QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.1

Example for predicting Skin Sensitization of mixture

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.

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 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, and
- 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

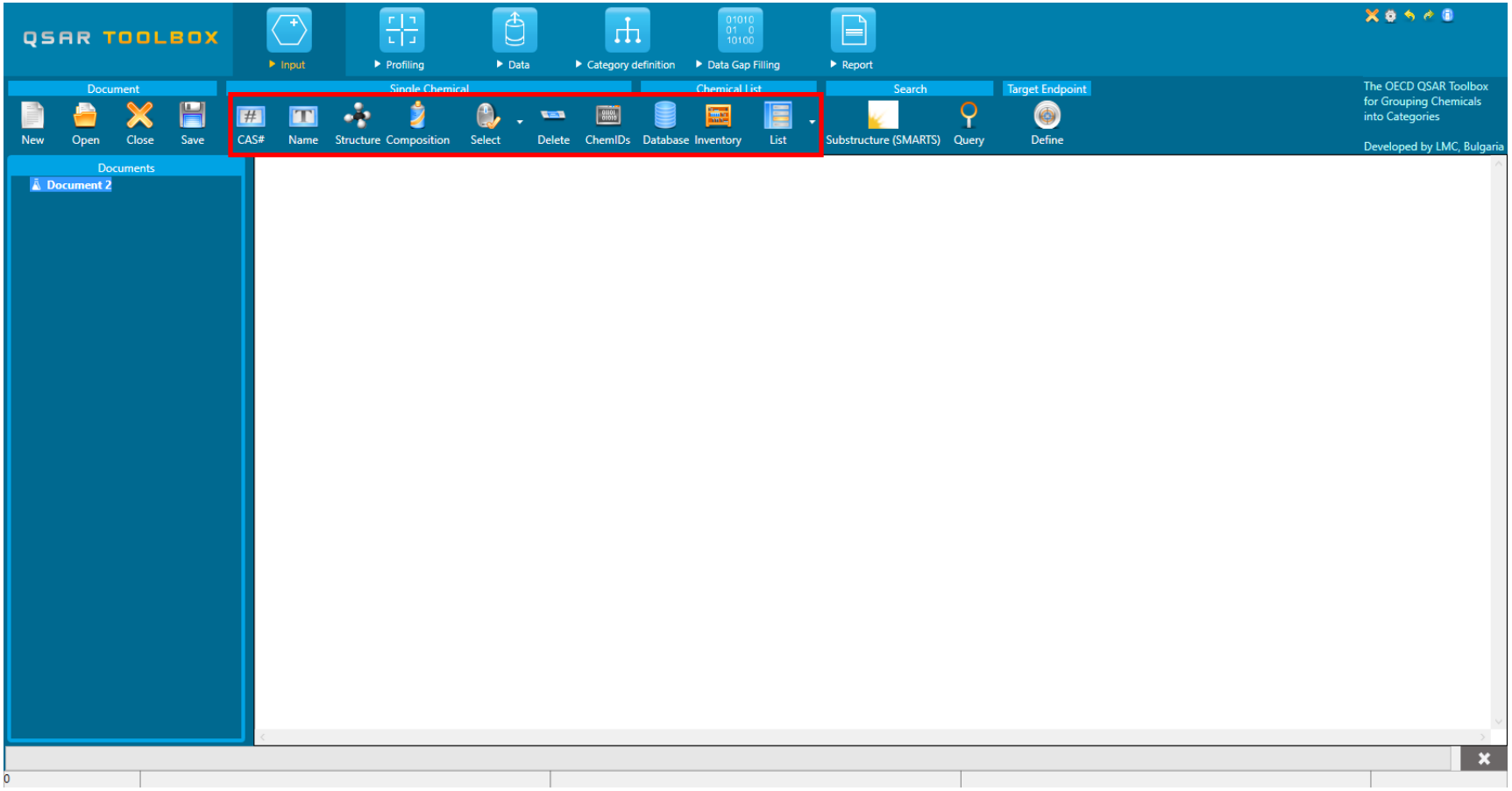
User alternatives for defining mixtures with known compositions:

- 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
- Load file with mixture

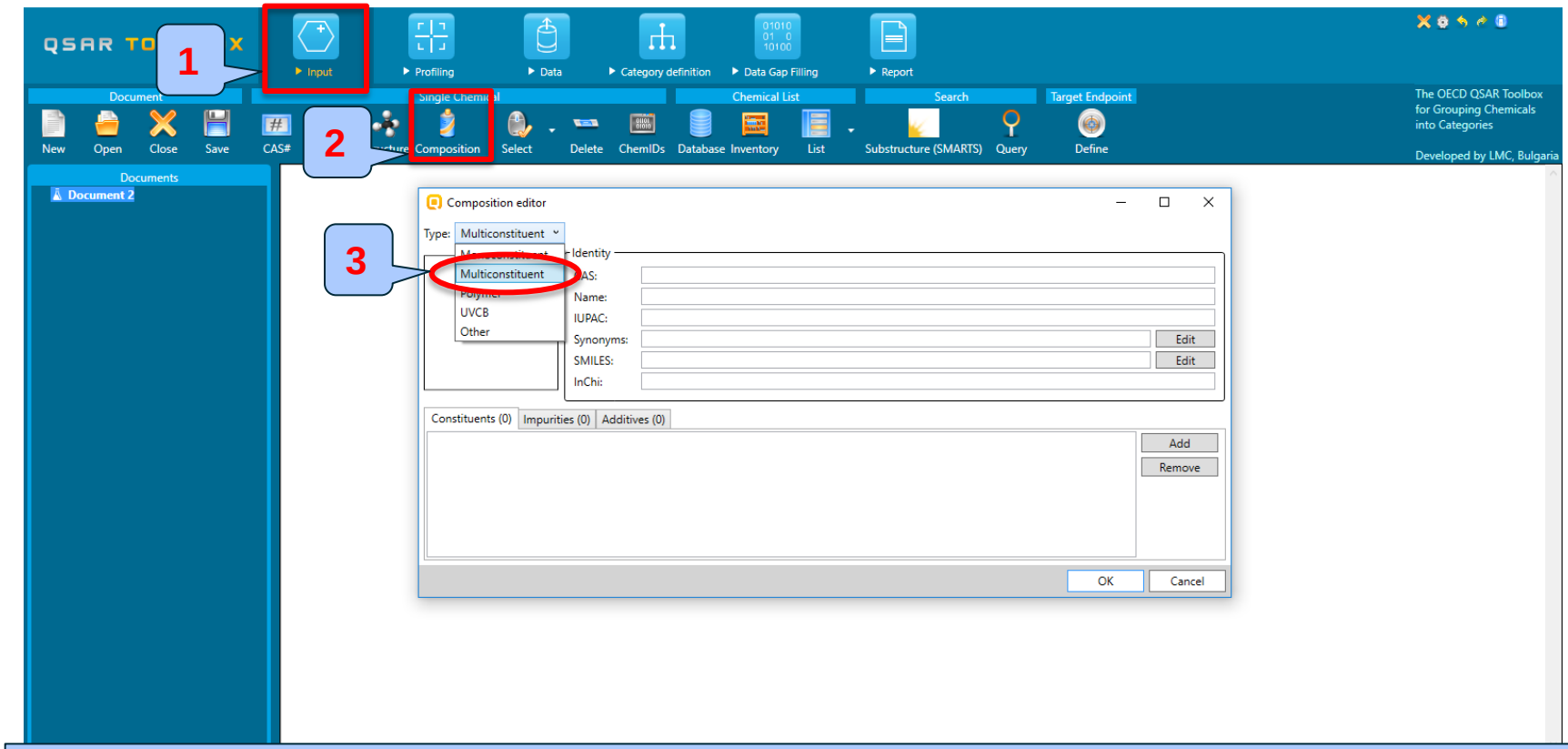
Getting Started

- 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

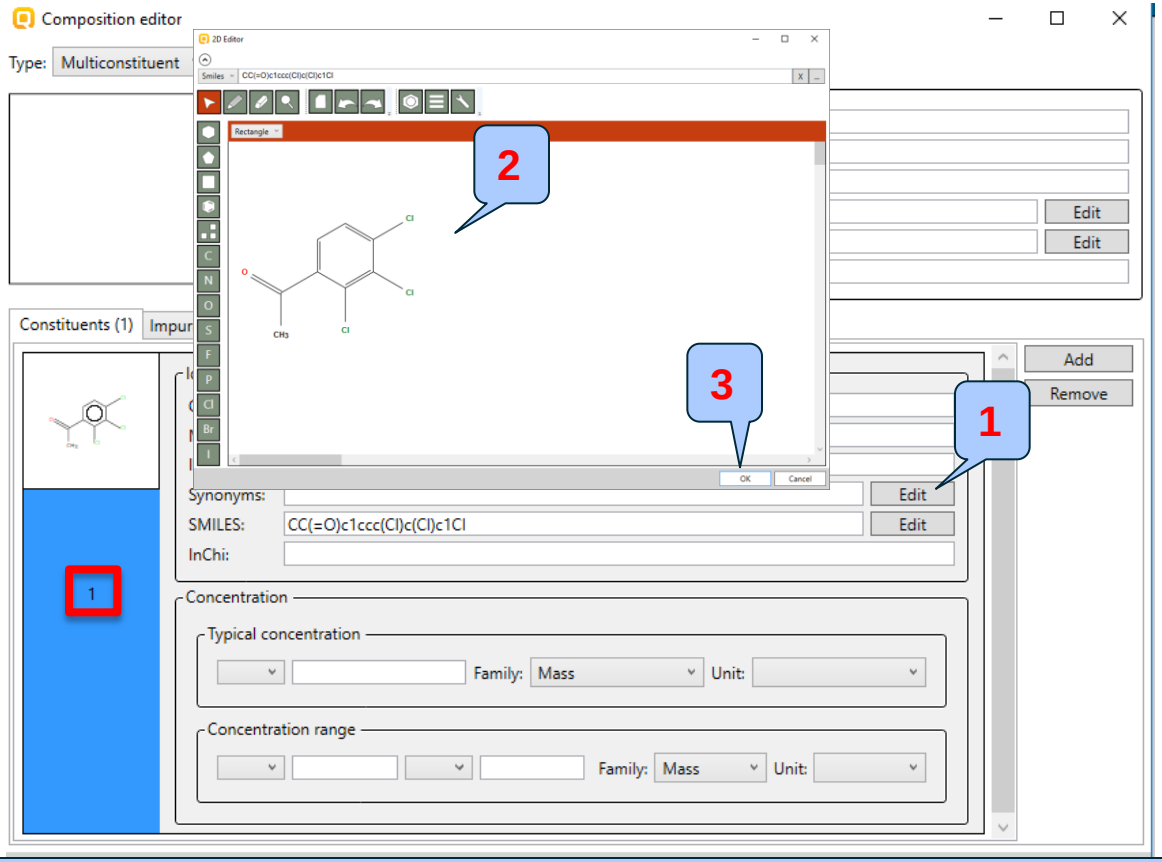


Chemical input
Input mixture



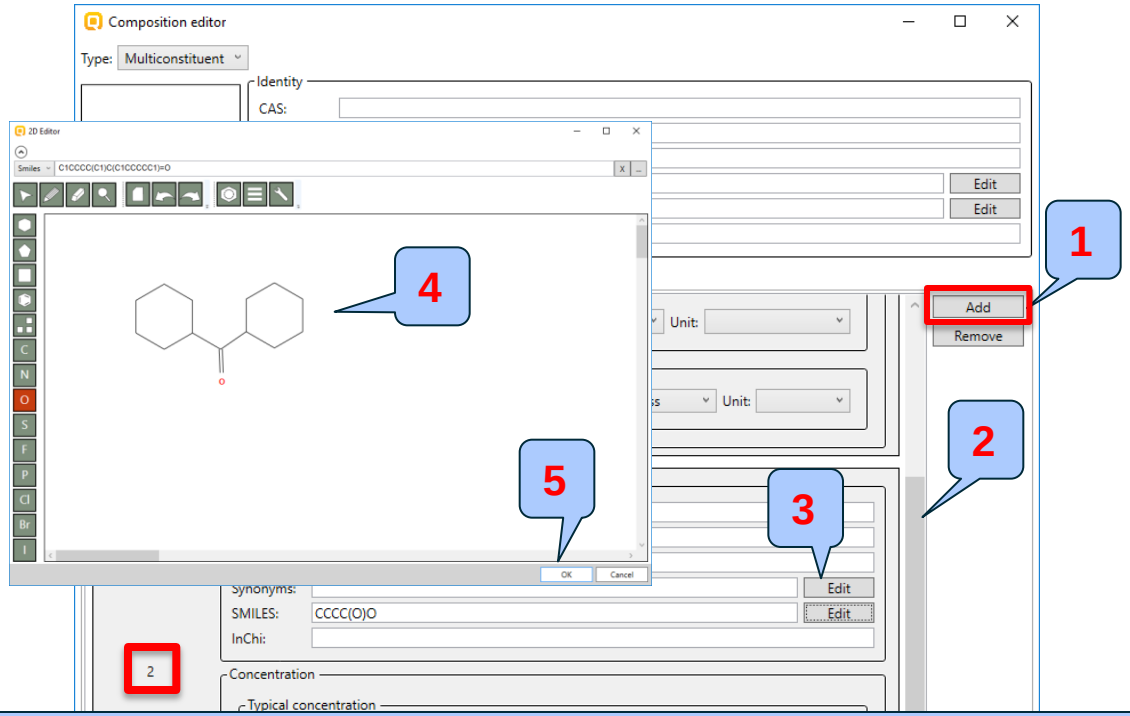
1. Select **Input**
2. Click on **Composition**
3. From composition editor select **Multiconstituent**

Chemical input
Input mixture



1. Click on **Edit** to open the 2D editor,
2. **Draw structure** of the first component (#1),
3. Confirm by clicking OK.

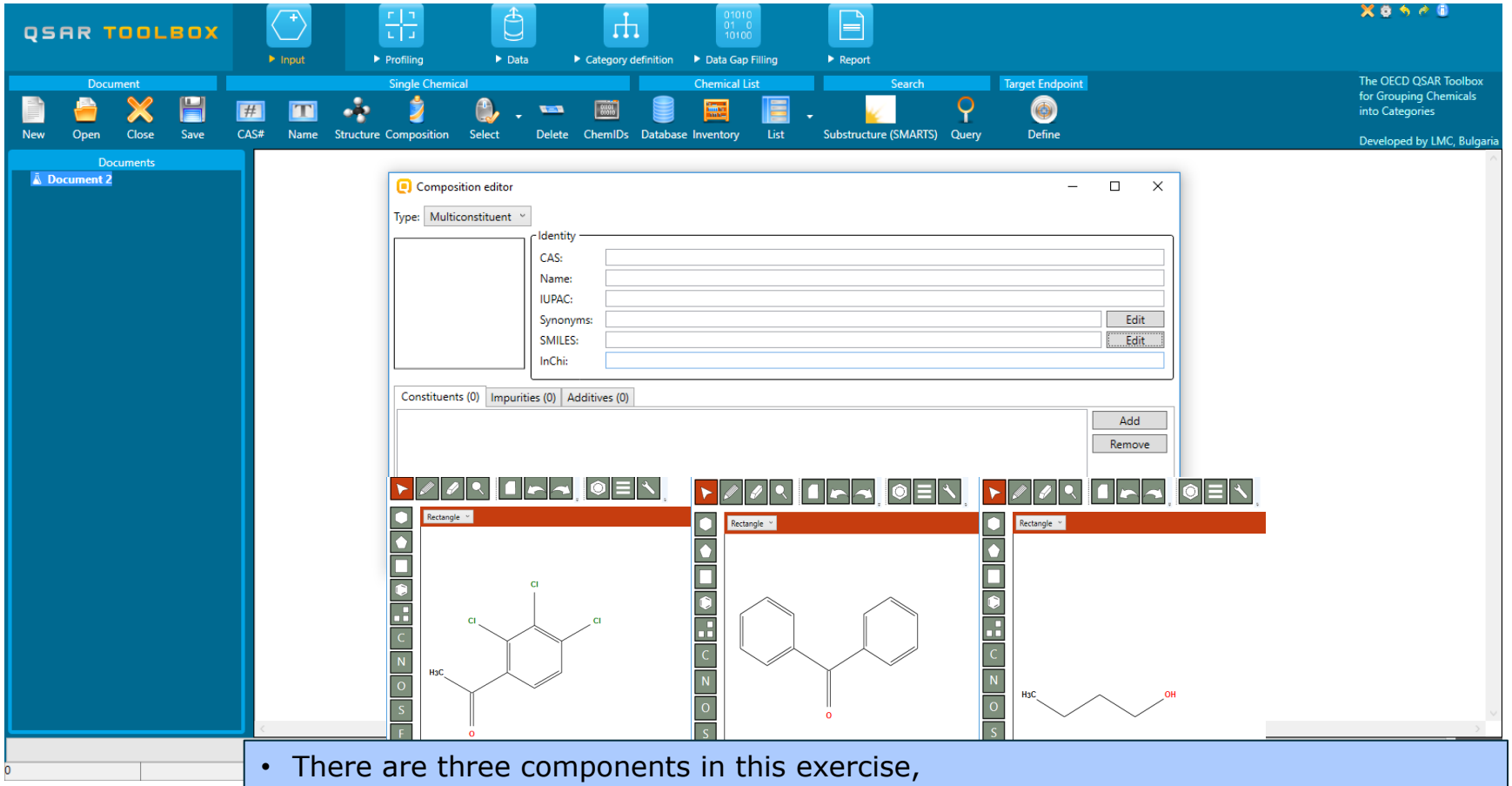
Chemical input
Input mixture



1. Click on **Add,**
2. Scroll down to add the second component (#2),
3. Click **Edit** to open the 2D editor,
4. Draw structure of the second component,
5. Confirm by OK.

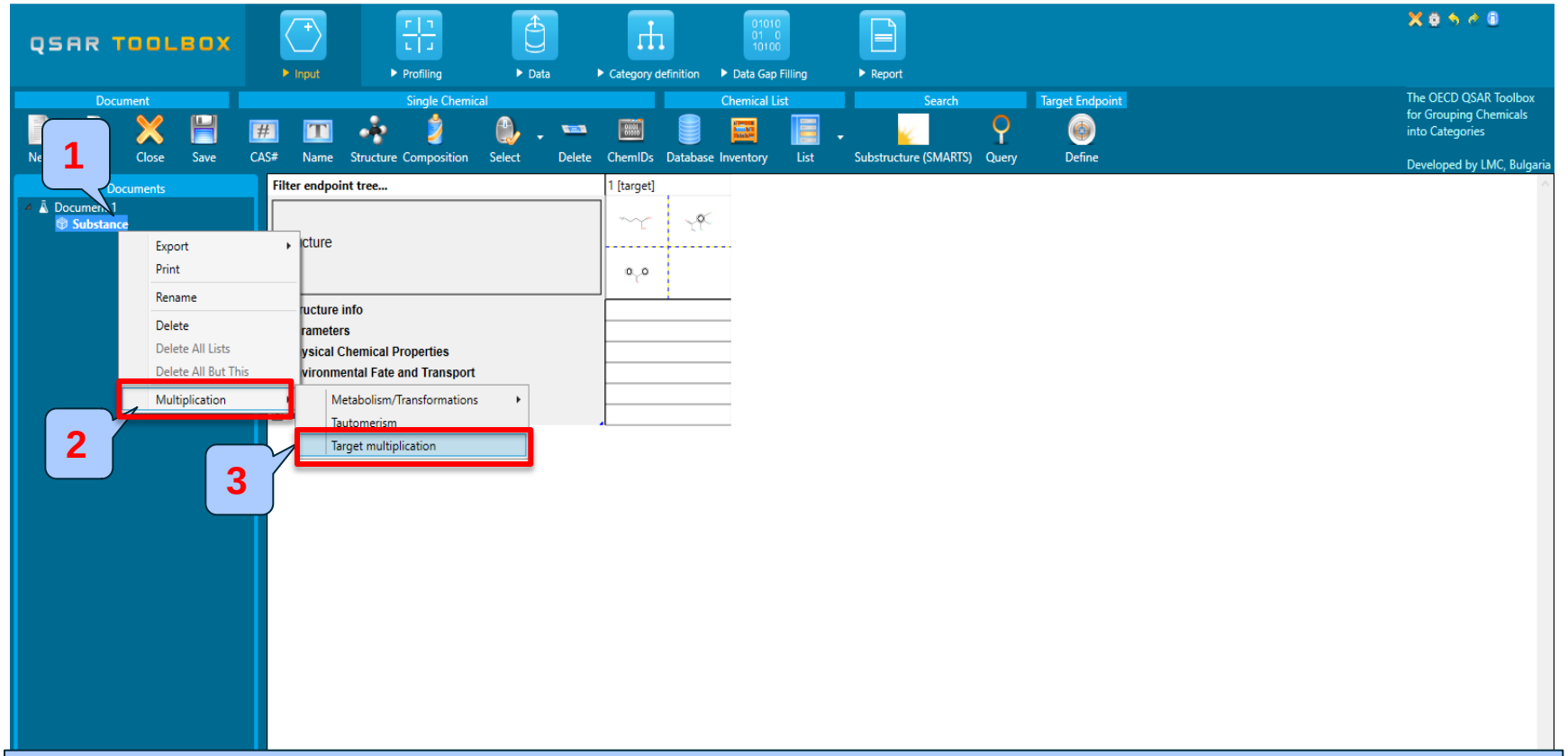
Use the same procedure to add as much components as needed (see next slide).

Chemical input
Input mixture



- There are three components in this exercise,
- Close the Composition editor by clicking **OK** and return to data matrix.

Chemical input
Input mixture

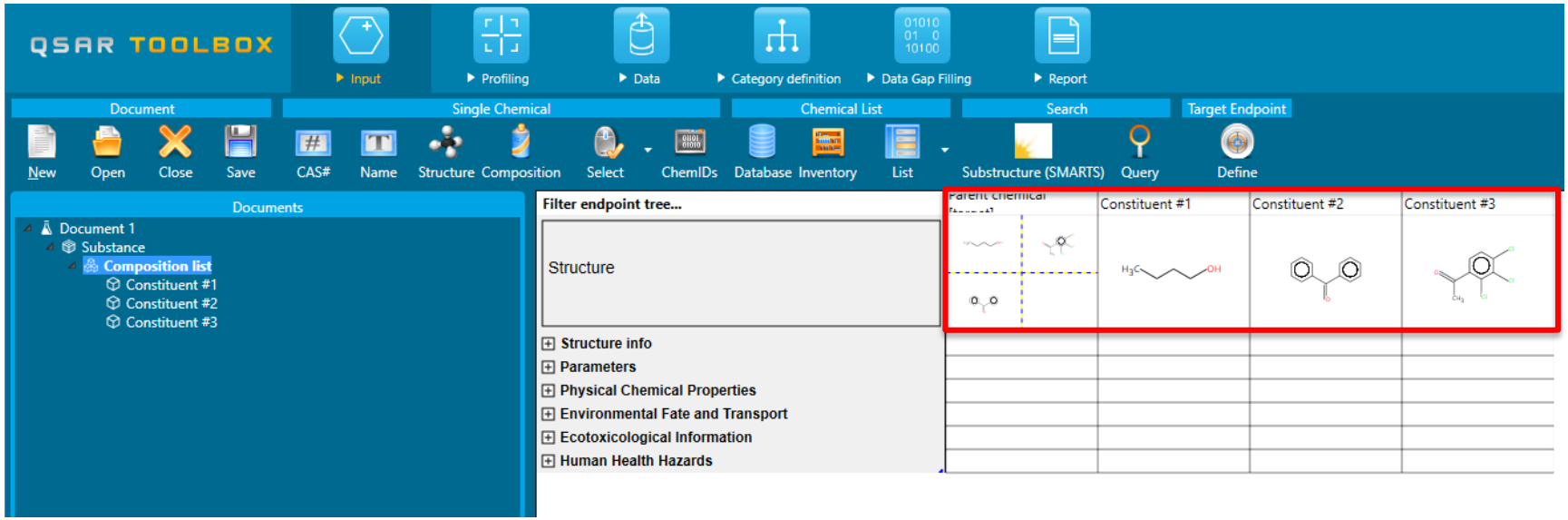


1. Make right mouse click on **Substance,**
2. Select **Multiplication,**
3. Click on **Target multiplication.**

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



All three constituents of the mixture can be treated as individual substances.

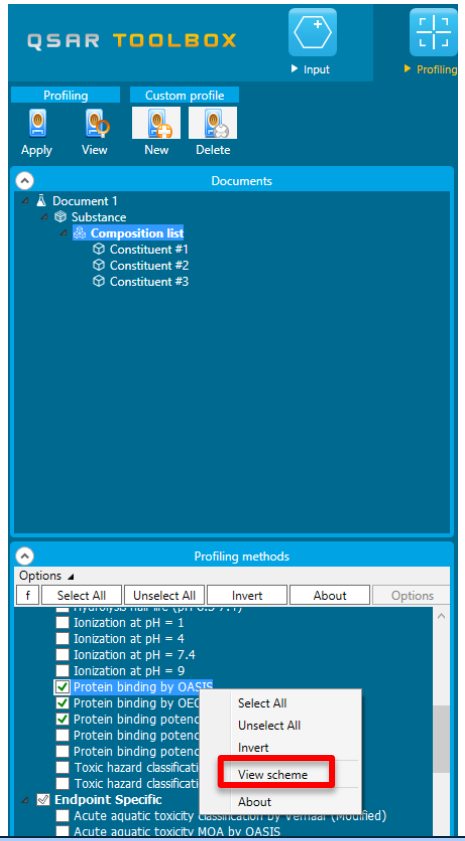
Outlook

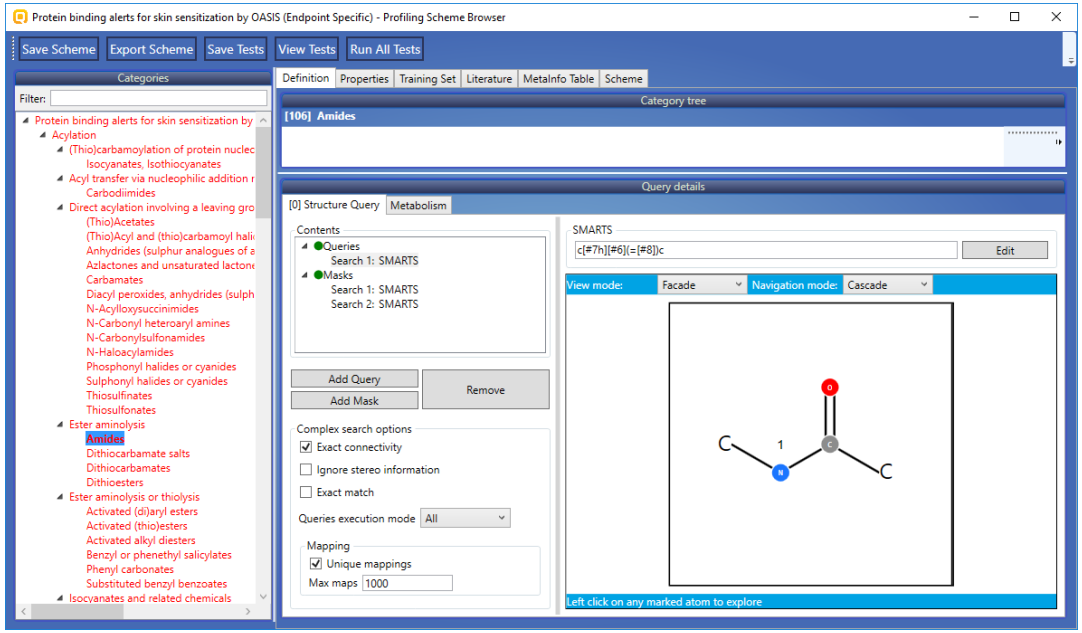
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - **Profiling**

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.
- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, Protein binding by OASIS and clicking on "View scheme" (see next screen shots).

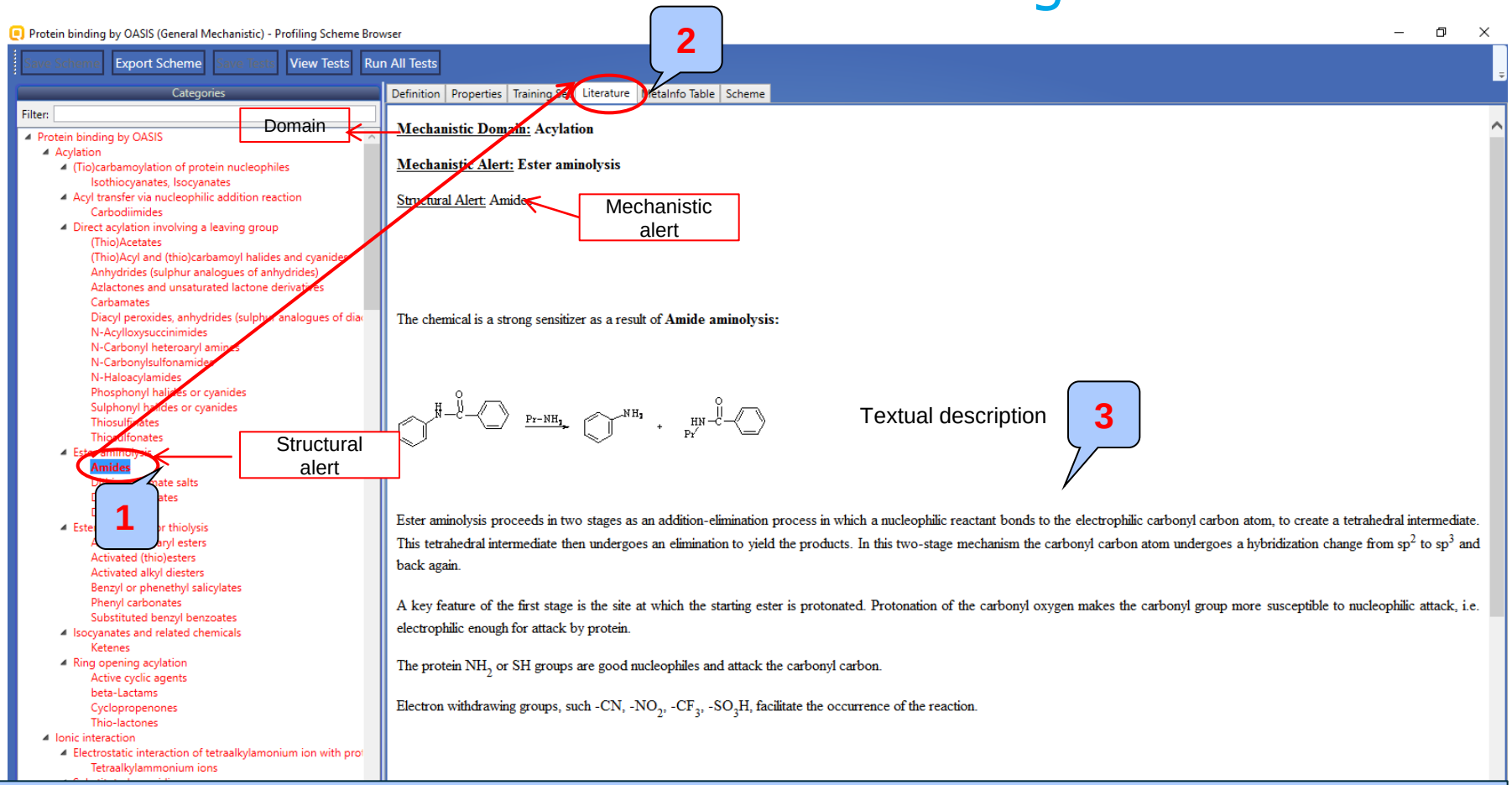
Profiling
Side-Bar to Profiling





Make right mouse click on selected profiler followed by "View scheme" to see defined structural boundaries associated with Amides in **Protein binding by OASIS** profile.

Profiling
Side-Bar to Profiling



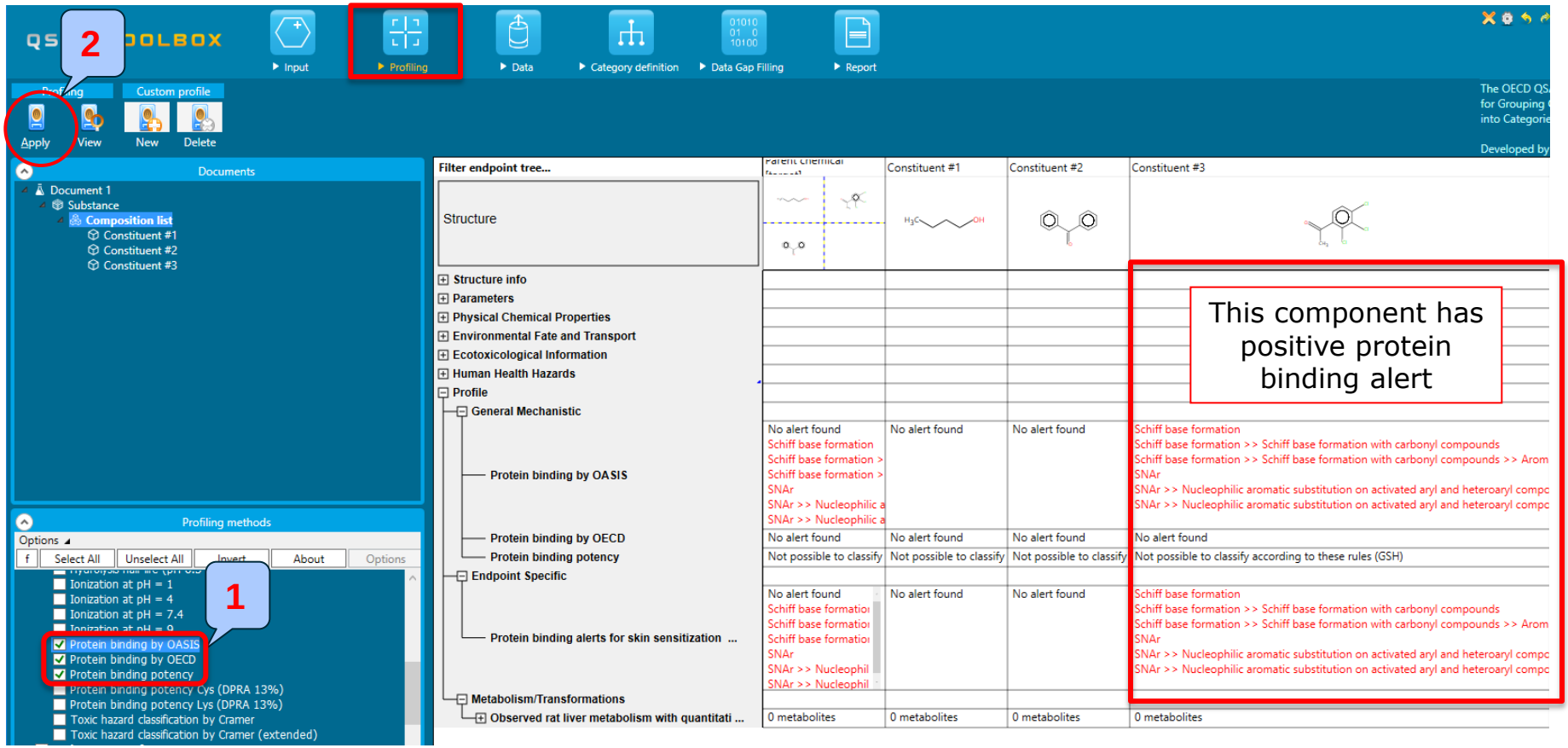
1. Click on **Amides**,
2. Select **Literature** to see mechanistic justification.

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

Profiling
Side-Bar to Profiling



1. Check the profilers related to the target endpoint;
2. Click **Apply**.

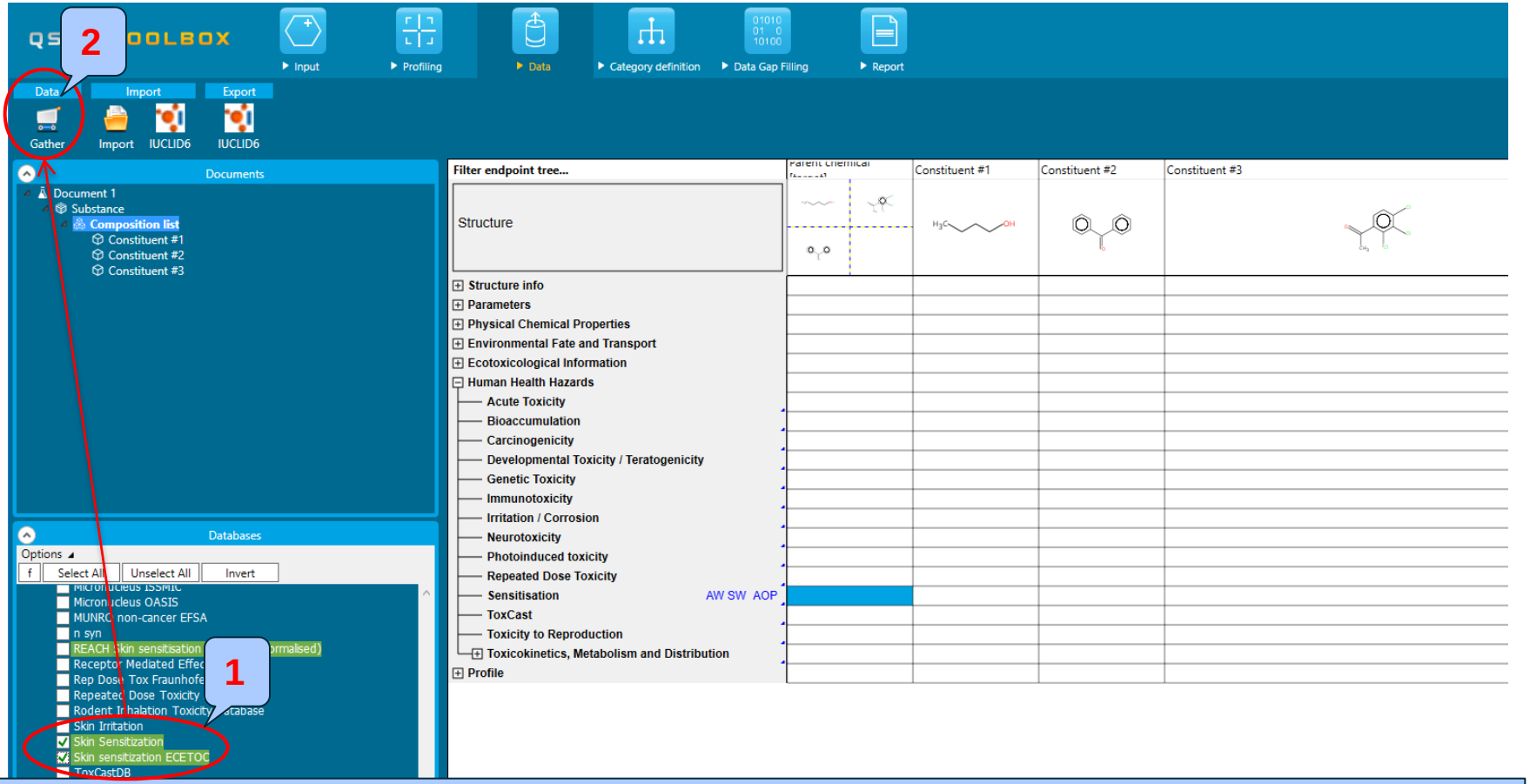
Outlook

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

Data

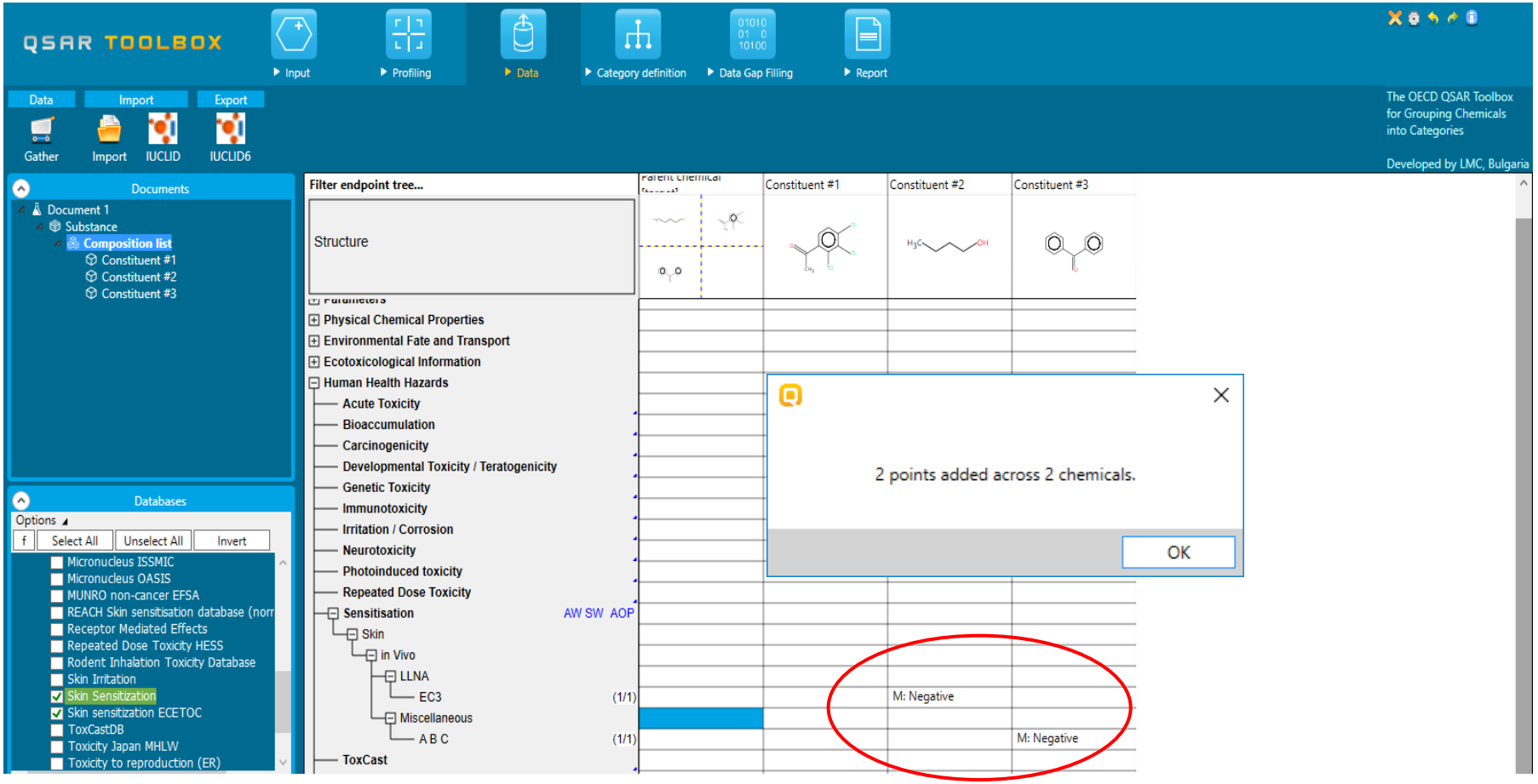
- "Data" refers to the electronic process of retrieving the environmental fate, eco-toxicity and toxicity data that are residing in the Toolbox.
- Data gathering can be executed in a global fashion (i.e. collecting all data of all endpoints) or on a more narrowly defined basis (e.g. collecting data for 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.

Data



1. Select databases related to the target endpoint;
2. Click on **Gather**

Data
Process of collecting data



Recap

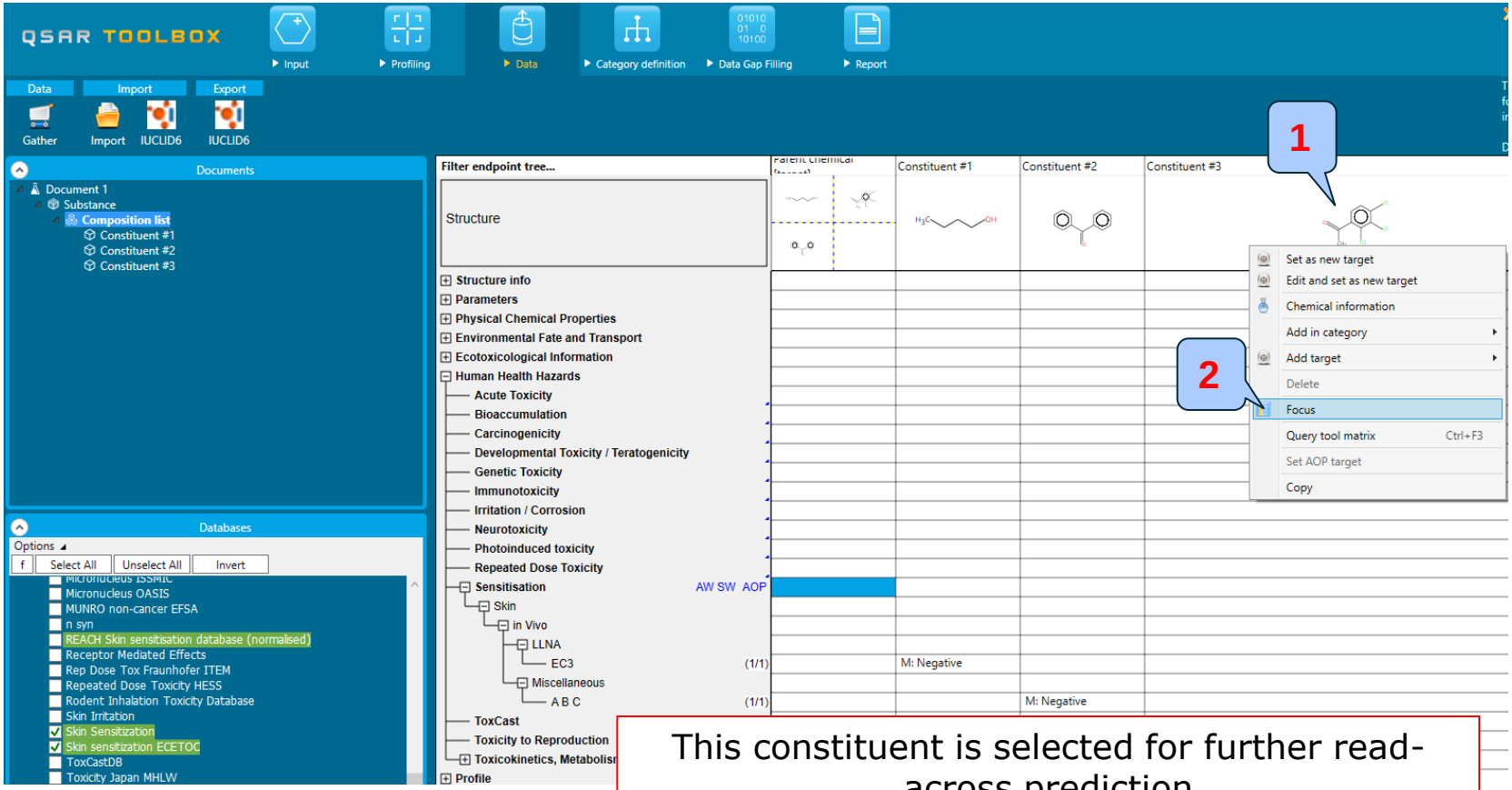
- You have entered the mixture with defined components.
- You have profiled the target chemical mixture and found no protein binding alerts for two of the mixture constituents. The third constituent has positive protein binding alerts and could elicit skin sensitization effect.
- Negative experimental 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 skin sensitization prediction of the mixture.

Outlook

- Background
- 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

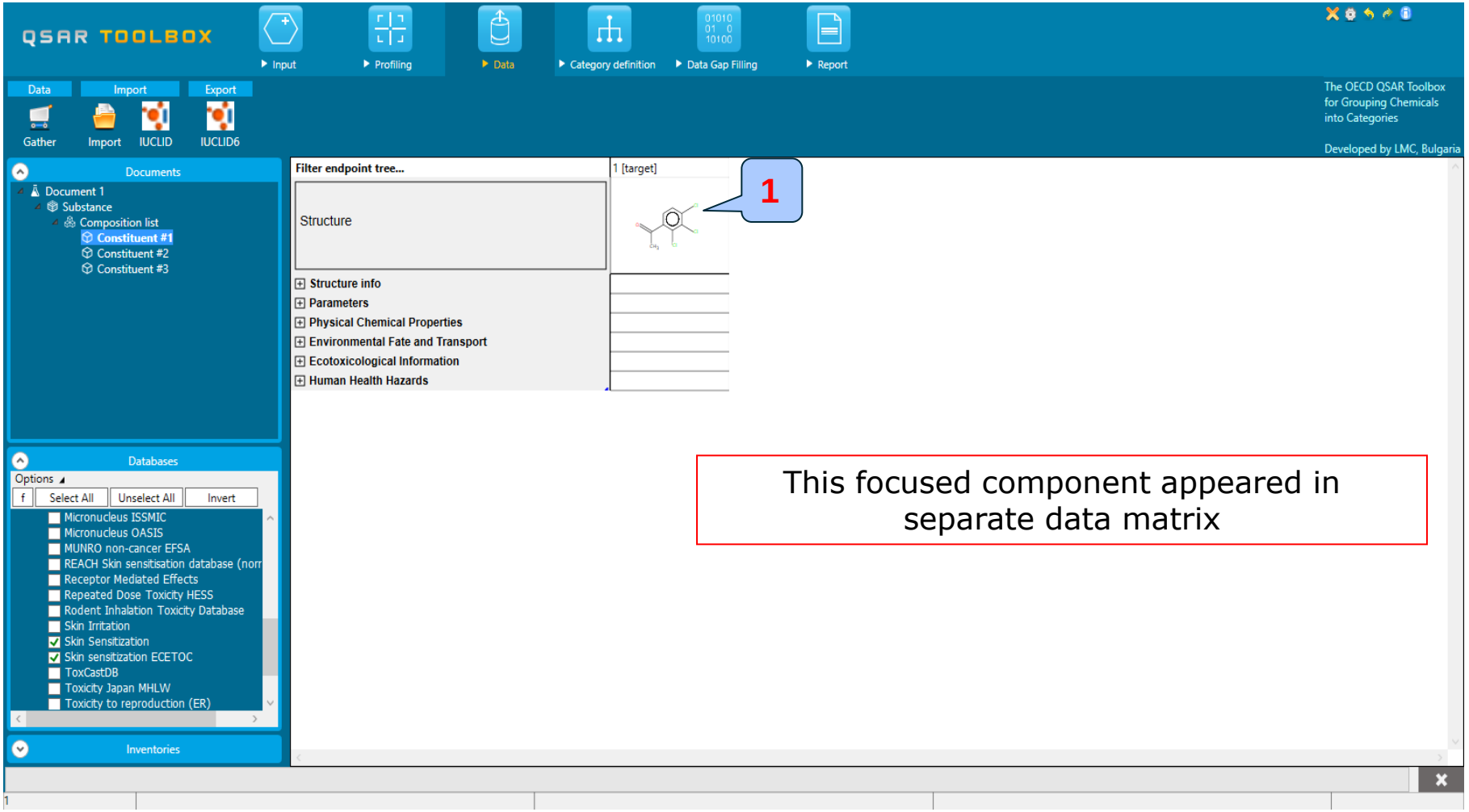


This constituent is selected for further read-across prediction

1. Right mouse click on 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
  - 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.

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

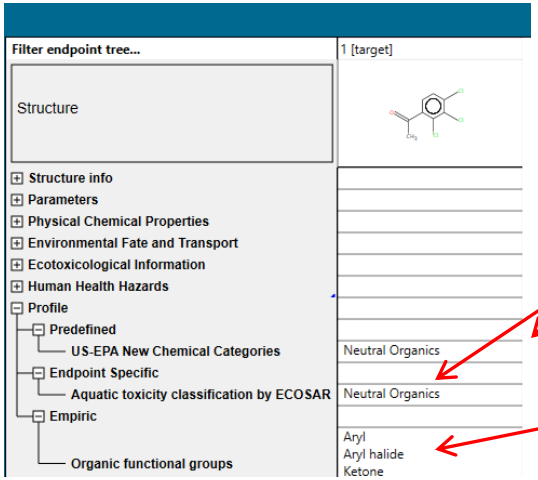
Phase I. Structure based

- 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.

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:

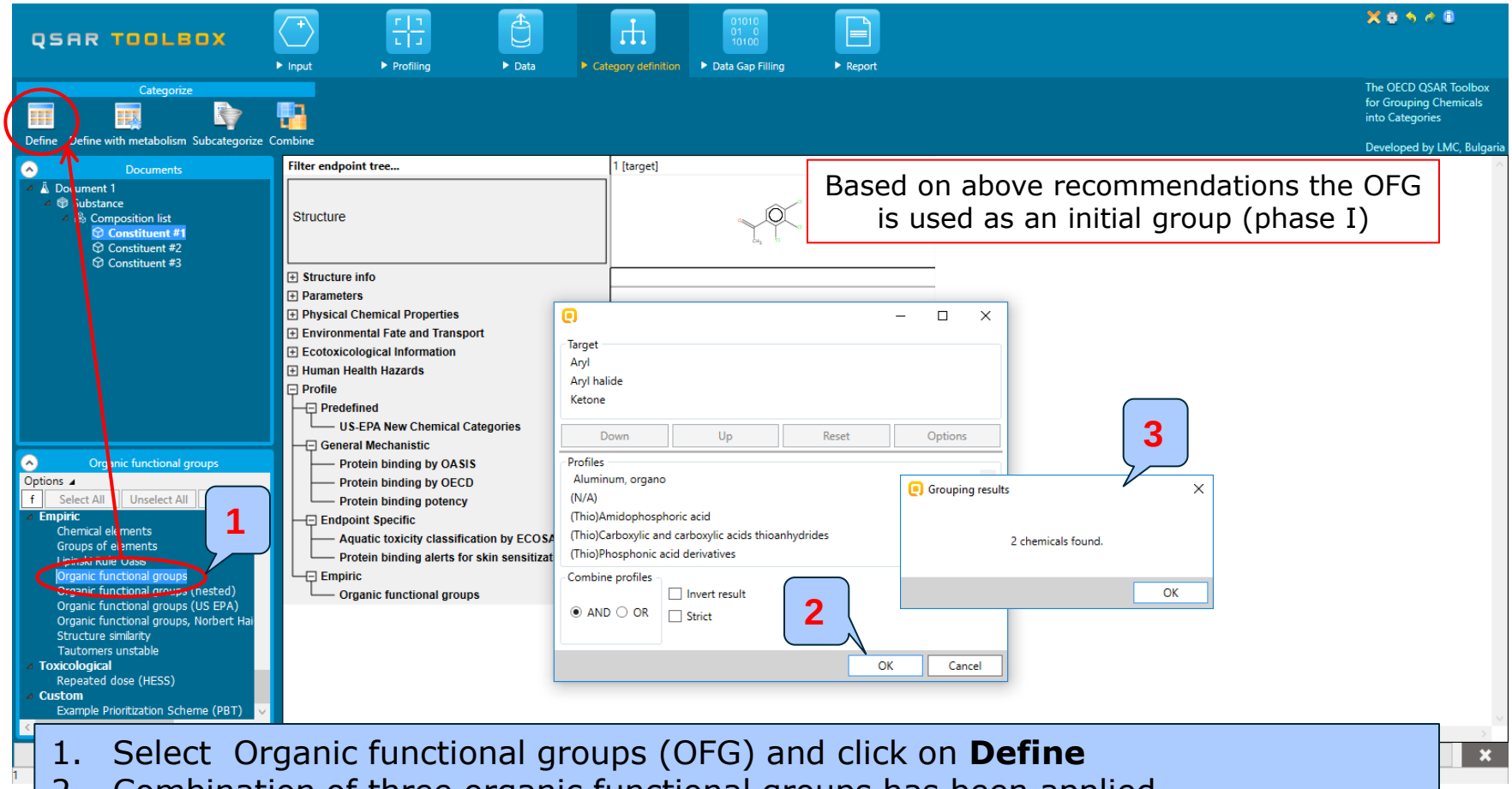
http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm

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

See next slides

Read across prediction of constituent without data

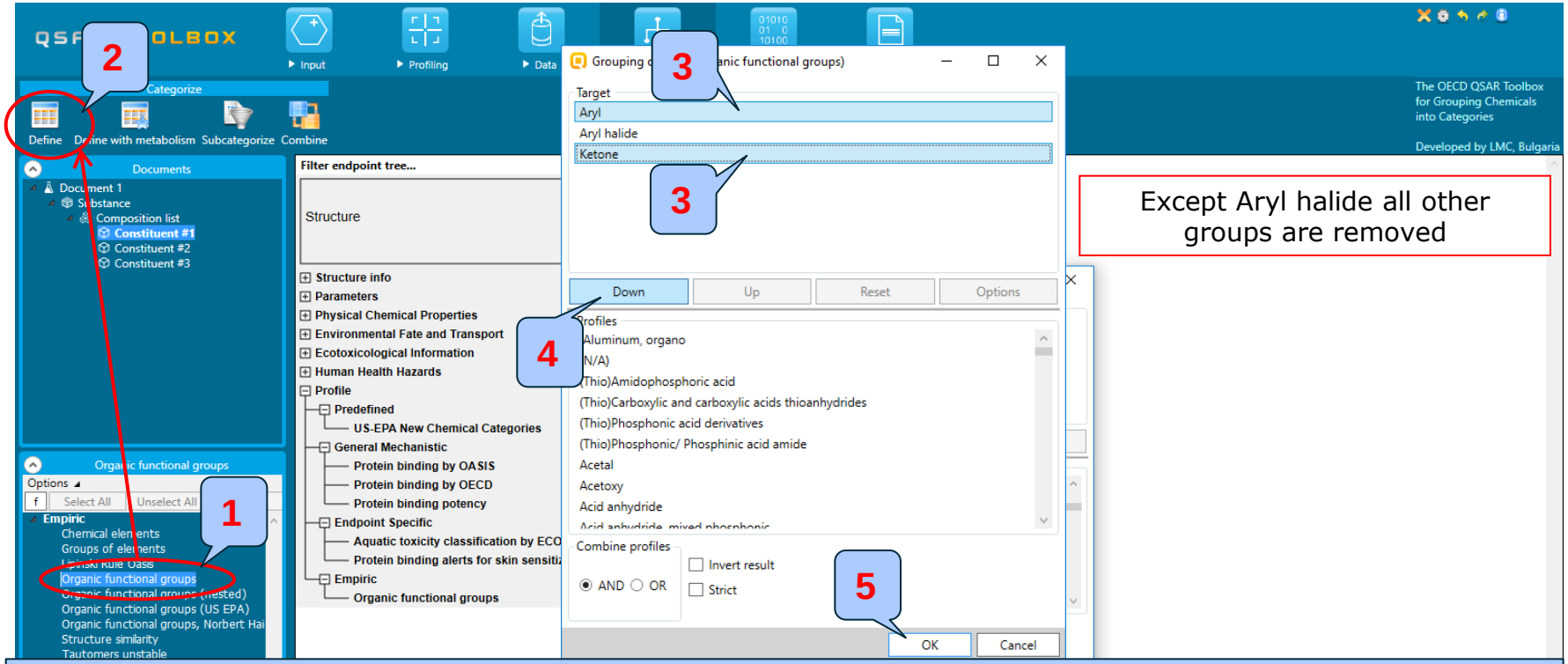
Define category by OFG



1. Select Organic functional groups (OFG) and click on **Define**
2. Combination of three organic functional groups has been applied.
3. Two chemicals have been found.

Read across prediction of constituent without data

Define category by OFG



To make a category based on Aryl halides only:

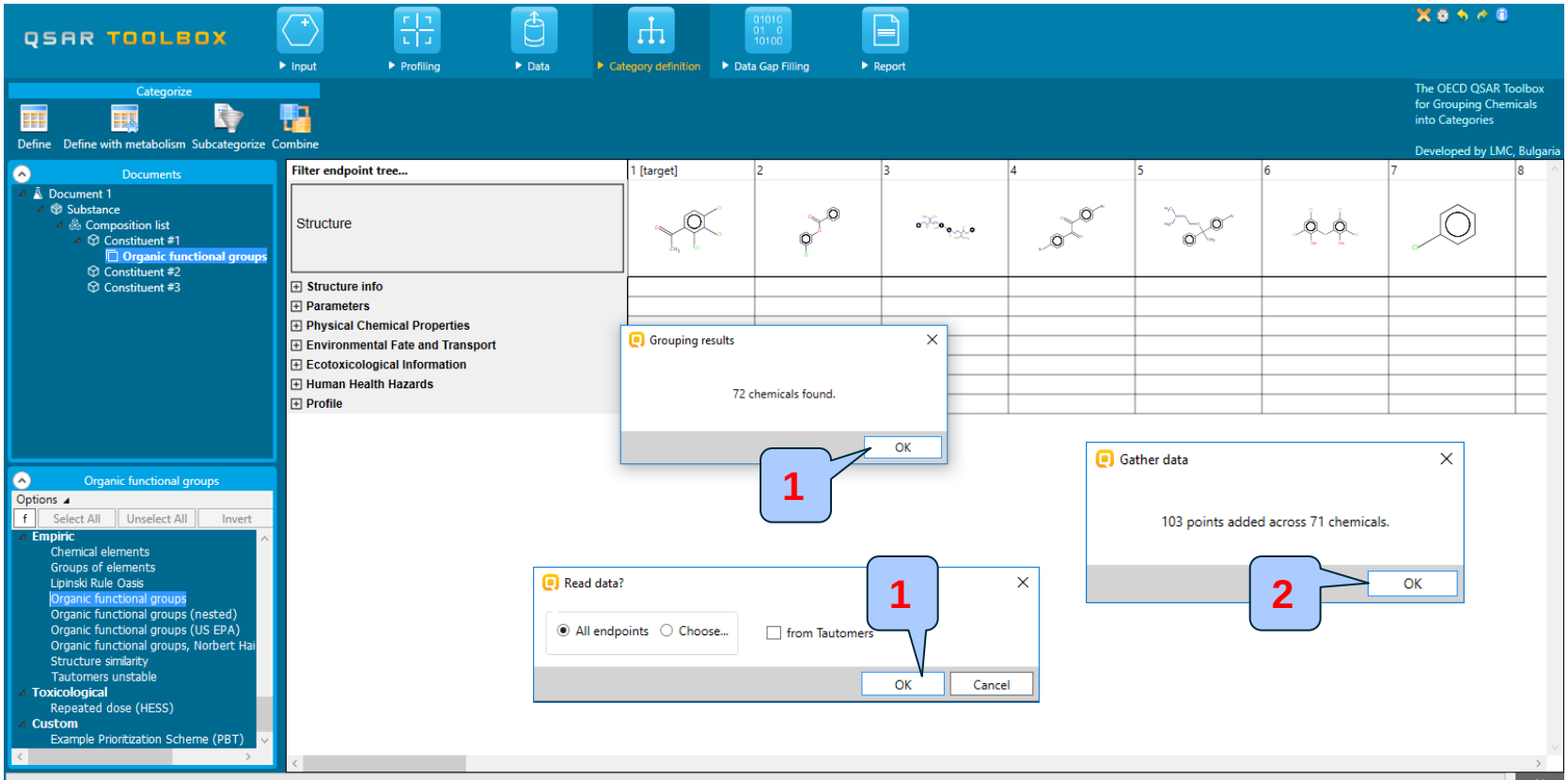
1. Select Organic functional groups,
2. Click **Define**,
3. Use Control from keyboard for subsequent selection of Aryl and Ketone,
4. Click **Down** to remove Aryl and Ketone; Aryl halides remain only,
5. Click **OK** to confirm.

Outlook

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

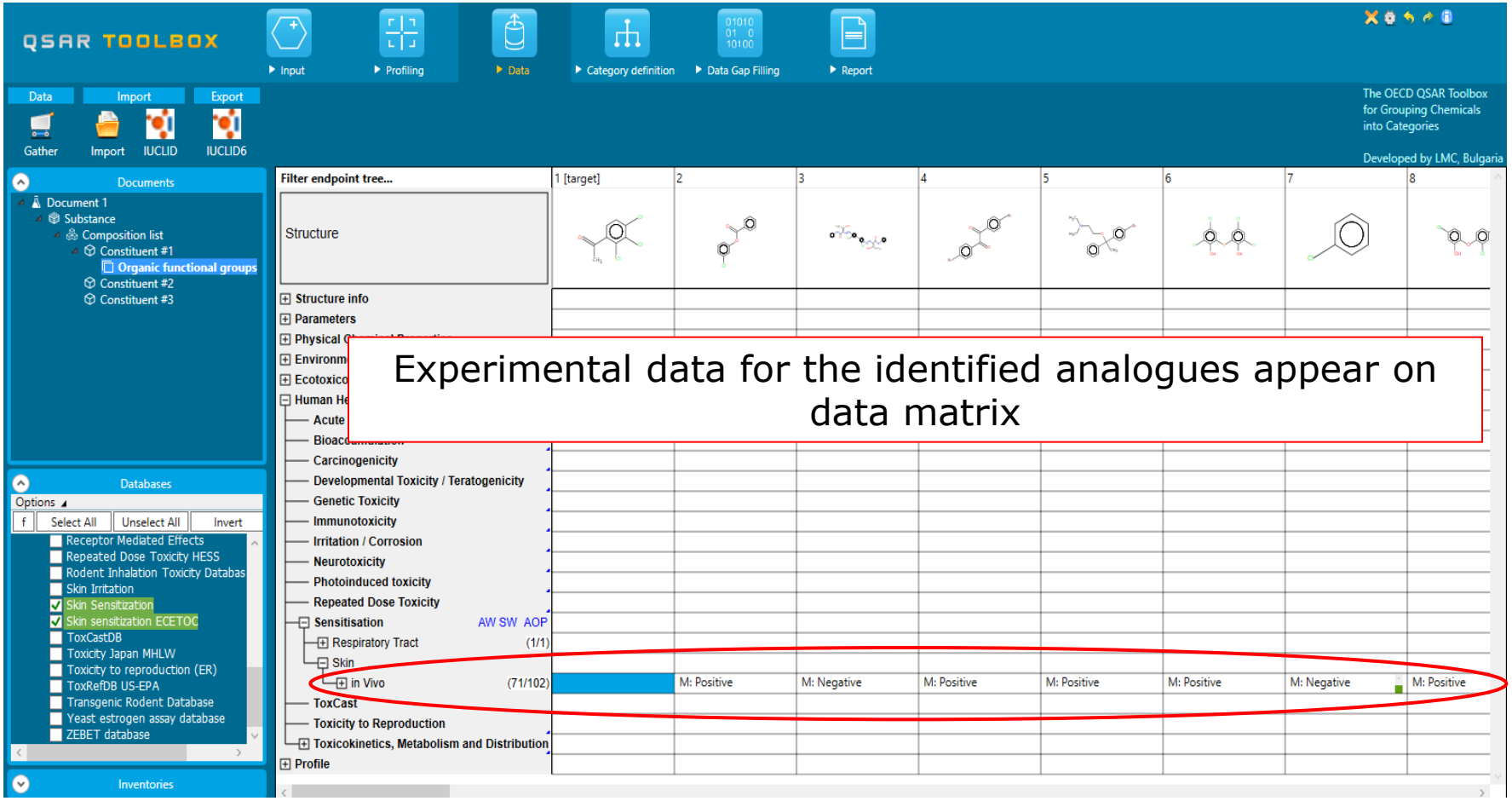
Gather data for analogues chemicals



1. Click **OK**,
2. Gather data

Read across prediction of constituent without data

Gather data for analogues chemicals

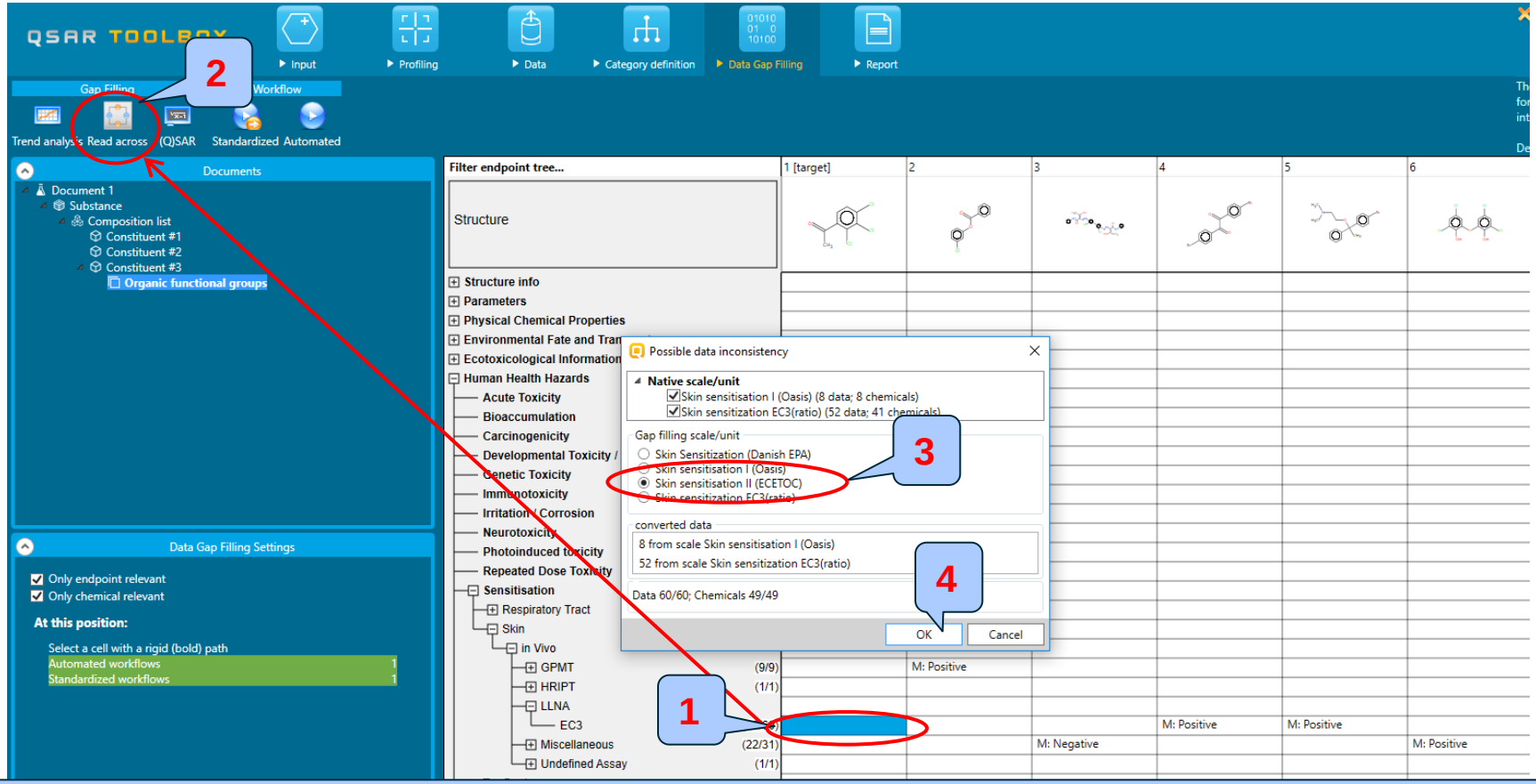


Outlook

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

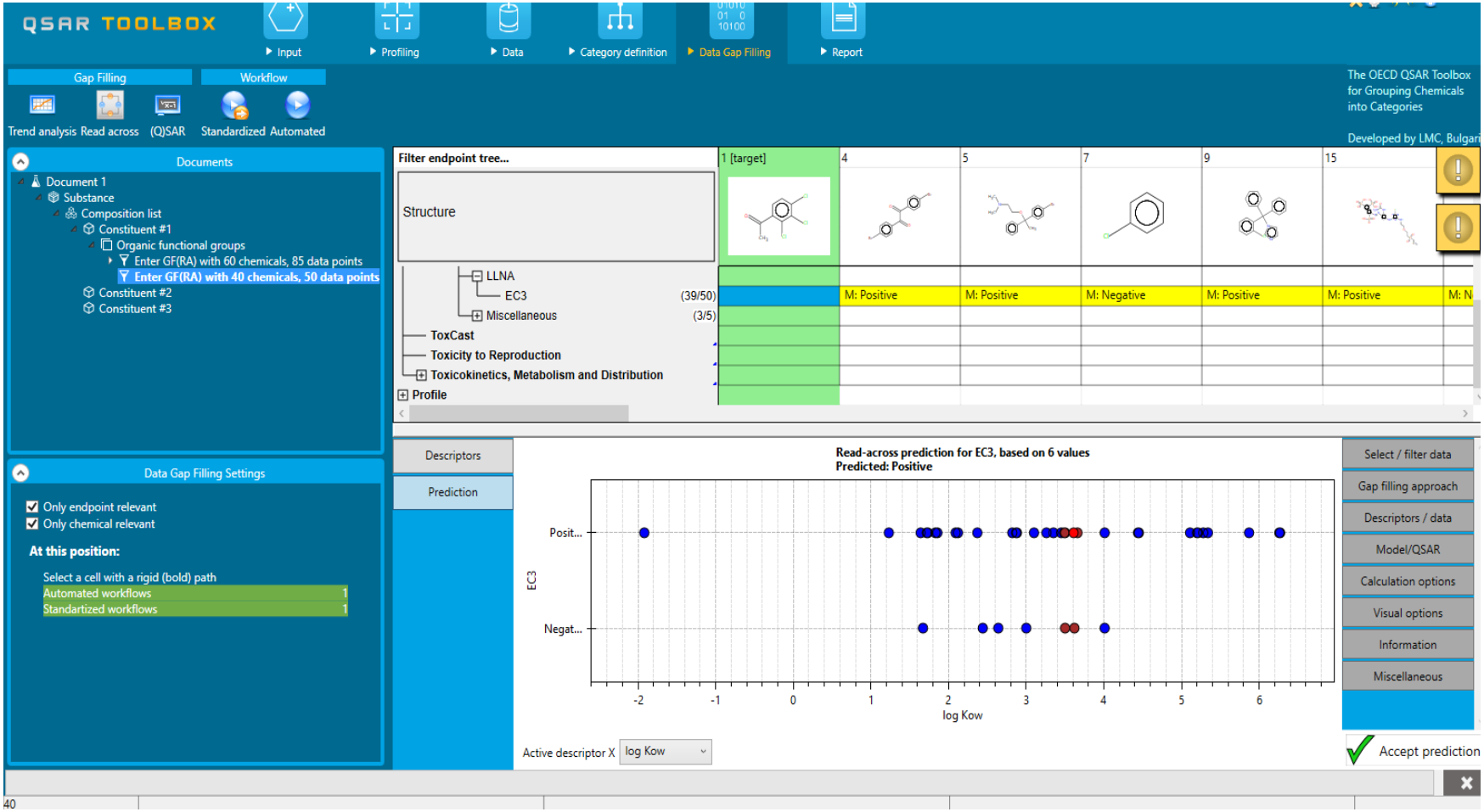
Apply read across



1. Click on the cell corresponding to Skin Sensitization in Vivo- LLNA-EC 3
2. Select **Read-across**
3. Select scale/unite **Skin sensitization II(ECETOC)**
4. Click **OK** (in this case we mix all endpoints and assays)

Read across prediction of constituent without data

Apply read across



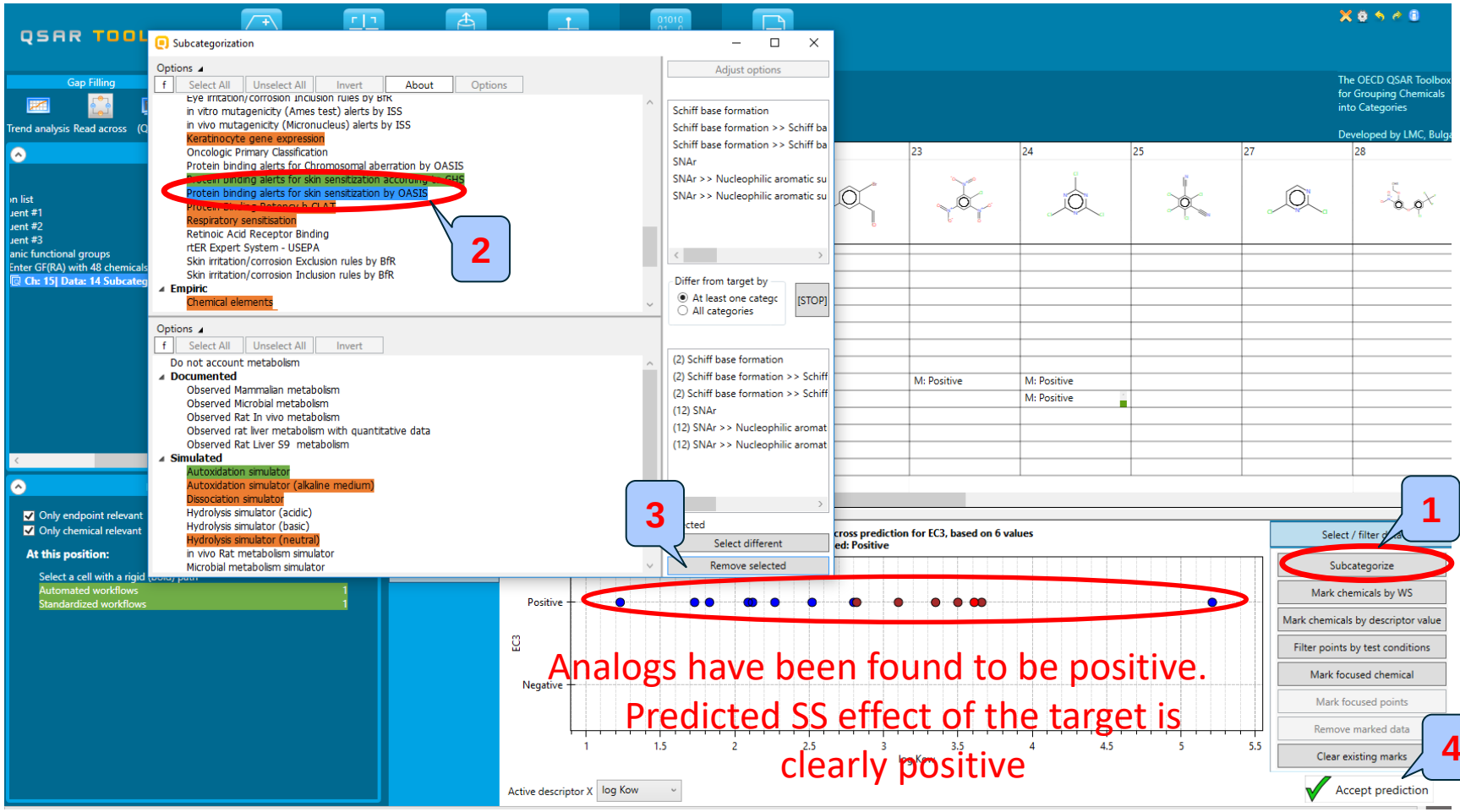
Read across prediction of constituent without data
Subcategorization

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

- These steps are summarized in the next screen shots.

Read across prediction of constituent without data

Subcategorization by Protein binding alert for SS



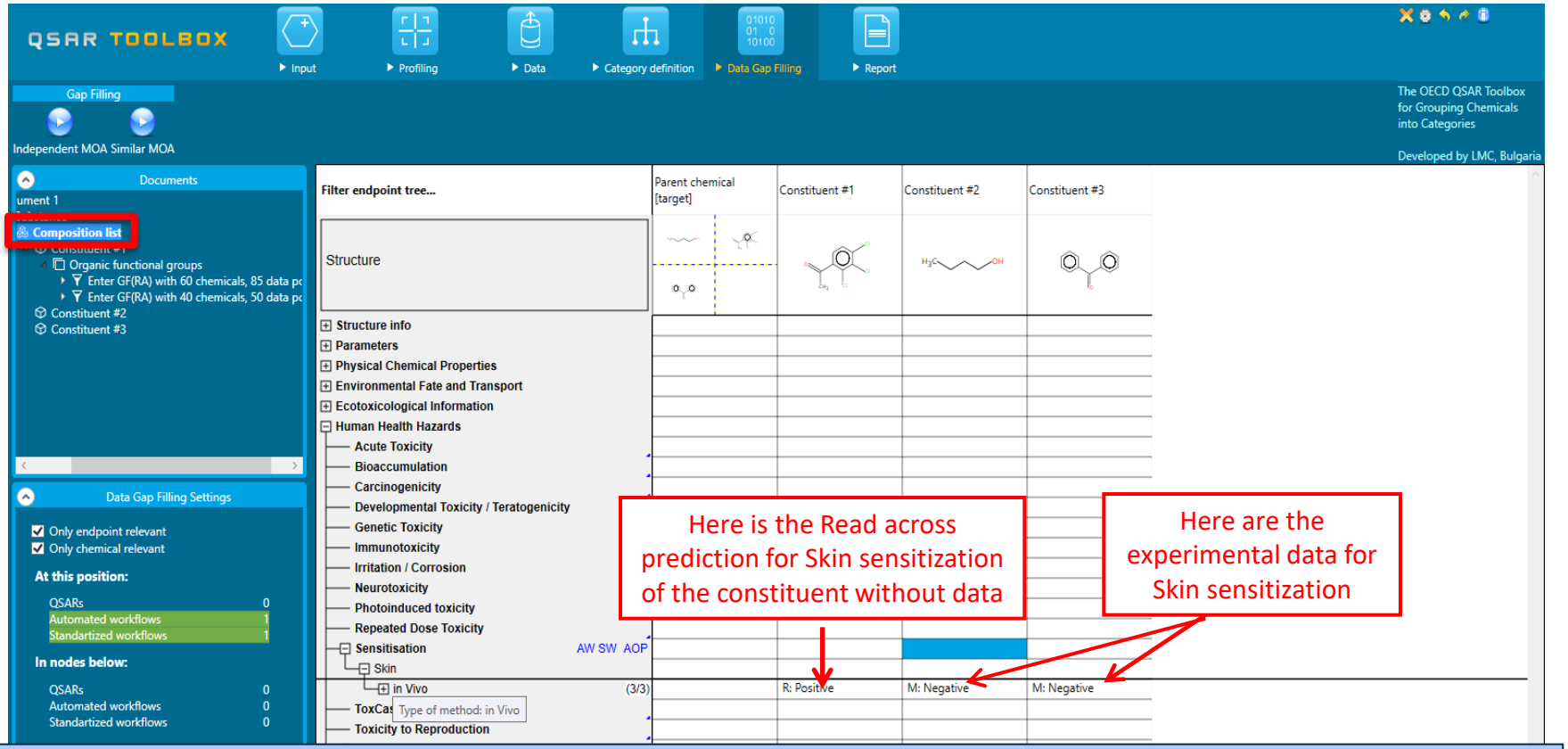
1. Select filter data/Subcategorize 2. Select **Protein binding alerts for skin sensitization by OASIS**. 3. **Remove selected** to eliminate dissimilar chemicals. 4. **Accept prediction** to return to data matrix.

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

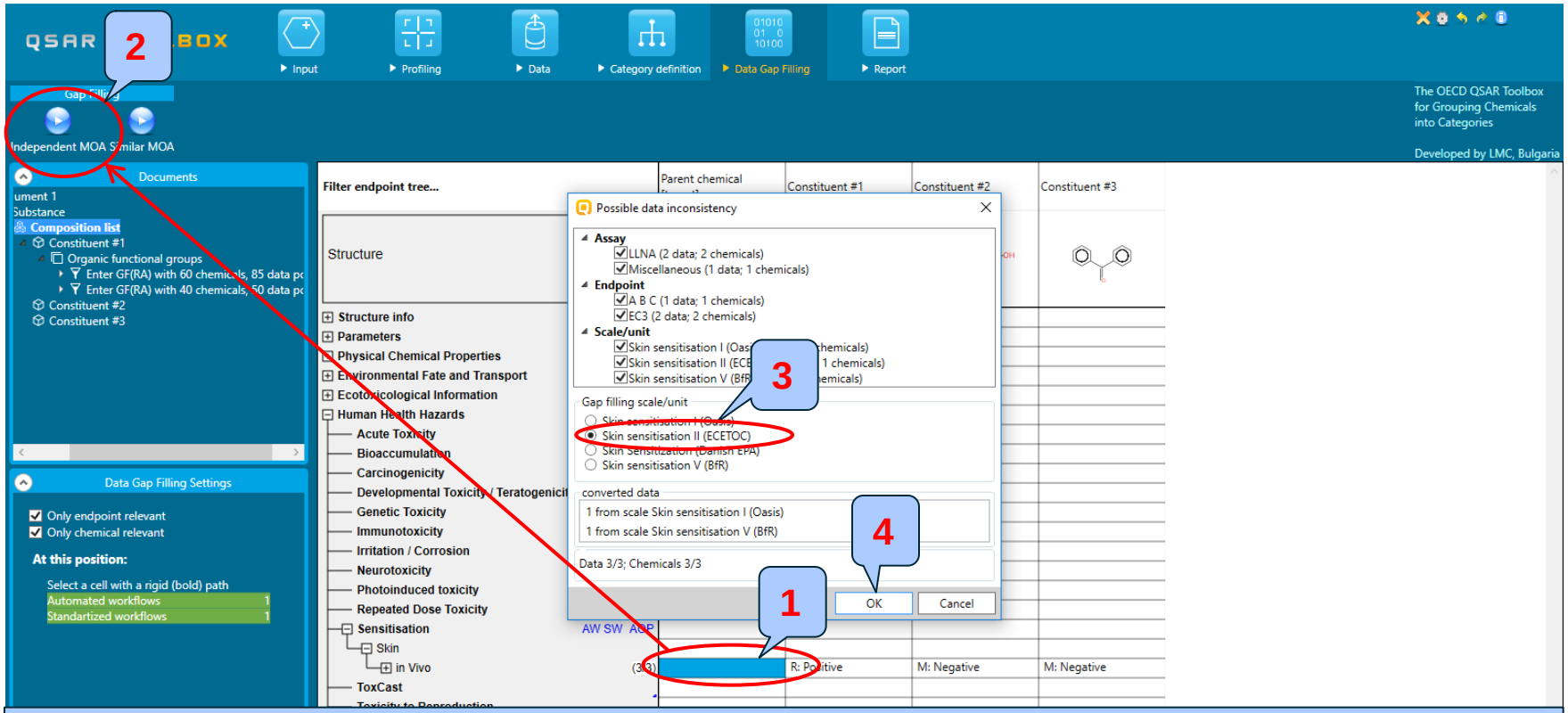


Select **Composition list** (highlighted); this will activate two mode of actions:

- Independent MOA;
- Similar MOA.

Filling data gap for skin sensitization of mixture

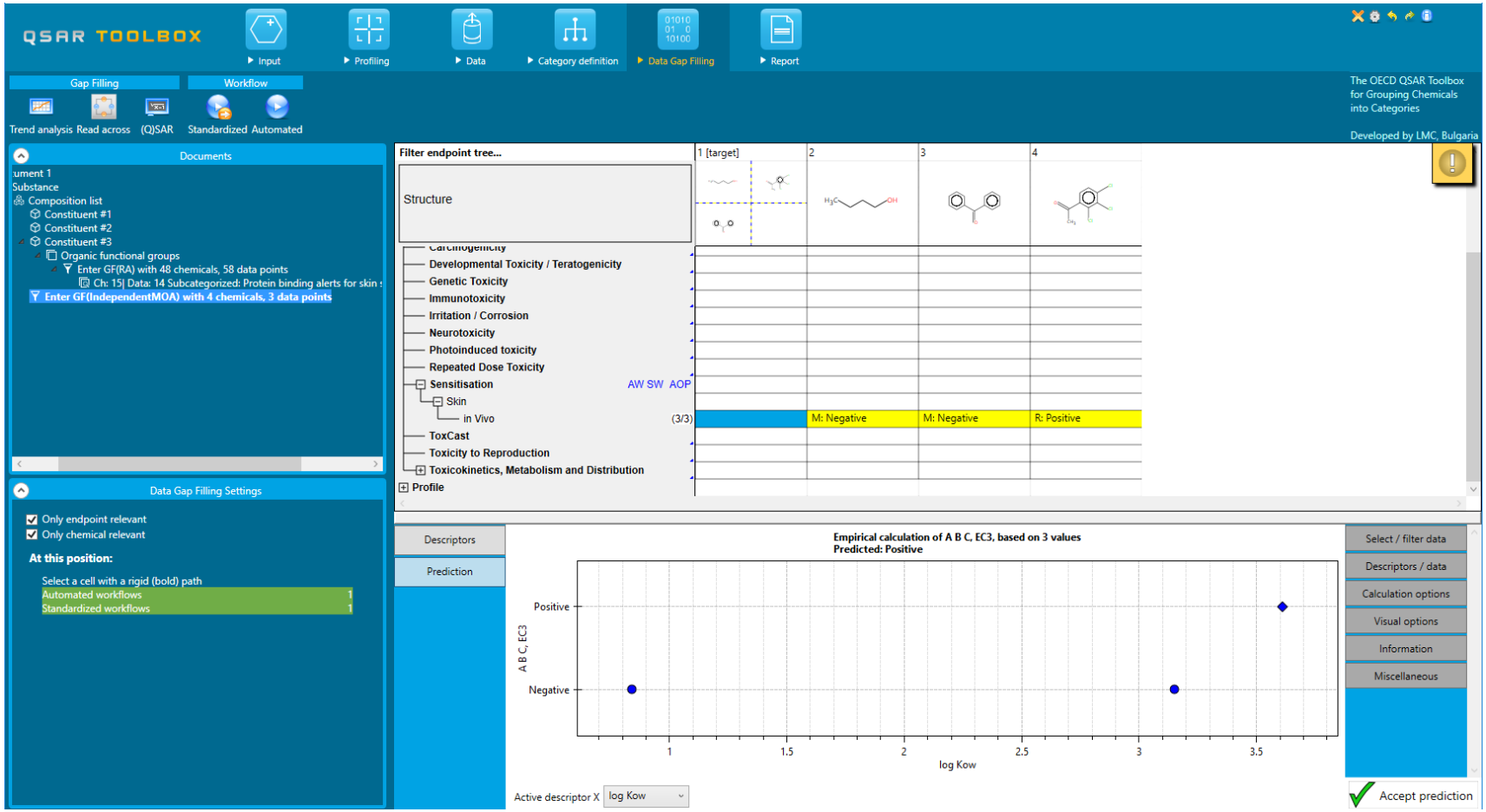
Applying Independent Mode of Action



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

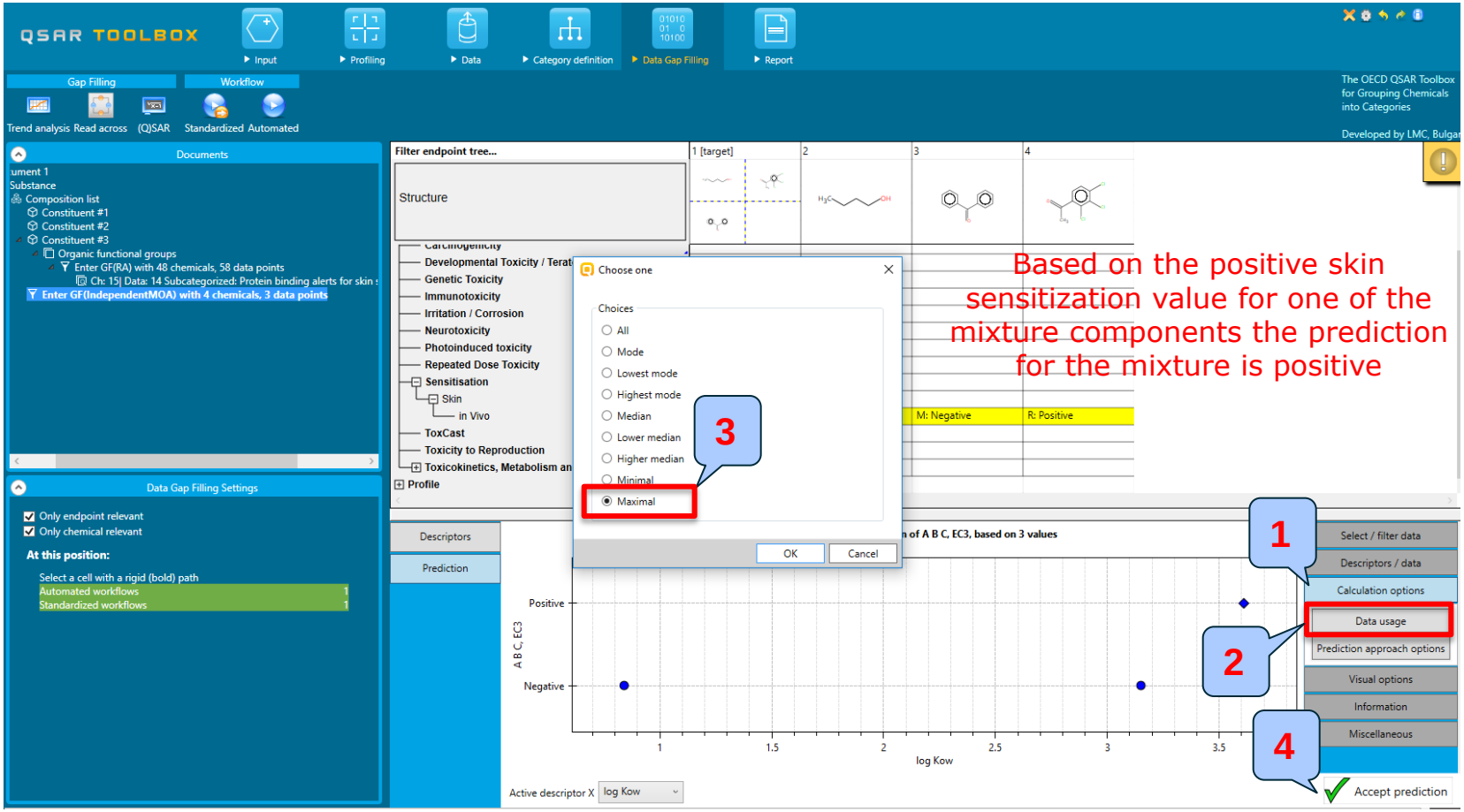
Applying Independent Mode of Action



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

Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action

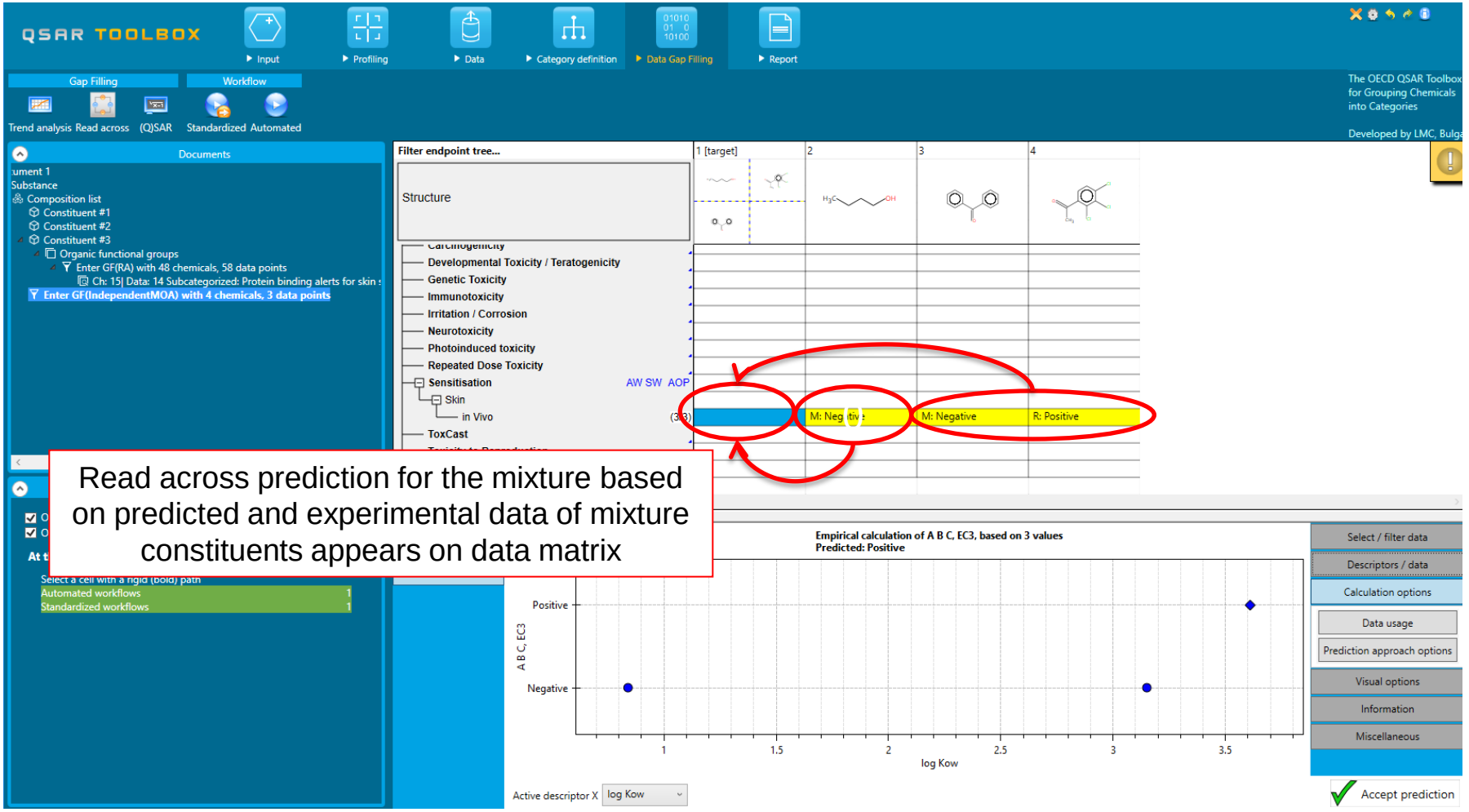


Consecutive steps:

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

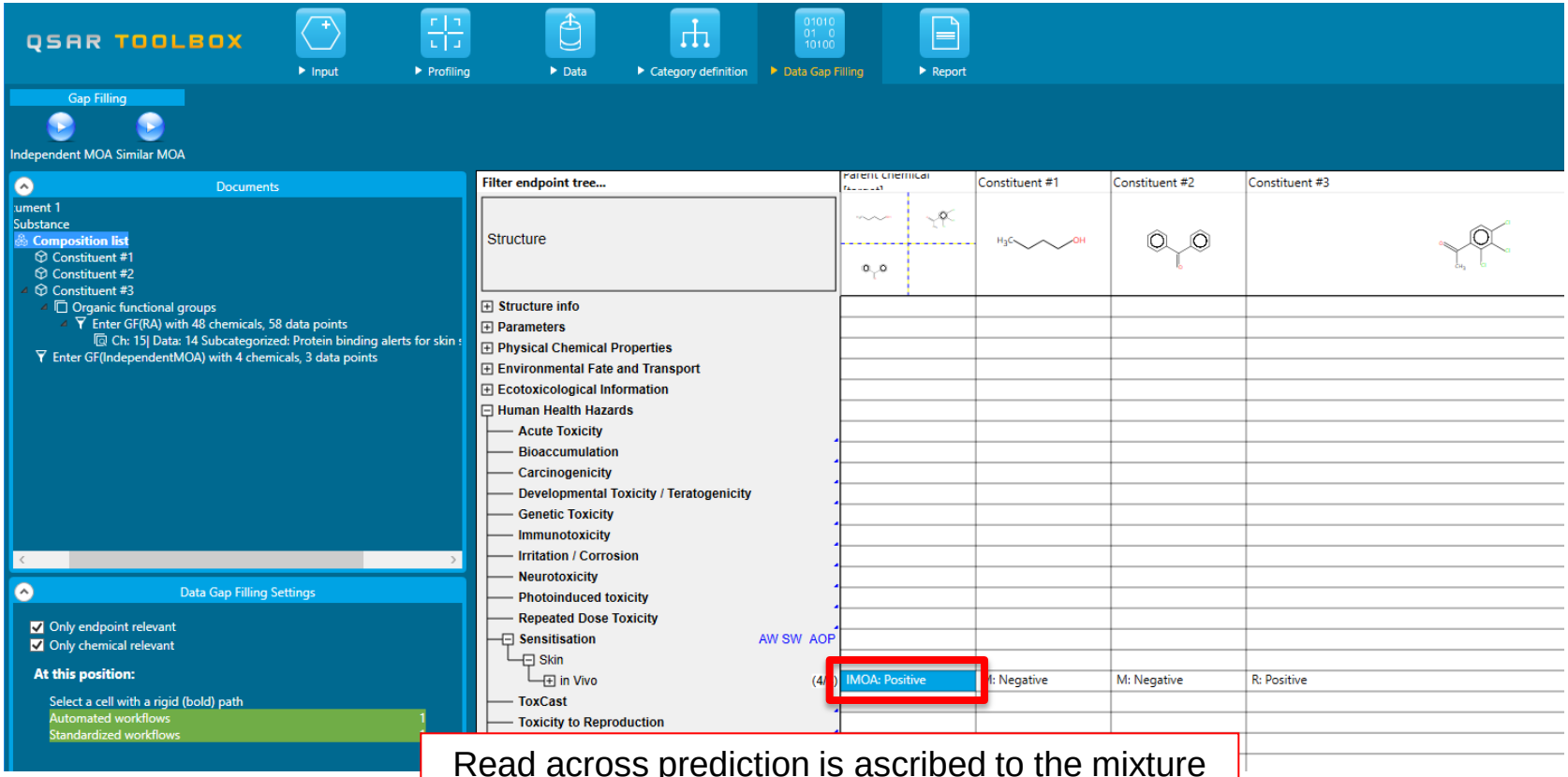
Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action



Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action



Read across prediction is ascribed to the mixture

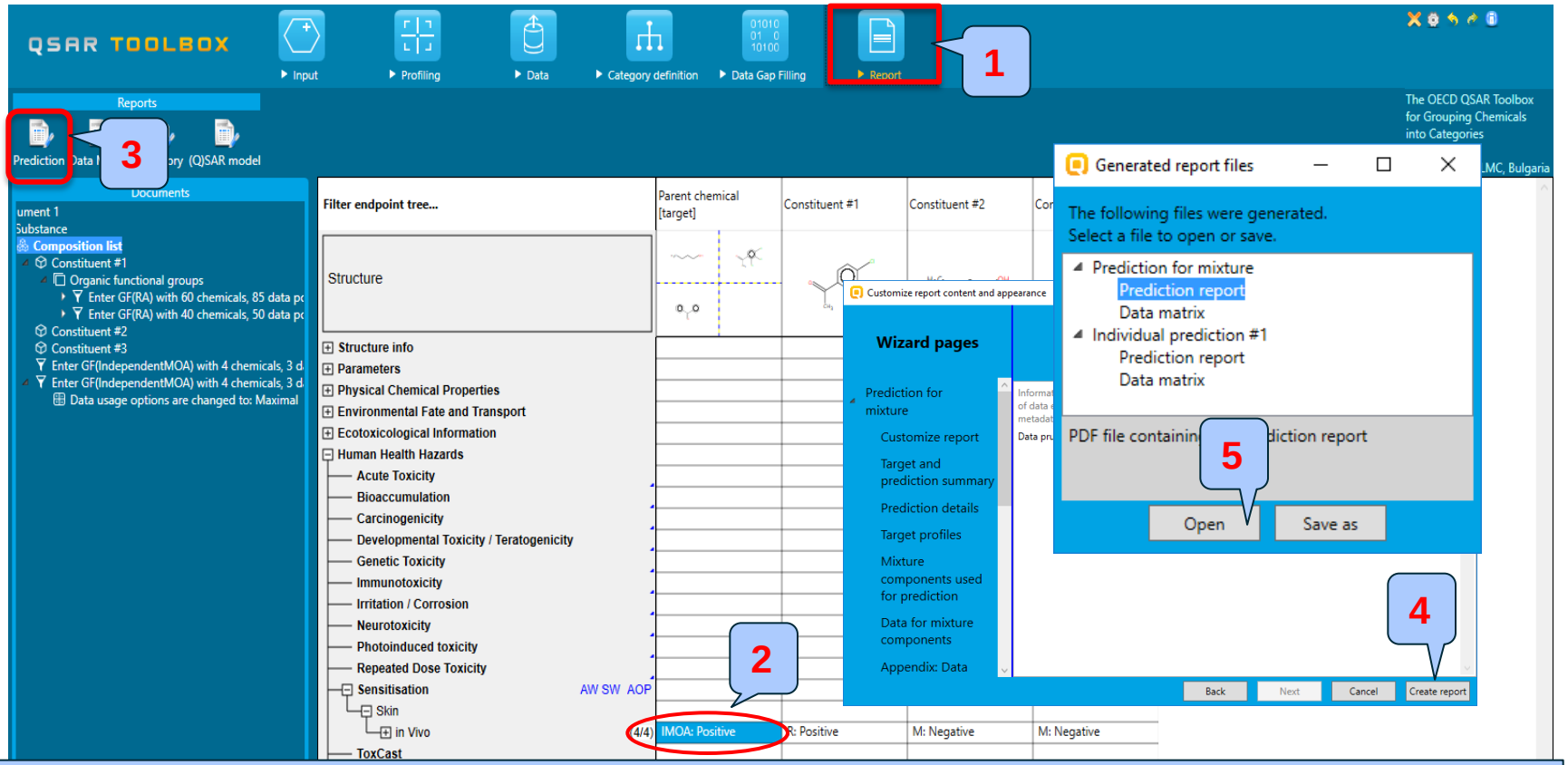
Outlook

- 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 user to generate a report on the predictions performed with the Toolbox.
- This module contains predefined report templates as well as a template editor which allows users to provide modifications.
- The report can be printed or saved in different formats.
- Generating the report is shown on next screenshots.

Report



1. Select **Report**
2. Click on the cell corresponding to Skin Sensitization/ in Vivo for mixture
3. **Prediction**
4. **Create report**
5. **Open** Prediction report

Report

QSAR Toolbox prediction for multicomponent substance

Based on observed and predicted data for mixture components

Date: 2 Aug 2017
Author(s):
Contact details:

Target information		
Structural information SMILES: CCCCO.CC(=O)c1ccc(Cl)c(Cl)c1Cl.O=C(c1ccccc1)c1ccccc1 Structure	**Numerical identifiers** EC#: N/A CAS#: Invalid CAS number: 0-00-0 Other: N/A	**Chemical names**

Toolbox report for mixture

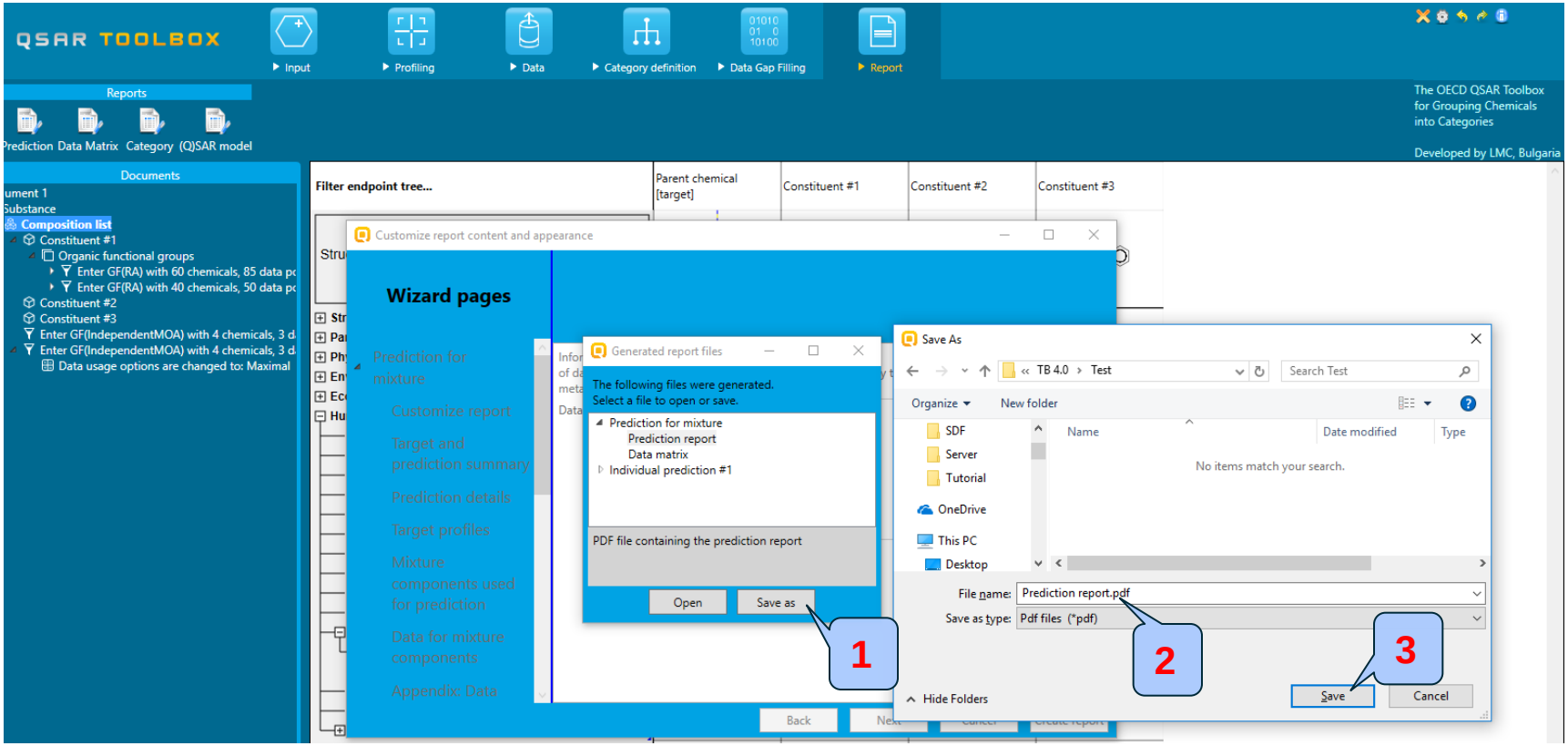
Outlook

- Background
- 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. Click on **Save as** button;
2. Define path and name of the pdf file;
3. Click **Save** button.