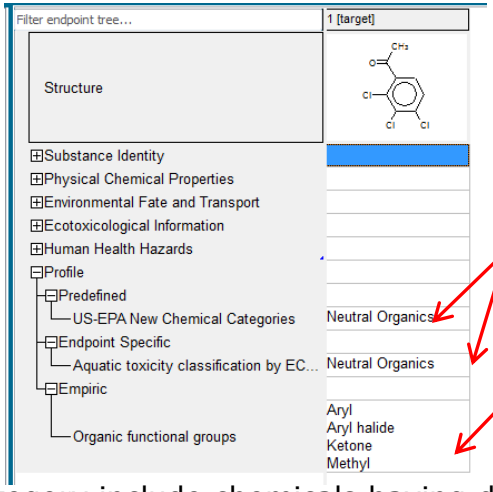
Phase I. Structure based

- 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

46 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 and classifications from structurally similar profilers 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 by OASIS v1.4.

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 LMC and OECD website:

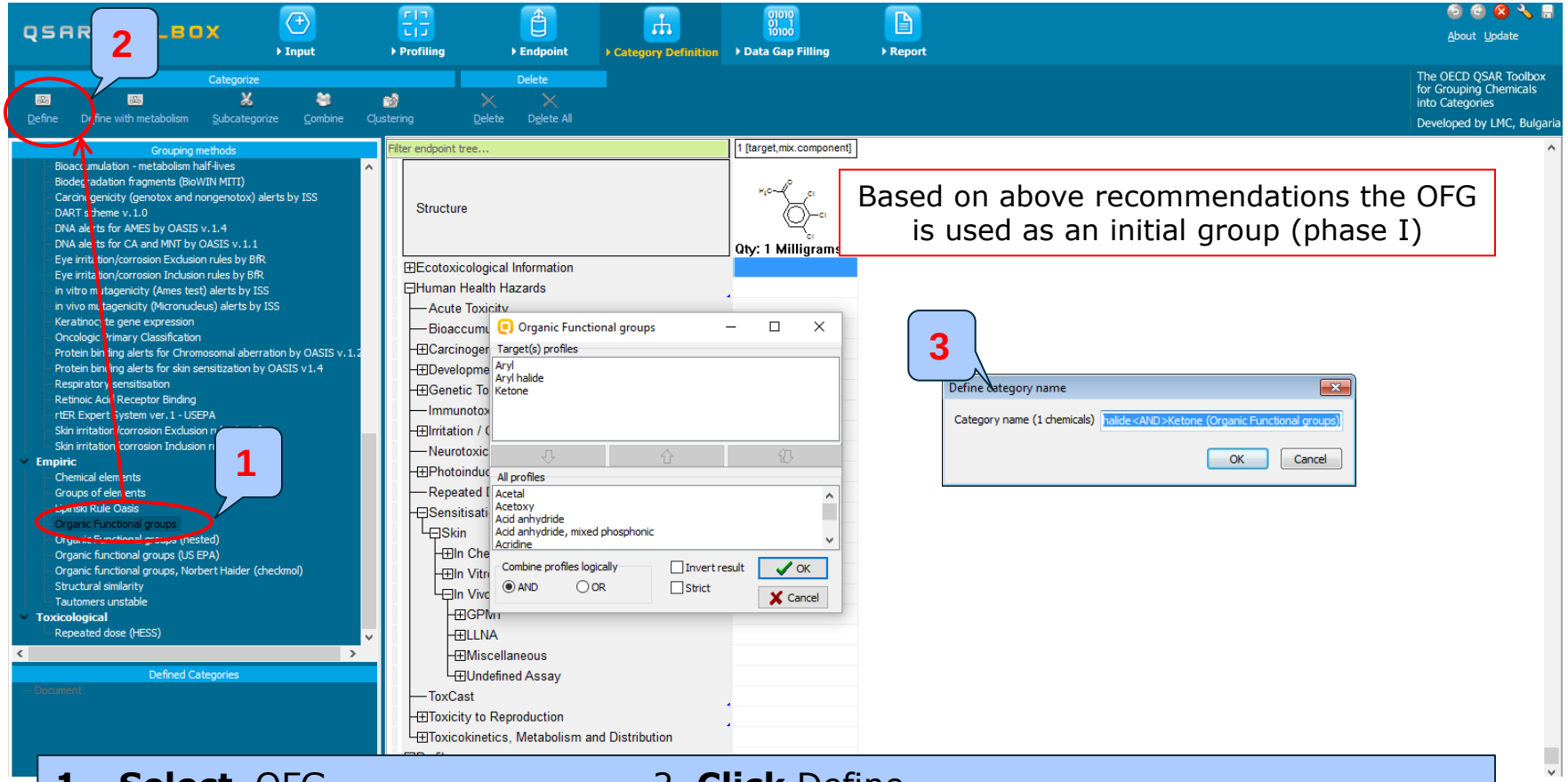
http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm

http://superhosting.oasis-lmc.org/products/software/toolbox/toolbox-support.aspx

See next slides

Read across prediction of constituent without data

Define category by OFG



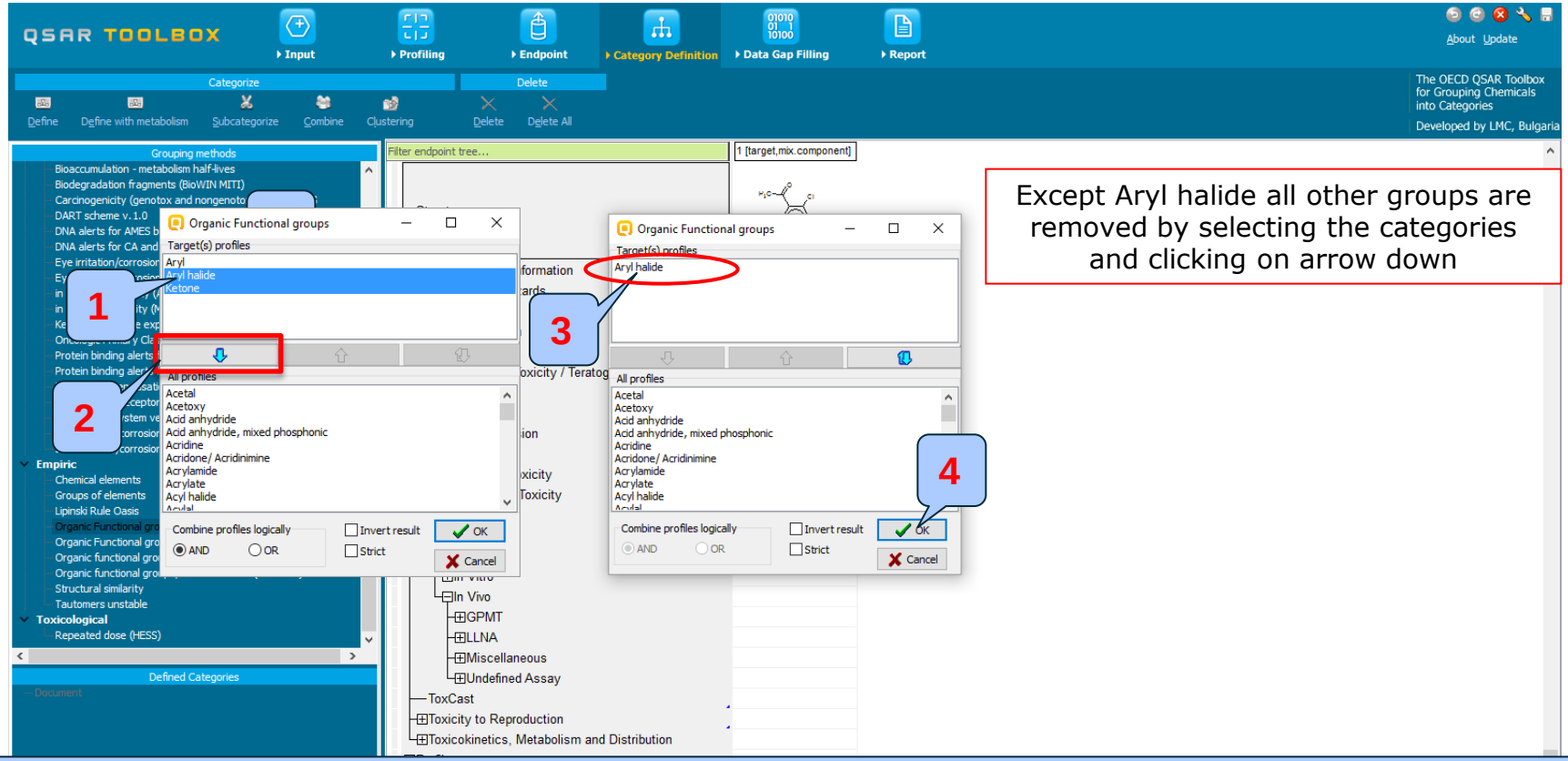
1. Select OFG 2. **Click** Define

Combination of three organic functional groups identified a single analogue (**3**).

In order to expand the category Aryl halide is used only. See next slide

Read across prediction of constituent without data

Define category by OFG



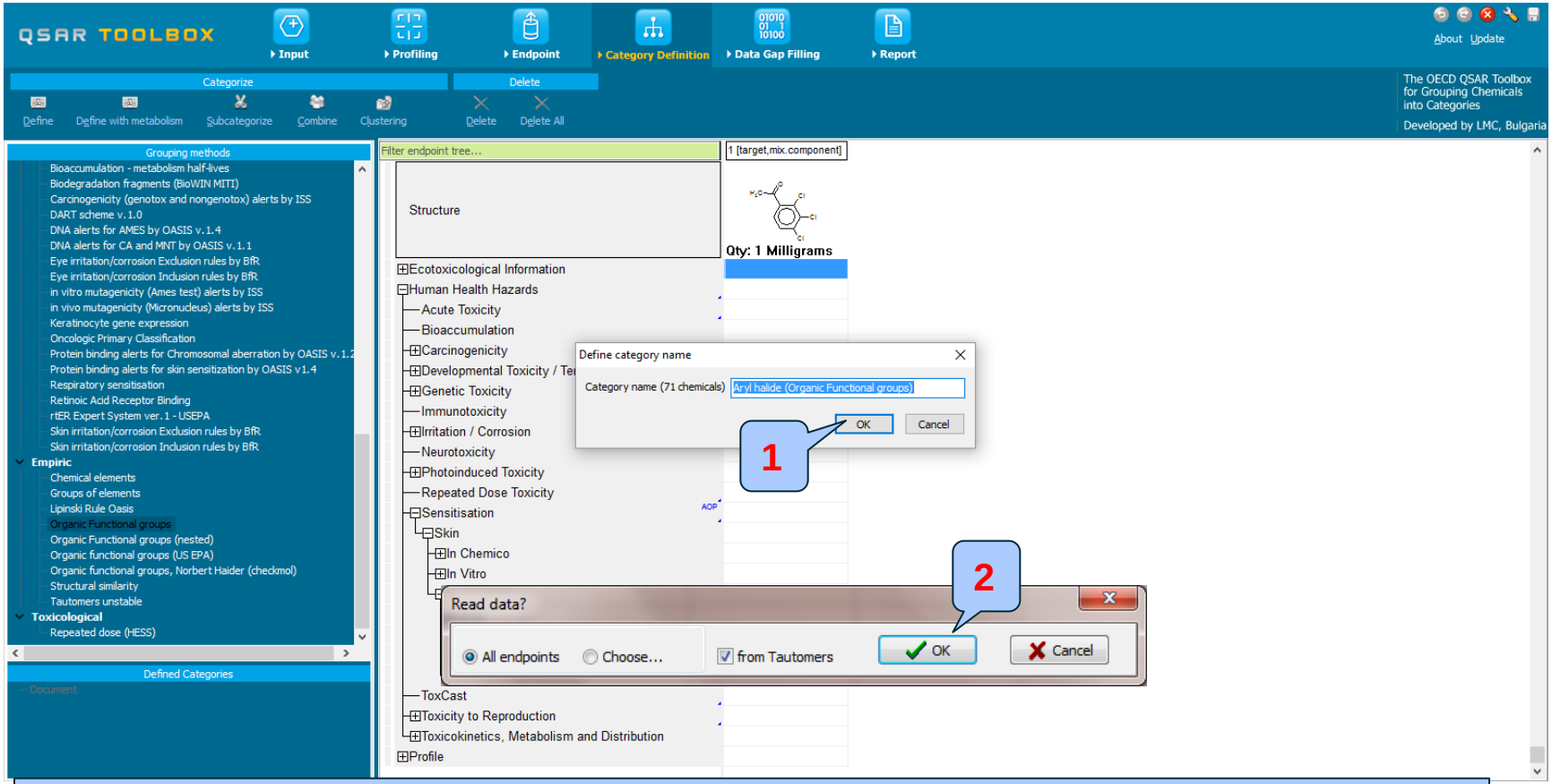
1. Select Aryl and Ketone (hold Ctrl button) and remove them by the arrow down 3. Aryl halide should be remained only 4. **Click** OK

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Read across prediction of constituent without data**
    - **Focus constituent without experimental data**
    - **Define category**
    - **Gather data for analogues**

Read across prediction of constituent without data

Gather data for analogues chemicals

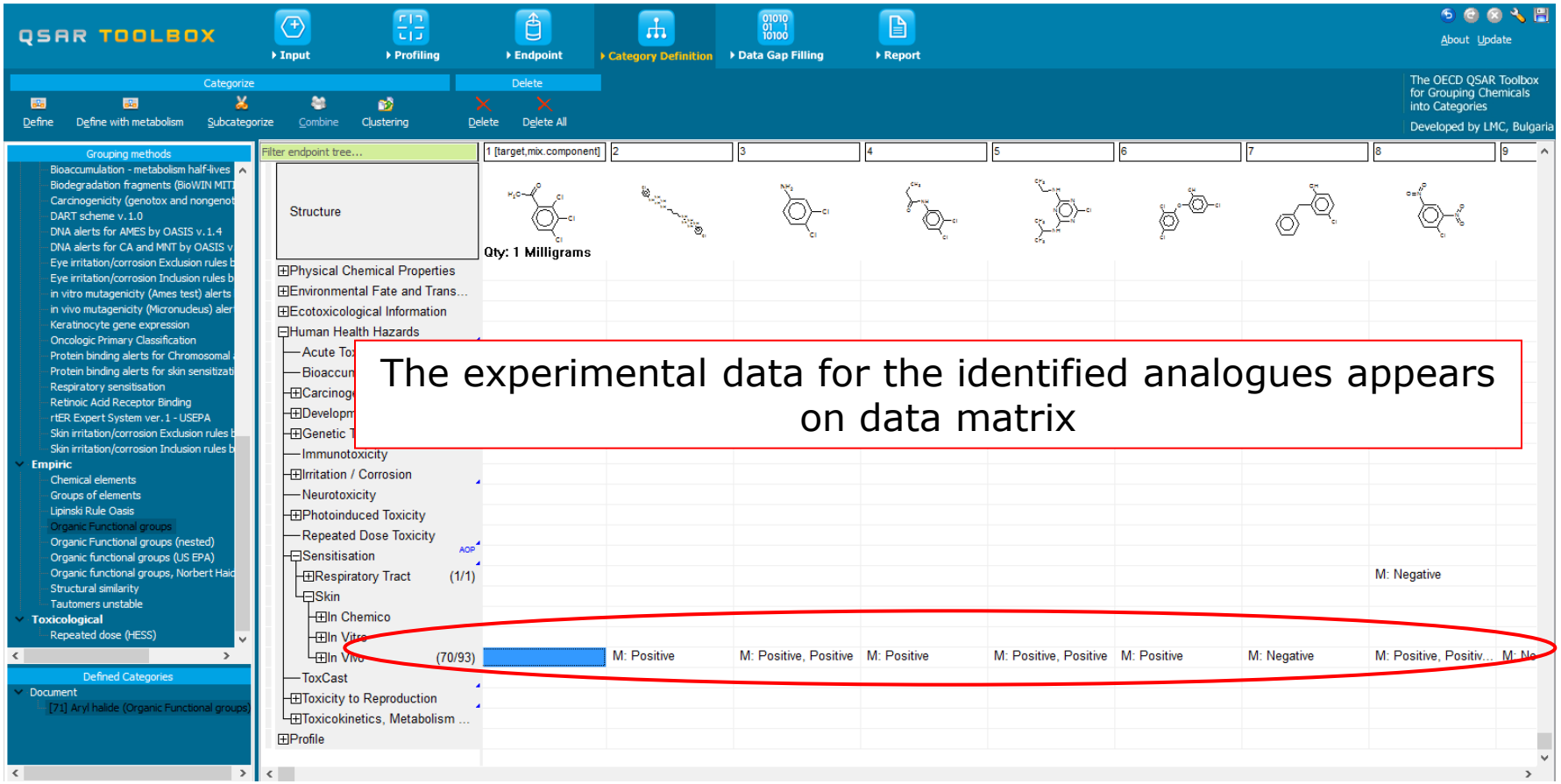


1. Click OK

2. **Click** OK in order to read data for all endpoints

Read across prediction of constituent without data

Gather data for analogues chemicals

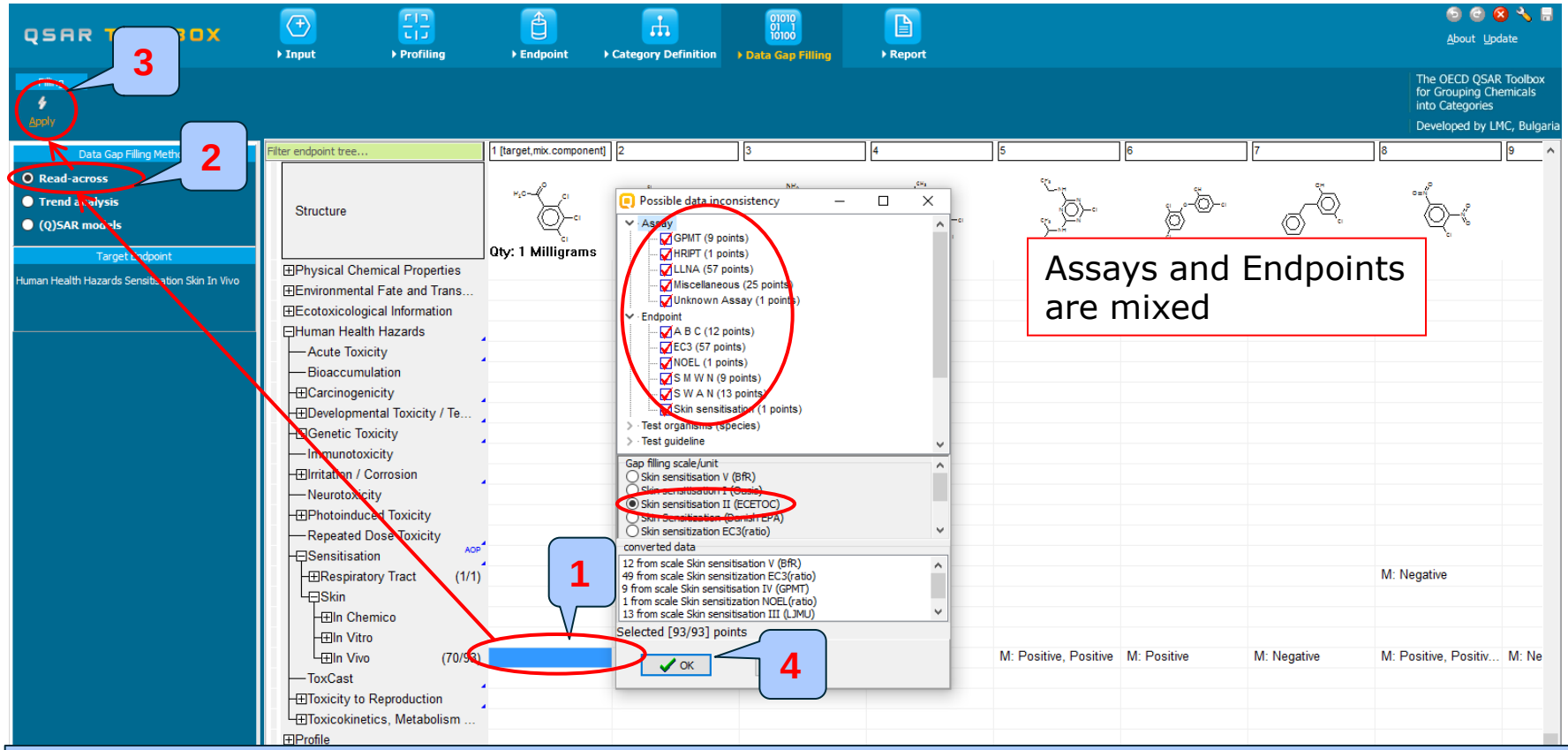


Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Read across prediction of constituent without data**
    - **Focus constituent without experimental data**
    - **Define category**
    - **Gather data for analogues**
    - **Apply read across**

Read across prediction of constituent without data

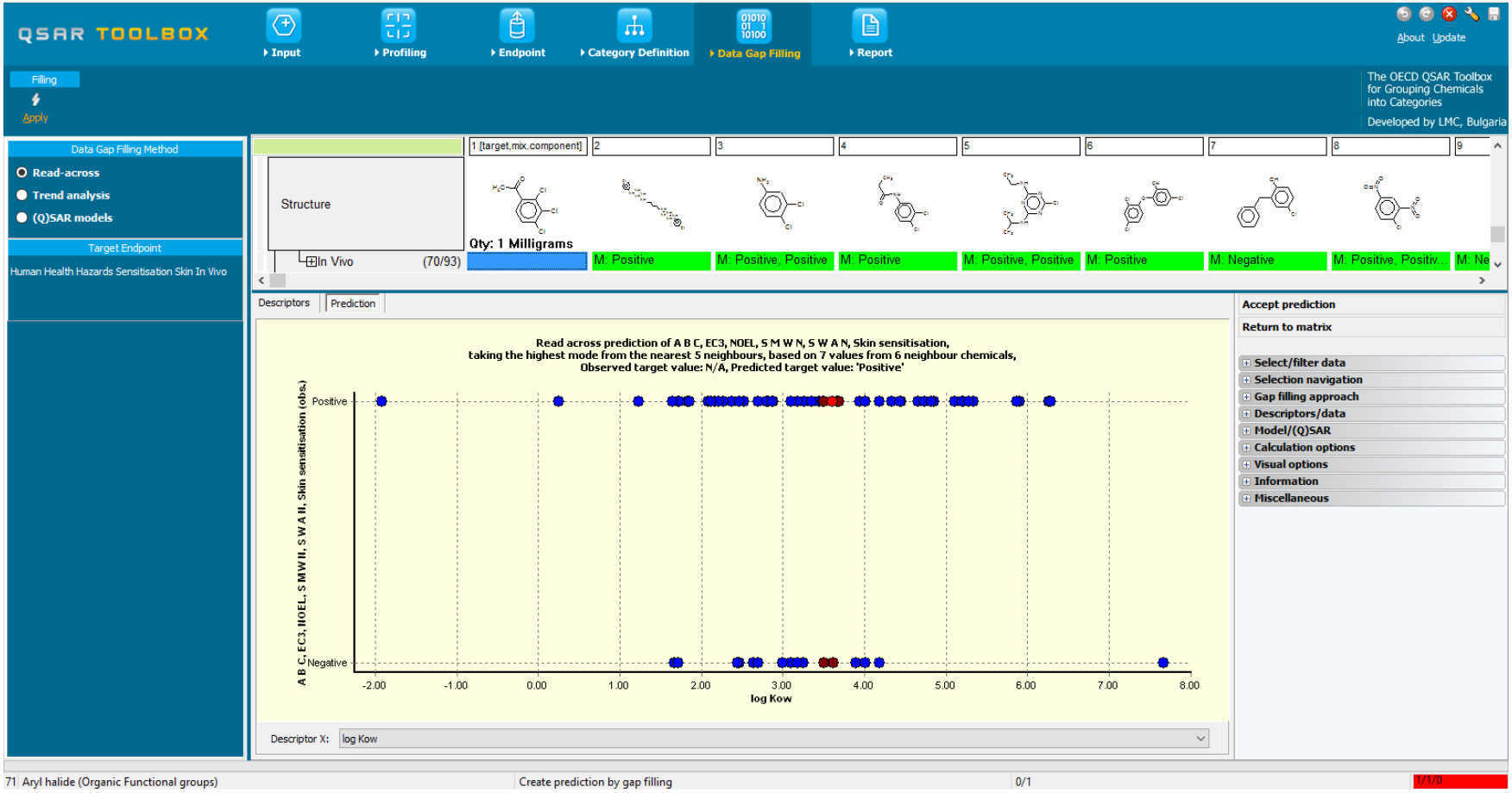
Apply read across



1. **Click** on the cell corresponding to Skin Sensitization in Vivo
2. **Select** Read-across
3. **Click** Apply
4. **Click** OK (in this case we mixed all endpoints and assays)

Read across prediction of constituent without data

Apply read across



Read across prediction of constituent without data
Subcategorization

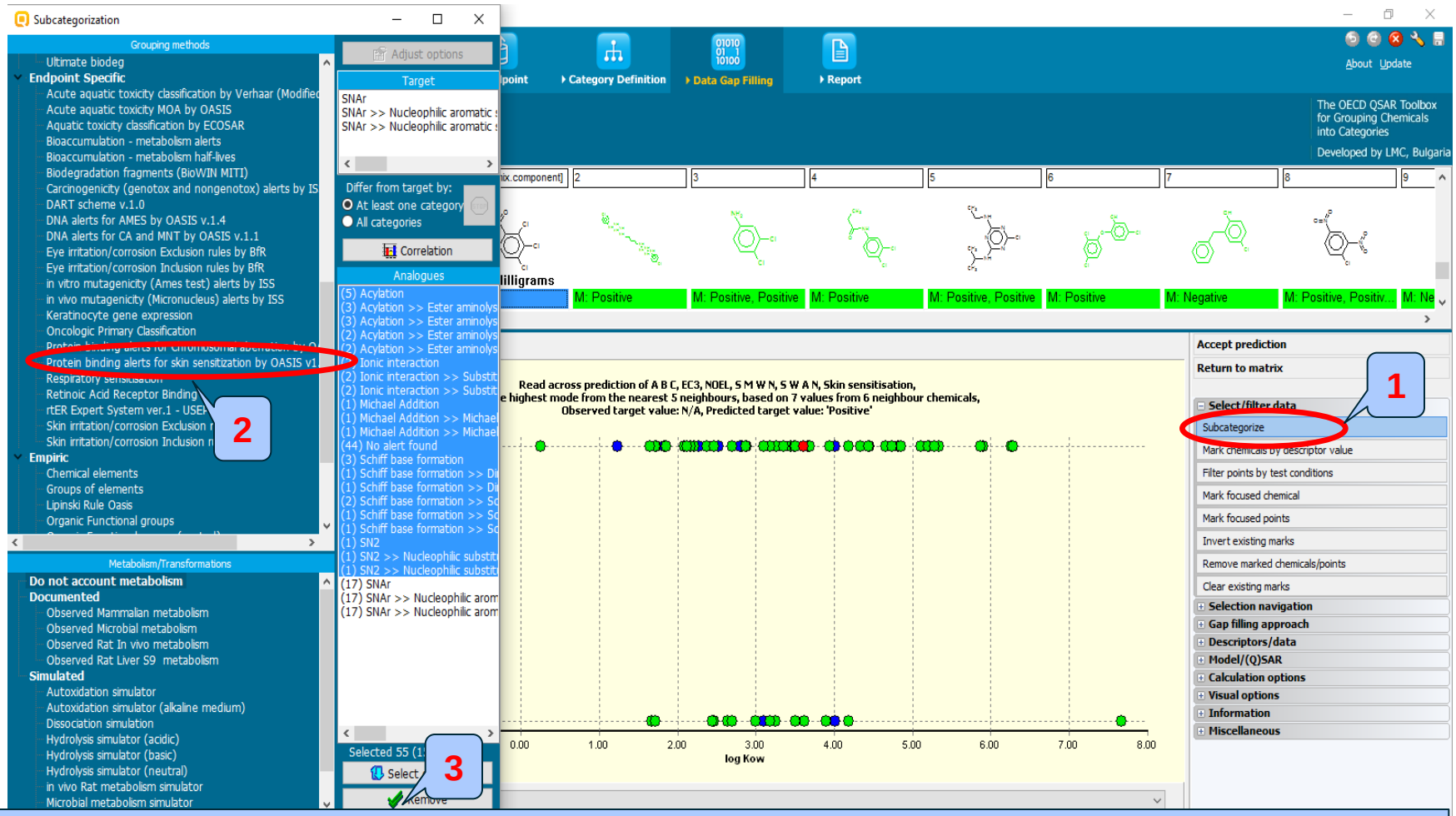
- The initial category could be refine by subcategorizing the analogues according to the following endpoint specific profiler (phase II, slide #37):

  - Protein binding alerts for skin sensitization by OASIS v1.4.

- These steps are summarized in the next screen shots.

Read across prediction of constituent without data

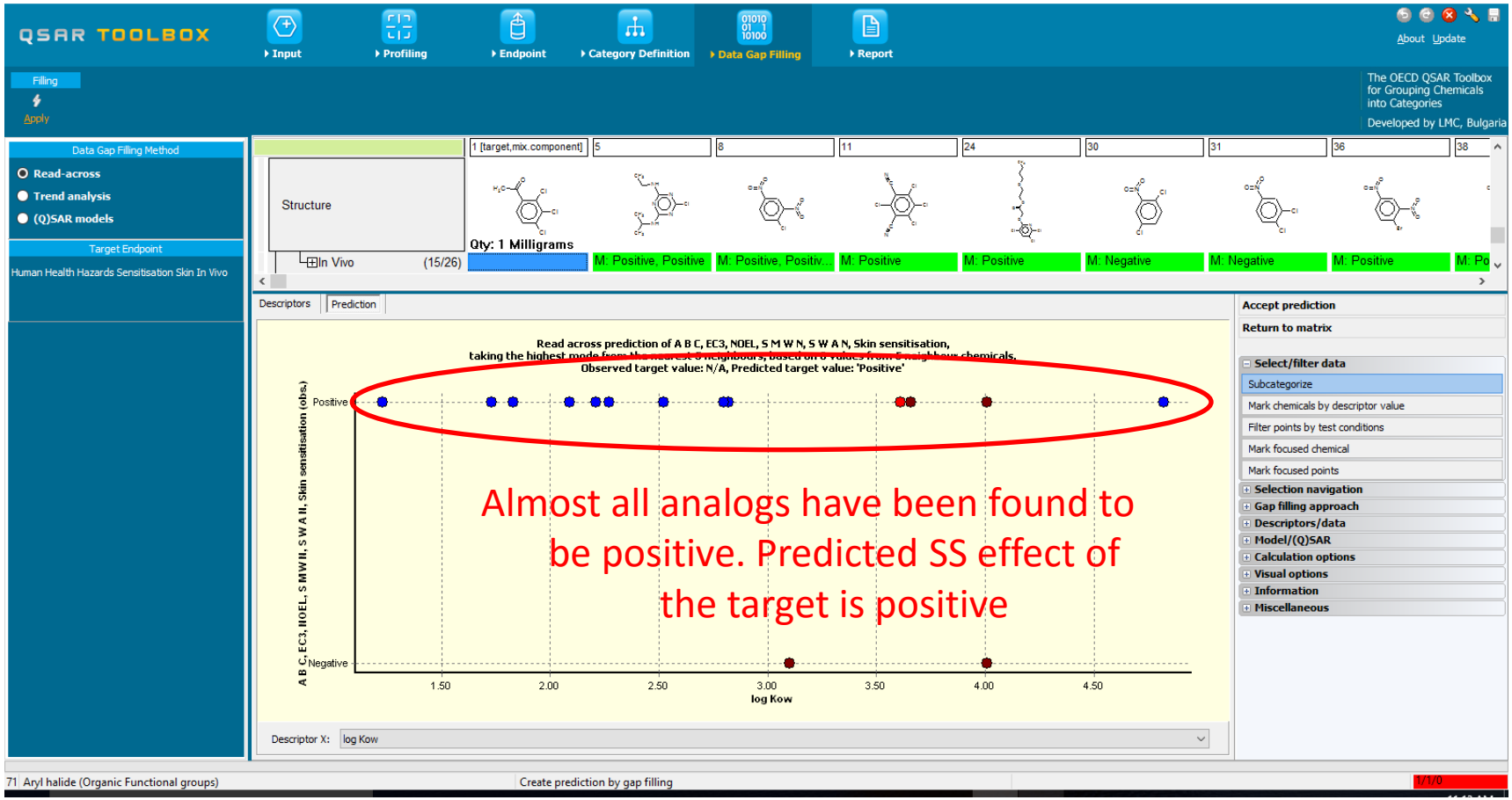
Subcategorization by Protein binding alert for SS



1. Select filter data/subcategorize 2. **Select** Protein binding alerts for SS by OASIS v1.4.
3. **Click** Remove to eliminate dissimilar chemicals.

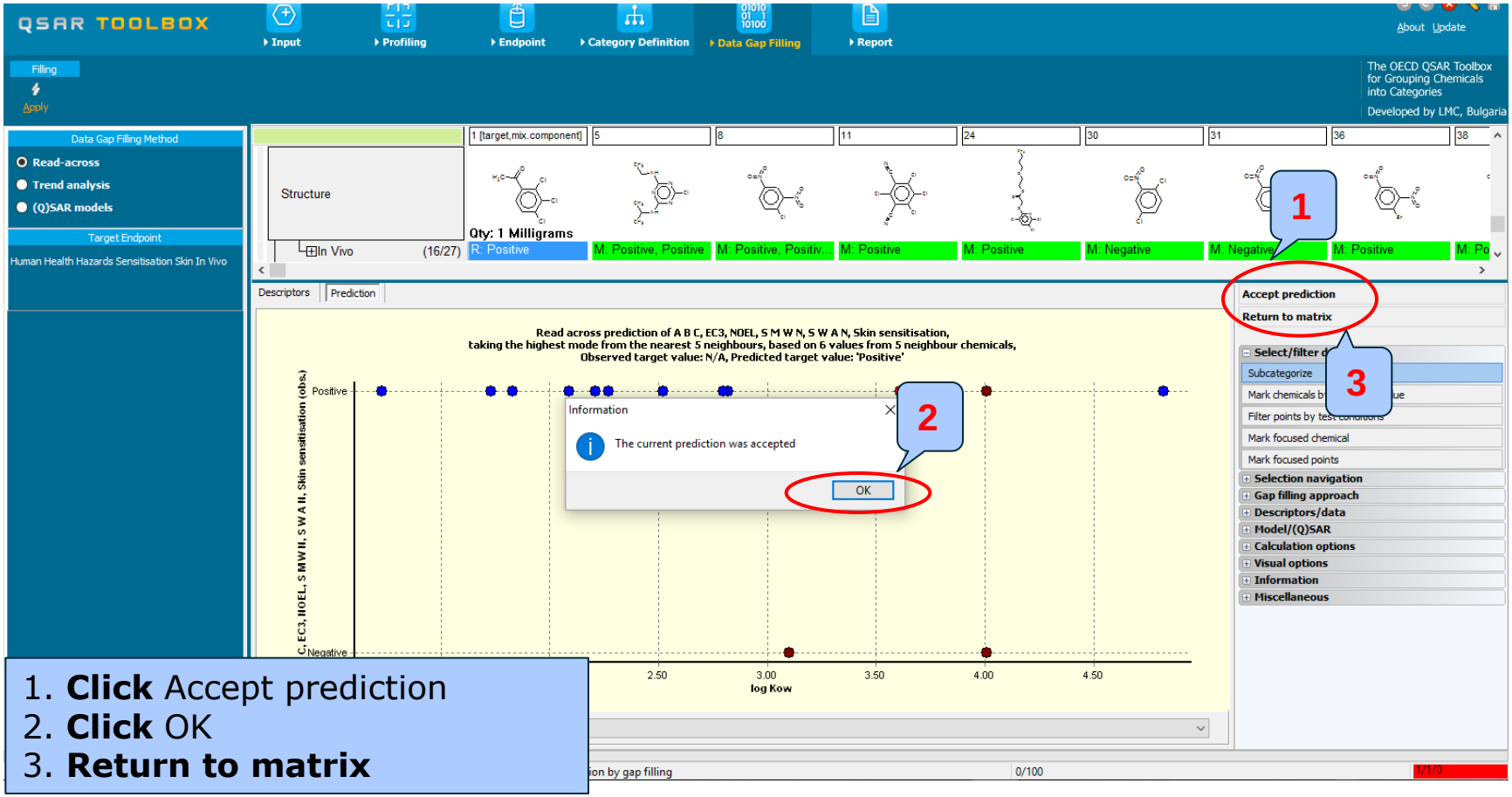
Read across prediction of constituent without data

Apply read across



Read across prediction of constituent without data

Apply read across



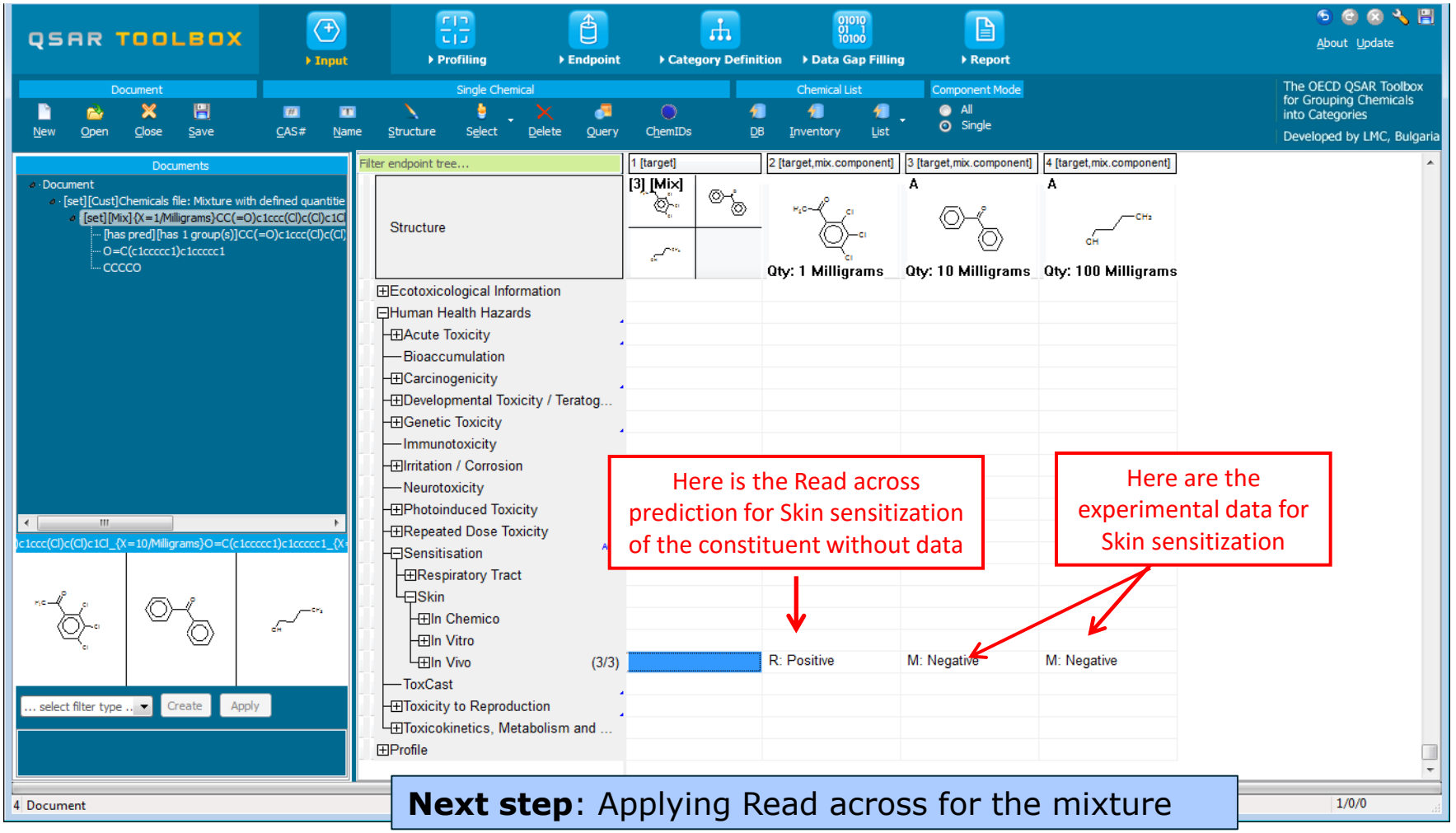
1. **Click** Accept prediction
2. **Click** OK
3. **Return to matrix**

Outlook

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

Filling data gap for skin sensitization of mixture

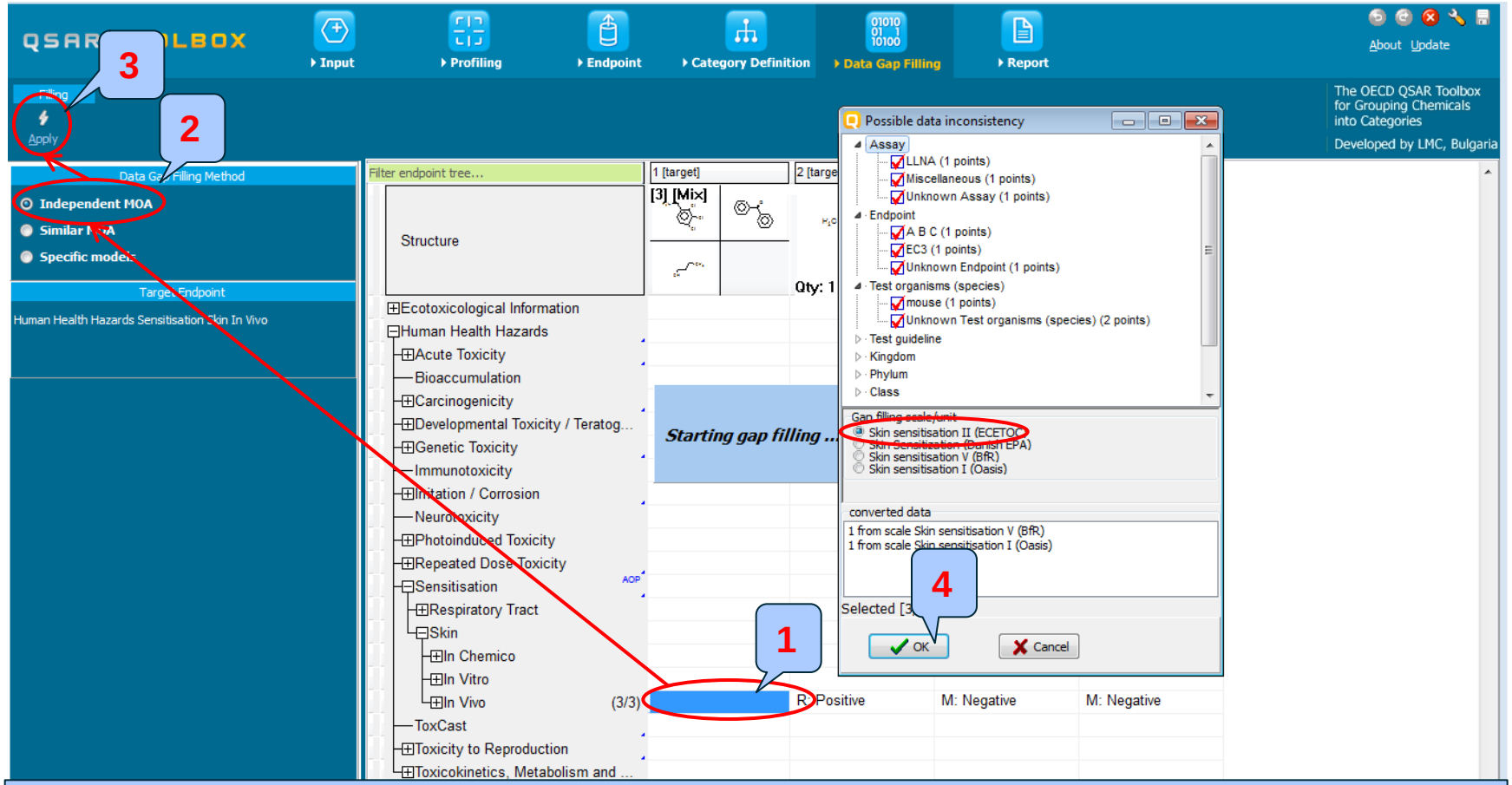
Applying Independent Mode of Action



Next step: Applying Read across for the mixture

Filling data gap for skin sensitization of mixture

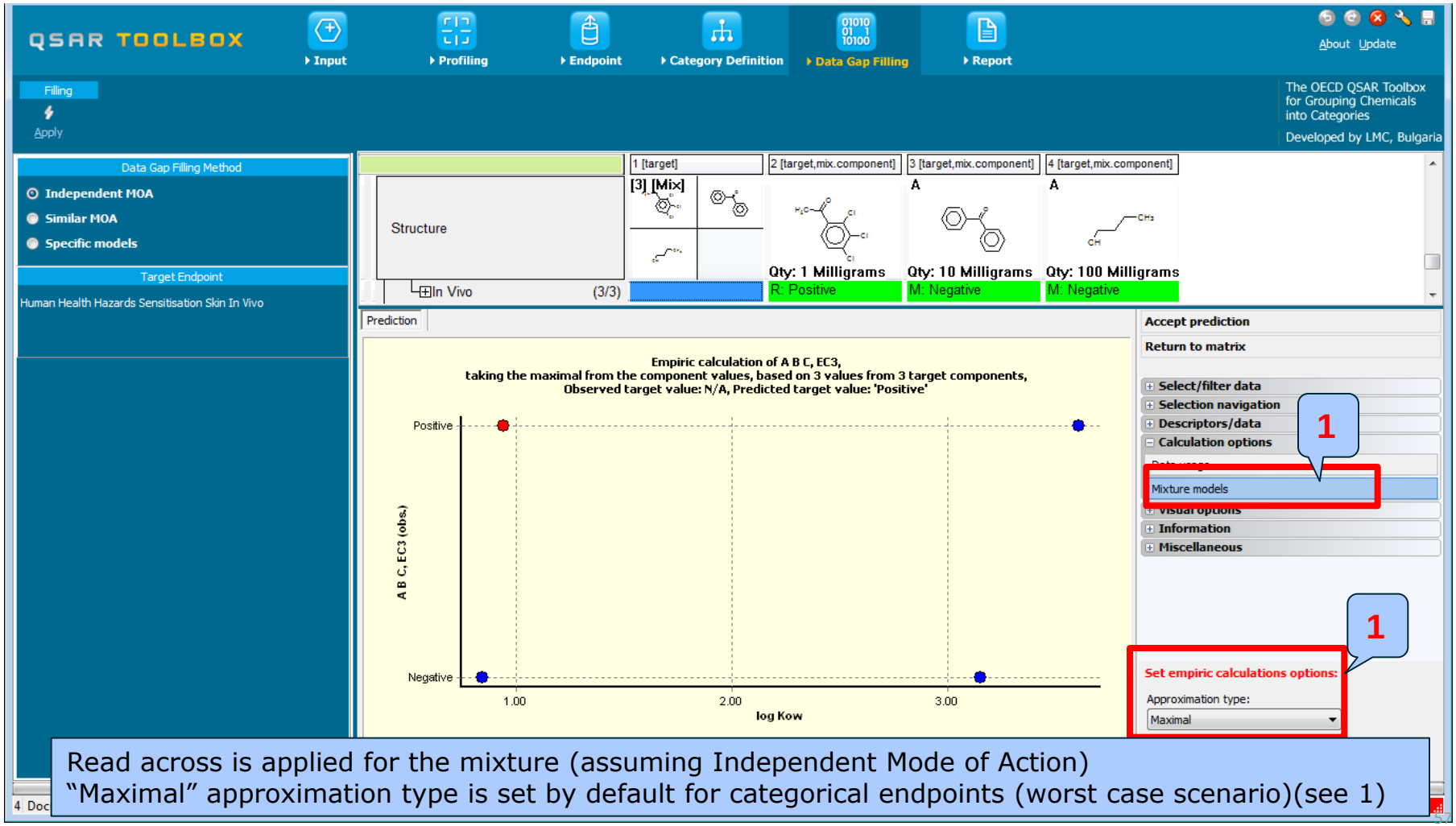
Applying Independent Mode of Action



1. **Click** on the cell corresponding to Skin Sensitization for mixture
2. **Select** Independent MOA
3. **Click** Apply
4. **Click** OK

Filling data gap for skin sensitization of mixture

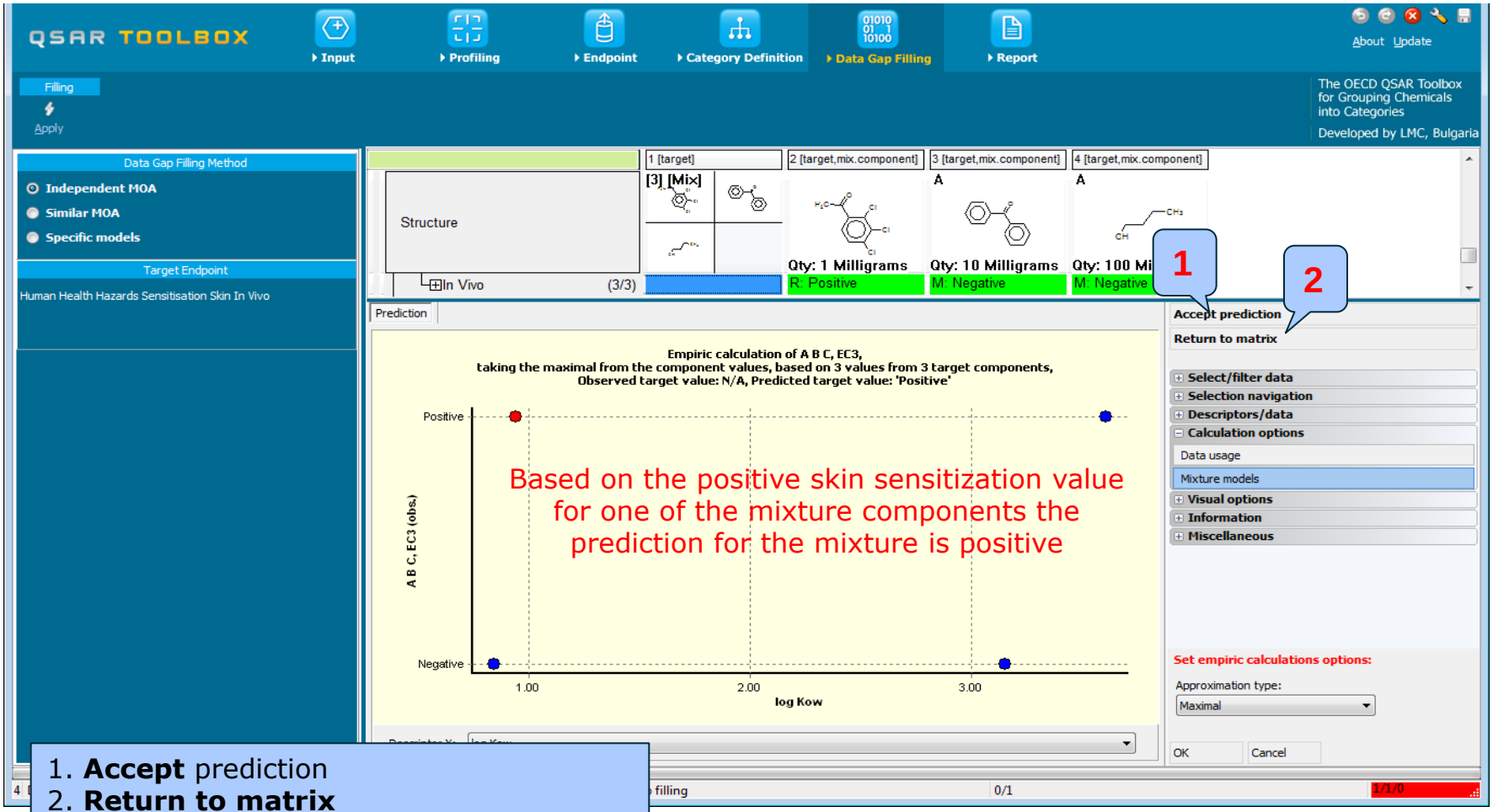
Applying Independent Mode of Action



Read across is applied for the mixture (assuming Independent Mode of Action)
"Maximal" approximation type is set by default for categorical endpoints (worst case scenario)(see 1)

Filling data gap for skin sensitization of mixture

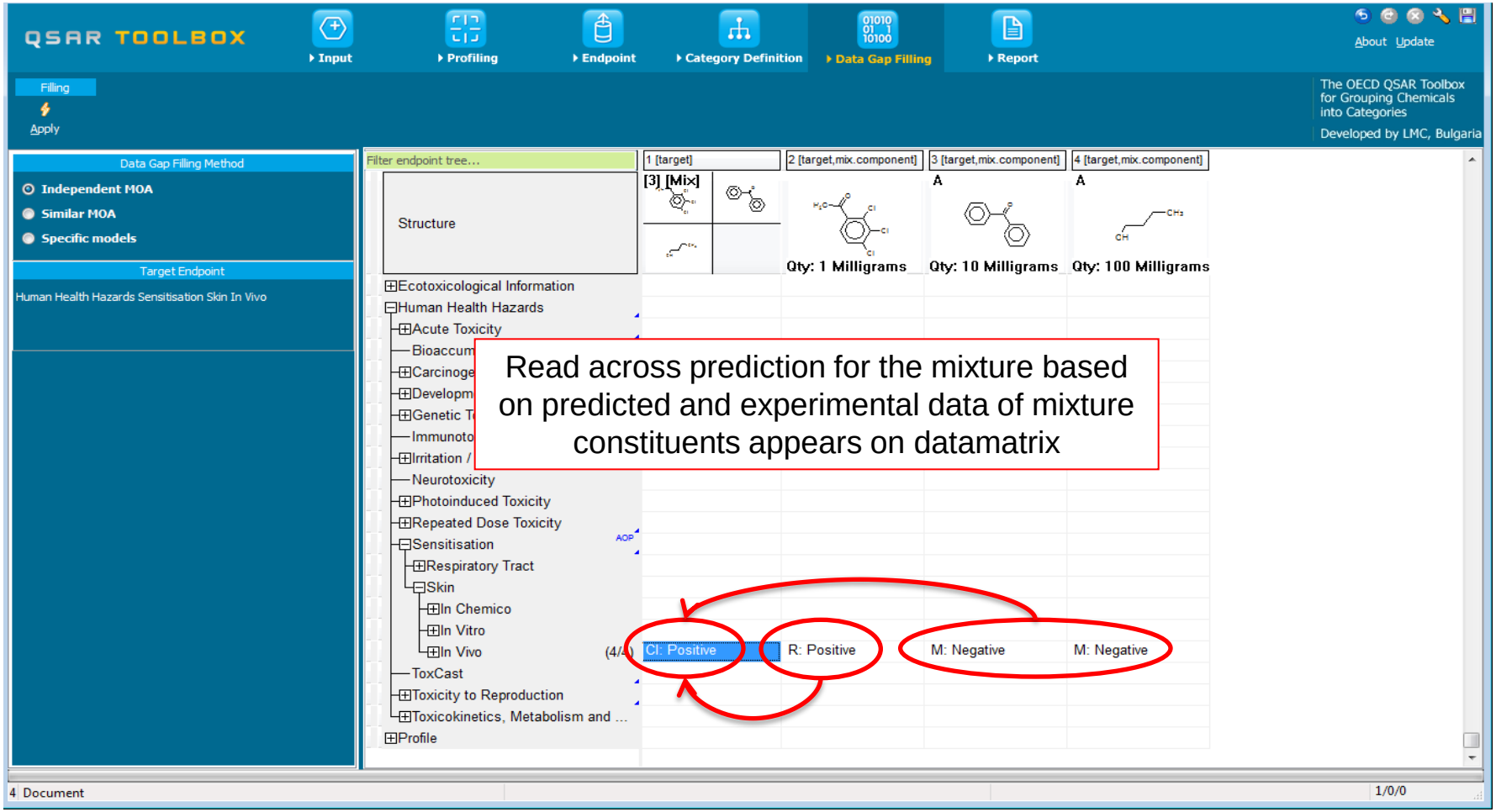
Applying Independent Mode of Action



1. **Accept** prediction
2. **Return to matrix**

Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action



Outlook

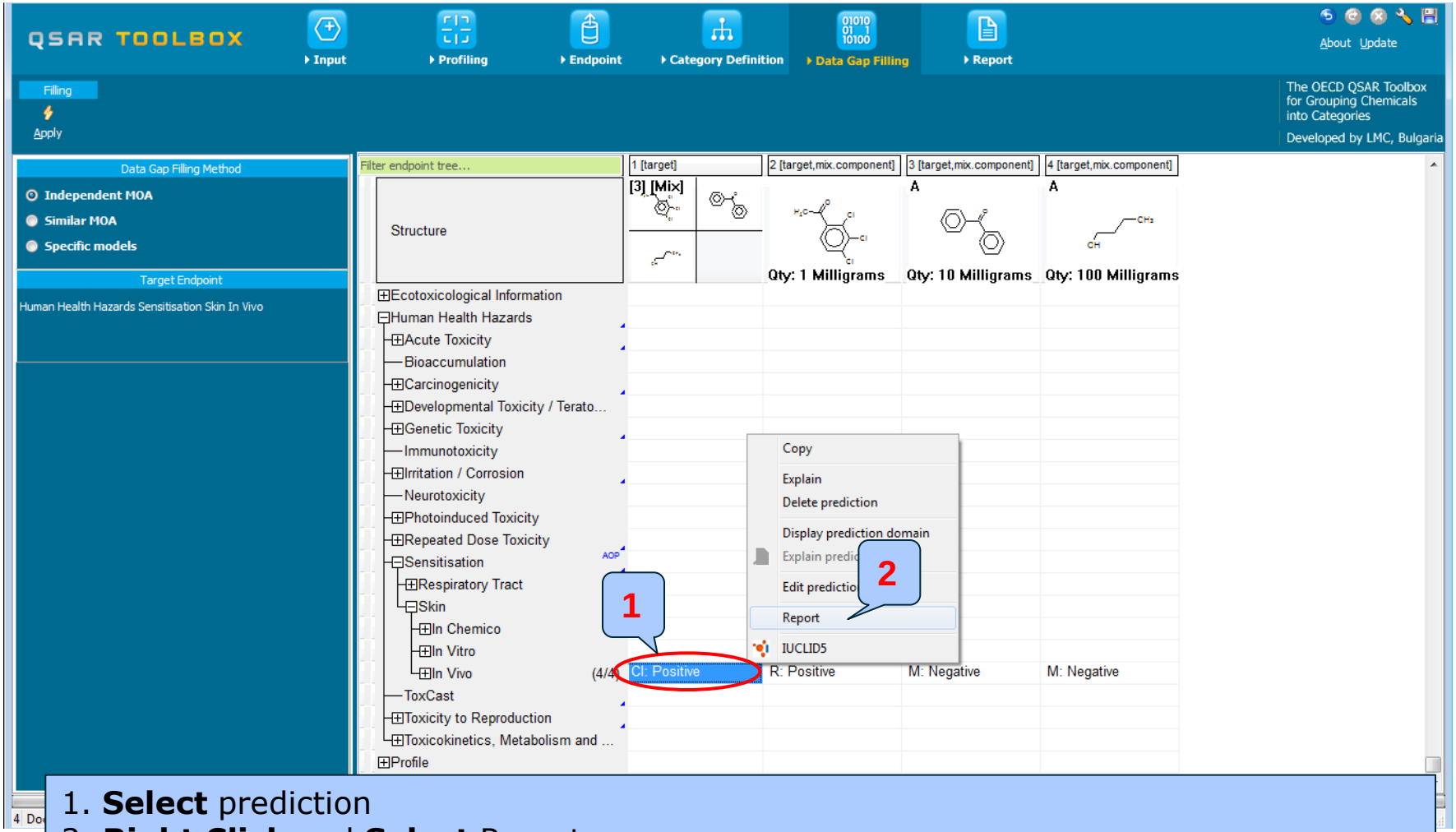
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Read across prediction of constituent without data
  - Filling data gap for skin sensitization of mixture
  - **Generating report for mixture**

Report

- Remember the report module allows you to generate a report on the predictions performed with the Toolbox. This module contains predefined report templates as well as a template editor with which users can define their own user defined templates. The report can then be printed or saved in different formats.

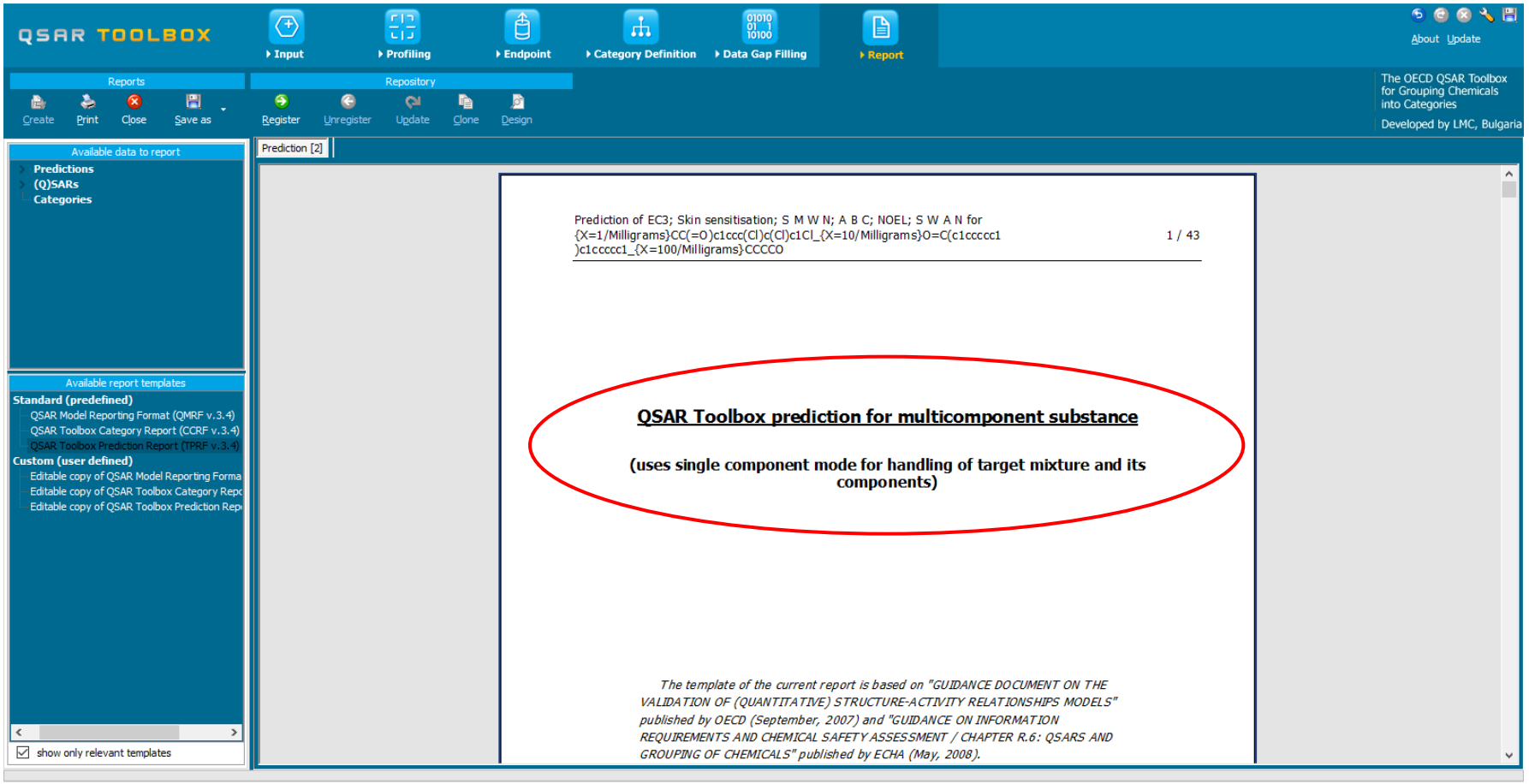
- Generating the report is shown on next screenshots

Report



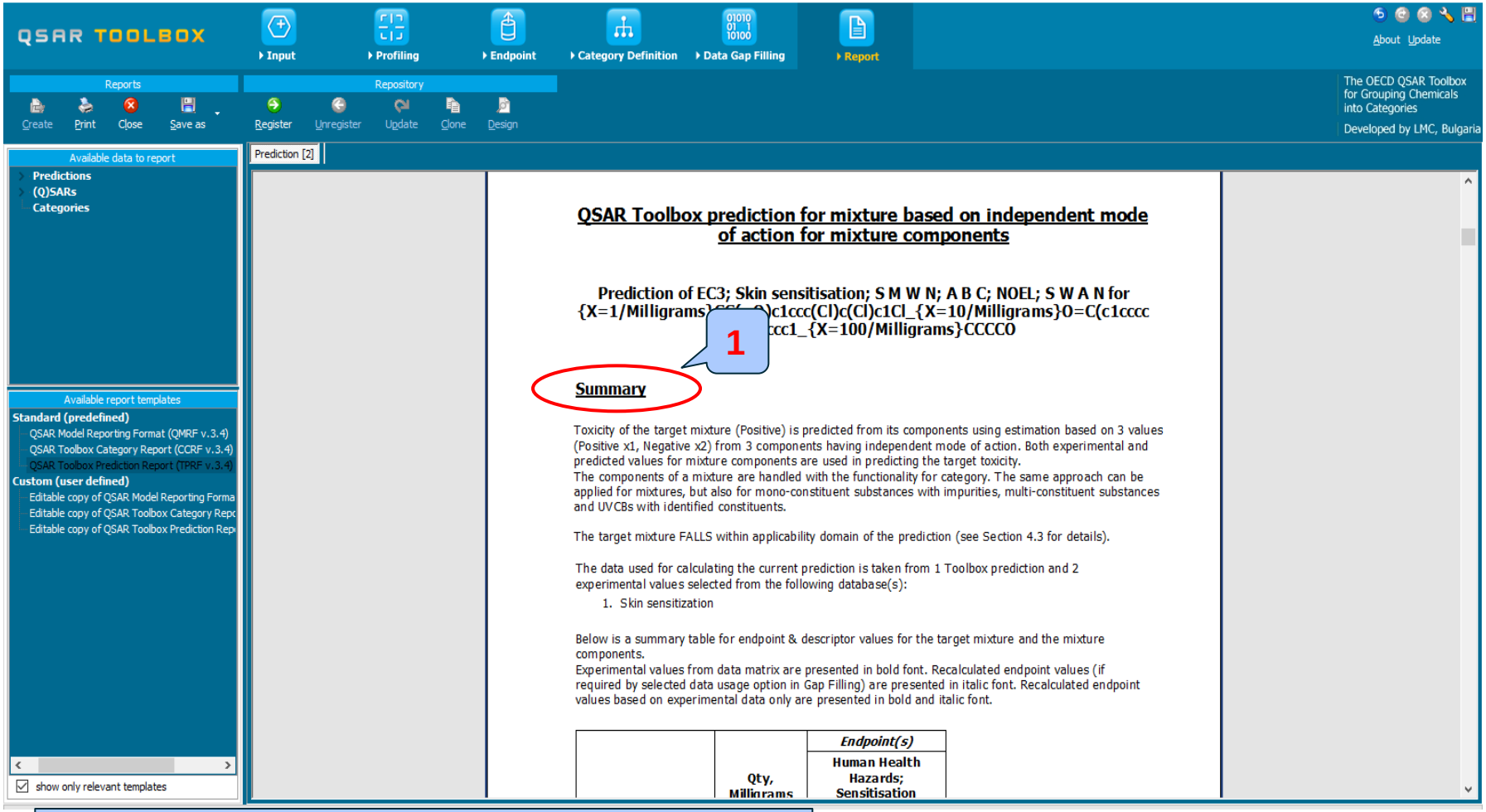
1. **Select** prediction
2. **Right Click** and **Select** Report

Report



Toolbox report for mixture

Report



1. Summary information for mixture prediction

Outlook

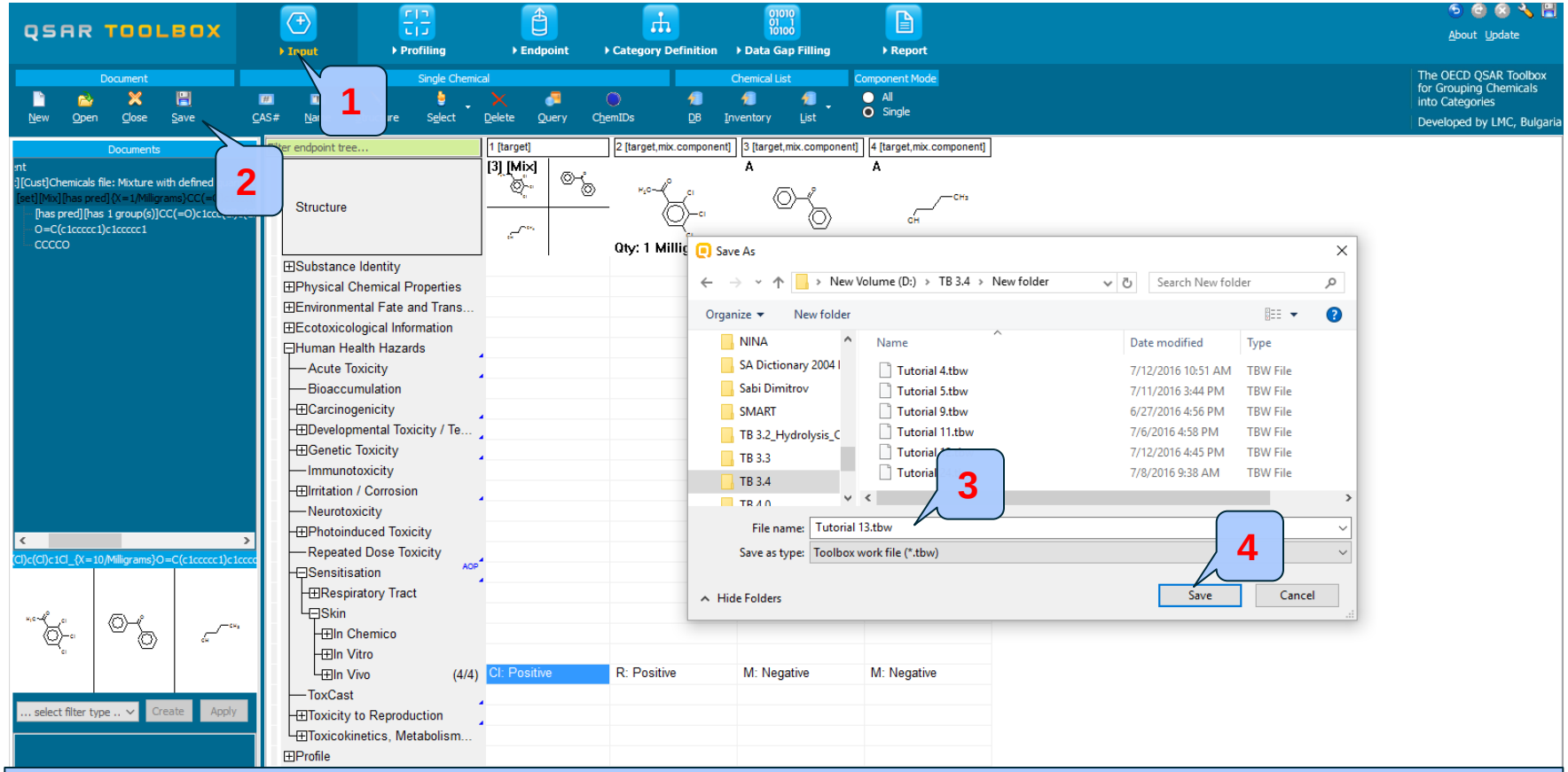
- Background
- Objectives
- The exercise
- Workflow
- **Save the prediction result**

Saving the prediction result

- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc, on the same computer. The 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).

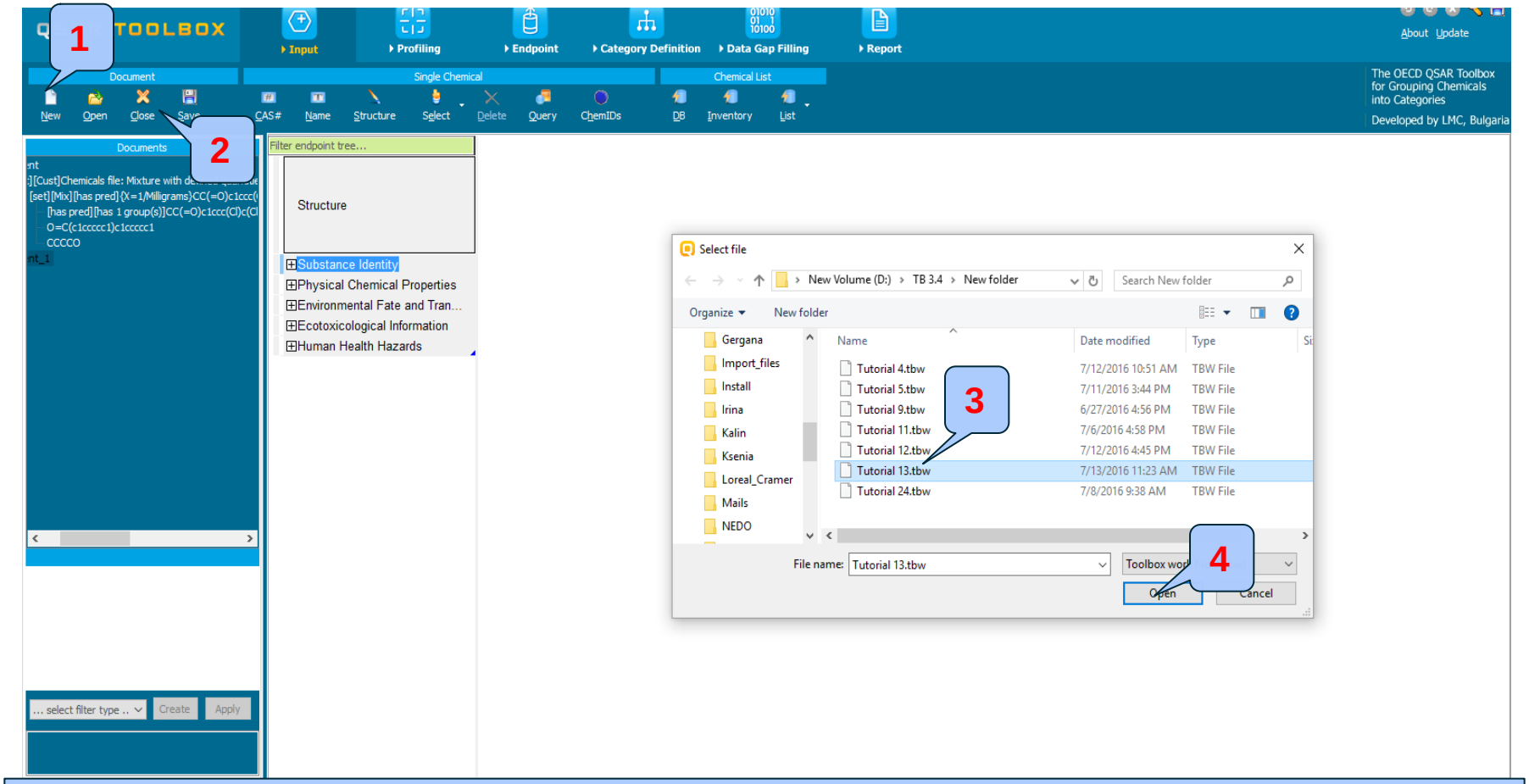
- Saving/Loading the file with TB prediction is shown on next screenshots

Saving the prediction result



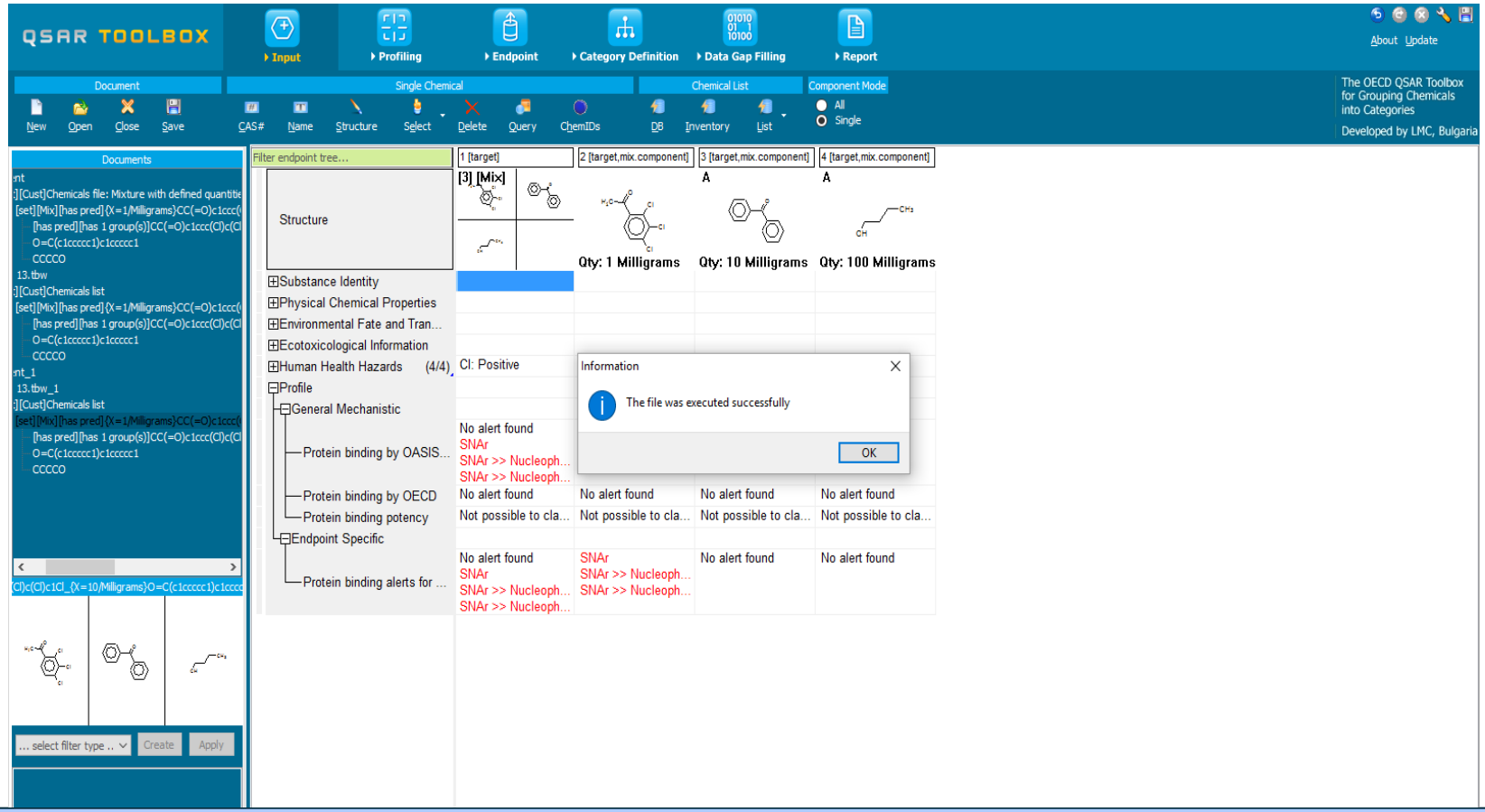
1. Go to Input section 2.**Click** on Save button 3. **Define** name of the file; 4. **Click** Save button

Open saved file



1. **Create** new document 2. **Click** Open; 3. **Find** and **select file**; 4. **Click** Open

Open saved file



1. The file is opened successfully 1. **Click** OK