QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD (Q)SAR Toolbox v.4.4.1

Example for predicting skin sensitisation potential of (*2E,6Z*)-2,6-nonadien-1-ol accounting for skin metabolism

Outlook

- **Background**
- Objectives
- The exercise
- Workflow
- Save/open the workflow

Background

- This is a step-by-step presentation designed to take the user through the Toolbox workflow for filling a data gap for skin sensitization of trans-2,cis-6-nonadienol accounting for its metabolic transformations in skin.

Outlook

- Background
- **Keywords**
- Objectives
- The exercise
- Workflow
- Save/open the workflow

Keywords

TARGET CHEMICAL - chemical of interest

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/open the workflow

Objectives

This presentation demonstrates a number of functionalities of the Toolbox, such as:

- Simulating skin metabolism of a target chemical;
- Identification of active metabolite;
- Searching analogues for a selected active metabolite;
- Filling the data gap for active metabolites by read across;
- Assigning prediction of a metabolite to the parent chemical.

Outlook

- Background
- Keywords
- Objectives
- **The exercise**
- Workflow
- Save/open the workflow

The Exercise

- In this exercise we will predict the skin sensitization potential of a target chemical: **trans-2,cis-6-nonadienol** [**CAS # 28069-72-9**].
- The target chemical will be checked for Protein binding alerts associated with skin sensitization.
- The available experimental data for the target chemical will be collected.
- Skin metabolism of target chemical will be accounted for.
- Read across prediction for the active metabolite will be applied.
- The predicted result of the metabolite will be transferred to the target chemical.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
- Save/open the workflow

Workflow

- **As you know the Toolbox has 6 modules which are typically used in sequence:**
  - Chemical Input
  - Profiling
  - Data
  - Category Definition
  - Data Gap Filling
  - Report
- **In this example we will use the modules in a different order, tailored to the aims of the example.**

Workflow

The following steps will be executed in the workflow of the current example:

- Input target chemical
- Define target endpoint
- Profile the target chemical
- Collecting experimental data for the target chemical
- Handling of skin metabolism of the target chemical
  - Multiplication of the target chemical
  - Profiling the set of metabolites
  - Focus on the active metabolite
  - Defining category for the active metabolite
  - Filling data gap of the active metabolite
- Transferring the prediction of active metabolite to the parent
- Reporting the prediction.

Outlook

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

Input
Overview

- This module provides the user with several means of entering the chemical of interest (i.e. 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.

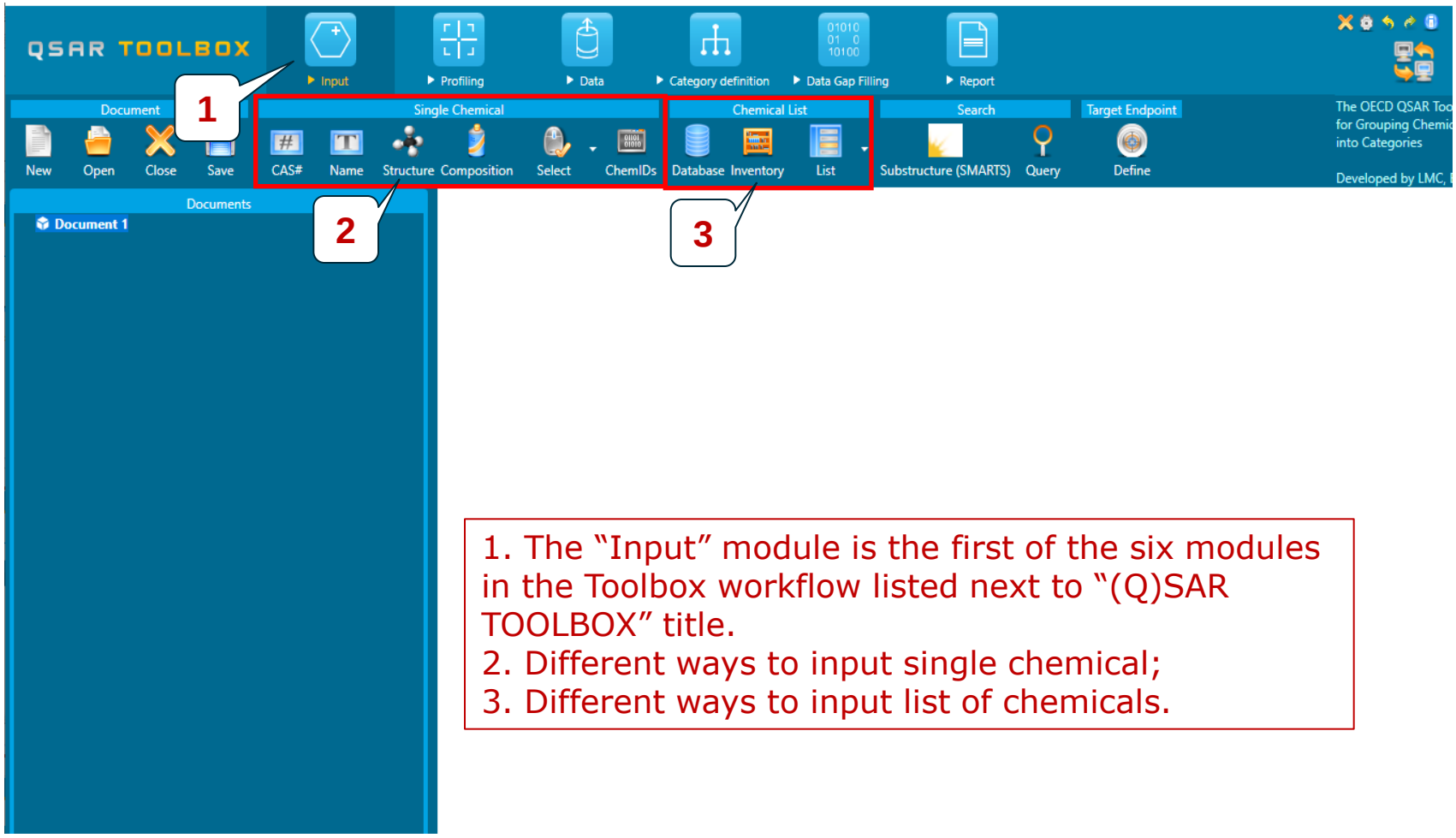
Input
Input chemical(s)

Alternative ways to input a chemical(s):

A. Single target chemical
- 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

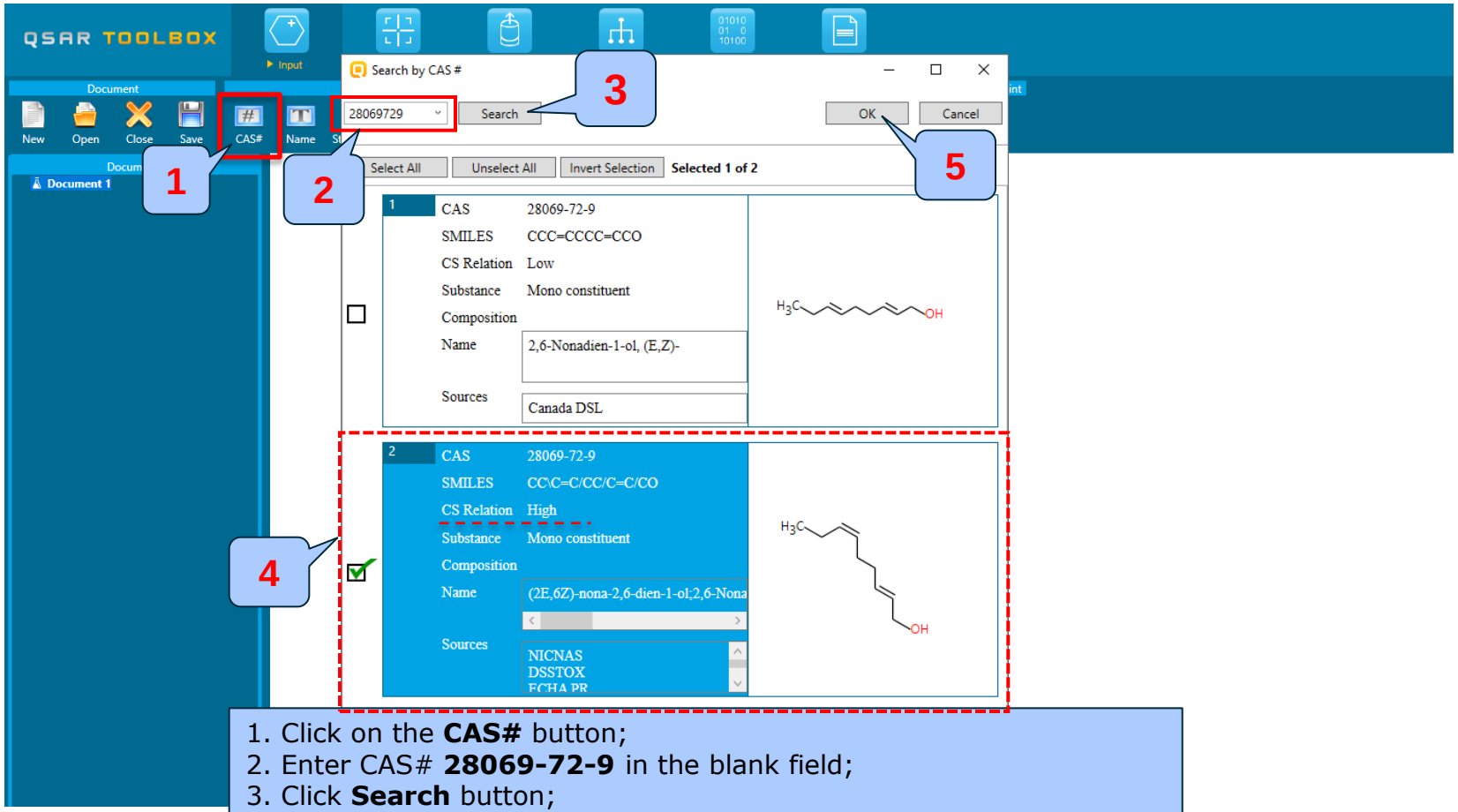
Input
Input Screen



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

Input

Input target chemical by CAS#



1. Click on the **CAS#** button;
2. Enter CAS# **28069-72-9** in the blank field;
3. Click **Search** button;
4. Select the second structure (with High CAS-SMILES relationship);
5. Click **OK**.

Outlook

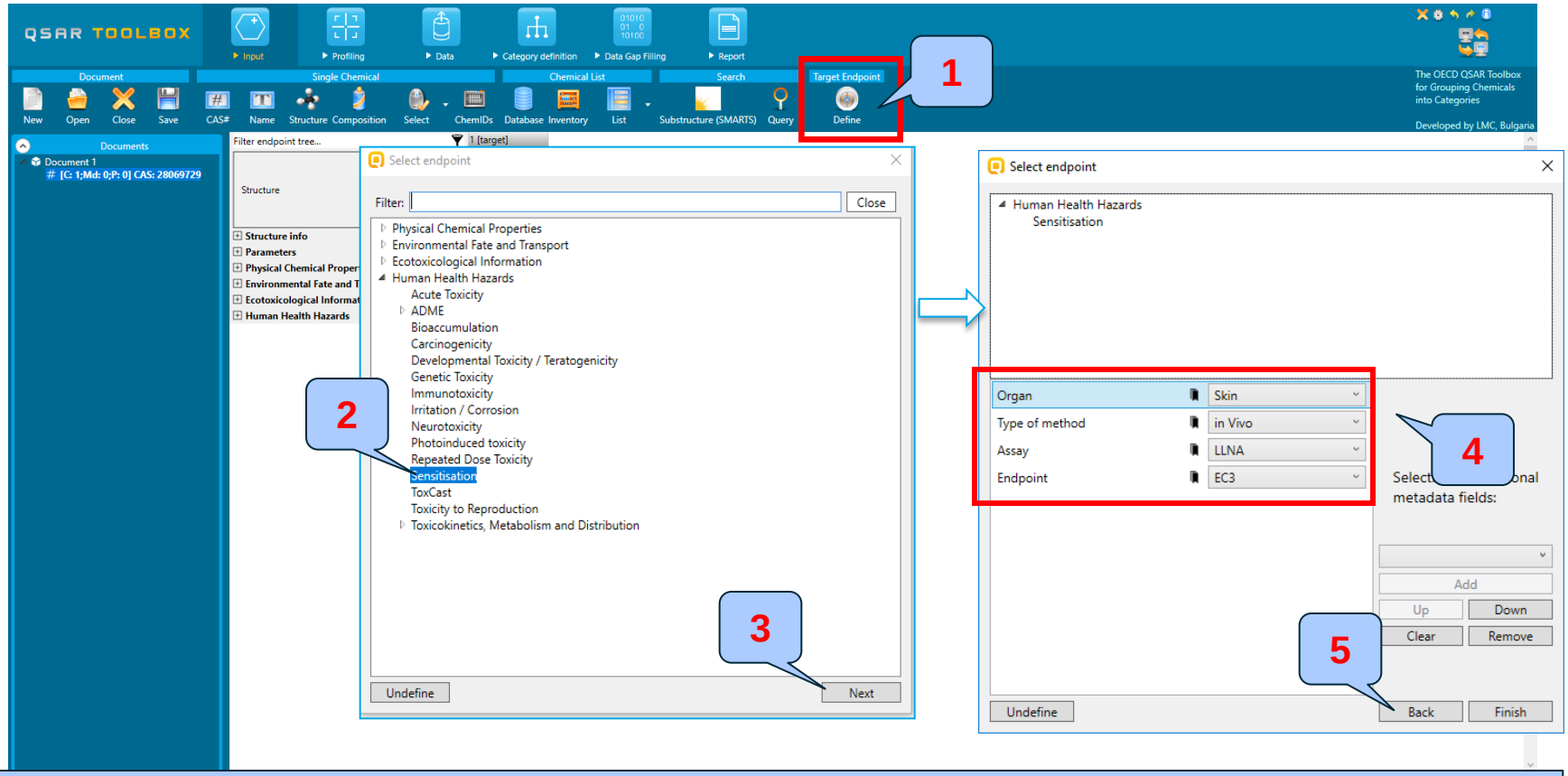
- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
    - **Define target endpoint**

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 (in green the most suitable and in orange – the plausible ones);
- 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.

Define target endpoint

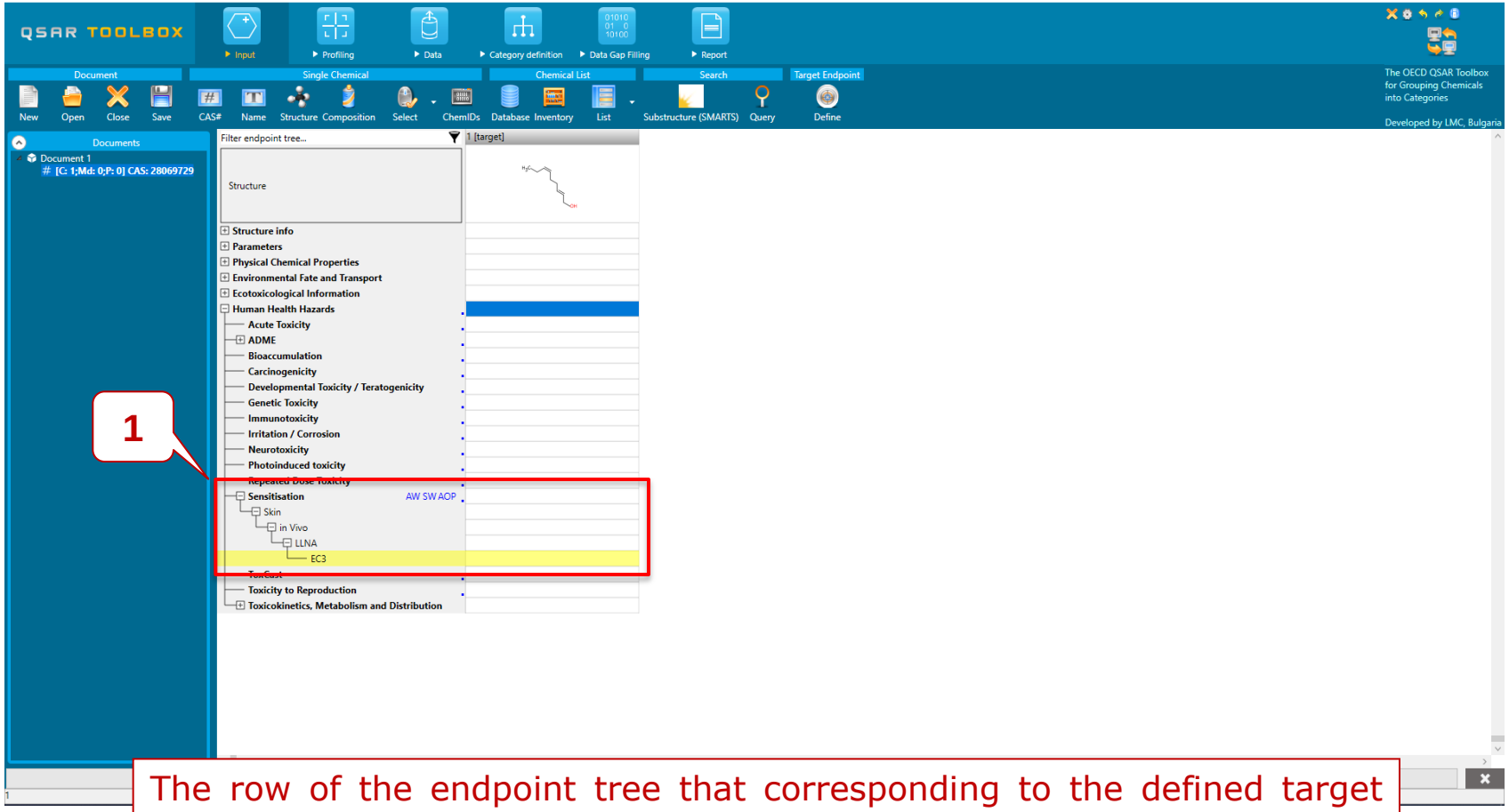
Definition of *in vivo Skin sensitization*, EC3



Click on the **Define** button (1). In the new window select the general endpoint "*Skin Sensitization*" (2) and click on **Next** (3).
Specify the endpoint using the drop-down menus as follows: For *Endpoint* select **EC3**; for *Assay* – **LLNA**; for *Type of method* - **in Vivo**; for *Organ* - "**Skin**" (4). Finally click on **Finish** (5).

Define target endpoint

Visualisation of the defined target endpoint



The row of the endpoint tree that corresponding to the defined target endpoint is highlighted in yellow (1)

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - **Profile the target chemical**

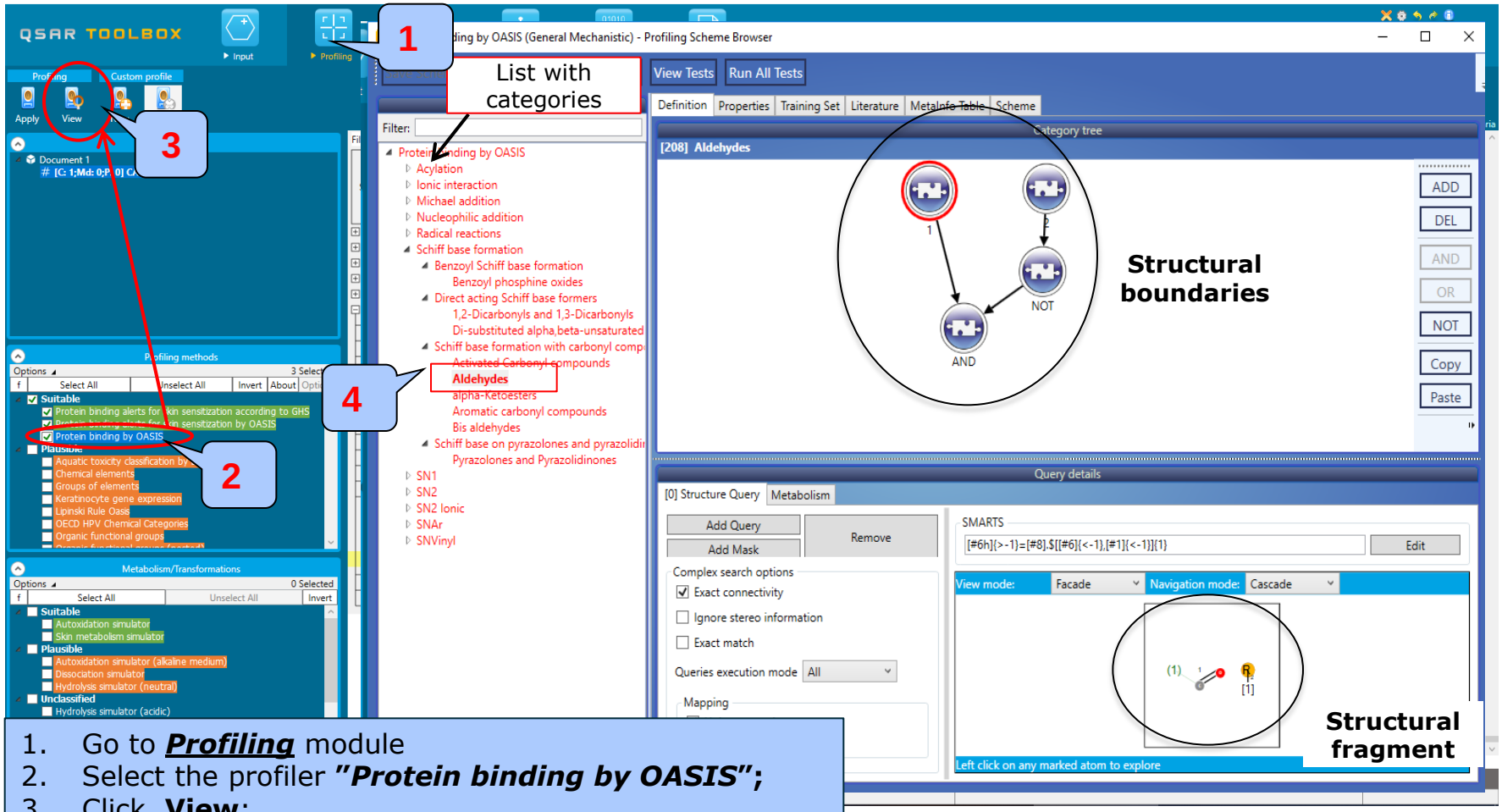
Profiling
Overview

- "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;
- "Profiling" module contains all the knowledge in the system coded in profiling schemes (profilers);
- "Profilers" are a collection of empirical and mechanism knowledge (expertly derived) which could be used to analyse the structural properties of chemicals;
- The "profilers" identify the affiliation of the target chemical(s) to preliminary defined categories (functional groups/alerts);
- The "Profiling" module contains also observed and simulated metabolisms/transformations, which could be used in combination with the profiling schemes;
- The outcome of the profiling determines the most appropriate way to search for analogues, but they are also useful for preliminary screening or prioritization of substances;
- The "profilers" are not (Q)SARs, i.e. they are not prediction models themselves;
- Based on the "profilers' relevancy" (determined by the defined target endpoint), the most suitable once are getting colour highlighted*.
- For most of the profilers, background information can be retrieved by selection of a profilers (for example, Protein binding by OASIS) and then click on "View'' (see the next slide).

*For more details regarding relevancy of the profilers see ppt: Example for predicting skin sensitization taking into account alert performance

Profiling

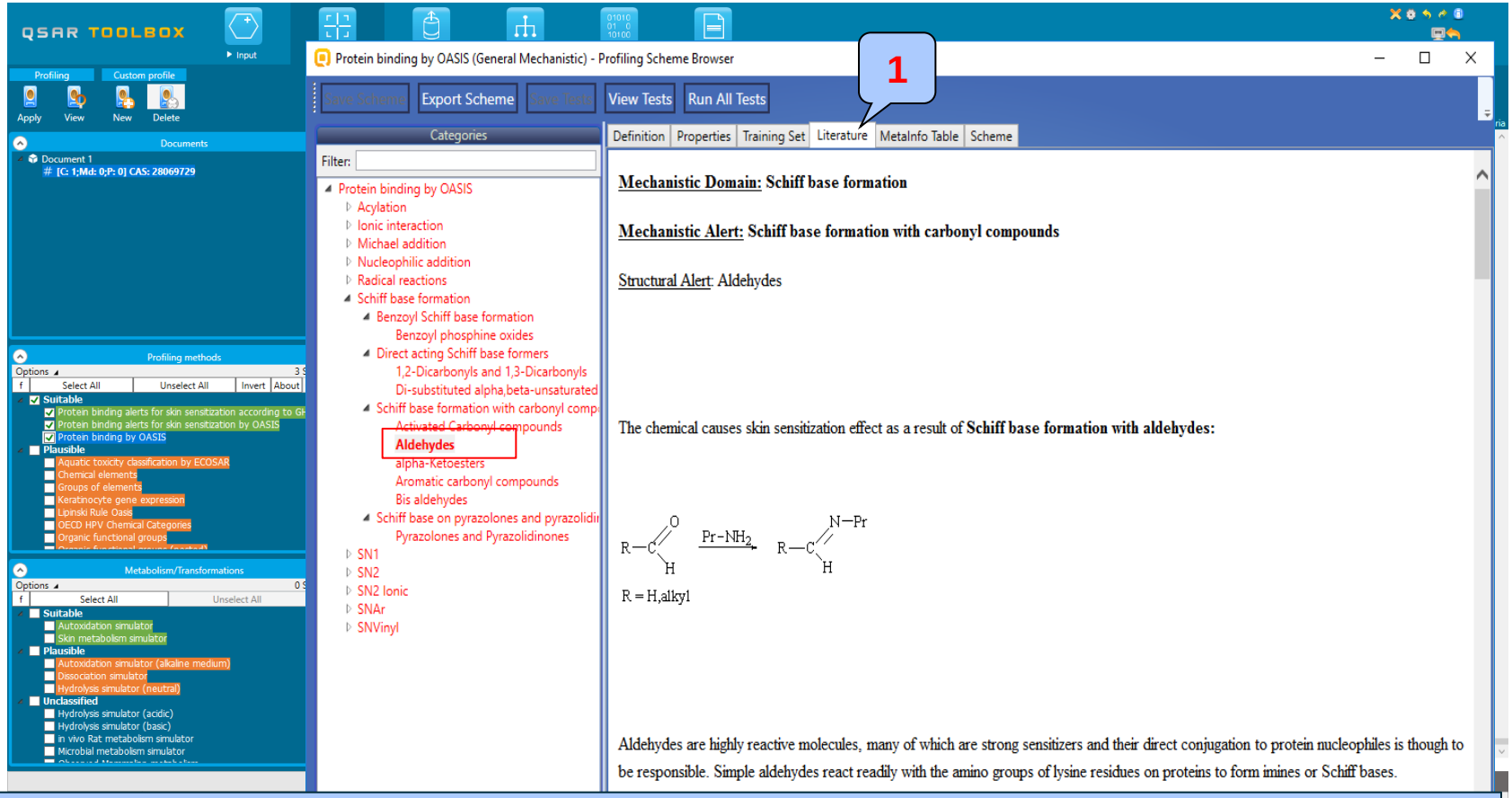
Profilers` background information



1. Go to ***Profiling*** module
2. Select the profiler ***"Protein binding by OASIS"***;
3. Click **View**;
4. Select the **Aldehydes** category.

Profiling

Profilers` background information



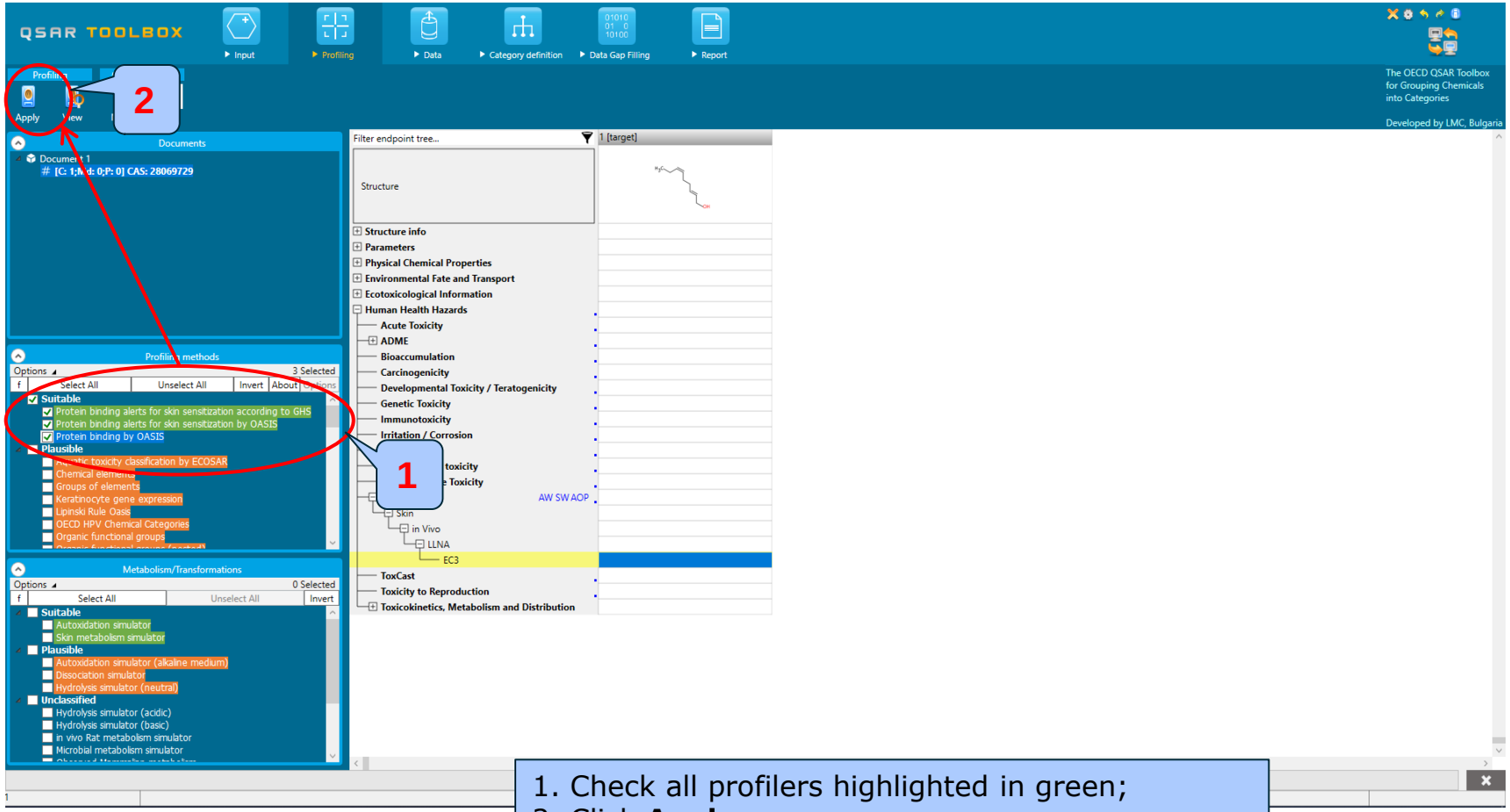
1. Go to the **Literature** tab in order to see detailed description of highlighted category (in this case "*Aldehydes*")

Profiling
Profiling the target chemical

- An option for colouring the profilers related to a given endpoint is implemented.
- The following profiling schemes are defined as relevant to the **Skin sensitization**:
  - *Protein binding by OASIS* – general mechanistic
  - *Protein binding alerts for skin sensitization according to GHS* – endpoint specific
  - *Protein binding alerts for skin sensitization by OASIS* - endpoint specific

Profiling

Profiling the target chemical



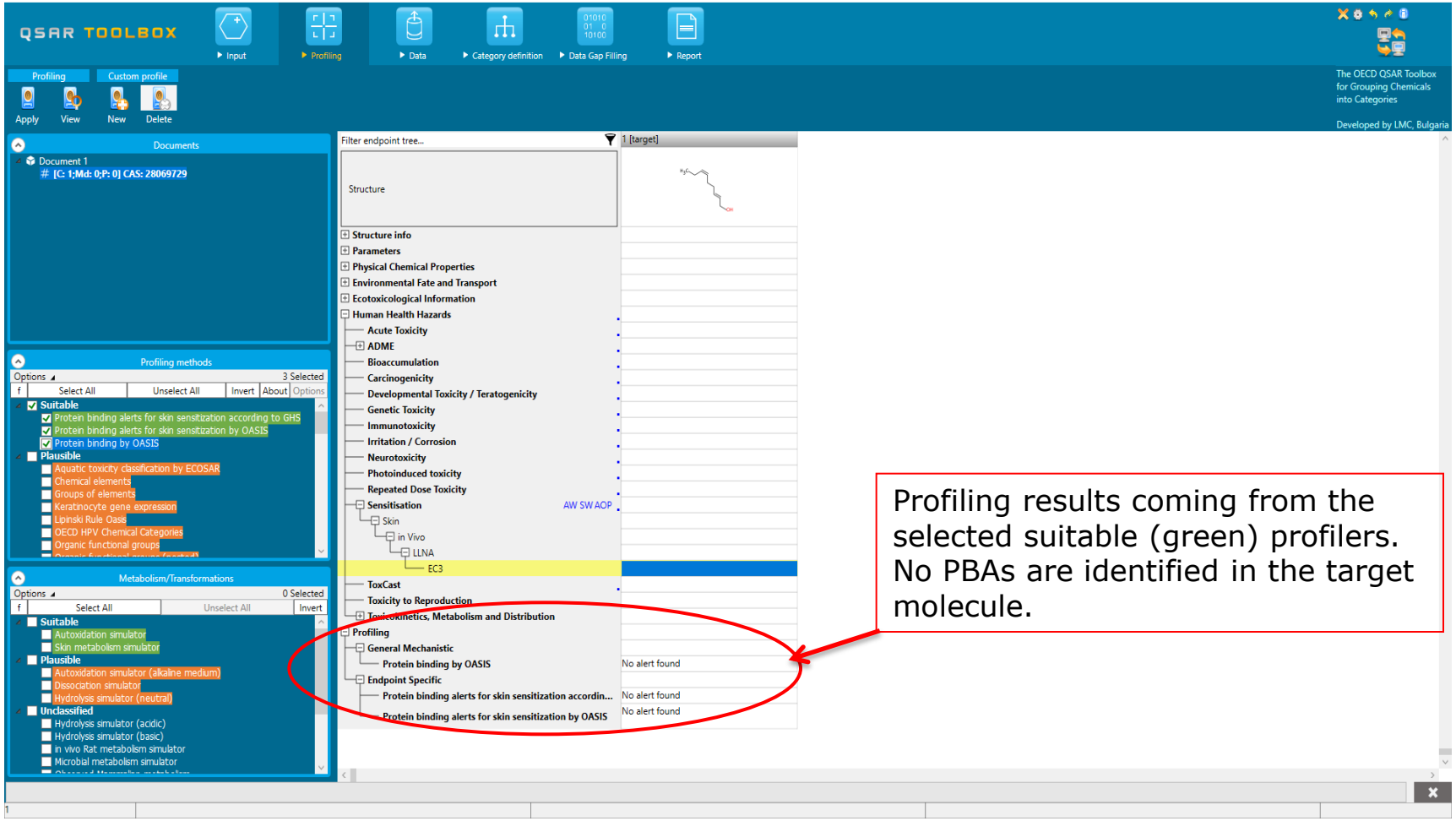
1. Check all profilers highlighted in green;
2. Click **Apply**.

Profiling
Profiling the target chemical

- The actual profiling will take up to several seconds depending on the number and type of profilers selected.
- The profiling result automatically appear as a last level of the endpoint tree (see next screenshot).
- In this case no protein binding alert (PBA) has been identified in the target substance (trans-2,cis-6-nonadienol).

Profiling

Profiling the target chemical



Profiling results coming from the selected suitable (green) profilers. No PBAs are identified in the target molecule.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - Profile the target chemical
  - **Collecting experimental data for the target chemical**

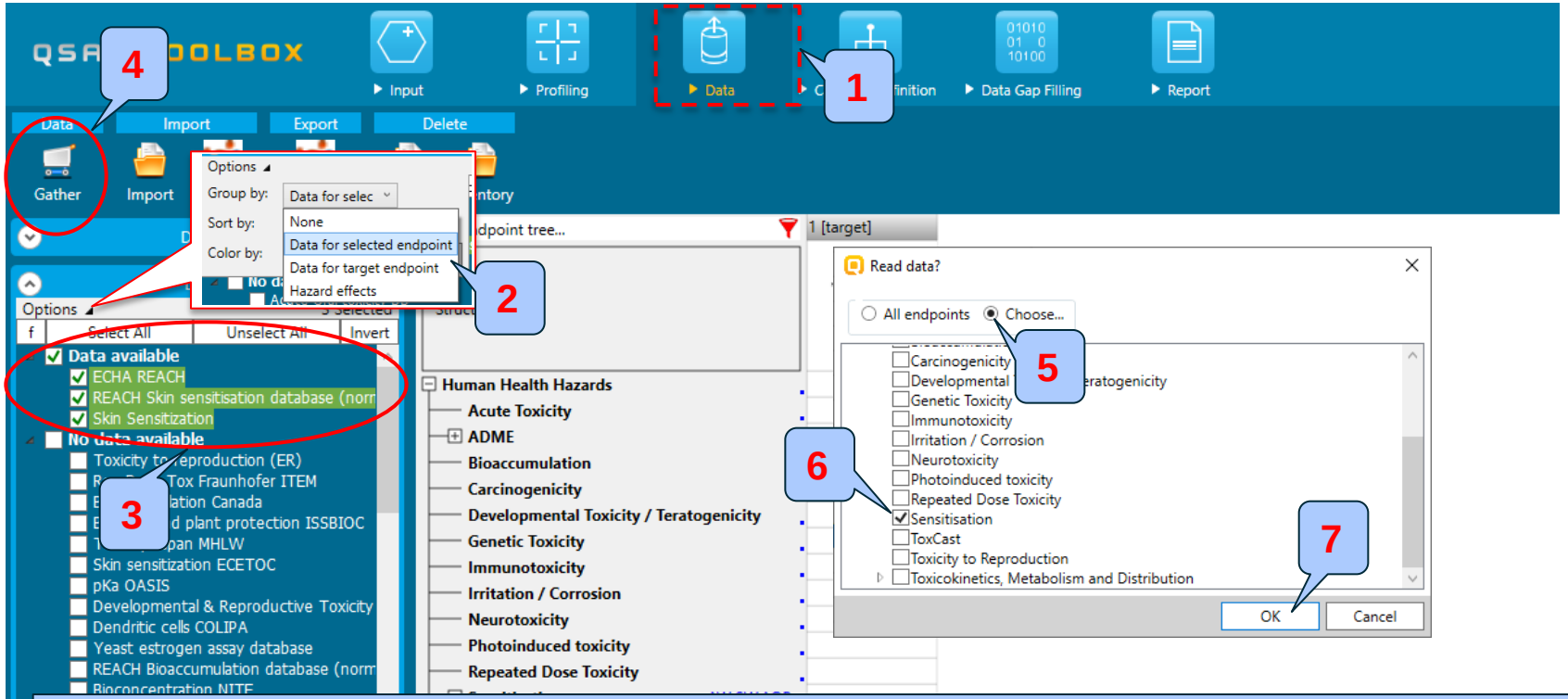
Data
Overview

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

Data
Case study

- In this example, we will limit the data gathering to a single toxicity endpoint (skin sensitization).
- In case of a defined endpoint, the databases containing data for this endpoint, will be highlighted in green. The user could select all or just some of them.

Data
Gather data



1. Go to the ***Data*** module;
2. Go to the *Options* and group the databases according to the *target endpoint*.;
3. Select all green highlighted databases (i.e. the databases containing data for the defined target endpoint);
4. Click **Gather**;
5. Select to **Choose** endpoint in the pop-up menu;
6. Select **Sensitization**;
7. Confirm by **OK**.

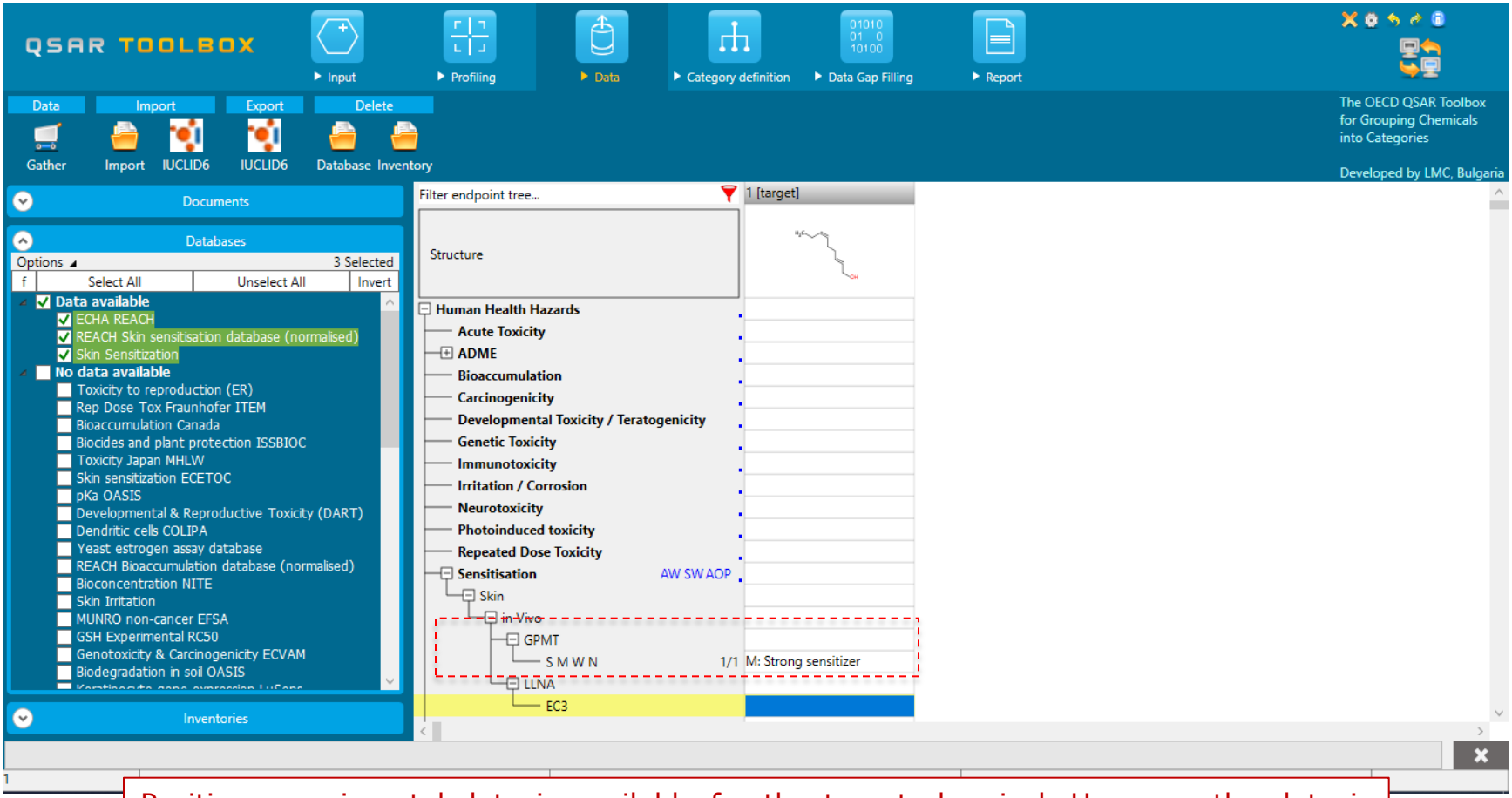
Data
Gather data

- Toxicity information on the target chemical is electronically collected from the selected dataset(s).

- It should be kept in mind that the search for analogues and available data will be performed only among the selected databases.

- In this example, Positive experimental data is available for the target chemical (see next screen shots)

Data
Available experimental data



Positive experimental data is available for the target chemical. However, the data is associated with the Guinea pig maximization test (GPMT). No data for EC3 is available for the target chemical in the selected databases.

Recap

- In the first module (*Input*) we:

  - entered the target chemical by CAS;
  -selected the chemical with High reliability of the CAS/SMILES correlation;
  - defined the target endpoint.

- In the second module (*Profiling*) we:

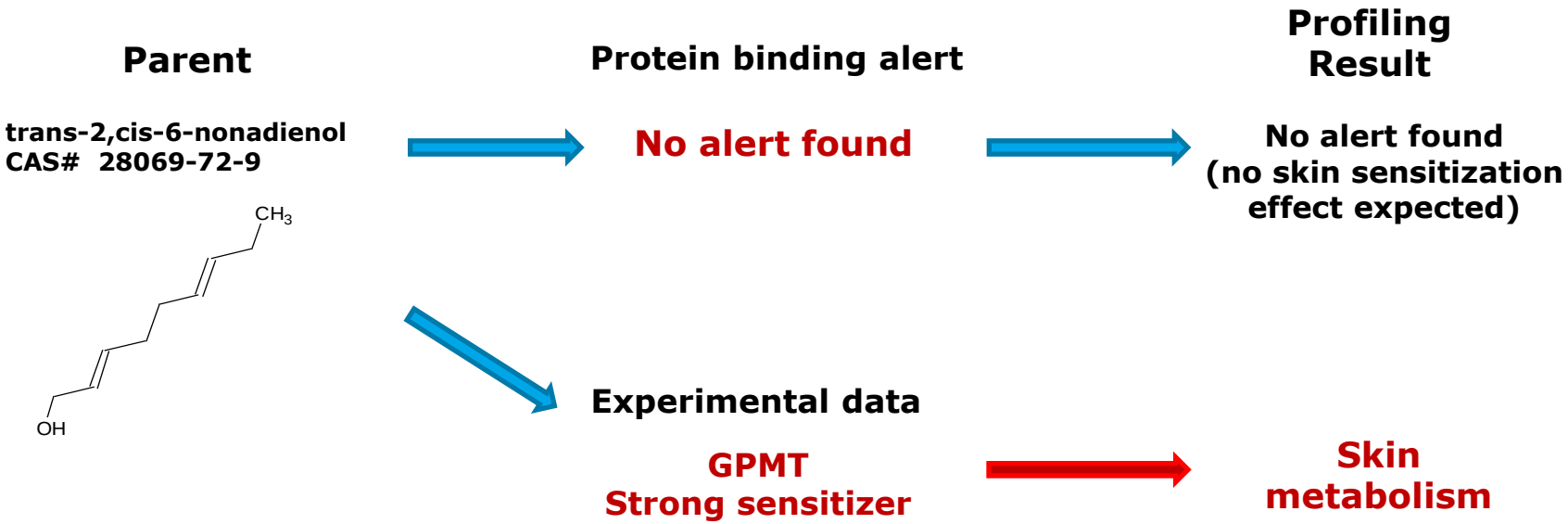
  - selected the profilers related to the target endpoint (highlighted in green);
  - saw that there is no PBA in target molecule.

- In the third module (*Data*), we:

  - selected the databases containing data for the defined target endpoint (the databases in green)
  - saw that there is no experimental data for the defined endpoint (EC3), however positive data for GPMT is available.

Recap

- The positive experimental data could be due to skin metabolism.



- The study continues with accounting for skin metabolism of target chemical (see next slides).

Outlook

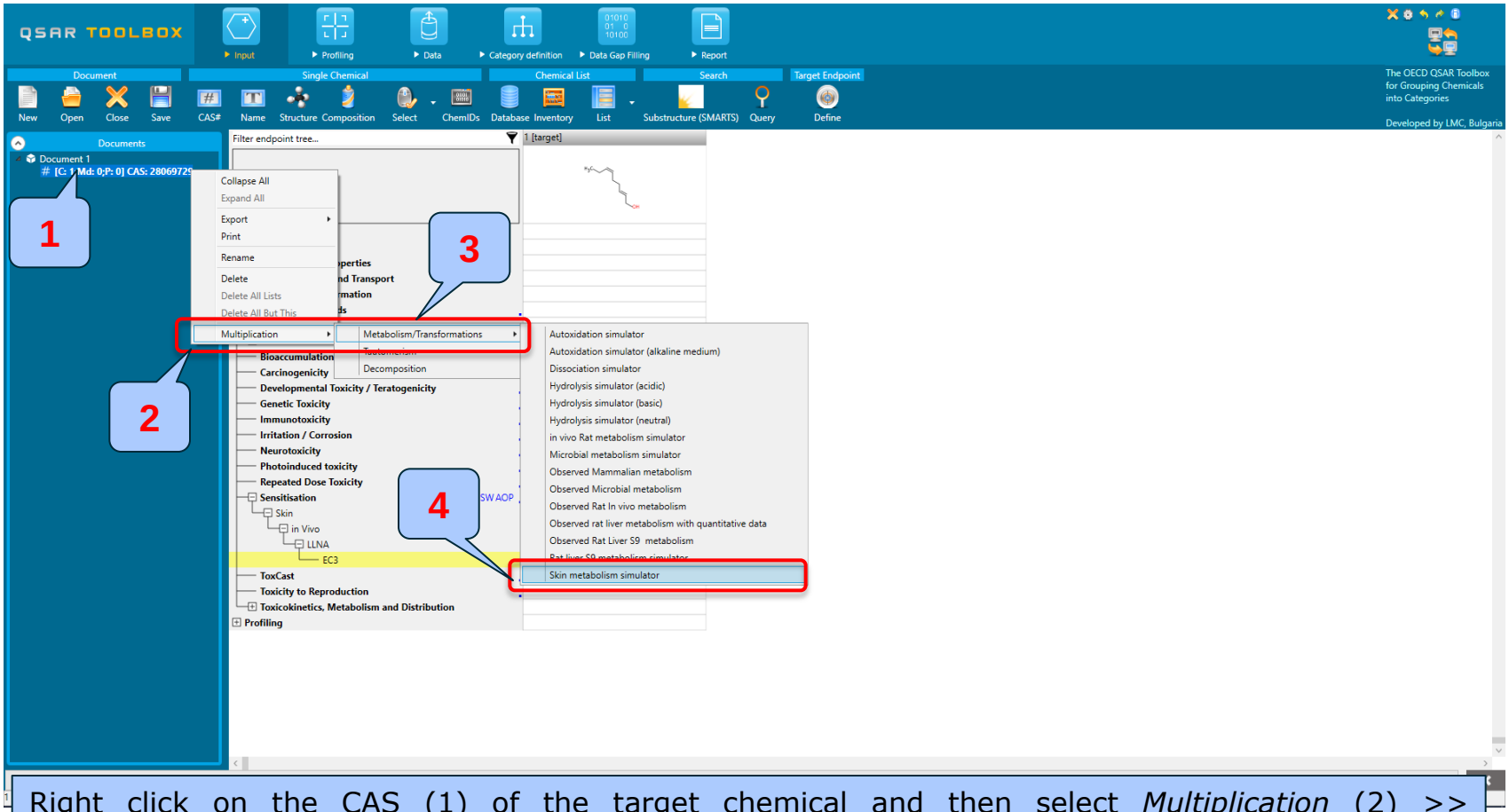
- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - Profile the target chemical
  - Collecting experimental data for the target chemical
  - **Handling of skin metabolism of target chemical**
    - **Multiplication of the target chemical**

Handling of skin metabolism of the target chemical

- In order to explore the skin metabolites of the target chemical, we need first to multiply it.
- Multiplication happens in the **Input module** and it represents a generation and visualization of all metabolites generated by a selected metabolic simulator. The generated metabolites appear on the data matrix along with the target as well as in a tree-like form.
- In the current example we will use the Skin metabolism simulator in order to multiply the target chemical.

Handling of skin metabolism of the target chemical

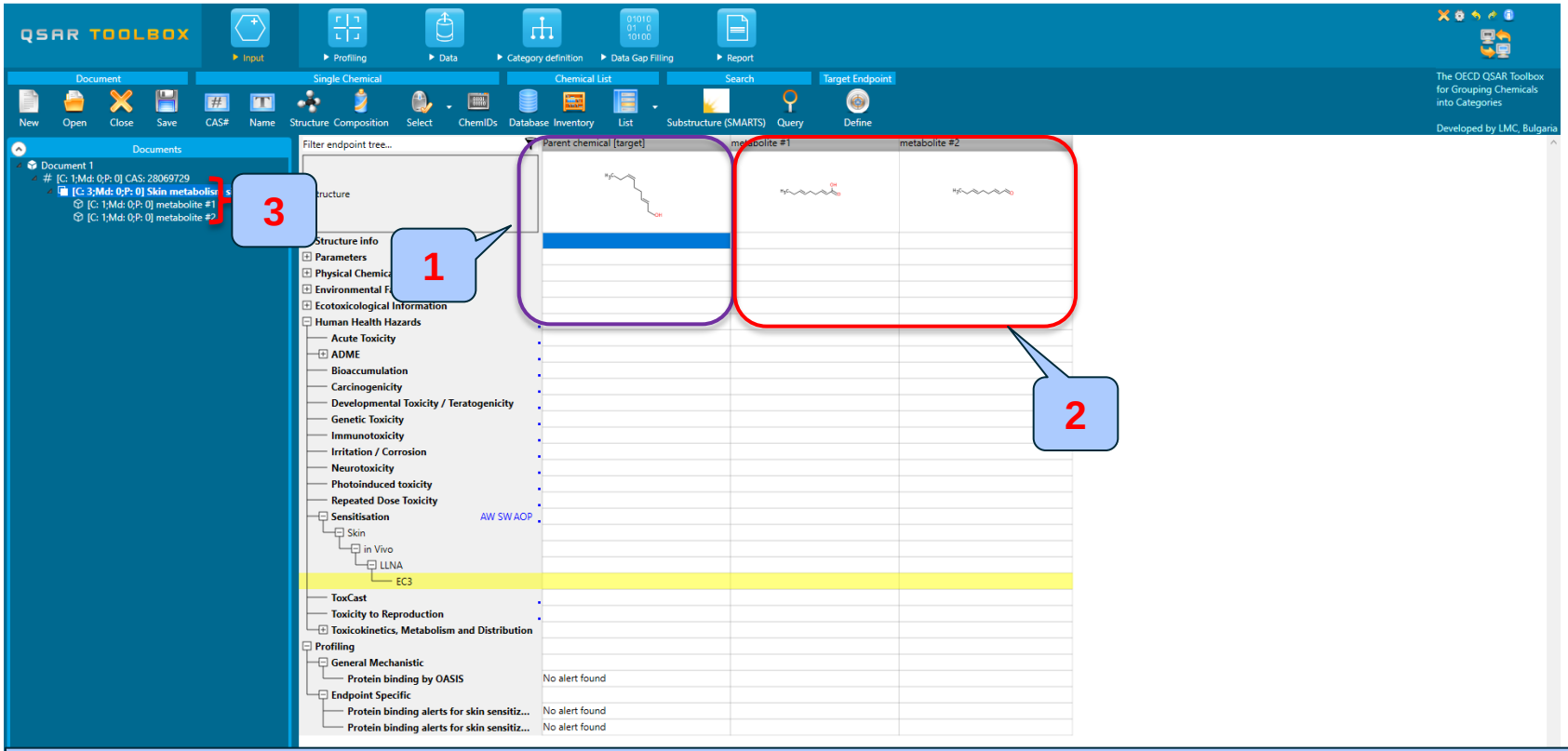
Multiplication of target chemical



Right click on the CAS (1) of the target chemical and then select *Multiplication* (2) >> *Metabolism/Transformations* (3) >> *Skin metabolism simulator* (4).

Handling of skin metabolism of the target chemical

Multiplication of target chemical



1. The first chemical is the **parent;**
2. Next to the parent are placed **Metabolites** of the parent chemical generated by the Skin metabolism simulator;
3. Generated metabolites appear in a tree-like form

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - Profile the target chemical
  - Collecting experimental data for the target chemical
  - **Handling of skin metabolism of the target chemical**
    - **Profiling the set of metabolites**

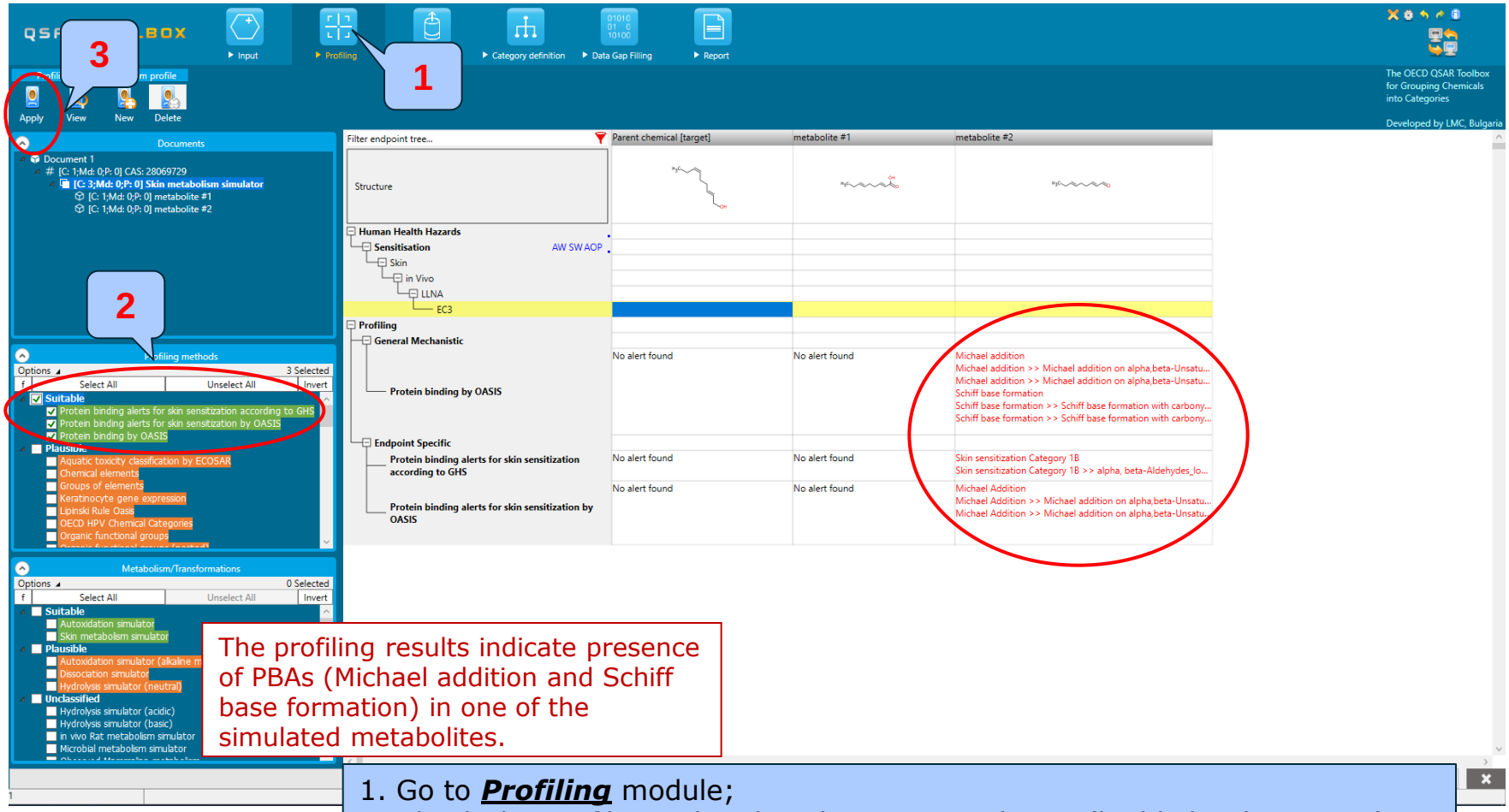
Handling of skin metabolism of the target chemical

Profiling the set of metabolites

- In this step we will check for available PBAs in the generated skin metabolites.
- As the metabolites are already on the data matrix, we could directly profile them with the relevant protein binding profiles.

Handling of skin metabolism of the target chemical

Profiling the set of metabolites



The profiling results indicate presence of PBAs (Michael addition and Schiff base formation) in one of the simulated metabolites.

1. Go to ***Profiling*** module;
2. Check the profilers related to the target endpoint (highlighted in green);
3. Click **Apply**

Handling of skin metabolism of the target chemical

Analysis of the profiling results

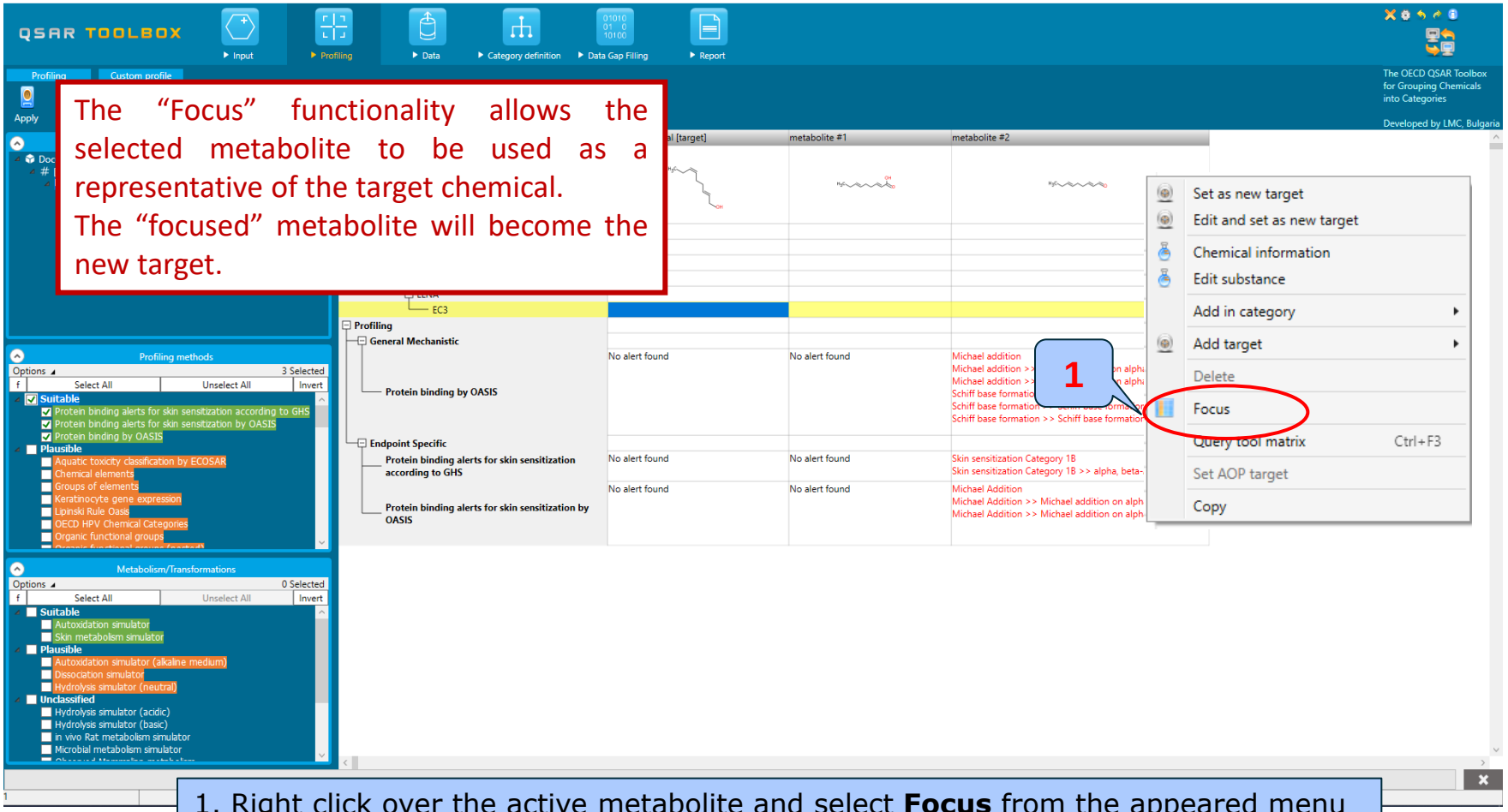
- The profiling results indicate no PBA is identified in the target molecule
- PBAs are identified in one of the two simulated skin metabolites. The identified PBAs are associated with two mechanisms of interaction with the skin proteins:
  1) Schiff base formation with carbonyl compounds
  2) Michael addition on alpha,beta-unsaturated carbonyl compounds
- This active metabolite (having PBAs) will be used for further read across analysis
- The next two parts of the exercise will be focused on the reactive metabolite and searching of similar analogues of this metabolite.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - Profile the target chemical
  - Collecting experimental data for the target chemical
  - **Handling of skin metabolism of the target chemical**
    - **Focus on the active metabolite**

Handling of skin metabolism of the target chemical

Focus of active metabolite



1. Right click over the active metabolite and select **Focus** from the appeared menu

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - Profile the target chemical
  - Collecting experimental data for the target chemical
  - **Handling of skin metabolism of the target chemical**
    - **Defining category for the active metabolite**

Category Definition
Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of "similarity" so that within a category data gaps can be filled by read-across.

Category Definition
Suitable Categorization/Assessment Phases

Suitable categorization phases:

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

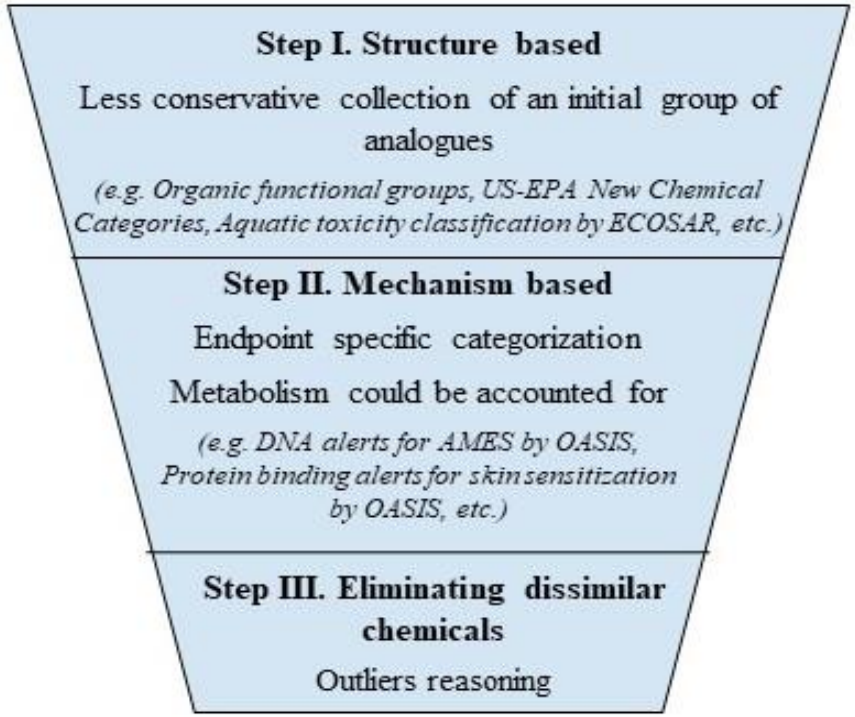
Performing categorization:

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 the suitable categorization phases is shown on next slide

Category Definition

Suitable Categorization/Assessment Phases



Broad grouping
Endpoint Non-specific

Subcategorization
Endpoint Specific

Subcategorization
Endpoint Non-specific

Note: As long as an acceptable level of structural and mechanistic similarity is achieved, it is not mandatory to follow all the stages described in the order given above; they can be executed differently or even skipped.
For example, in case the target chemical or any of its metabolites interact with biomacromolecules via a clearly defined mechanism relevant for the endpoint to predict, Stage I could be skipped and Stage II to be used for the primary categorization step.

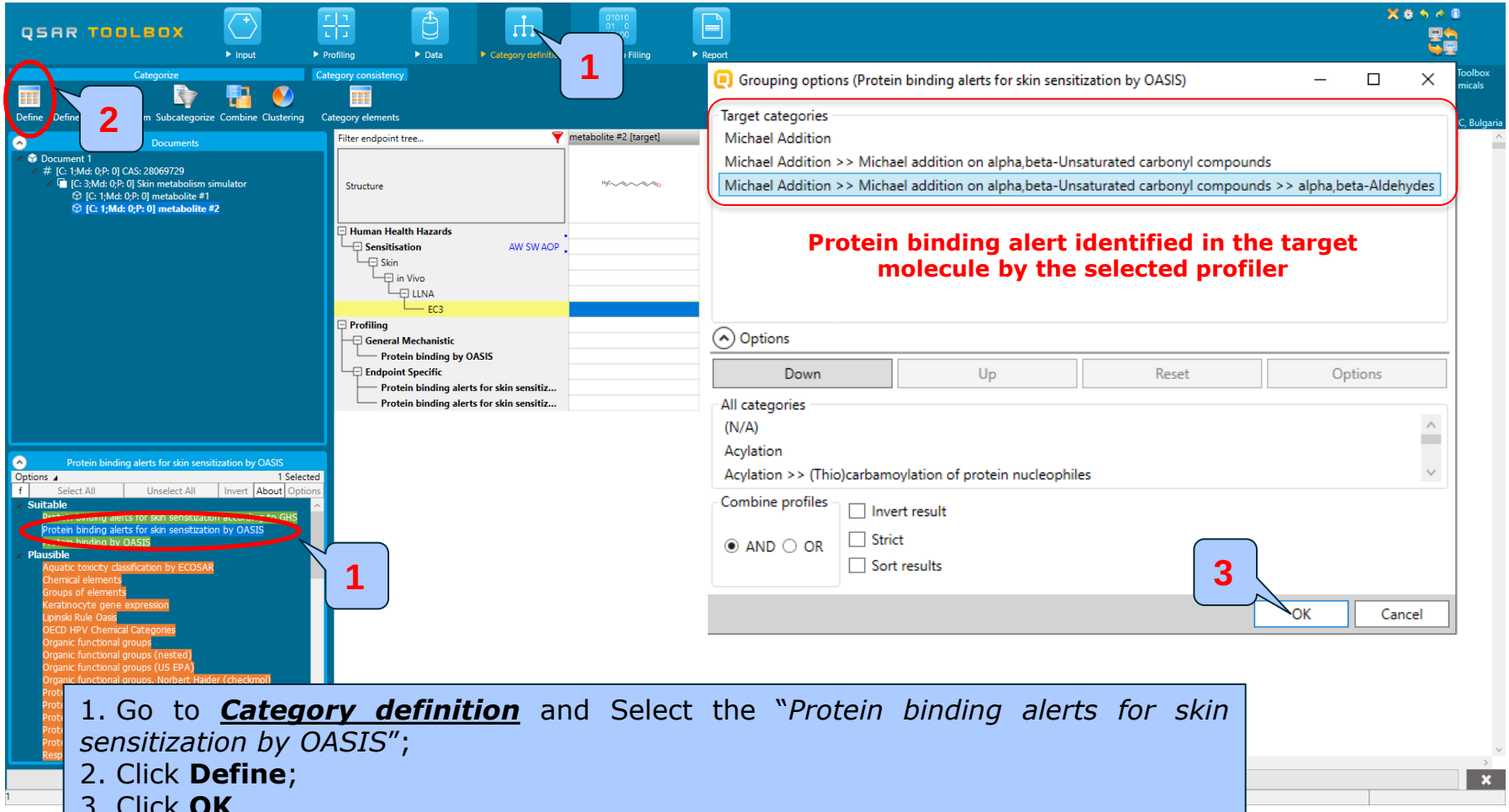
Handling of skin metabolism of the target chemical

Defining category for the active metabolite

- In this exersice PBA (Michael Addition >> Michael addition on alpha,beta-Unsaturated carbonyl compounds >> alpha,beta-Aldehydes) is identified in the active metabolite.
- The identified PBA will be used for searching of analogues acting via the same mechanism (Stage II)
- Searching for similar analogues is accomplished in the selected databases (see slide 32)

Handling of skin metabolism of the target chemical

Defining category for the active metabolite



1. Go to ***Category definition*** and Select the "*Protein binding alerts for skin sensitization by OASIS*";
2. Click **Define**;
3. Click **OK**.

Handling of skin metabolism of the target chemical

Defining category for the active metabolite

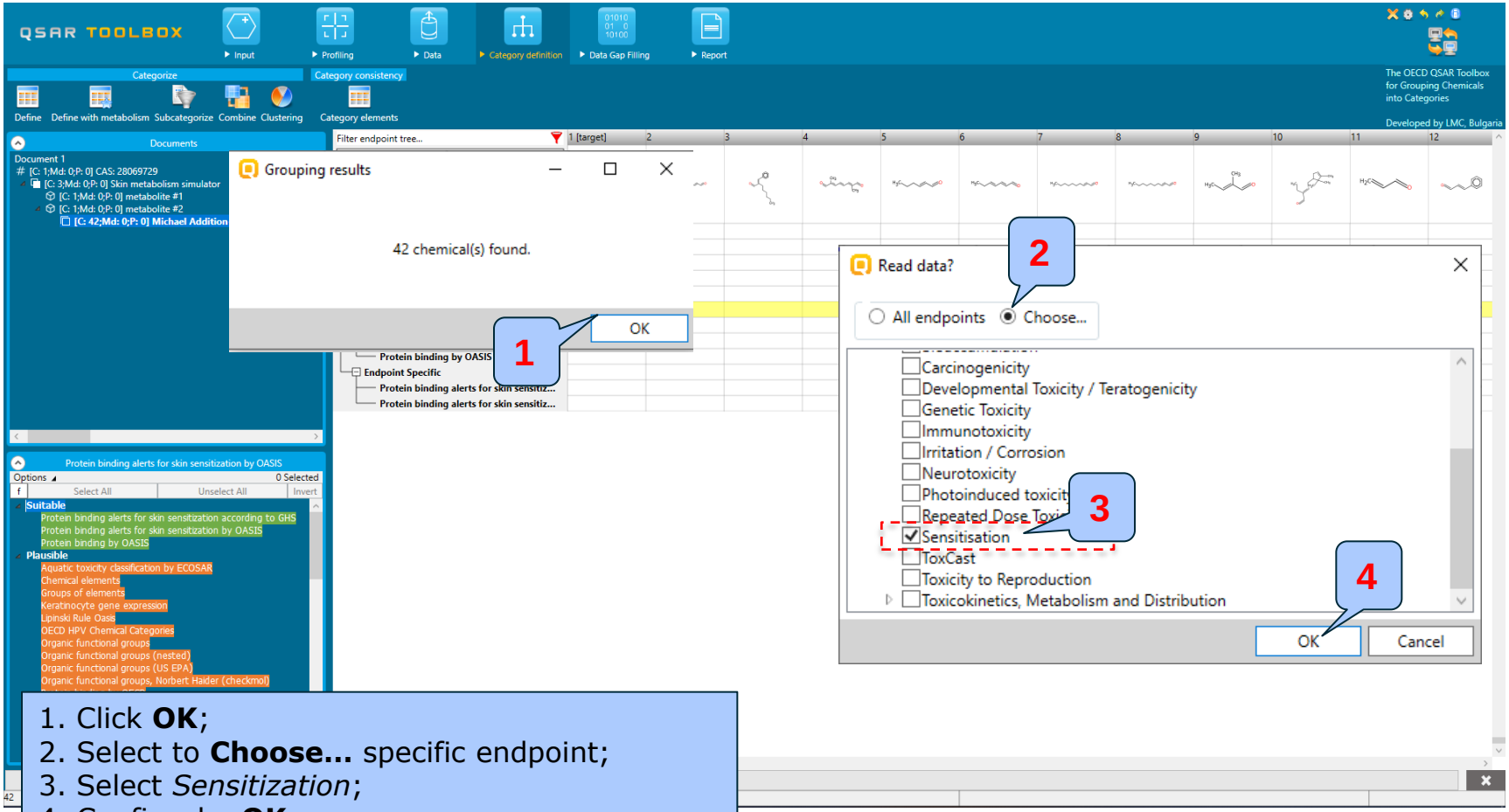
- The Toolbox now identifies all chemicals corresponding to the *alpha, beta-Aldehydes* alert (according to the *Protein binding alerts for skin sensitization by OASIS* profiler) listed in the skin sensitization databases.

- 42 analogues including the target chemical are identified in the selected databases (see next slide). They form a mechanistic category named "*alpha, beta-Aldehydes*", which will be used for further data gap filling.

- The experimental data for analogues in the category appears on data matrix

Handling of skin metabolism of the target chemical

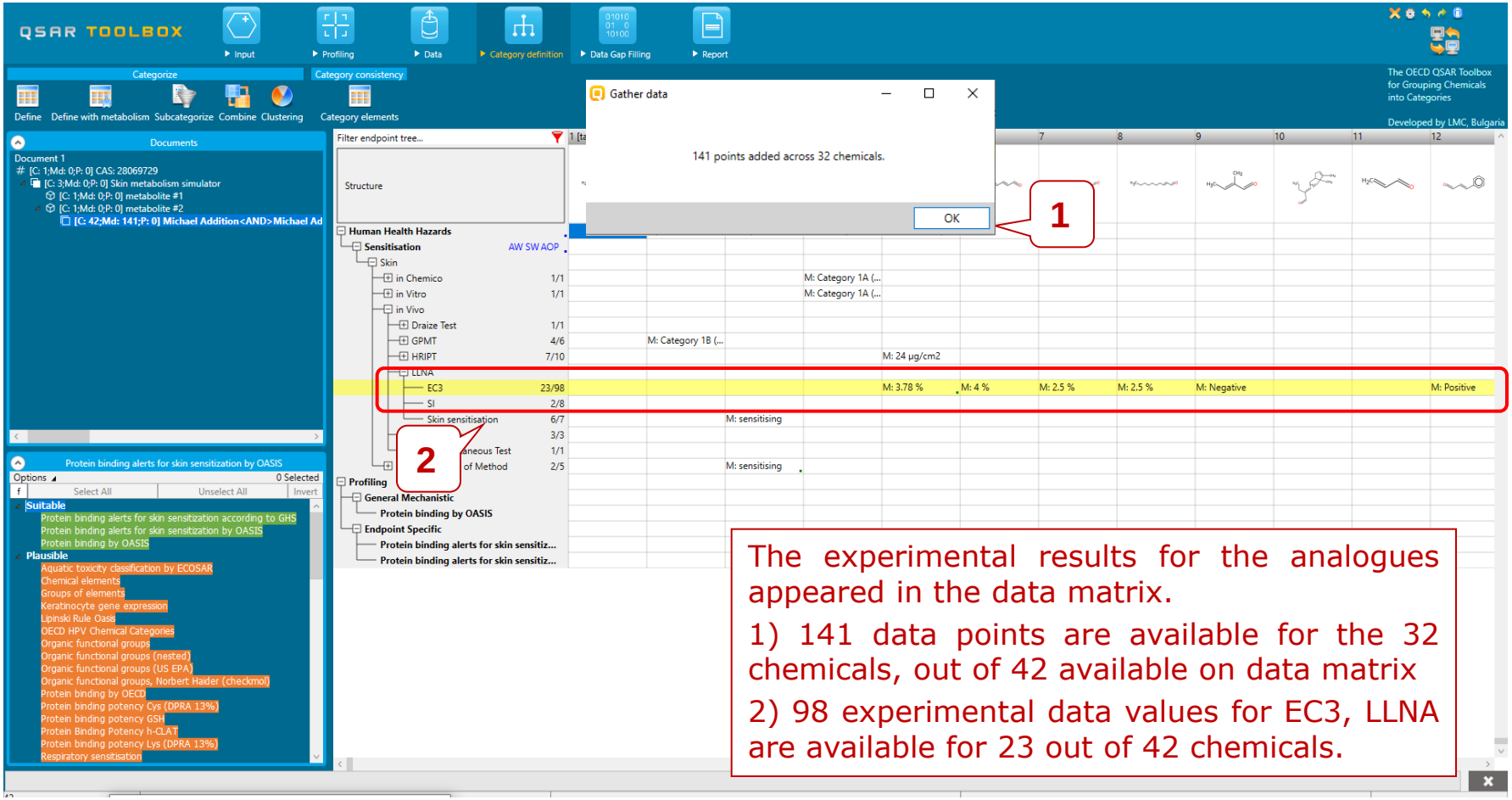
Defining category for the active metabolite



1. Click **OK**;
2. Select to **Choose...** specific endpoint;
3. Select *Sensitization*;
4. Confirm by **OK**.

Handling of skin metabolism of the target chemical

Defining category for the active metabolite



The experimental results for the analogues appeared in the data matrix.

1) 141 data points are available for the 32 chemicals, out of 42 available on data matrix

2) 98 experimental data values for EC3, LLNA are available for 23 out of 42 chemicals.

Recap

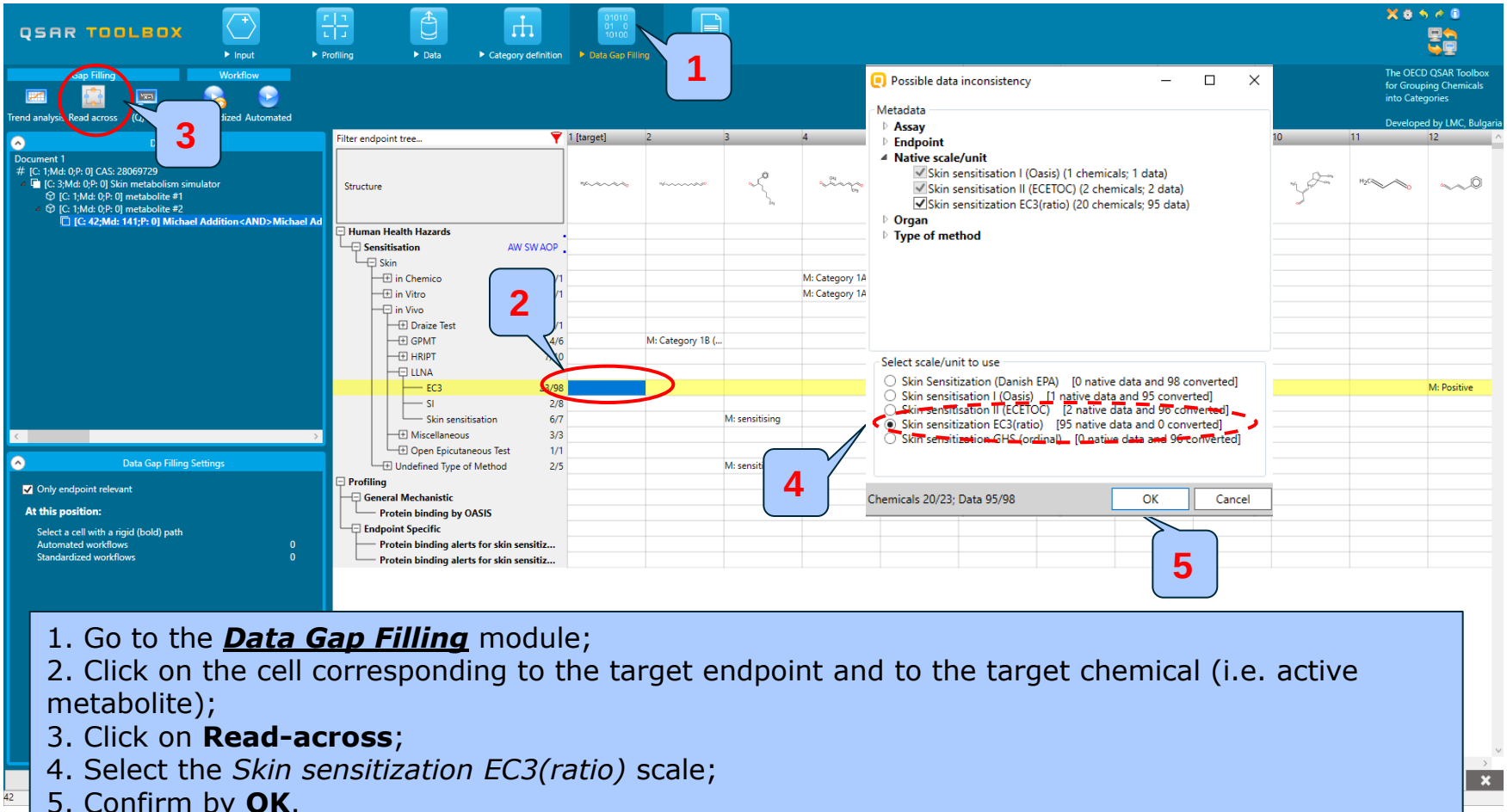
- In this case the *alpha, beta-Aldehydes* category of the *Protein binding alerts for skin sensitization by OASIS* profiler is used for categorization purposes.
- The defined category consists of 41 analogues along with the target chemical
- The available experimental data for these 41 analogues have been collected from the endpoint-relevant skin sensitization databases.
- Experimental data values for the target endpoint (EC3, LLNA) are available for 23 out of 41 analogues.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - Profile the target chemical
  - Collecting experimental data for the target chemical
  - **Handling of skin metabolism of the target chemical**
    - **Filling data gap of the active metabolite**

Handling of skin metabolism of the target chemical

Filling data gap of the active metabolite



1. Go to the ***Data Gap Filling*** module;
2. Click on the cell corresponding to the target endpoint and to the target chemical (i.e. active metabolite);
3. Click on **Read-across**;
4. Select the *Skin sensitization EC3(ratio)* scale;
5. Confirm by **OK**.

Sidebar on the data scales

- The available data in the databases could be represented in different scales depending on the database` donators.
- The skin sensitisation data could be: 1) categorical (for example: positive; negative; weak sensitizer; strong sensitizer, etc.) or 2) numerical (e.g. EC3 values in percentage).
- The main purpose of the scales is to unify all data available in the Toolbox databases for a certain endpoint by making conversions between them.
- The default scale for Skin Sensitisation is "*Skin Sensitisation ECETOC*". It converts all skin data into dichotomous scale: Positive/Negative.
- In the current example we will use the potency scale - *Skin sensitization EC3(ratio)* scale in order to see the final prediction in %.

Handling of skin metabolism of the target chemical

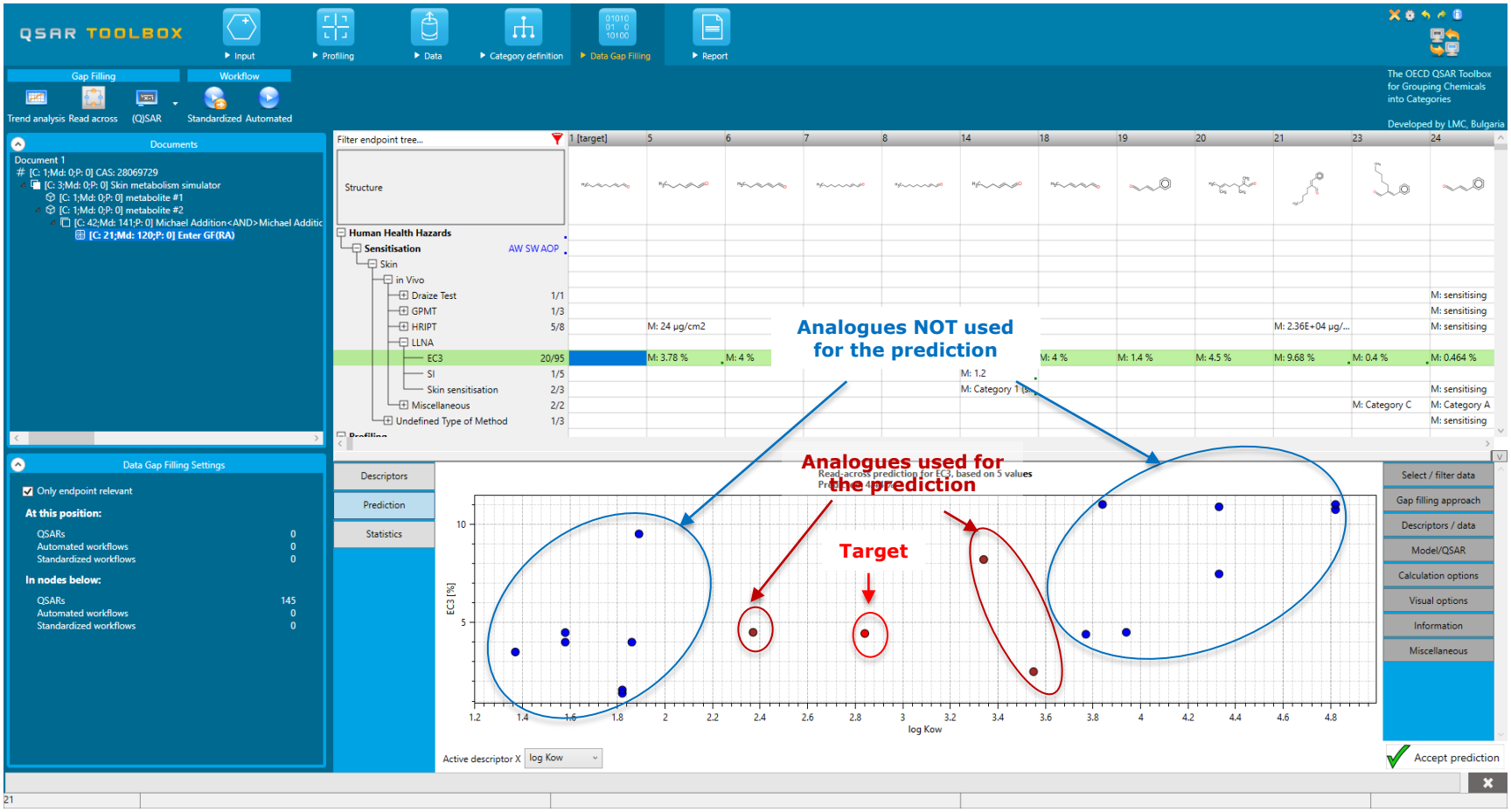
Filling data gap of the active metabolite

- The resulting plot is experimental results of all analogues (Y axis) according to a descriptor (X axis) with the default descriptor of log Kow (see next screen shot).
- The **RED** dot represents predicted result for the target chemical;
- The **BROWN** dots represent the experimental results available for the analogues that are used for the read-across. By default the closest five analogues are used for the prediction;
- The **BLUE** dot represents the experimental results available for the analogues that are not used for read-across.

Note: In some case one brown/blue point could correspond to more than one chemical. This holds for structures (e.g. different isomers) having the same X axis descriptor value (e.g. logKow) and the same Y axis descriptor value (e.g. positive).

Handling of skin metabolism of the target chemical

Filling data gap of the active metabolite



Filling data gap of the active metabolite
Subcategorization

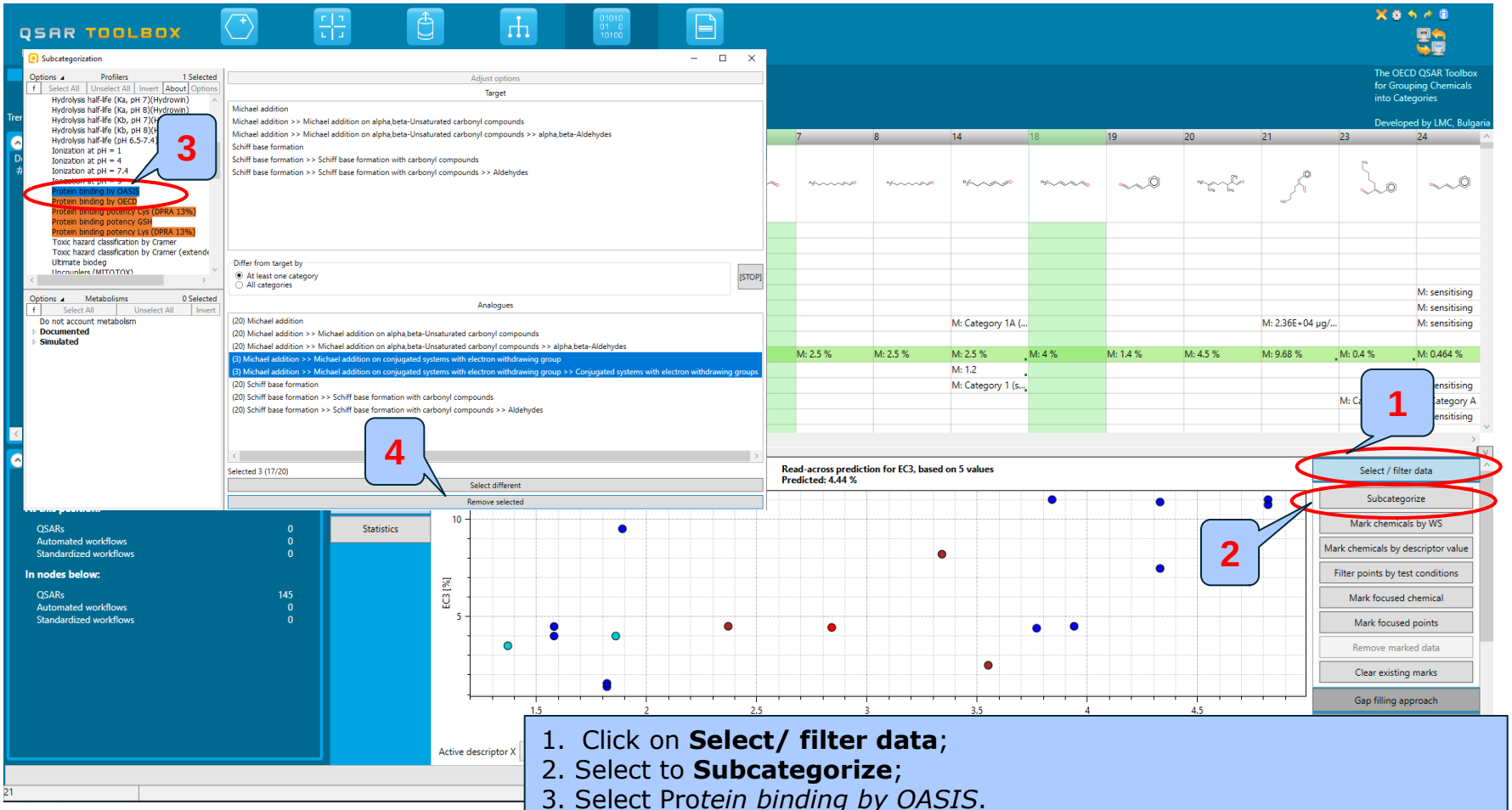
- The category could be further narrowed-down by using of some of the endpoint-specific or structure-based profilers. This aim to reduce the uncertainty among the category members.

- The following profilers will be used for subcategorization:
  - Protein binding by OASIS
  - Organic functional groups

- These steps are summarized in the next screen shots.

Filling data gap of the active metabolite

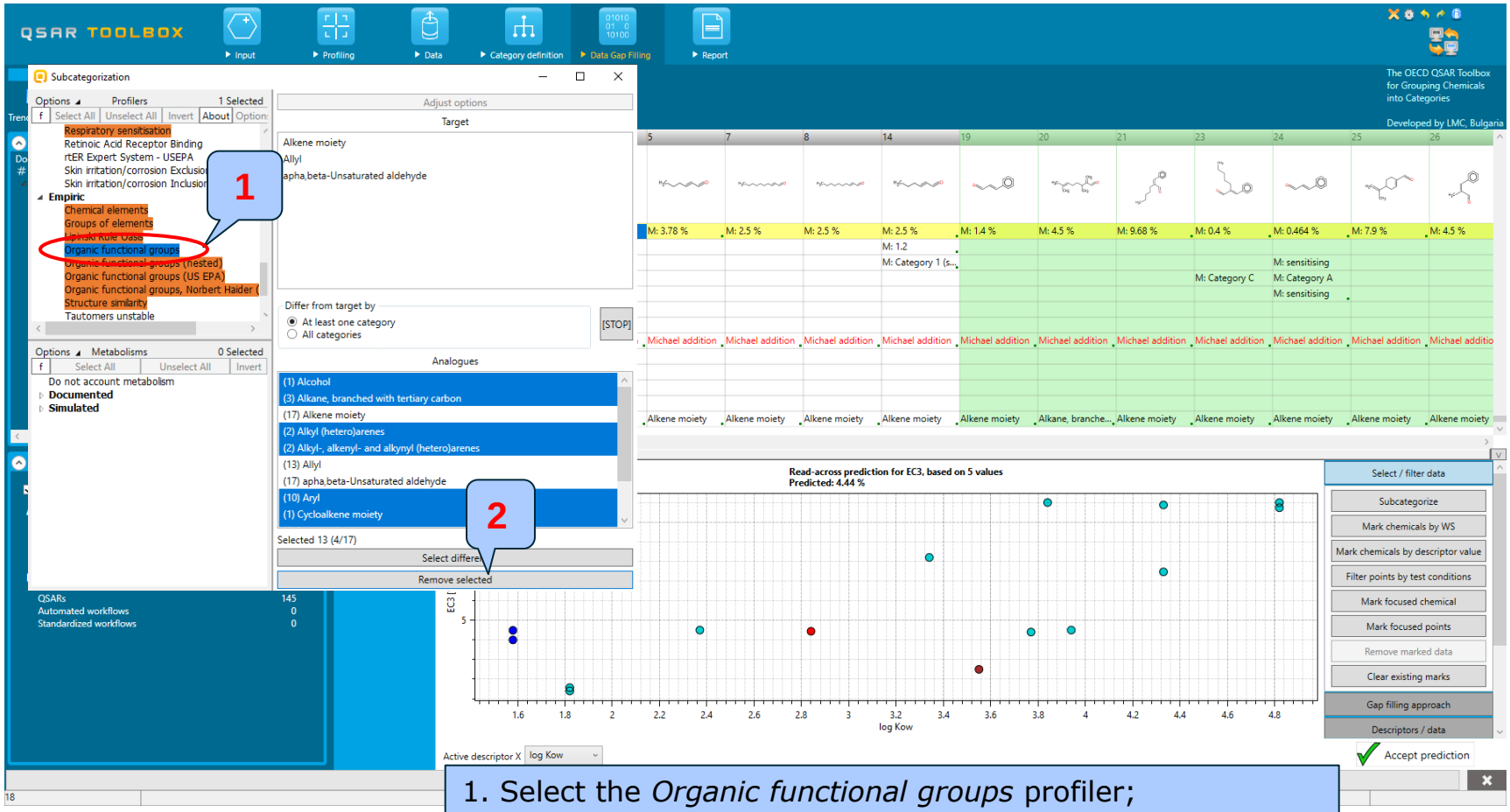
Subcategorization 1: Protein binding by OASIS



1. Click on **Select/ filter data**;
2. Select to **Subcategorize**;
3. Select Pro*tein binding by OASIS*.
4. Eliminate the dissimilar analogues with respect to the selected profiler.

Filling data gap of the active metabolite

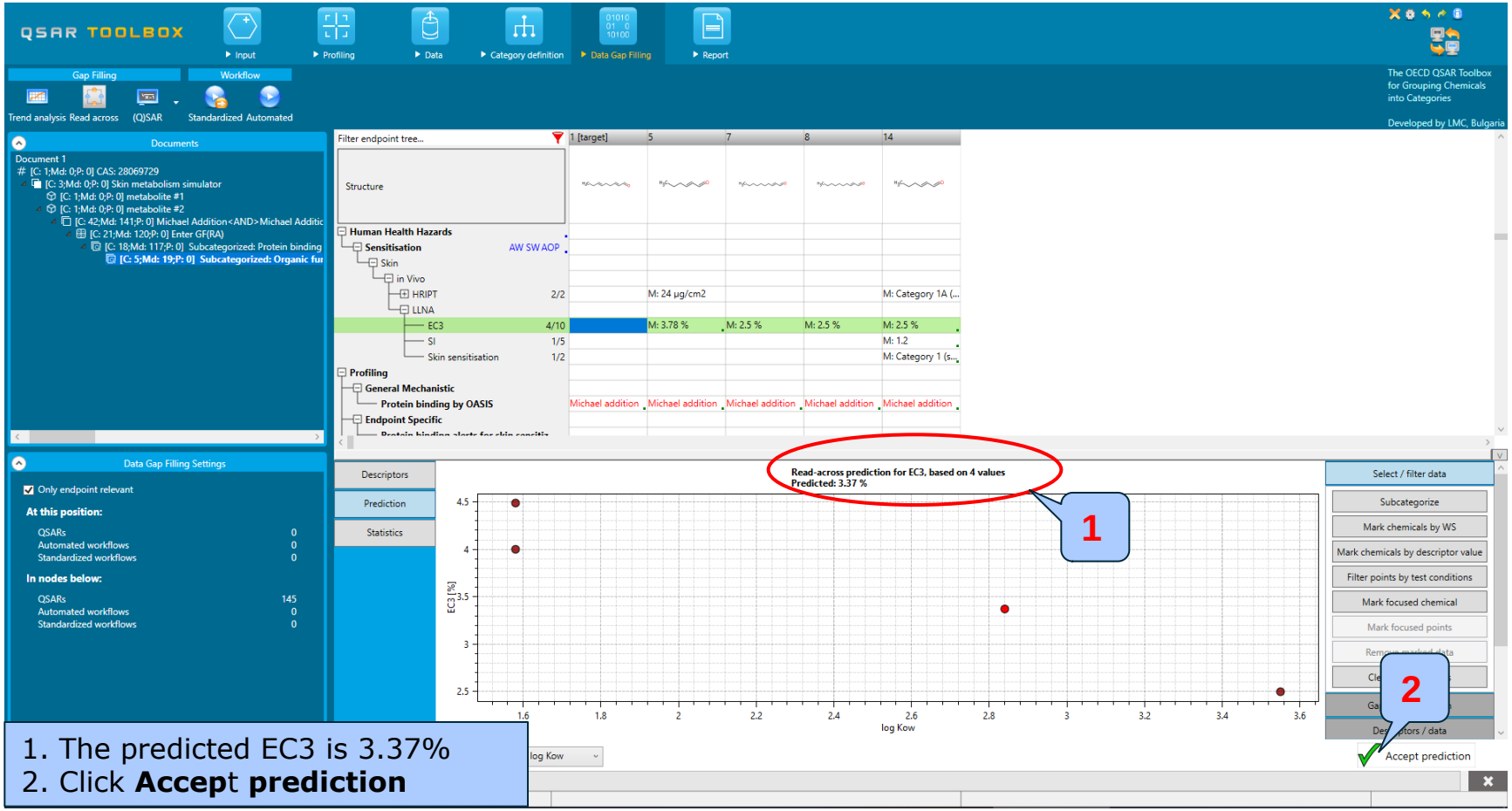
Subcategorization 2: Organic functional groups



1. Select the *Organic functional groups* profiler;
2. Remove the dissimilar chemicals

Filling data gap of the active metabolite

Results after the subcategorizations



1. The predicted EC3 is 3.37%
2. Click **Accept prediction**

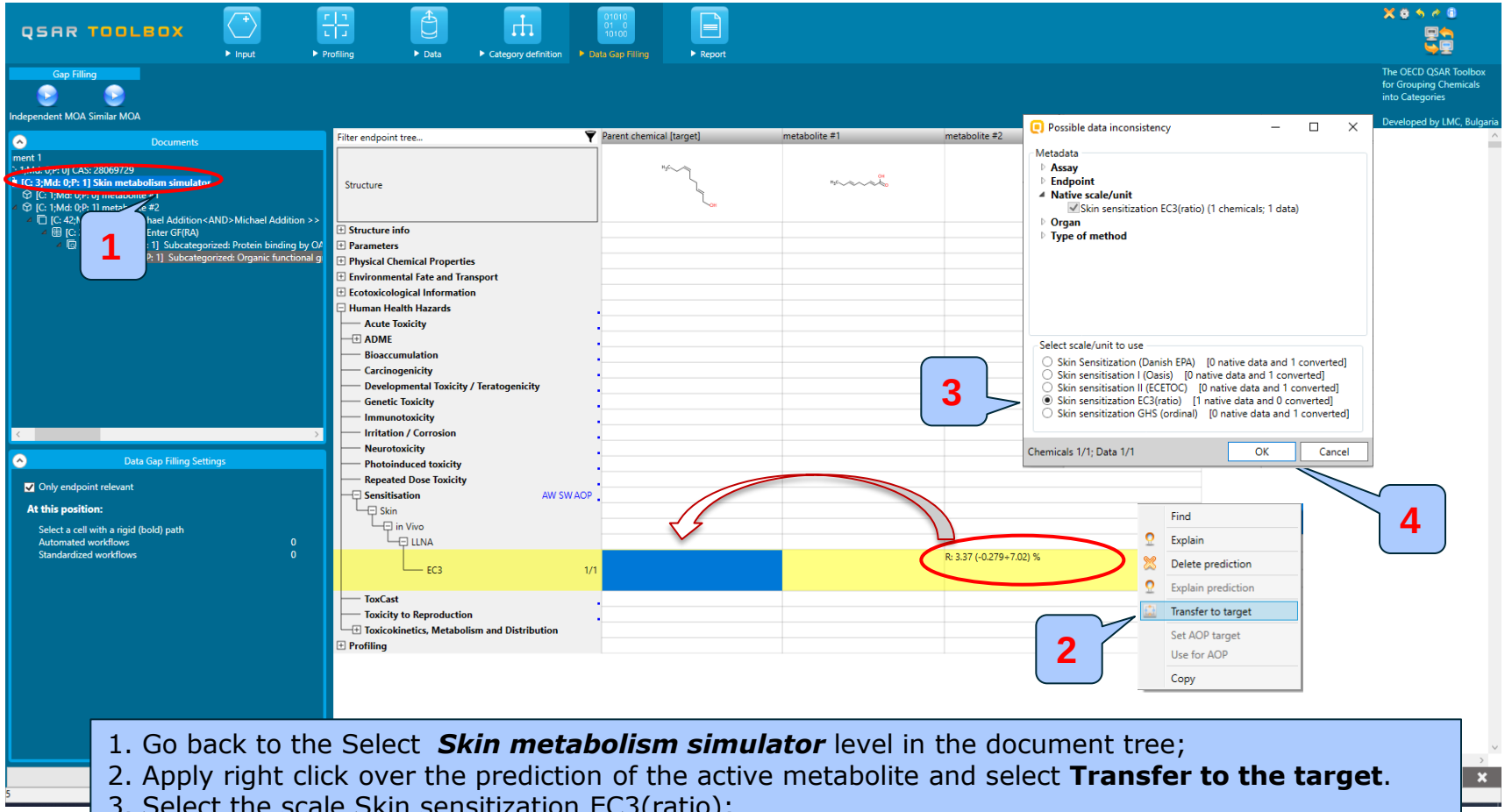
Data gap filling for active metabolite
Interpreting Read-across

- In this example, all analogues have the same PBAs.
- All analogues have positive skin sensitization experimental data.
- The sensitising potential of the target chemical is predicted similar to those of the used analogues.
- The obtained prediction is further transferred to the parent chemical.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - Profile the target chemical
  - Collecting experimental data for the target chemical
  - Handling of skin metabolism of the target chemical
  - **Transferring the prediction of active metabolite to the parent**

Transferring the prediction of active metabolite to the parent



1. Go back to the Select ***Skin metabolism simulator*** level in the document tree;
2. Apply right click over the prediction of the active metabolite and select **Transfer to the target**.
3. Select the scale Skin sensitization EC3(ratio);
4. Confirm by **OK**.

The data will be automatically transferred from the metabolite to the parent.

Recap

- The target chemical **trans-2,cis-6-nonadienol** has been entered into the system.
- It has been profiled by Protein binding profilers; no protein binding alert has been found for target chemical.
- Positive experimental data has been retrieved for target chemical.
- Skin metabolism of target chemical is investigated. One of simulated skin metabolites has protein binding alerts (**α,β-unsaturated aldehydes**).
- This metabolites is used for further read across analysis.
- No experimental data for the selected metabolite has been found, so the category of similar analogues has been defined.
- The initial group of analogues was defined by *Protein binding alerts for skin sensitization by OASIS* profiler.
- 41 analogues including the target chemical are identified; they form a mechanistic category ":**α,β-unsaturated aldehydes**", which was used for gap filling.
- Read-across is used for filling the data gap of the active metabolite.
- The initial category has been refined by some endpoint-specific and structure-related subcategorizations.
- The active metabolite has been predicted LLNA positive (EC3=3.37%).
- The positive prediction for reactive metabolite has been transferred to the parent chemical.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input target chemical
  - Profile the target chemical
  - Collecting experimental data for the target chemical
  - Handling of skin metabolism of the target chemical
  - Transferring the prediction of active metabolite to the parent
  - **Reporting the prediction**

Report
Overview

- The Report module allows you to generate a report on the predictions performed within the Toolbox.
- This module contains a predefined report template with automatically populated sections as well as manually editable sections, where the users could add some additional/custom information.
- The Prediction report generates three files:

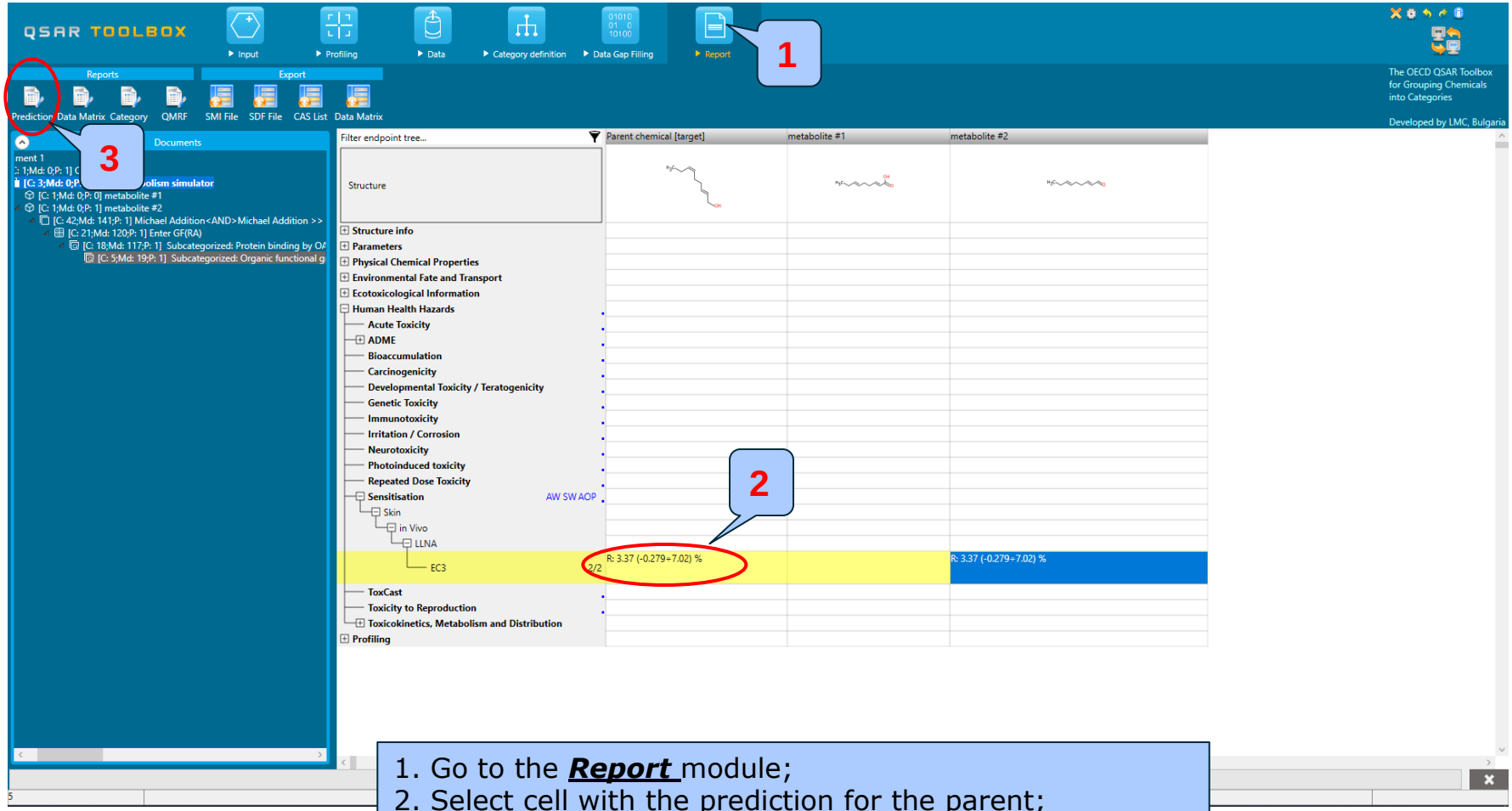
  - *Prediction report* - a PDF file containing information for the target and how the prediction is obtained.

  - *Category report* - a PDF file containing information for the consistency of the final category (target plus used analogues)

  - *Data matrix* - a MS Excel file containing chemicals used for prediction along with their data for selected parameters, profiles and endpoint tree positions.

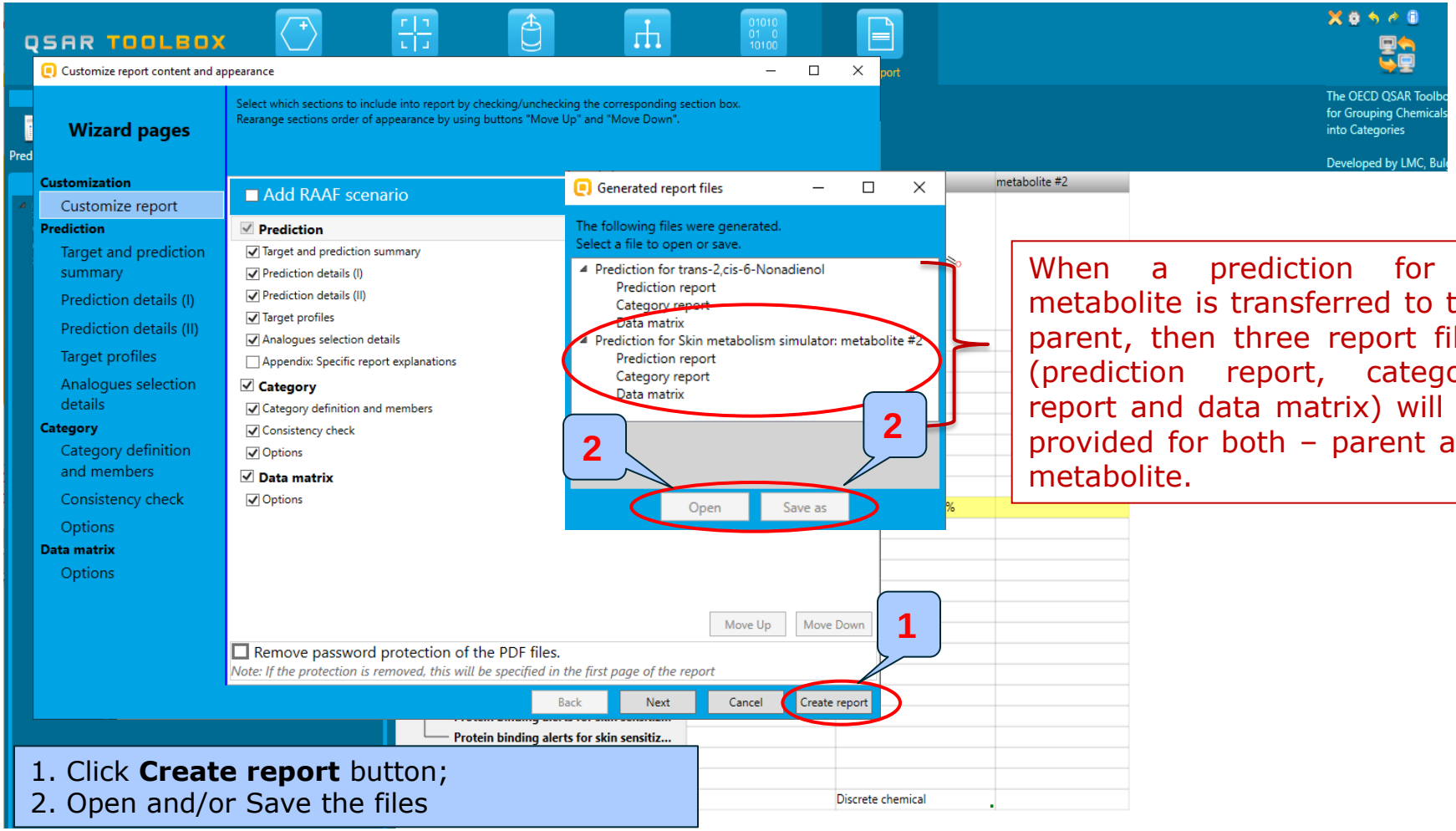
- The generated reports can then be saved or just opened.

Report



1. Go to the ***Report*** module;
2. Select cell with the prediction for the parent;
3. Click on the **Prediction** button

Report



When a prediction for a metabolite is transferred to the parent, then three report files (prediction report, category report and data matrix) will be provided for both – parent and metabolite.

1. Click **Create report** button;
2. Open and/or Save the files

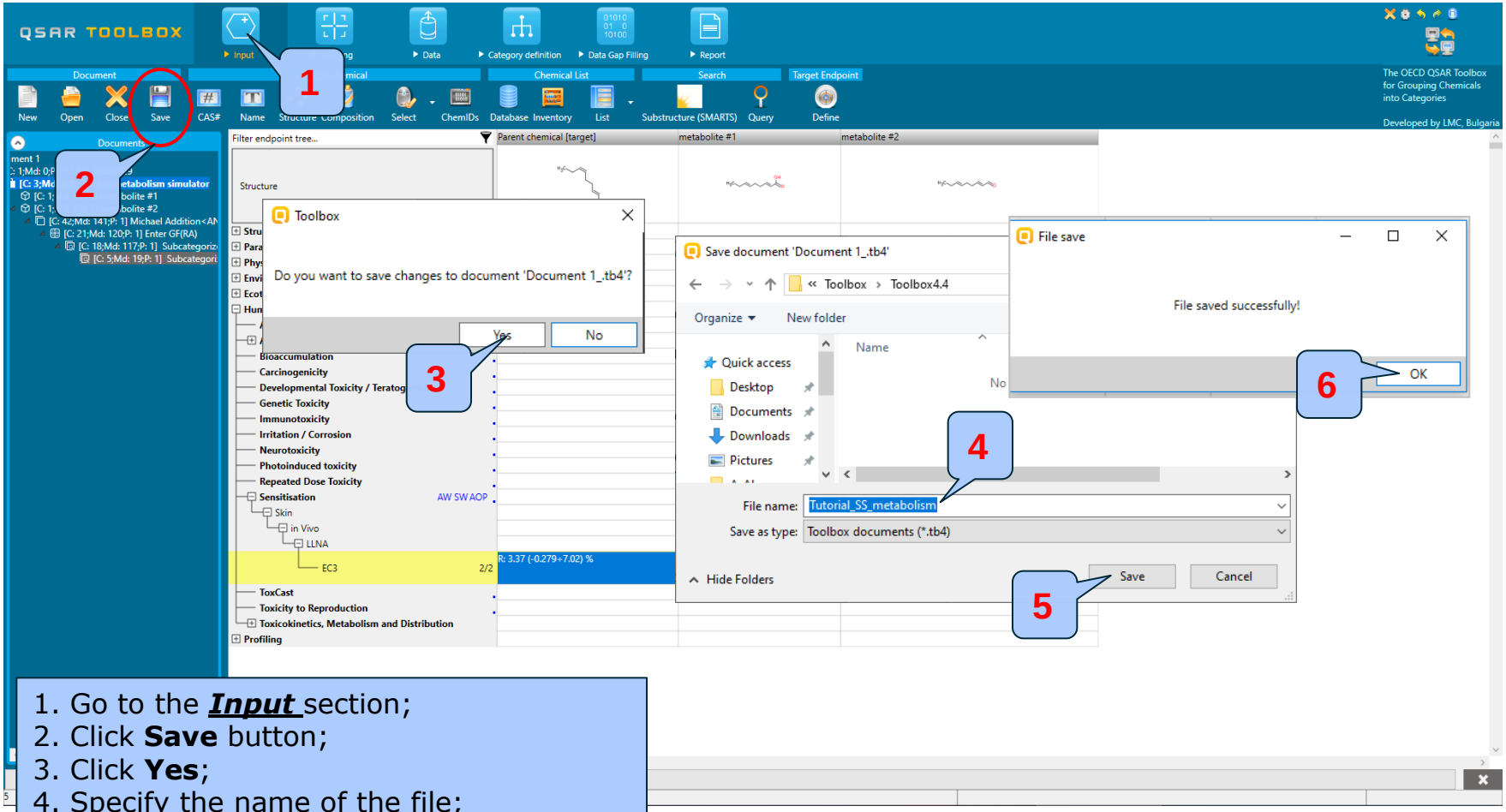
Outlook

- Background
- Keywords
- Objectives
- The exercise
- Workflow
- **Save/open the workflow**

Saving the prediction result
Overview

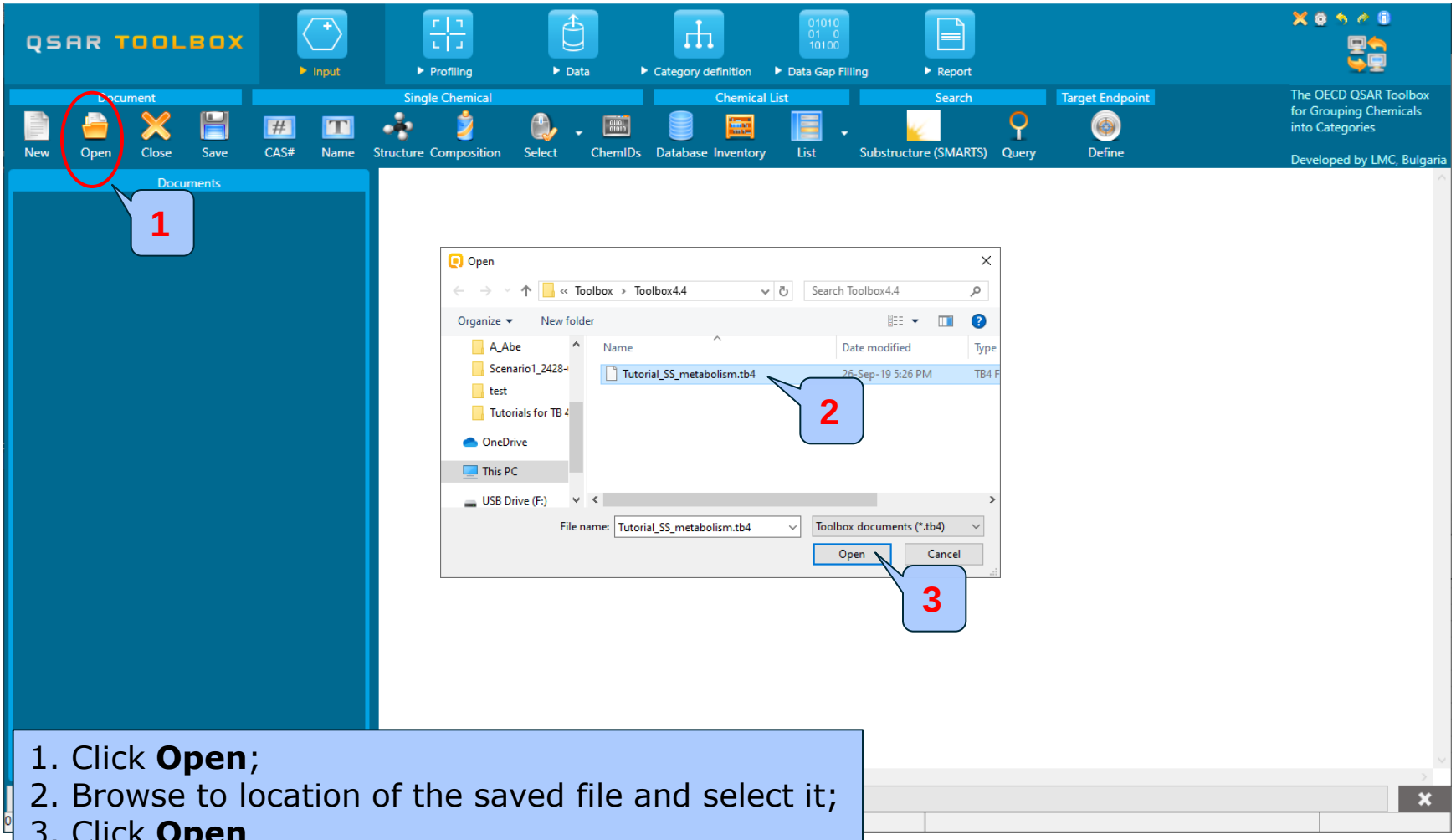
- The Save/Load functionalities allow storing/restoring of Toolbox documents including loaded chemicals, experimental data, profiles, predictions, etc.
- These functionalities are implemented based on saving the sequence of actions performed in a Toolbox document and later executing of these actions in the same sequence of steps in order to get the same result(s).
- Saving/Loading of a Toolbox file is shown on the next screenshots.

Saving the prediction result



1. Go to the ***Input*** section;
2. Click **Save** button;
3. Click **Yes**;
4. Specify the name of the file;
5. Click on the **Save** button;
6. Click **OK**.

Open the saved file



1. Click **Open**;
2. Browse to location of the saved file and select it;
3. Click **Open**