QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.2

Evaluating alert performance accounting for a metabolism

Outlook

- **Background**
- Objectives
- Specific Aims
- Alert performance
- The exercise
- Workflow

Background

- This is a step-by-step presentation designed to take the Toolbox user through the workflow of a data gap filling exercise and assessing of the outcome.

Outlook

- Background
- **Objectives**
- Specific Aims
- Alert performance
- The exercise
- Workflow

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Define a target endpoint;
- Relevancy of profiles and data availability;
- Define the primary group by accounting for a metabolism;
- Calculation of an alert performance (AP) accounting for a metabolism;
- Searching of analogues accounting for metabolism;

Outlook

- Background
- Objectives
- **Specific Aims**
- Alert performance
- The exercise
- Workflow

Specific Aims

- To familiarize the user with the Alert performance (AP) functionality;
- To introduce to the user the calculation of AP accounting for a metabolism;
- To explain to the Toolbox user the rationale behind each step of the exercise.

Outlook

- Background
- Objectives
- Specific Aims
- **Alert performance**
- The exercise
- Workflow

Alert performance

- Alerts are the main category-building units of many profiling schemes (profilers) and their definition is based on a theoretical knowledge and empirical observations.
- The alert performance is estimated based on the distribution of the chemicals having (a) specific alert(s) across the available experimental data for a defined endpoint.
- AP is suitable to be applied for endpoints for which the experimental data exists as potency categories (e.g. Positive, Equivocal, Negative; Strong, Weak, Non sensitizer, etc.).
- The outcome of the estimation provides percent of the Positive and Negative performance and the number of chemicals used to evaluate the performance.

Outlook

- Background
- Objectives
- Specific Aims
- Alert performance
- **The exercise**
- Workflow

The Exercise

- In this exercise we will predict the Skin sensitization potential of *1,3-Propanediamine, N-(3-aminopropyl)* [CAS# 56-18-8], which will be the "target" chemical.
- We will preliminary define the target endpoint.
- The category will be defined with accounting for a metabolism.
- The alert performance will be evaluated for the alerts found in the package a *parent & metabolites*.
- The prediction itself will be made by "read-across".
- The alert performance item generated for the report will be shown.

Outlook

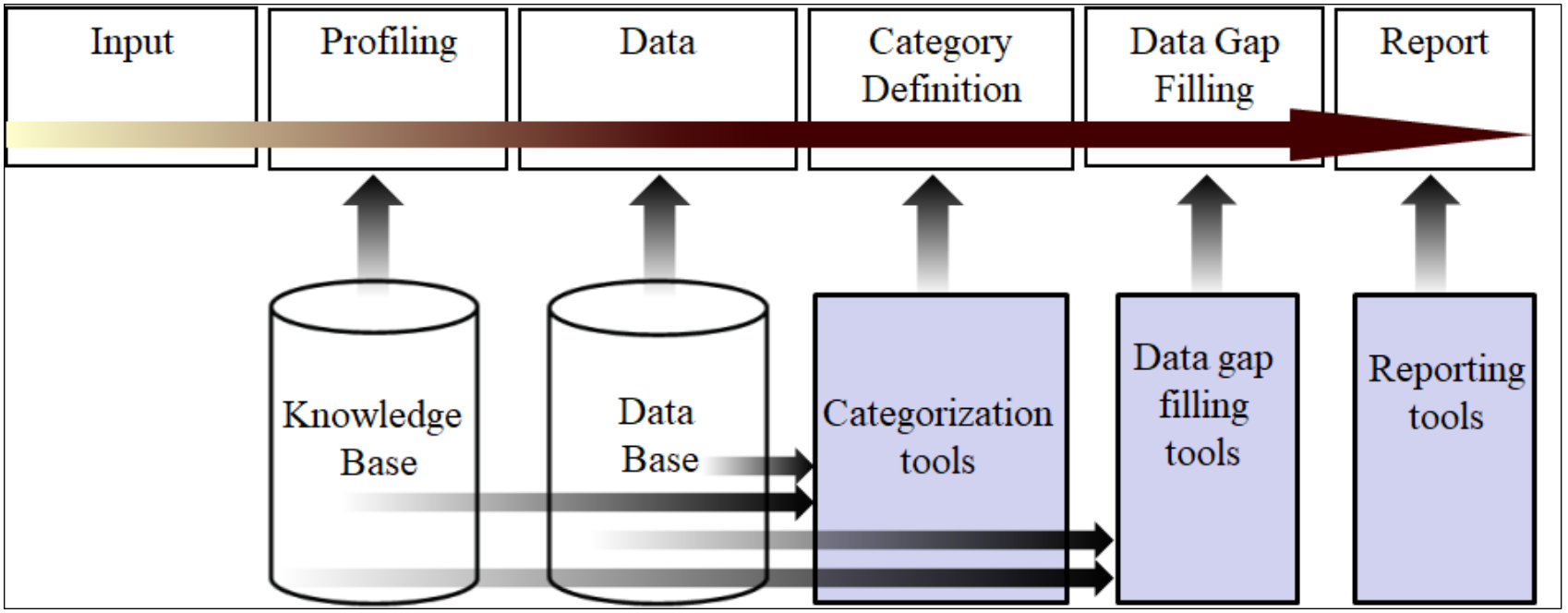
- Background
- Objectives
- Specific Aims
- Alert performance
- The exercise
- **Workflow**

Workflow

- **The Toolbox has six modules which are used in a sequential workflow:**
  - Input
  - Profiling
  - Data
  - Category Definition
  - Data Gap Filling
  - Report
- **We will go through all of them with the exercise**

Workflow

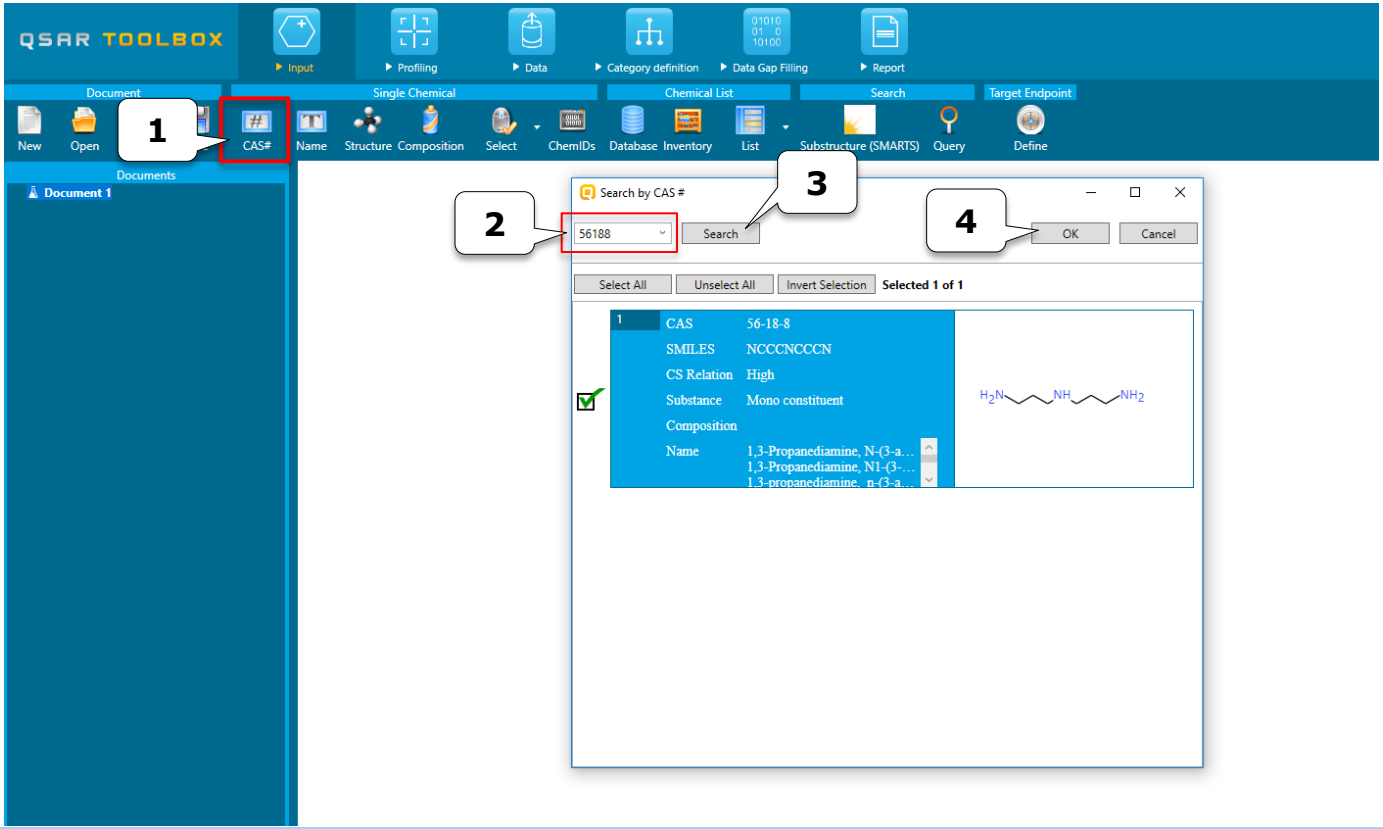
Scheme illustrating the Toolbox workflow



Input
Overview

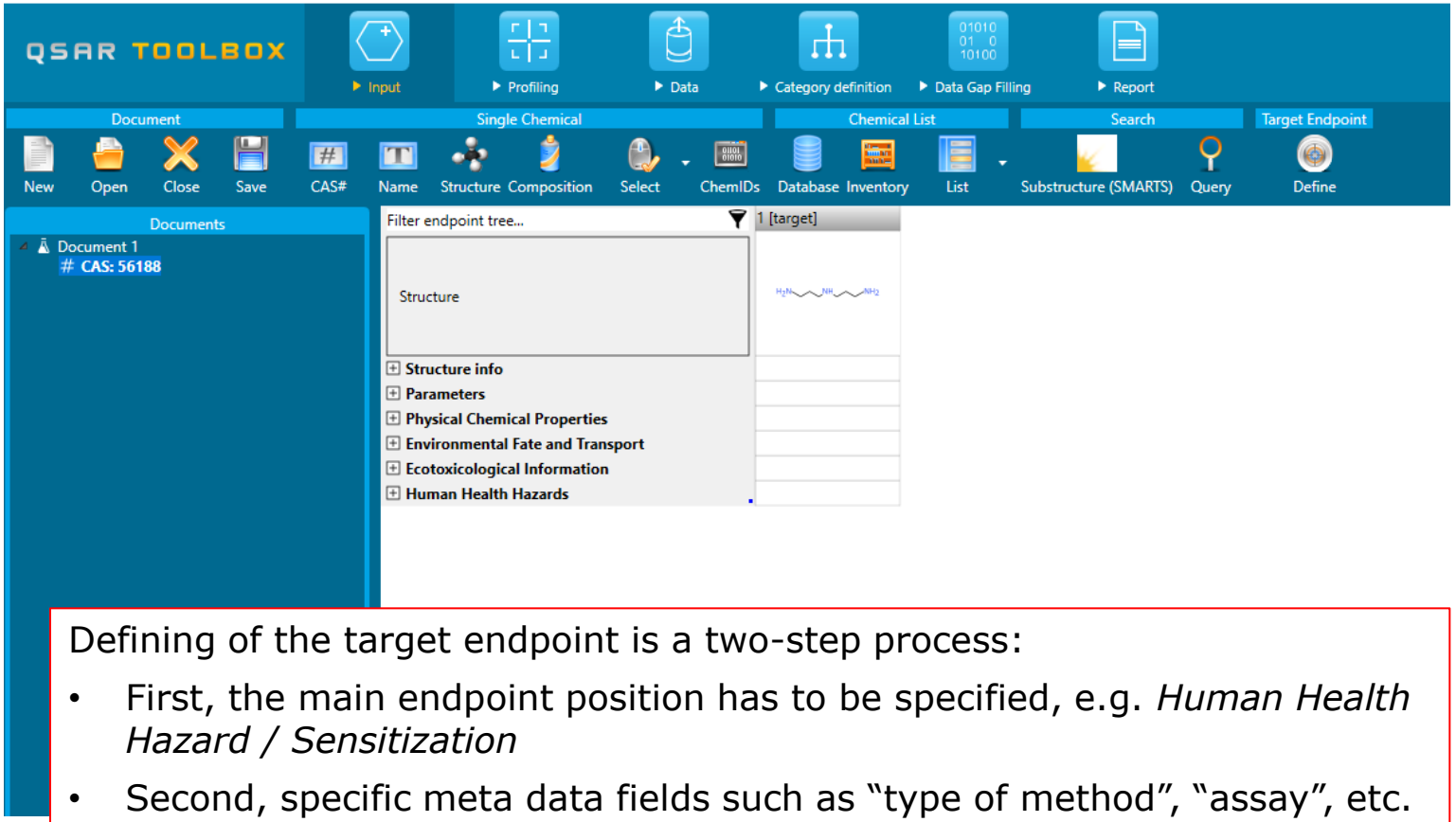
- This module provides the user with several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on a chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.
- Other key functionalities such as “Define a target endpoint” are also placed in the Input module

Input
Entering a Chemical by CAS



Click on **CAS#** icon (1) in the *Input* module. *Search by CAS #* dialogue appears. Type CAS number **56-18-8** in the field (2) and click the **Search** button (3). Confirm by **OK** (4).

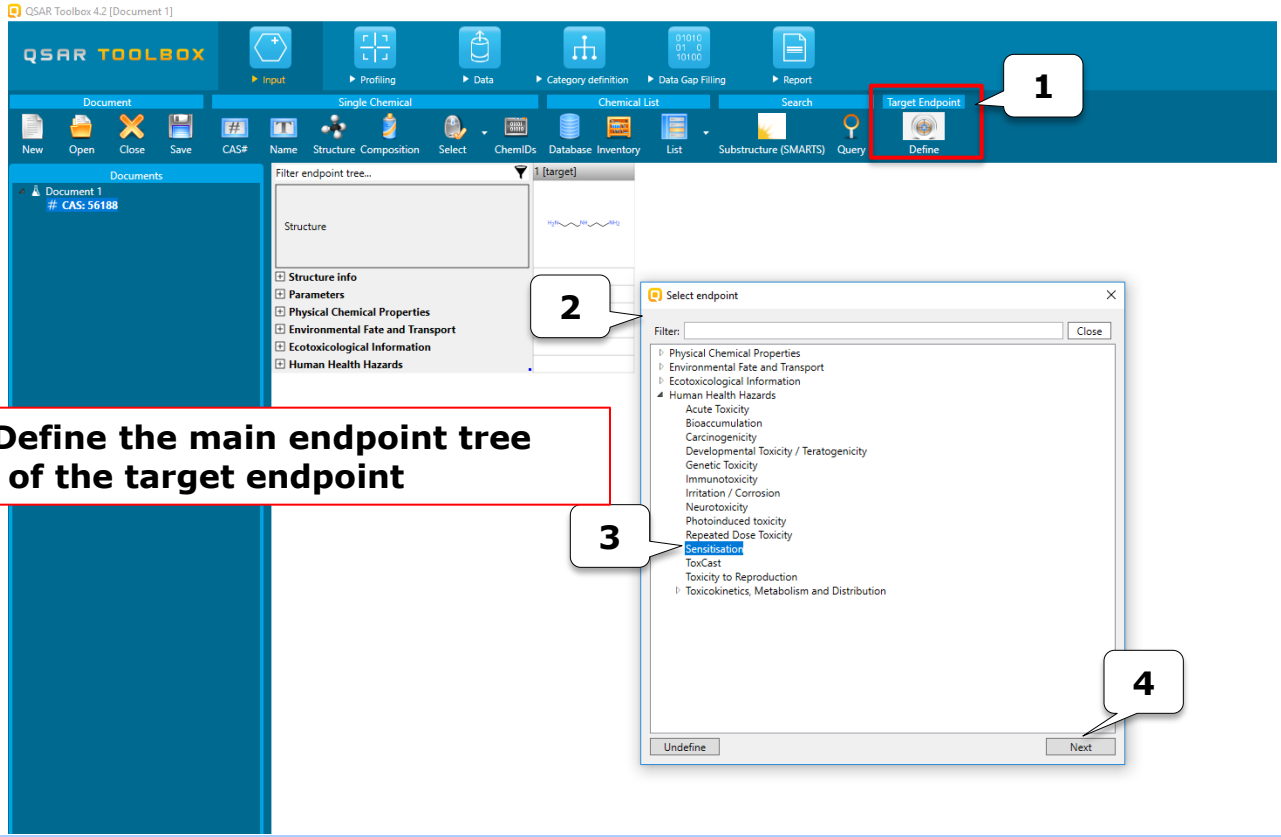
Input
Define the target endpoint



Defining of the target endpoint is a two-step process:

- First, the main endpoint position has to be specified, e.g. *Human Health Hazard / Sensitization*
- Second, specific meta data fields such as "type of method", "assay", etc. Related to the main endpoint tree position has to be defined

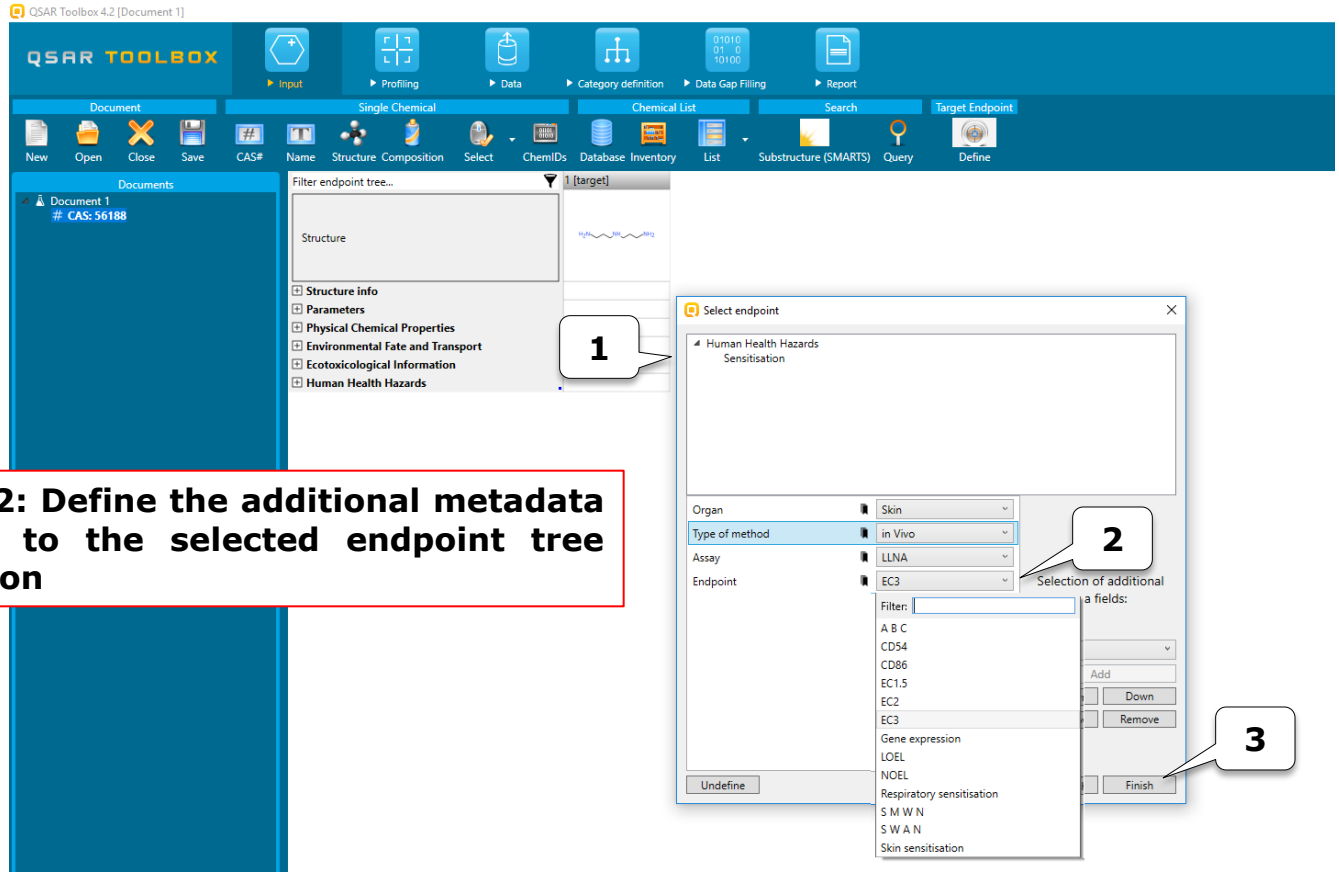
Input
Define the target endpoint



Step 1: Define the main endpoint tree position of the target endpoint

(1) Click on the **Define** icon; **(2)** "Select endpoint" dialogue appears where select "***Sensitization***" **(3)** and click on **Next (4)**

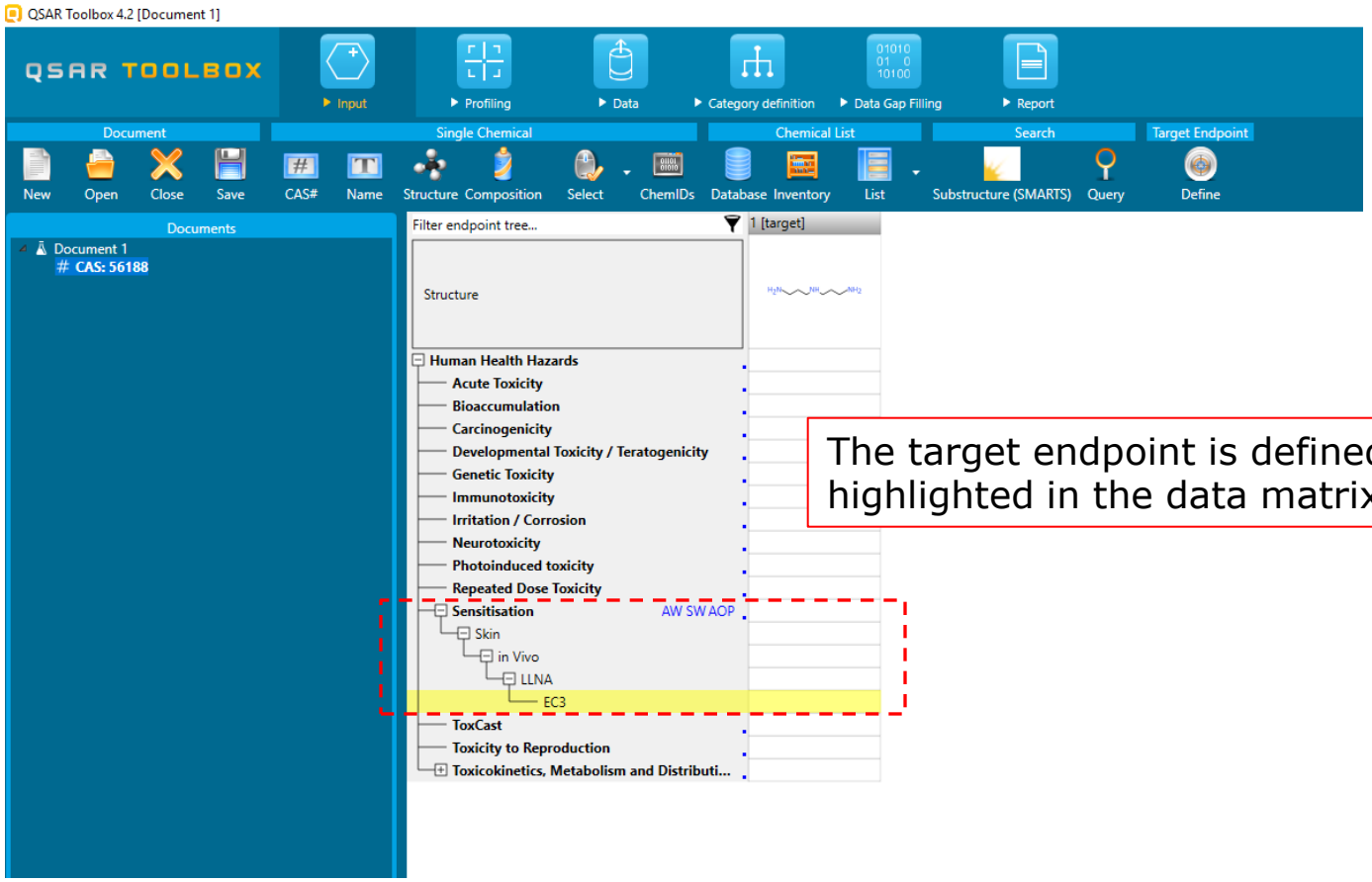
Input
Define the target endpoint



Step 2: Define the additional metadata fields to the selected endpoint tree position

(1) A new dialogue for defining additional details to the selected target endpoint appears;
(2) From the drop-down menus select the specific information for the metadata fields as follows: Endpoint is **EC3**; Organ is **Skin**; Type of method is **in Vivo;** Assay is **LLNA**
(3) Click on **Finish**

Input
Define target endpoint



The target endpoint is defined and highlighted in the data matrix.

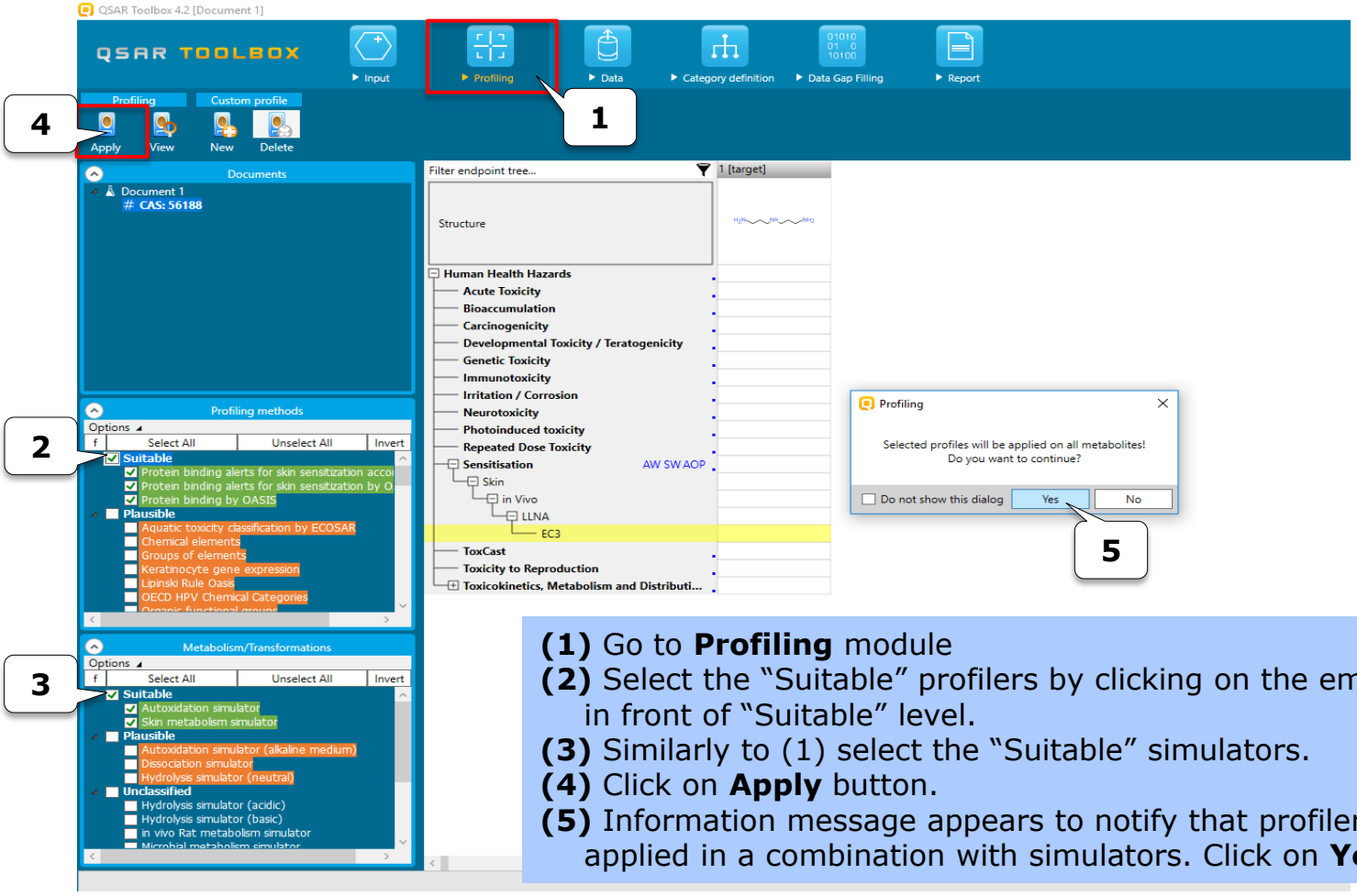
Profiling
Overview

- "Profiling" refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- The available information includes likely mechanism(s) of action, as well as observed or simulated metabolites.
- Based on the target endpoint the profilers and metabolism simulator are grouped by their relevancy for application*.

*More details regarding the profilers' grouping by relevancy could be seen in: *Example for predicting skin sensitization taking into account alert performance*

Profiling

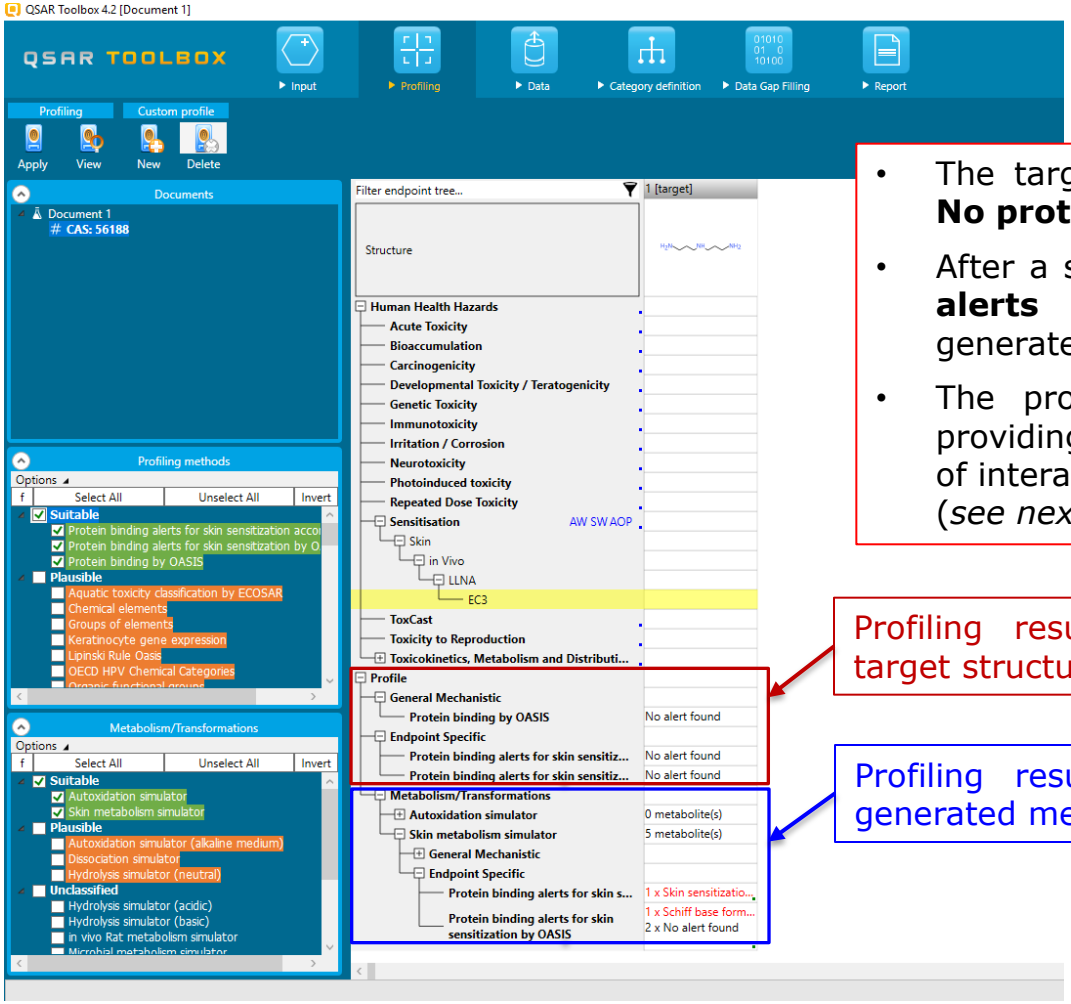
Profiling the target



(1) Go to **Profiling** module
(2) Select the "Suitable" profilers by clicking on the empty box in front of "Suitable" level.
(3) Similarly to (1) select the "Suitable" simulators.
(4) Click on **Apply** button.
(5) Information message appears to notify that profilers will be applied in a combination with simulators. Click on **Yes**.

Profiling

Profiling the target



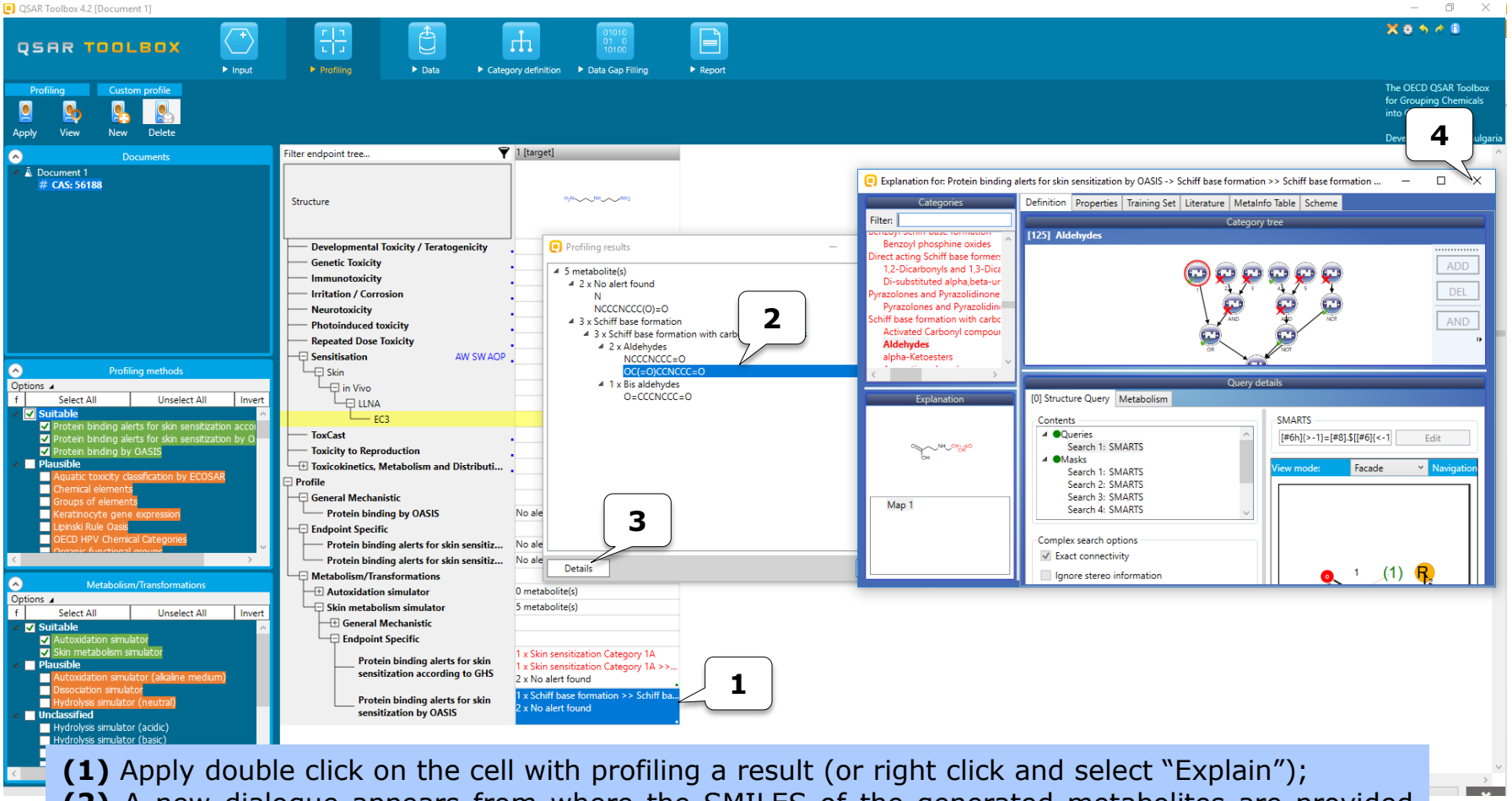
- The target chemical is profiled as having **No protein binding alert** as a parent
- After a skin metabolism, **protein binding alerts are identified** for some of generated skin metabolites
- The profiling result could be explained providing more details for the mechanism of interaction and an additional information (*see next page*)

Profiling result of the target structure

Profiling result of the generated metabolites

Profiling

Profiling the target



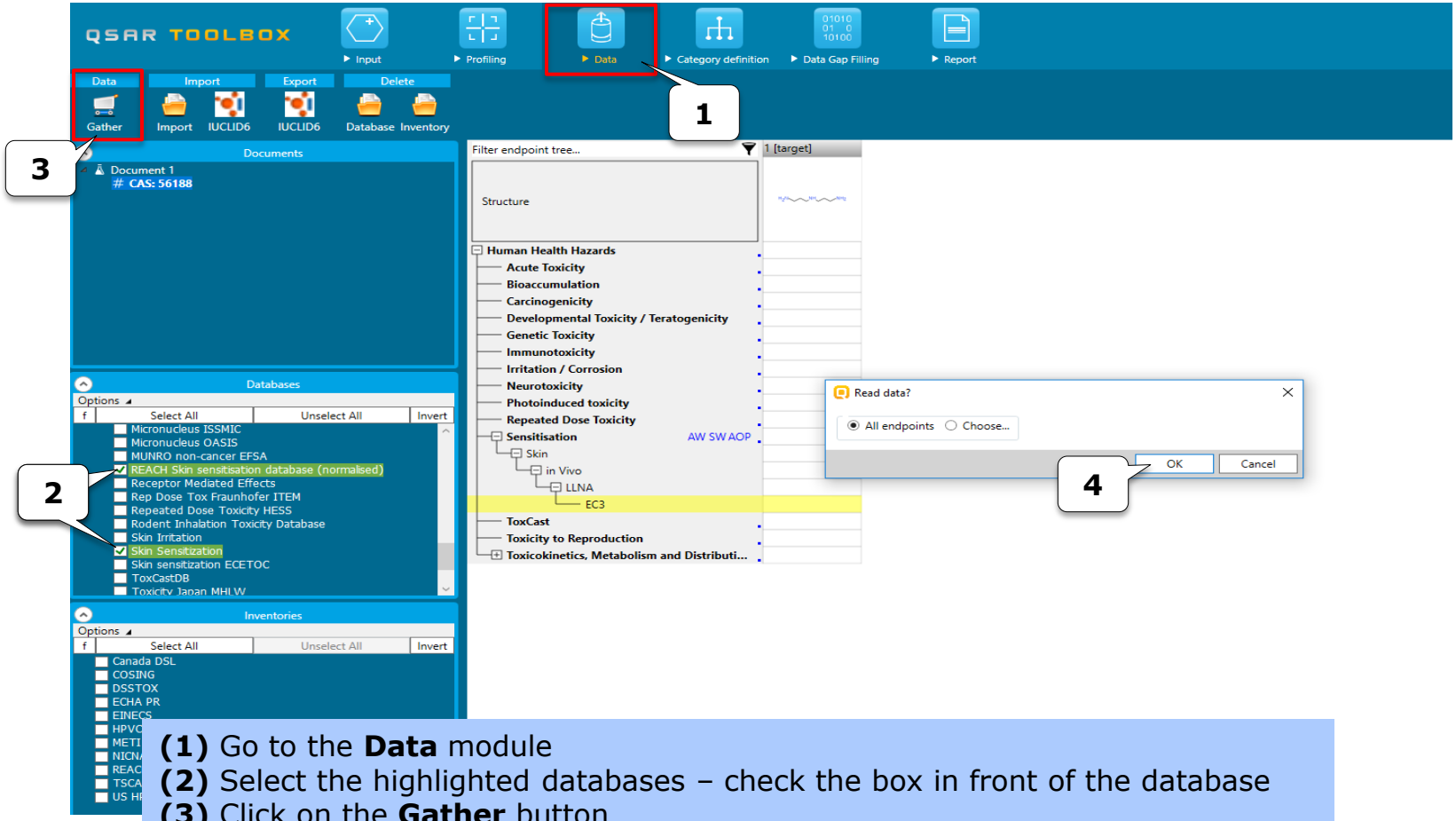
(1) Apply double click on the cell with profiling a result (or right click and select "Explain");
(2) A new dialogue appears from where the SMILES of the generated metabolites are provided along with the respective profiling result for each of the SMILES. Select the illustrated SMILES and provide a double click on it or click on **(3) Details**. The explanation result of the selected metabolite is shown in a new dialogue.; **(4)** Close the explanation window.

Data
Overview

- "Data" refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.
- The 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 a limited number of endpoints).
- Once the target endpoint is defined, the system highlights the databases where the data for the defined endpoint could be found

Data

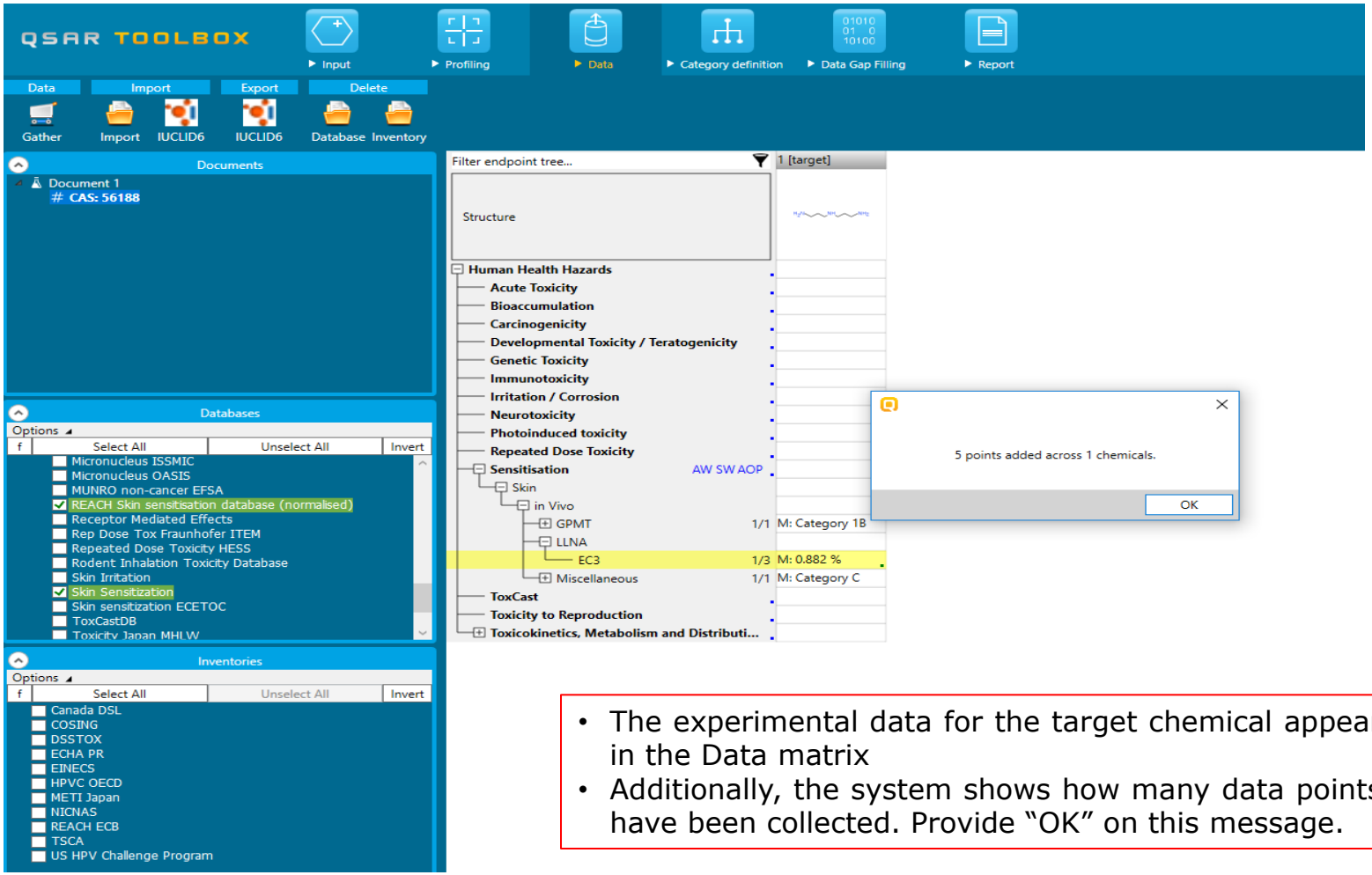
Collecting the experimental data



(1) Go to the **Data** module
(2) Select the highlighted databases – check the box in front of the database
(3) Click on the **Gather** button
(4) A message appears asking to Read data/All endpoints. Click on **OK**

Data

Collecting the experimental data



- The experimental data for the target chemical appear in the Data matrix
- Additionally, the system shows how many data points have been collected. Provide "OK" on this message.

Category Definition
Overview

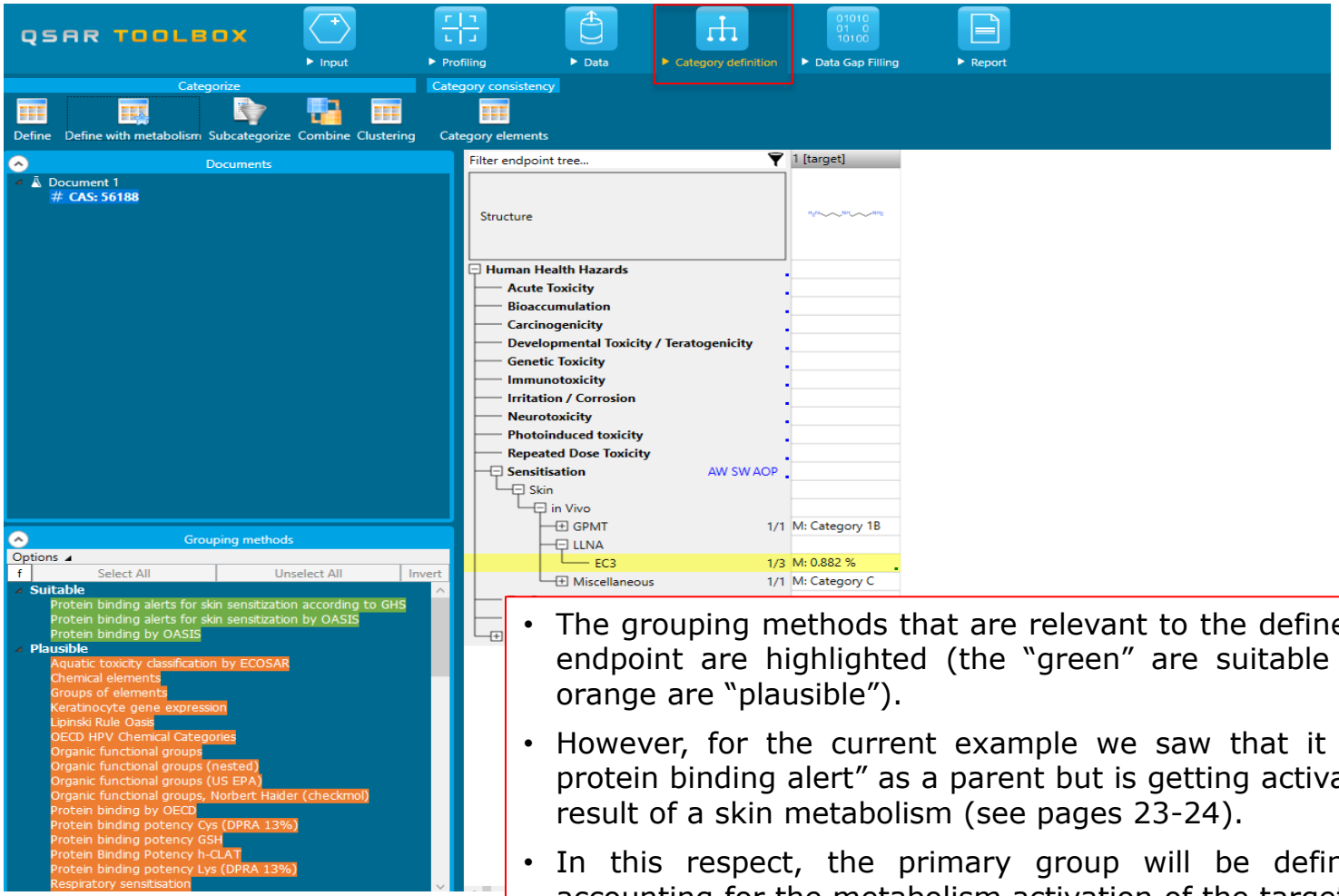
- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- This is the critical step in the workflow.
- Several options are available in the Toolbox to assist the user in refining the category definition.

Category Definition
Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of "similarity" so that within a category data gaps can be filled by a read-across.
- The knowledge implemented in the system as profilers appear here as grouping methods.
- For example, starting from a target chemical for which a specific protein binding mechanism is identified, analogues can be found which can bind by the same mechanism and for which experimental results are available.
- Relevant to the defined target endpoint profilers are highlighted and could be used to define the category.
- The group could be defined with accounting for a metabolism in the cases where an activation as a result of a metabolism is found for the target structure

Category Definition

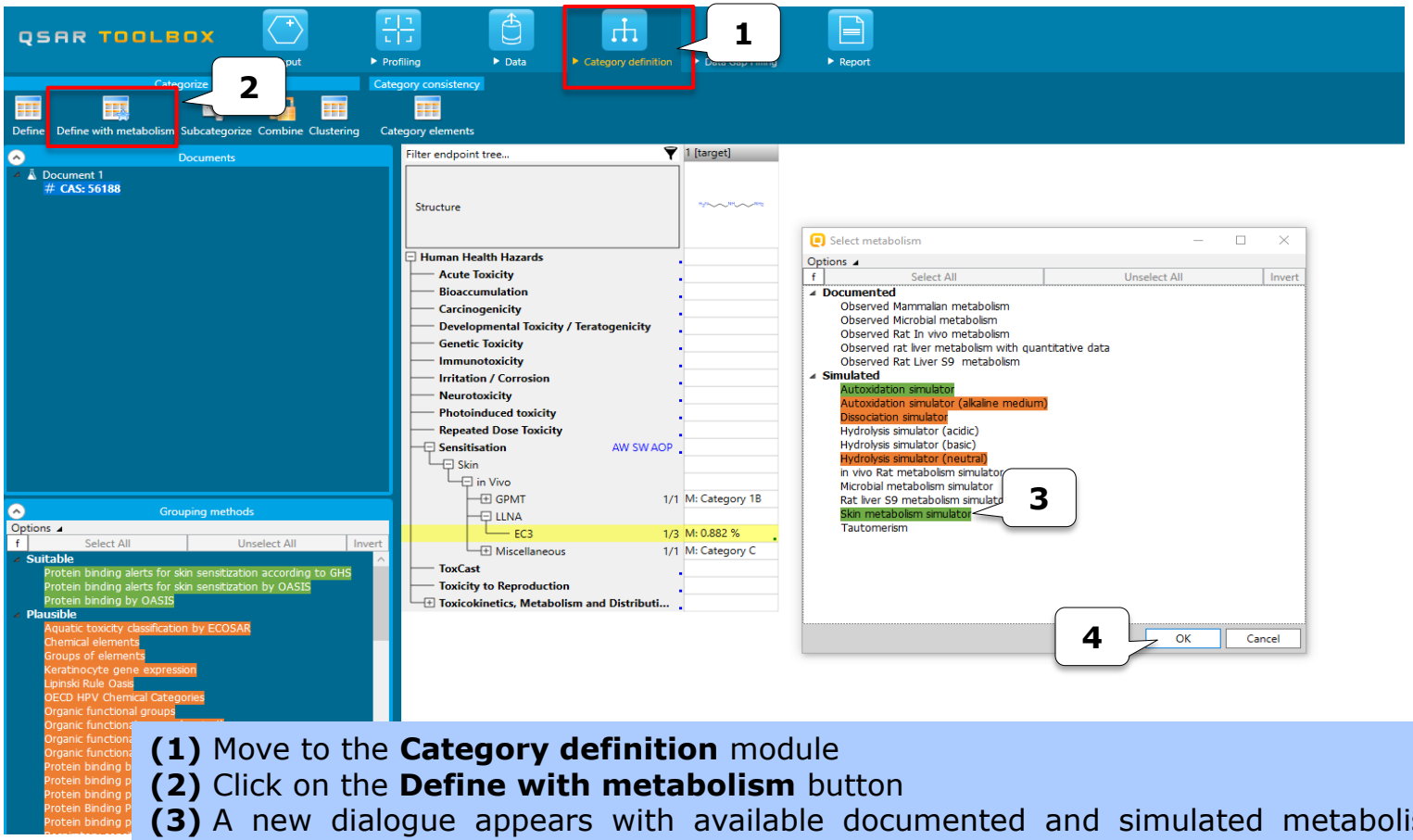
Define the category accounting for a metabolism



- The grouping methods that are relevant to the defined target endpoint are highlighted (the "green" are suitable and the orange are "plausible").
- However, for the current example we saw that it has "No protein binding alert" as a parent but is getting activated as a result of a skin metabolism (see pages 23-24).
- In this respect, the primary group will be defined with accounting for the metabolism activation of the target.

Category Definition

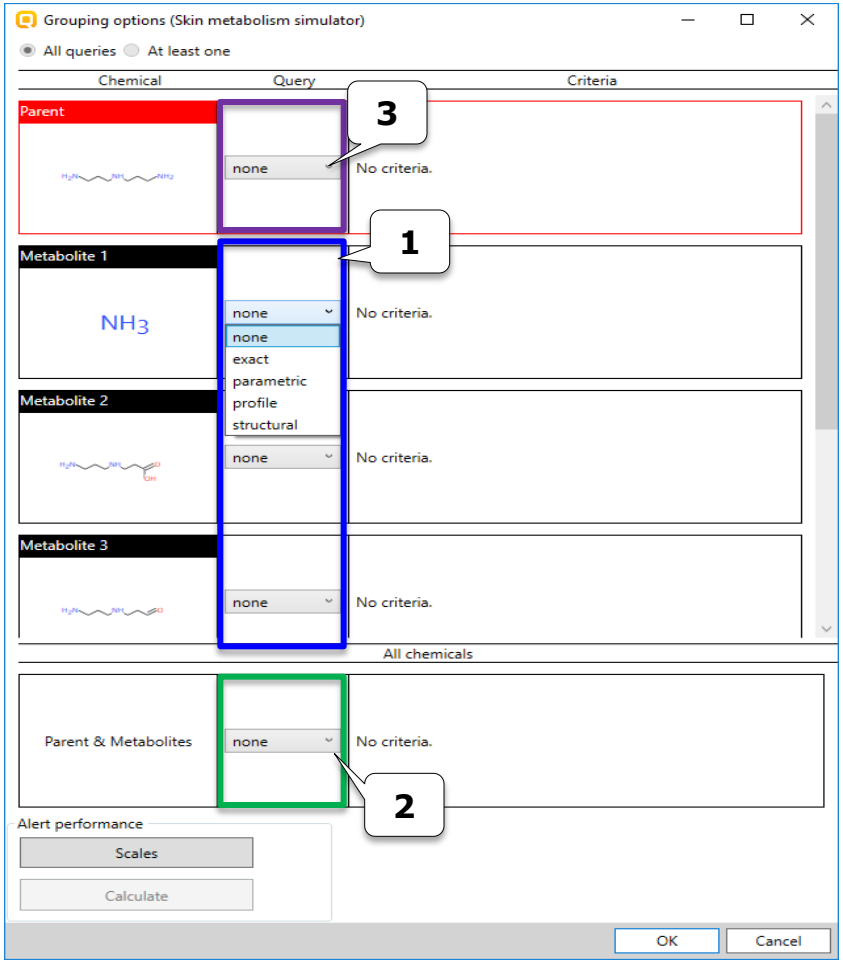
Define the group with accounting for a metabolism



(1) Move to the **Category definition** module
(2) Click on the **Define with metabolism** button
(3) A new dialogue appears with available documented and simulated metabolisms; Select ***Skin metabolism simulator***
(4) Click **OK** to confirm
A new dialogue appears which is explained in details on the next page

Category Definition

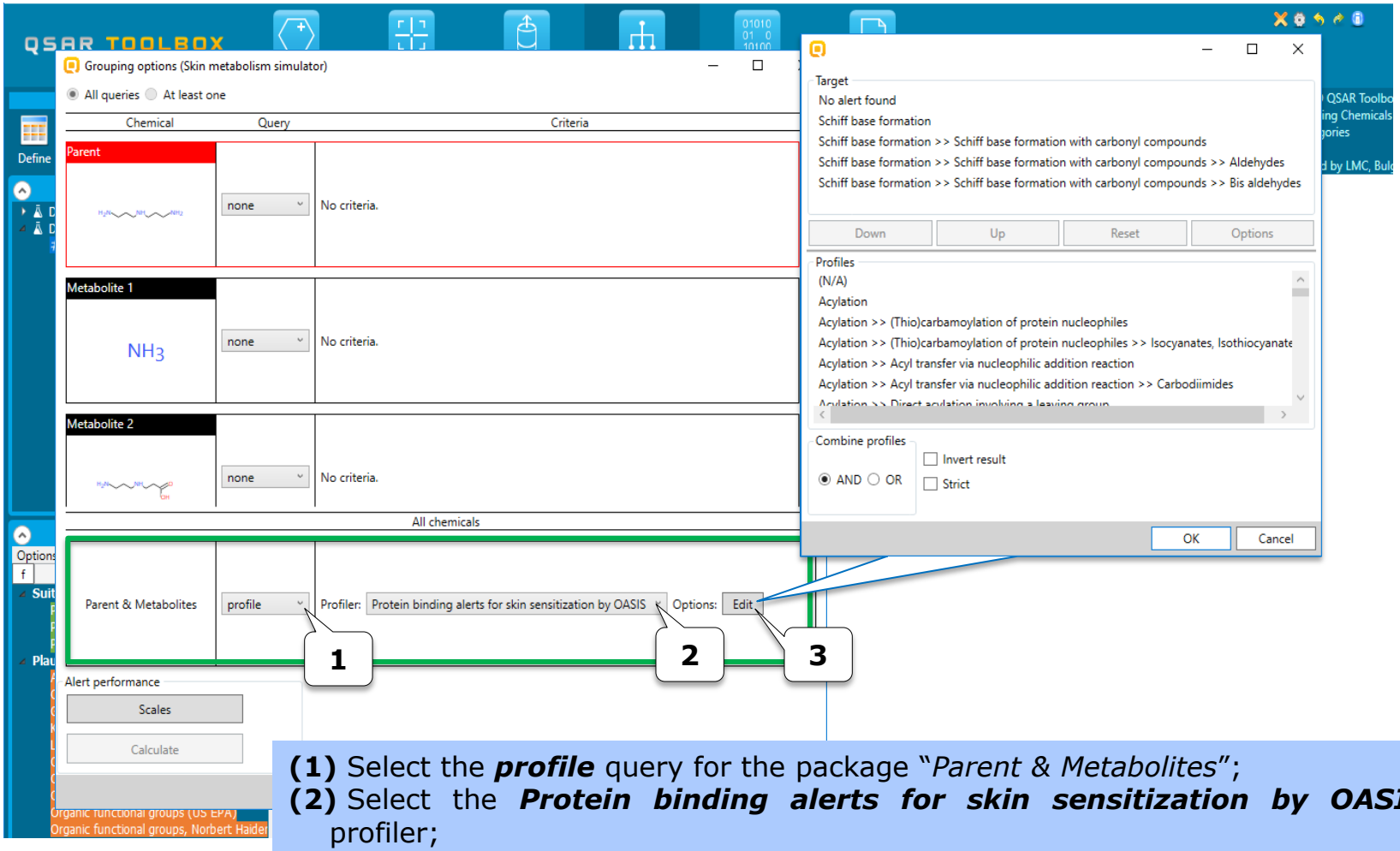
Define the group with accounting for a metabolism



- The newly appeared window shows the parent and all generated metabolites produced by the selected metabolic simulator (*Skin sensitization simulator* in the current example).
- The user is able to set a searching criteria for each of the metabolites (1) or for the whole package "*Parent & Metabolites*" (2).
- Query for the parent could be also defined as an addition (3). It is not possible to define searching criteria for the parent, only.
- The following queries could be set:
  - *None* – default options; no criteria is set
  - *Exact* – provides opportunity to search for metabolites in the analogues having exact to the specified metabolite structure
  - *Parametric* – to have a specific value or a range of variation of a defined parameter (a list with all parameters currently available in the Toolbox is provided)
  - *Profile* – to have a specific category by selected profiler (a list with all profilers is provided)
  - *Structural* – to have a specific similarity based on the atom centered fragments
- Calculation of the AP will take into account all defined criteria

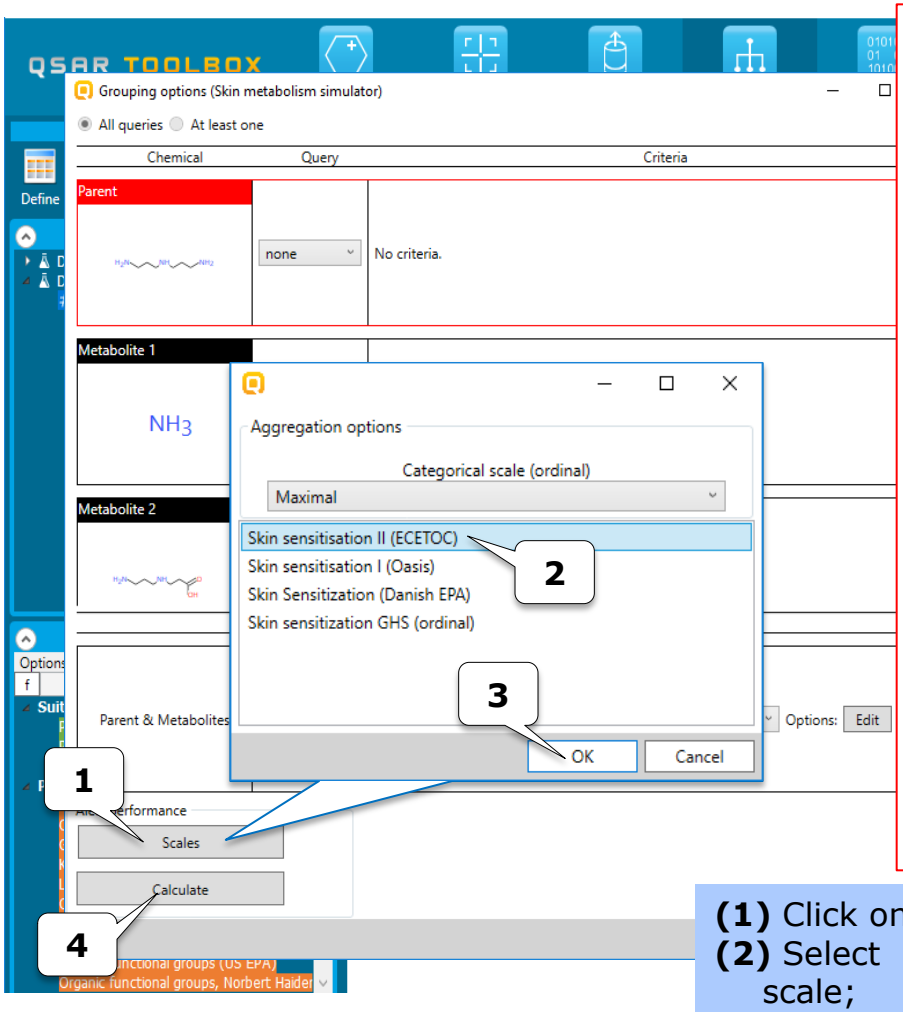
Category Definition

Define the group with accounting for a metabolism



(1) Select the ***profile*** query for the package "*Parent & Metabolites*";
(2) Select the ***Protein binding alerts for skin sensitization by OASIS*** profiler;
(3) Click **Edit** in order to see all identified alerts in the parent and its metabolites

Category Definition
Calculation of the alert performance



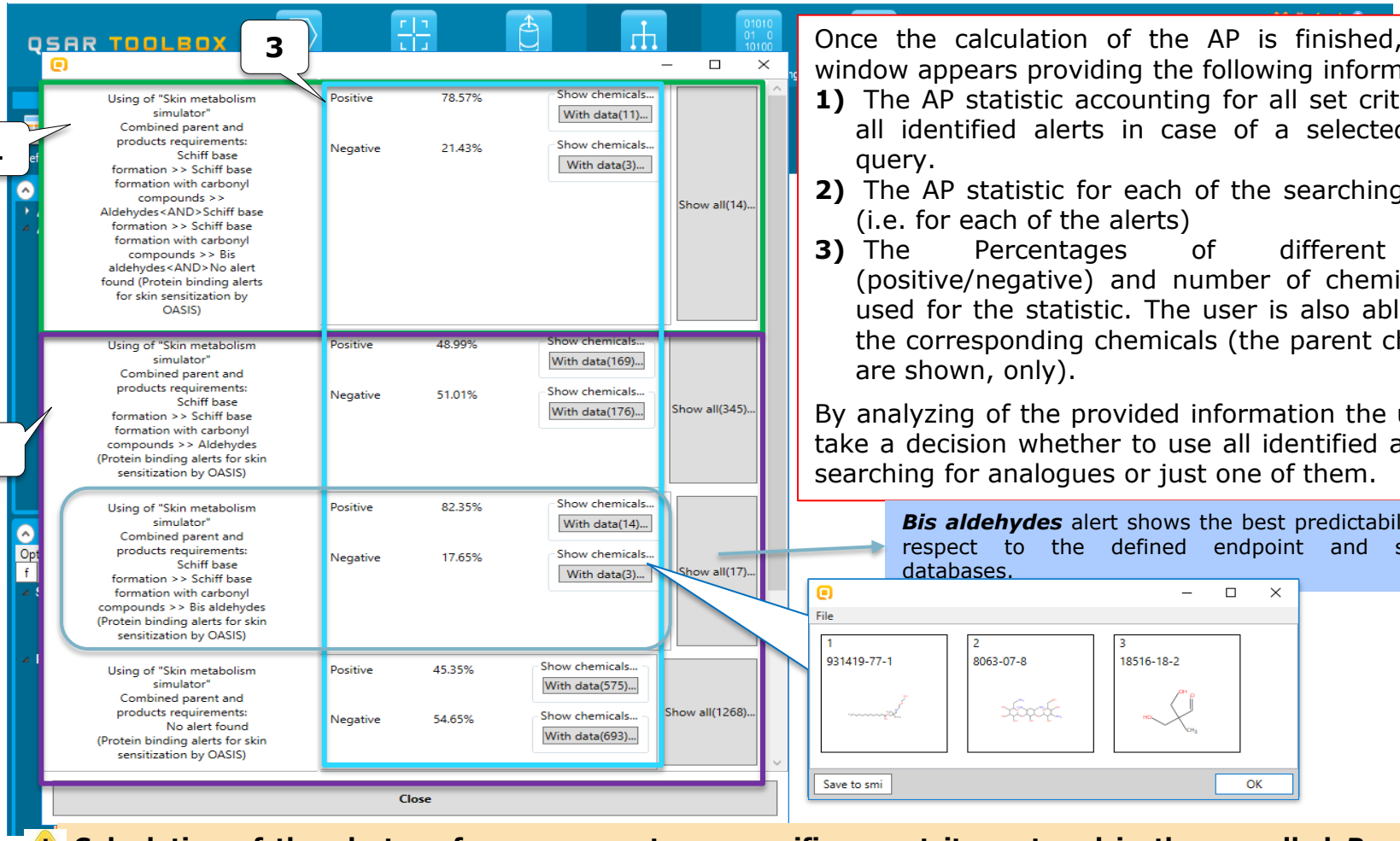
The alert performance results depend on:

- **The defined target endpoint** – The AP is endpoint-dependent. SS, EC3 is defined in the current example.
- **Selected databases** – The AP results will be based on the chemicals presenting in the selected databases. Skin sensitization and REACH Skin sensitization database (normalized) are selected.
- **Selected scale** – the available scales vary based on the defined target endpoint. For Skin sensitization the most appropriate scale is *Skin sensitization II (ECETOC)*. This is a dichotomous scale that converts the data into positive/negative. In this way the experimental data in different scales could be combined in order to provide the full AP statistic.
- **Mode** – the mode takes a role when a chemical from the selected databases has more than one experimental data that could be converted to the selected scale. The *Maximal* mode (the worst case scenario) is set by default (e.g. if a chemical has simultaneously positive and negative data, only the positive data will be taken when calculate AP.

(1) Click on **Scales**;
(2) Select the ***Skin sensitization II (ECETOC)*** scale;
(3) Confirm by **OK**;
(4) Click **Calculate**

Category Definition

Calculation of the alert performance



Once the calculation of the AP is finished, a new window appears providing the following information:

1) The AP statistic accounting for all set criteria and all identified alerts in case of a selected *profile* query.

2) The AP statistic for each of the searching criteria (i.e. for each of the alerts)

3) The Percentages of different data (positive/negative) and number of chemicals are used for the statistic. The user is also able to see the corresponding chemicals (the parent chemicals are shown, only).

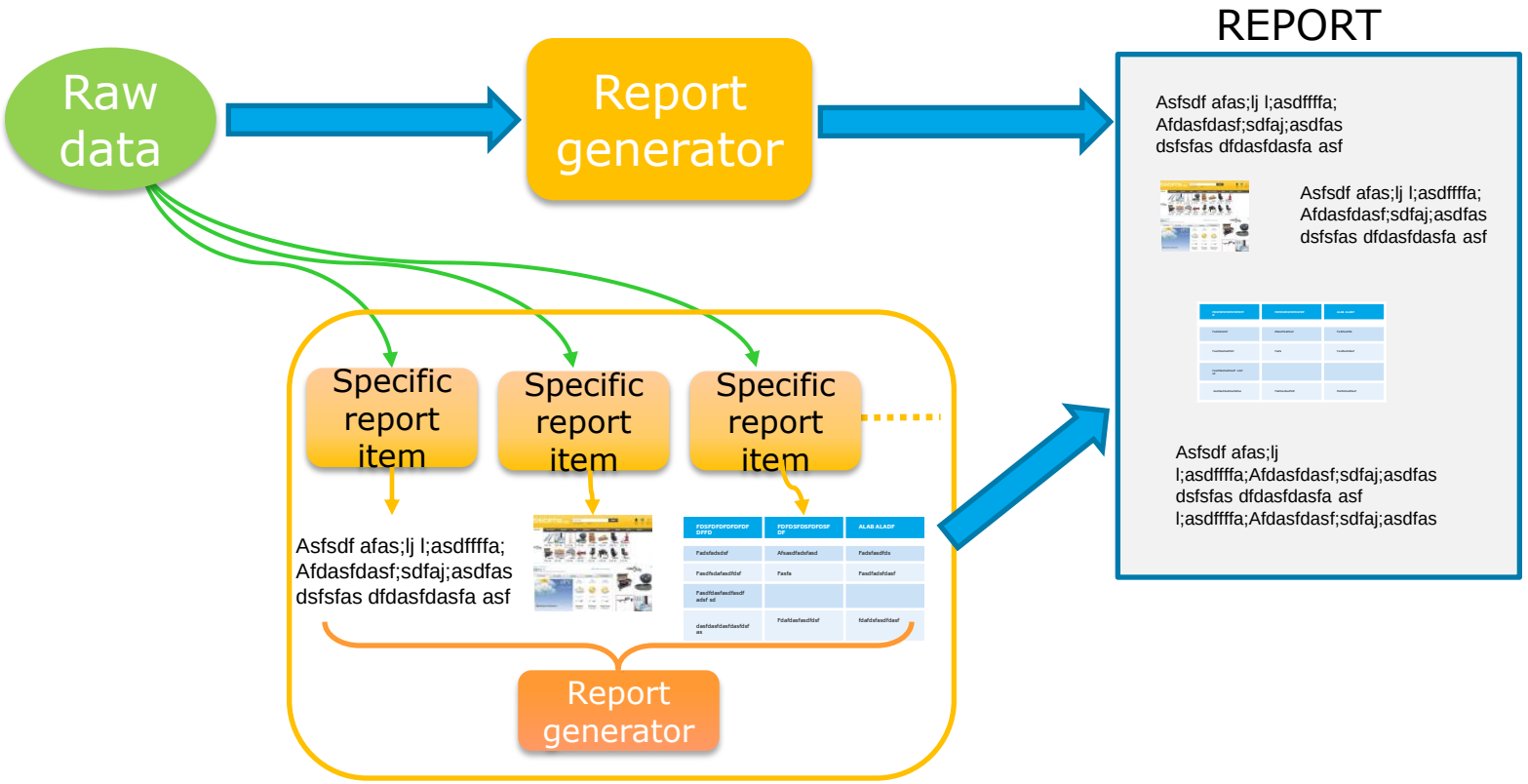
By analyzing of the provided information the user can take a decision whether to use all identified alerts for searching for analogues or just one of them.

Bis aldehydes alert shows the best predictability with respect to the defined endpoint and selected databases.

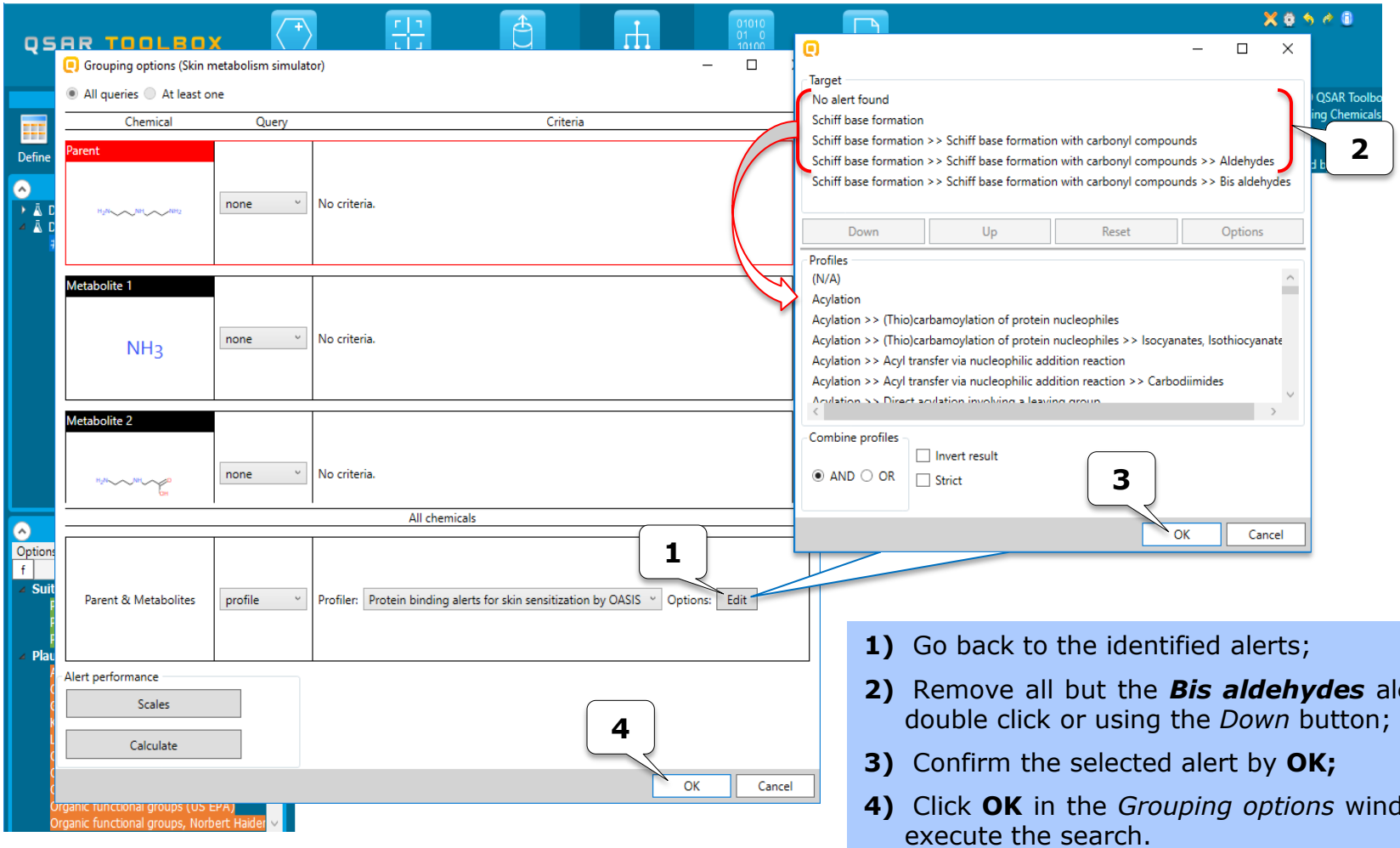
Calculation of the alert performance creates a specific report item stored in the so-called *Report basket*.

Sidebar on the Report basket

- The specific report items are collected during the workflow or from external modeling sources.
- All items are stored in the "*Report basket*" and can be used in the report to support or justify the consistency of a category.



Category Definition
Calculation of alert performance



1) Go back to the identified alerts;

2) Remove all but the ***Bis aldehydes*** alert by double click or using the *Down* button;

3) Confirm the selected alert by **OK;**

4) Click **OK** in the *Grouping options* window to execute the search.

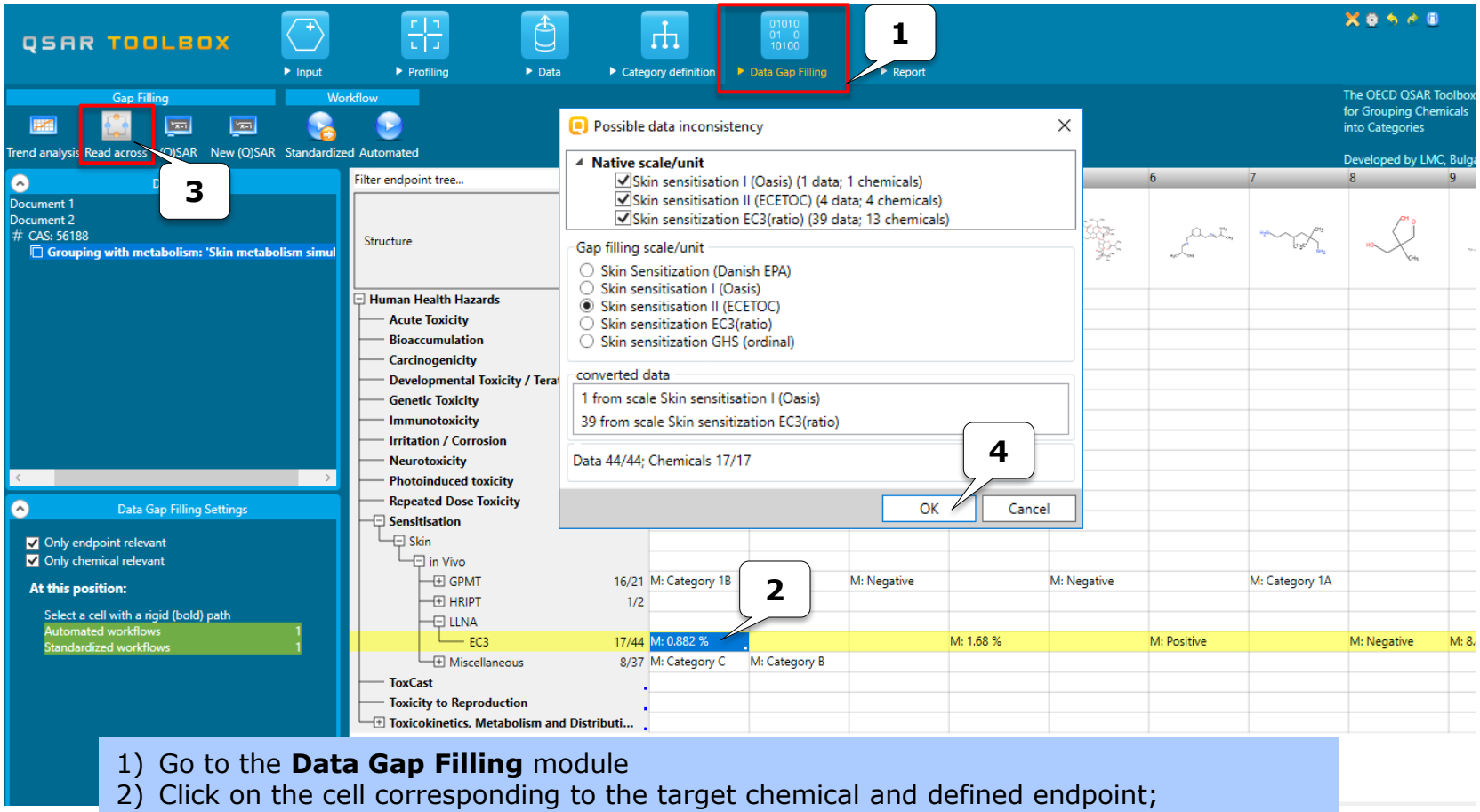
Data Gap Filling
Overview

- "Data Gap Filling" module give access to five different data gap filling tools:
  - Read-across
  - Trend analysis
  - (Q)SAR models
  - Standardized workflow
  - Automated workflow
- Depending on the situation, the most relevant data gap mechanism should be chosen, taking into account the following considerations:
  - The read-across is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitisation or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore the read-across is recommended for "quantitative endpoints" (e.g., 96h-LC50 for fish) if only a low number of analogues with experimental results are identified.
  - Trend analysis is the appropriate data-gap filling method for "quantitative endpoints" (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.
  - "(Q)SAR models" can be used to fill a data gap if no adequate analogues are found for a target chemical.

  Additionally, two workflows - Standardized and Automated have been developed in order to facilitate the users work. Once started, they follow the implemented logic and finish with prediction. The general differences between the two types of workflows are represented on the next slide.

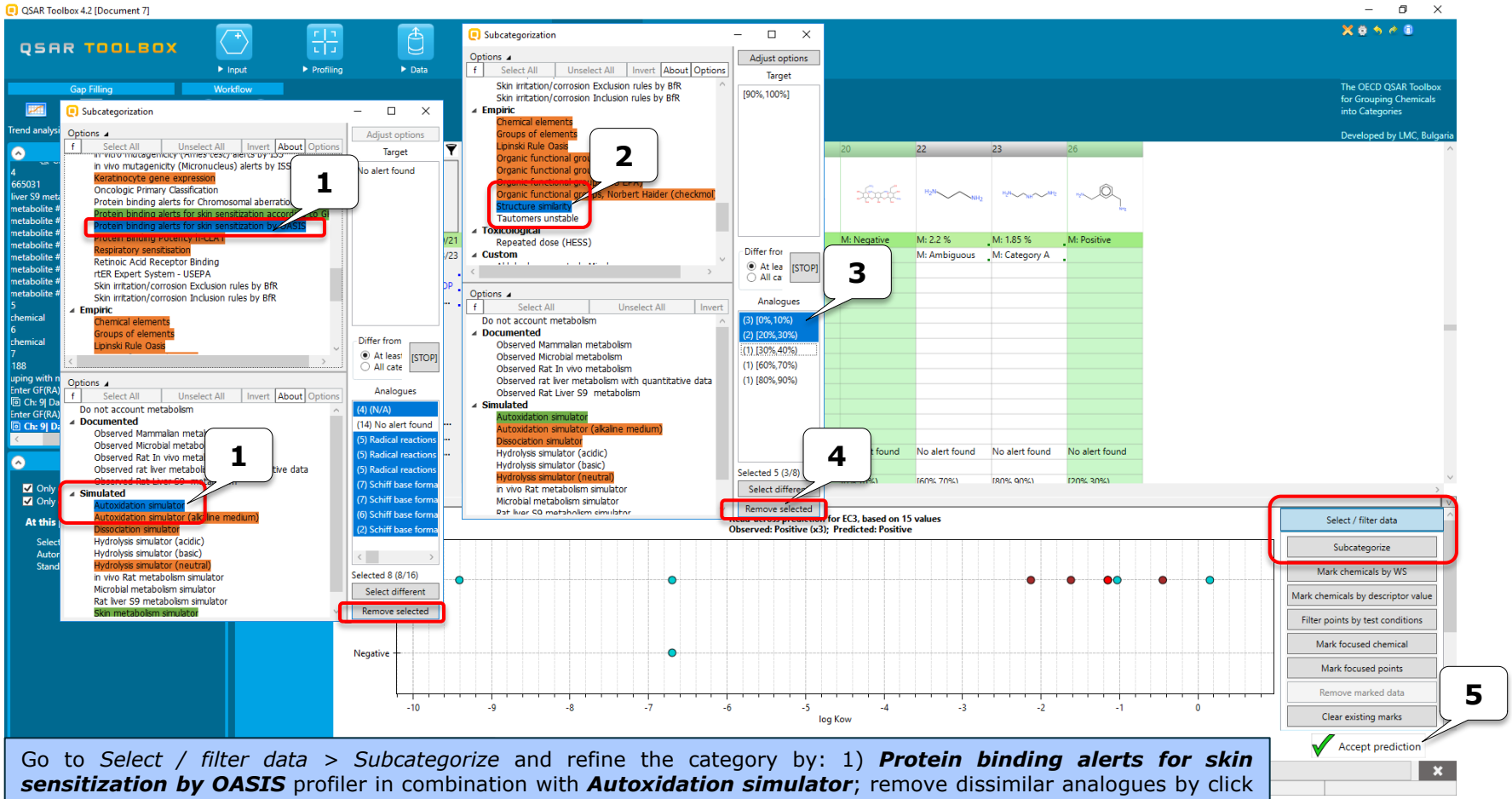
In this example we will use the manual read-across approach.

Data Gap Filling
Apply Read across



1) Go to the **Data Gap Filling** module
2) Click on the cell corresponding to the target chemical and defined endpoint;
3) Click the **Read across** button;
4) Skin sensitization II (ECETOC) scale is selected by default. Confirm by **OK**.

Data Gap Filling
Apply Read across

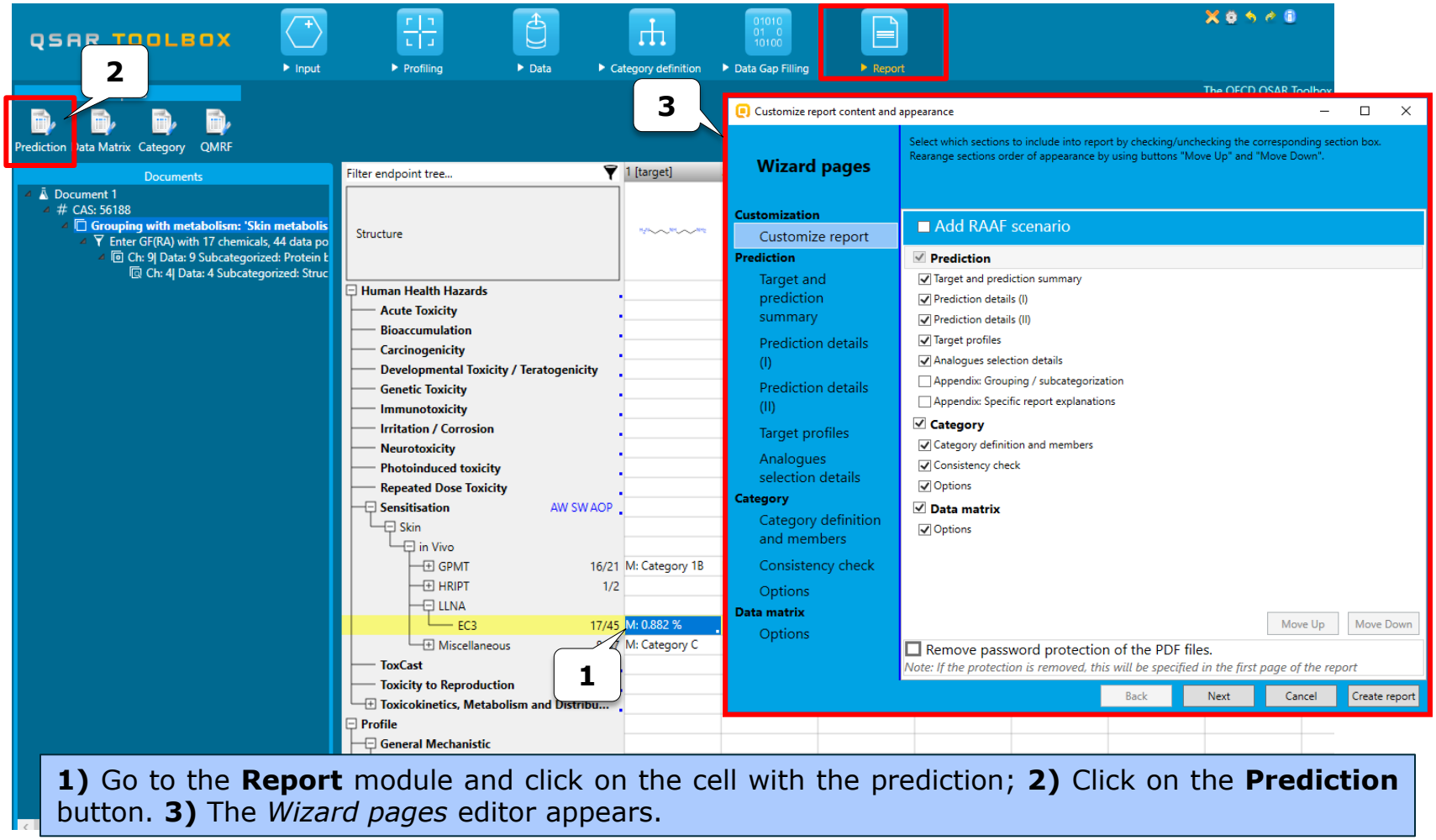


Go to *Select / filter data* > *Subcategorize* and refine the category by: 1) ***Protein binding alerts for skin sensitization by OASIS*** profiler in combination with ***Autoxidation simulator***; remove dissimilar analogues by click on "Remove selected" button 2) ***Structural similarity; 3)*** select analogues with similarity less than 30% to the target chemical (by hold Ctrl button); 4) Click *Remove selected*. Three analogues remain. 5) **Accept the prediction**.

Report
Overview

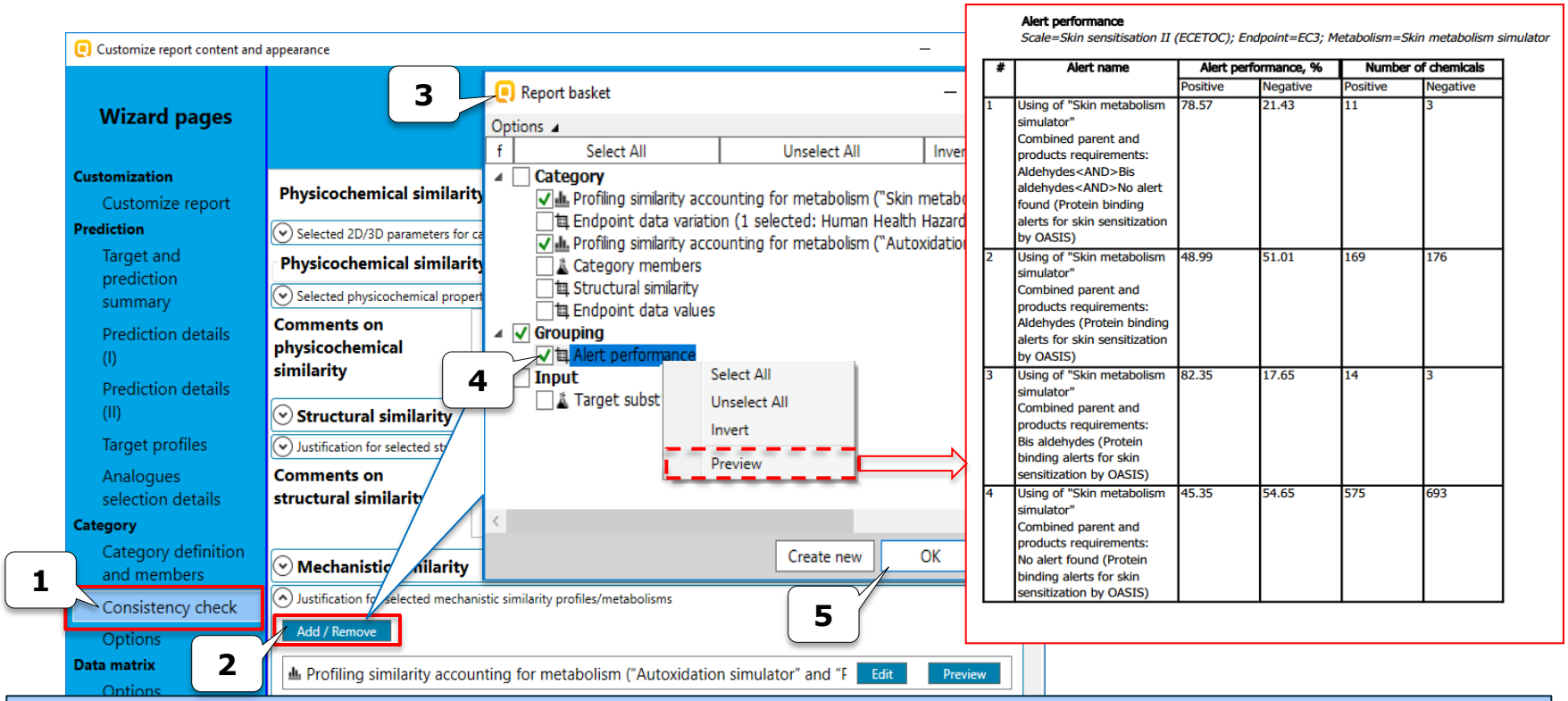
- The report module could generate a report on any of predictions performed with the Toolbox.
- The report module contains a predefined report template which the users can customize.
- Three type of report files are generated:
  - *A Prediction report* – containing information for the target
  - *A Category report* – containing information for the final category (target plus used analogues)
  - *A Data matrix* – containing information for the analogues used for the prediction.
- Additionally a specific RAAF scenario could be chosen. Selection of one of the scenarios will append automatically the related assessment elements (AE) related to the corresponding report sections.
- The *Report basket* (and Alert performance item, respectively) could be used for supporting information to the appropriate category elements or RAAF AE.

Report
Generating a prediction report



1) Go to the **Report** module and click on the cell with the prediction; **2)** Click on the **Prediction** button. **3)** The *Wizard pages* editor appears.

Report
Generating a prediction report



Alert performance

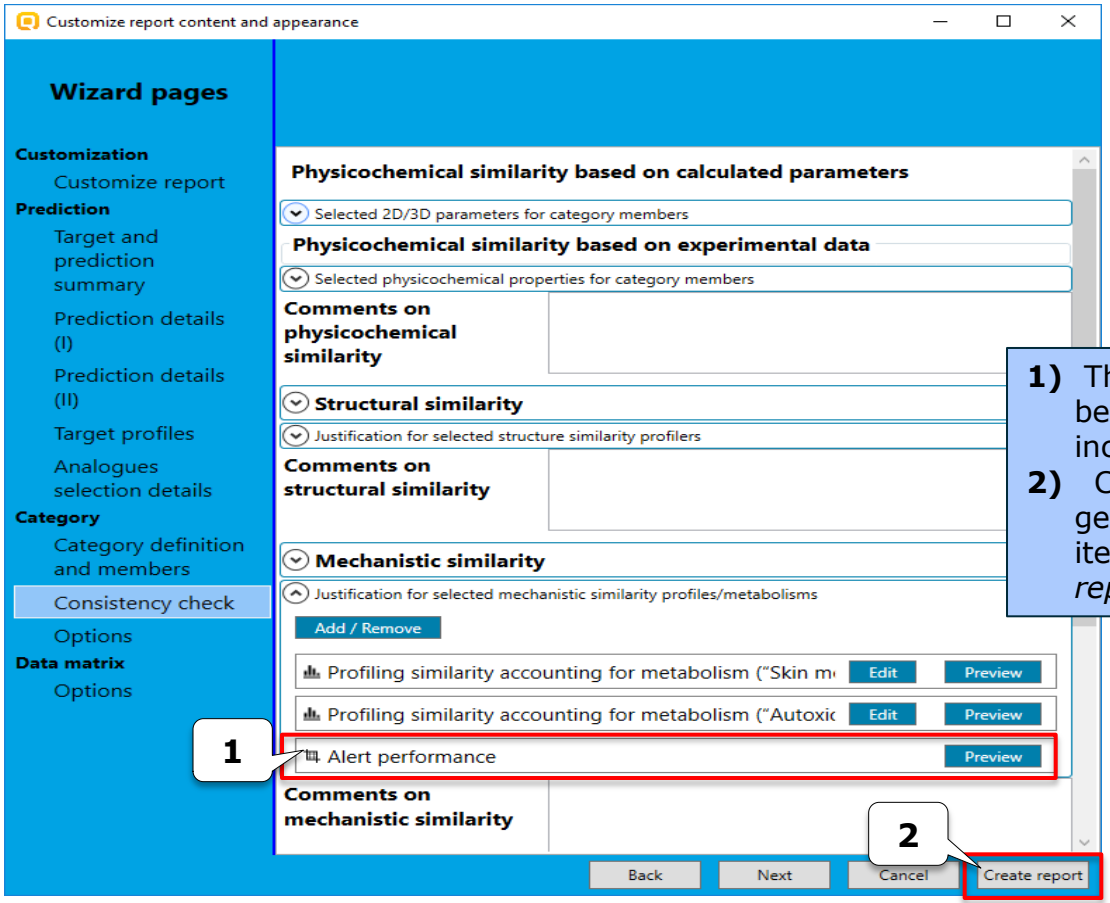
Scale=Skin sensitisation II (ECETOC); Endpoint=EC3; Metabolism=Skin metabolism simulator

#	Alert name	Alert performance, %		Number of chemicals	
		Positive	Negative	Positive	Negative
1	Using of "Skin metabolism simulator" Combined parent and products requirements: Aldehydes<AND>Bis aldehydes<AND>No alert found (Protein binding alerts for skin sensitization by OASIS)	78.57	21.43	11	3
2	Using of "Skin metabolism simulator" Combined parent and products requirements: Aldehydes (Protein binding alerts for skin sensitization by OASIS)	48.99	51.01	169	176
3	Using of "Skin metabolism simulator" Combined parent and products requirements: Bis aldehydes (Protein binding alerts for skin sensitization by OASIS)	82.35	17.65	14	3
4	Using of "Skin metabolism simulator" Combined parent and products requirements: No alert found (Protein binding alerts for skin sensitization by OASIS)	45.35	54.65	575	693

1) Go to the ***Consistency check*** section of the report; **2)** Click the **Add/Remove** button below the *Justification for selected mechanistic similarity profiles/metabolisms*. **3)** The *Report basket* appears. **4)** Check the box corresponding to the *Alert performance.* Right click over it and select preview to see the stored information. **5)** Finally confirm the selection by OK.

⚠ **If the Alert performance is calculated more than once by setting different searching criteria, information for the latest calculation will be stored in the Report basket.**

Report
Generating a prediction report



1) The Alert performance item appears below the other automatically included items.

2) Click the **Create report** button to generate the report files. The AP item will be included in the *Category report* file.

Congratulation

- You have now been introduced to the defining of a target endpoint;
- You have now been introduced to the calculation of the alert performance accounting for a metabolism;
- You have now been introduced to the Report basket;
- Note proficiency comes with practice.