QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.4.1

Evaluating alert performance accounting for the metabolic activation of chemicals

Outlook

- **Background**
- Keywords
- 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 taking into account the performance of the alerts identified in the target structure or in its metabolites.

Outlook

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

Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options

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).

ALERT PERFORMANCE – Performance of an alert to predict the target endpoint based on the chemicals in the selected databases having the same alert and experimental data available for them. The alert performance (AP) is also of help for selection the most appropriate alert(s) for defining categories.

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

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Define a target endpoint;
- Illustration of relevancy of the profilers and data availability;
- Define the primary group (i.e. searching for analogues) by accounting for a metabolism;
- Calculation of an alert performance (AP) accounting for the metabolic activation of the chemicals;

Outlook

- Background
- Keywords
- 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 the metabolic activation of the chemicals;
- To explain to the Toolbox user the rationale behind each step of the exercise.

Outlook

- Background
- Keywords
- 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 can 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 calculate the performance.

Outlook

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

The Exercise

- In this exercise the skin sensitization potential of *1,3-Propanediamine, N-(3-aminopropyl)* [CAS# 56-18-8] (i.e. the target chemical) will be predicted.
- The target endpoint (i.e. Skin sensitization) will be preliminary defined.
- Category will be defined accounting for the skin metabolism of the chemicals.
- Alert performance (AP) 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

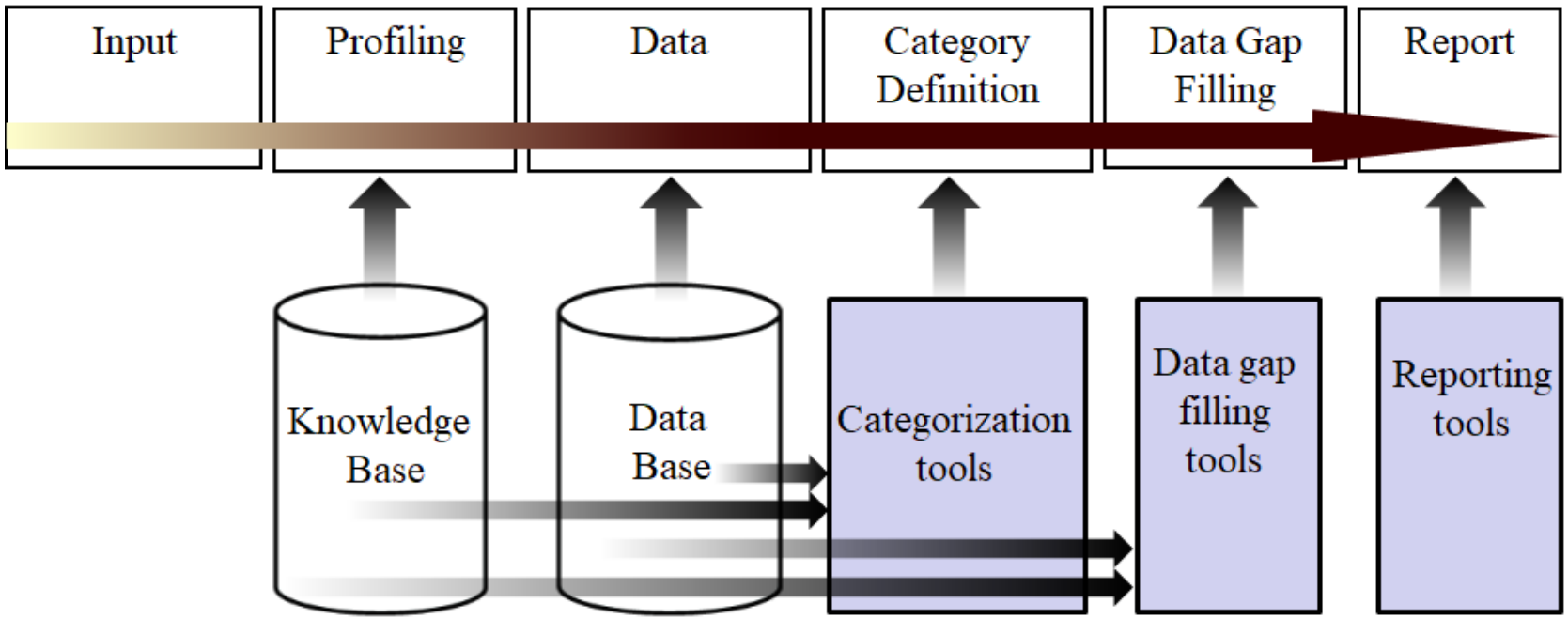
- Background
- Keywords
- Objectives
- Specific Aims
- Alert performance
- The exercise
- **Workflow**

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

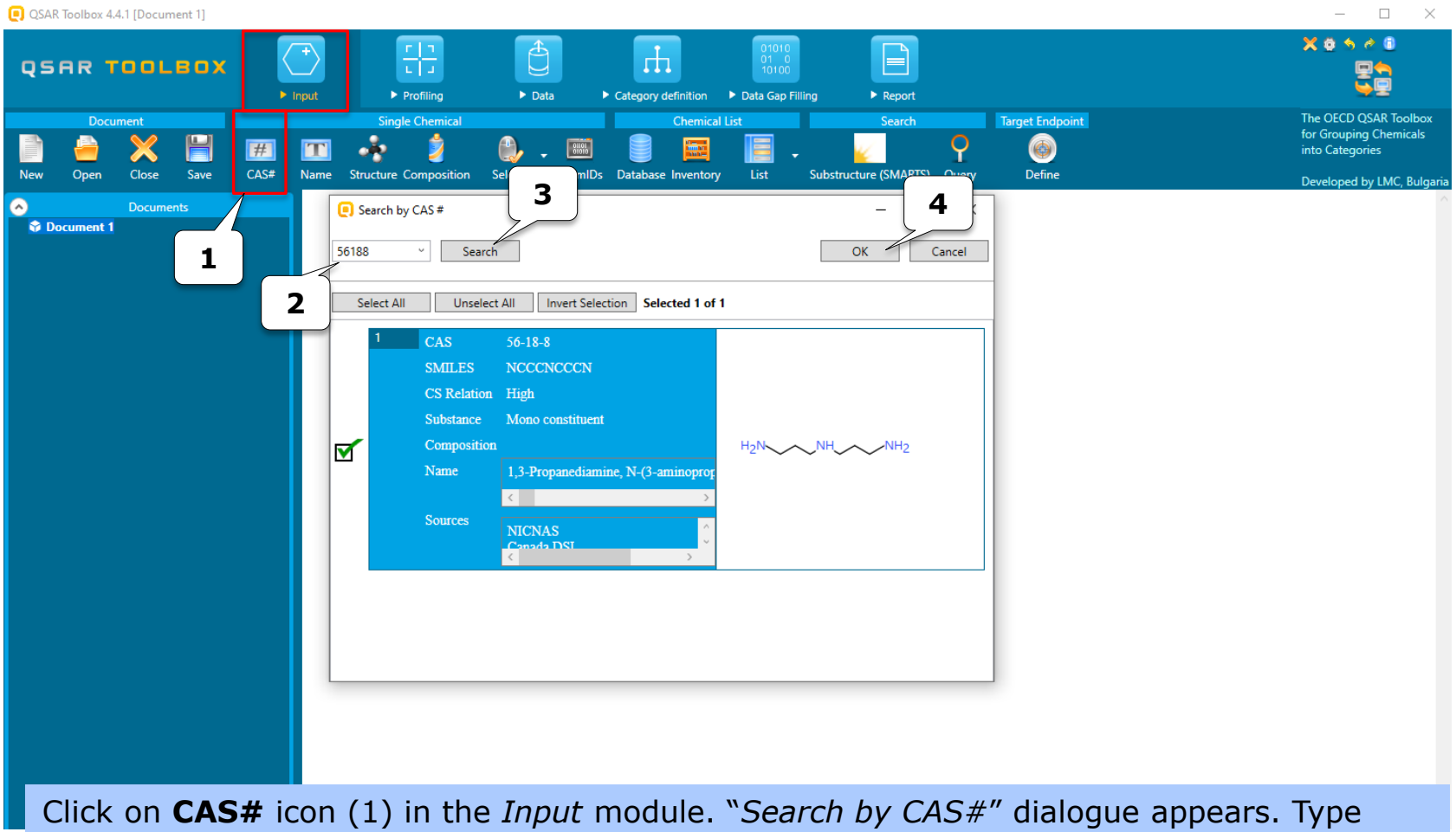
Scheme illustrating the Toolbox workflow



Input
Overview

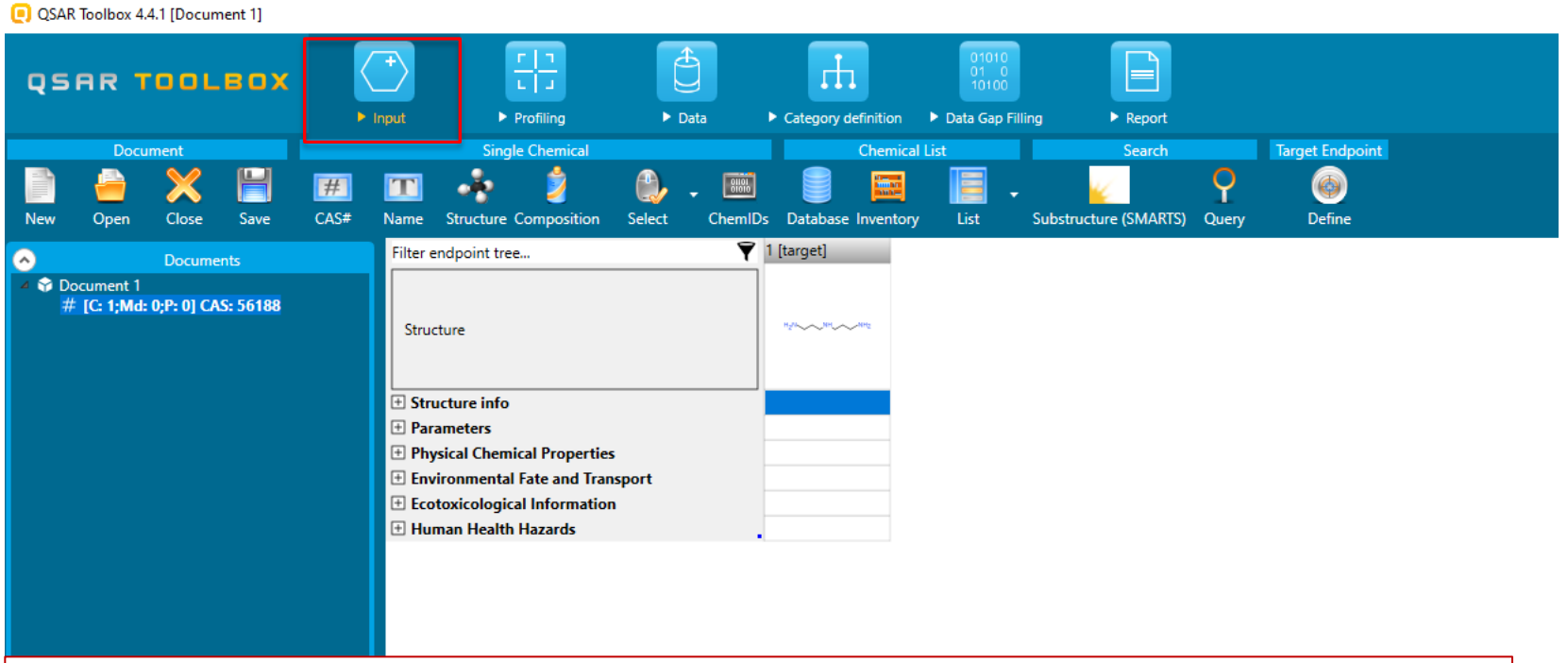
- This module provides the user with several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on the chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.
- Other key functionalities such as “Define a target endpoint” are also placed in the Input module.
- Possibility to enter the chemical via additional chemical ID (e.g. EC number) is also available since version 4.3 of the Toolbox.

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).

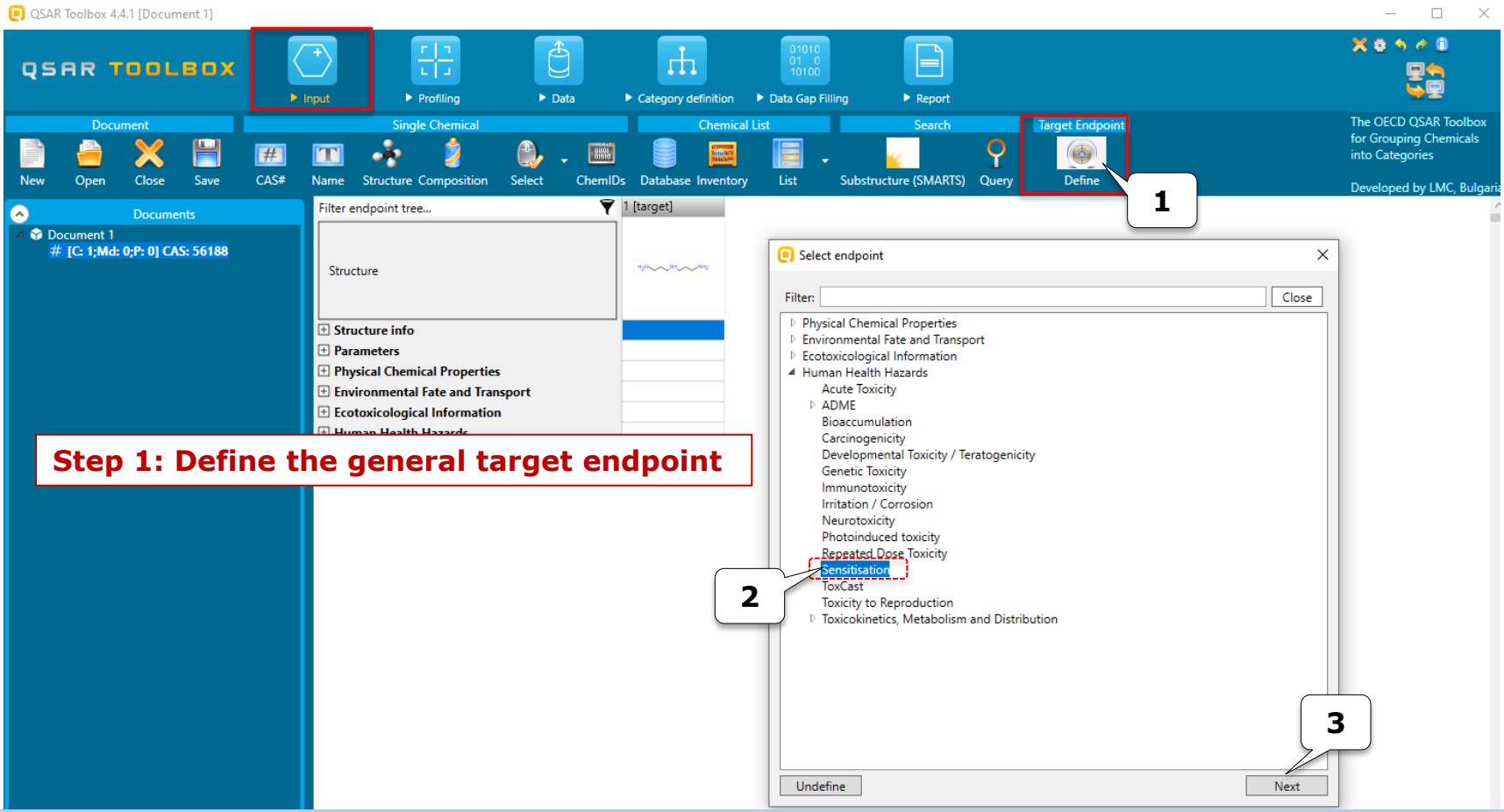
Input
Define the target endpoint



Defining of the target endpoint is needed for calculation of AP later on. This is a two-step way process:

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

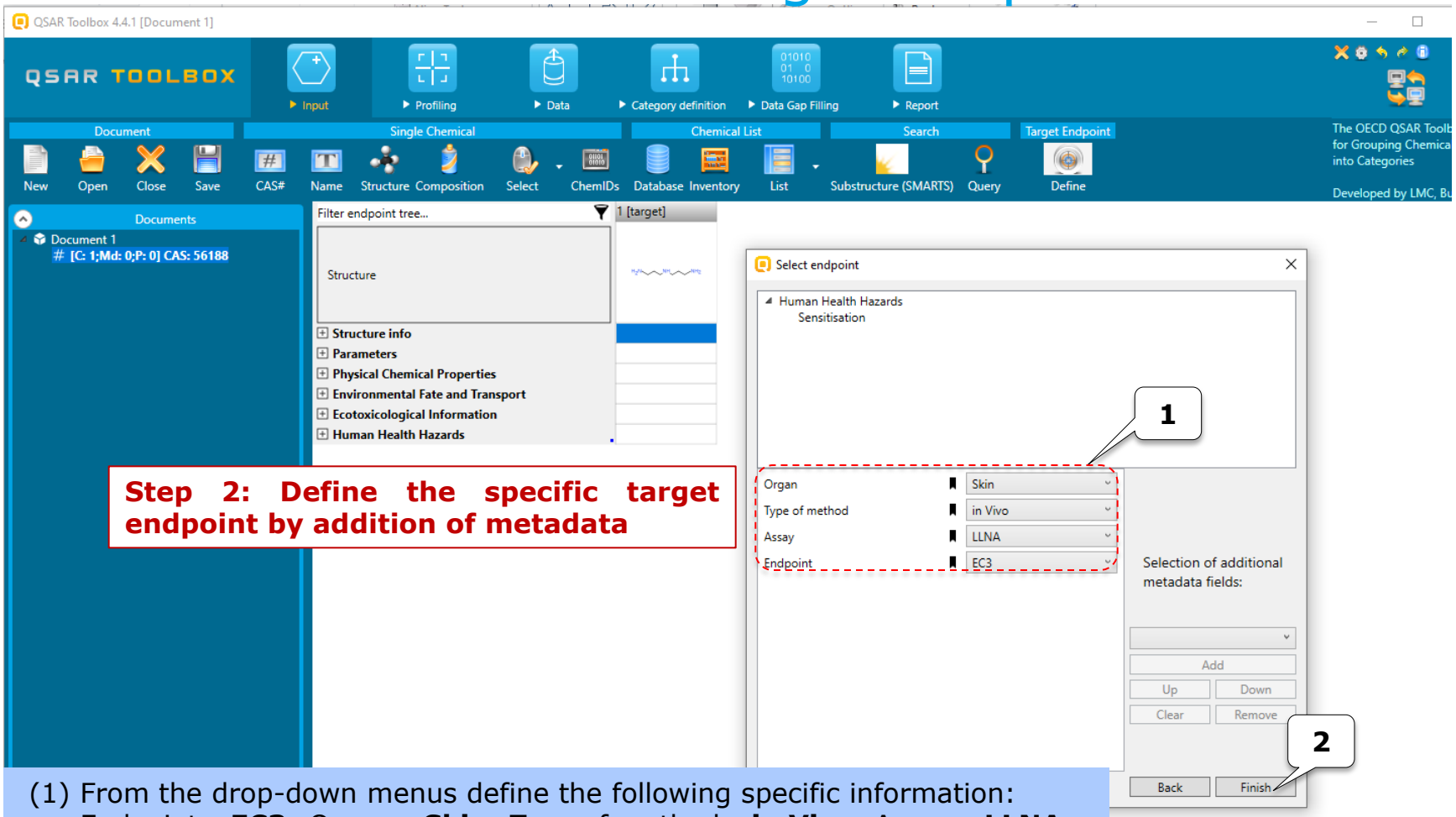
Input
Define the target endpoint



Click on the **Define** icon (1). The "Select endpoint" dialogue appears. First select the general target endpoint - ***"Sensitisation"*** (2) located under the Human Health Hazard and then click on **Next** (3)

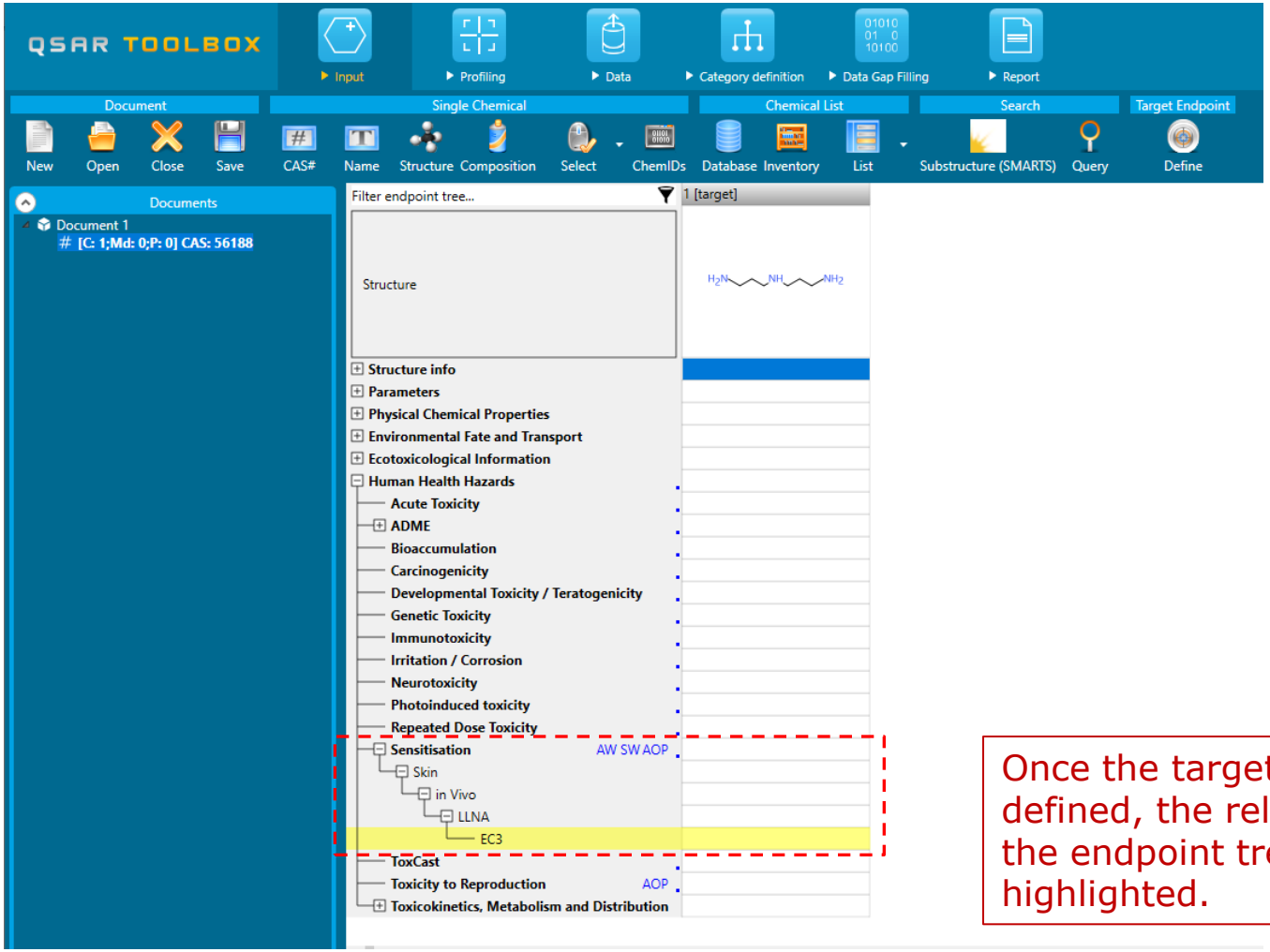
Input

Define the target endpoint



(1) From the drop-down menus define the following specific information:
Endpoint - **EC3**; Organ - **Skin**; Type of method - **in Vivo;** Assay - **LLNA**
(2) Click on **Finish.**

Input
Define target endpoint



Once the target endpoint is defined, the related row of the endpoint tree is highlighted.

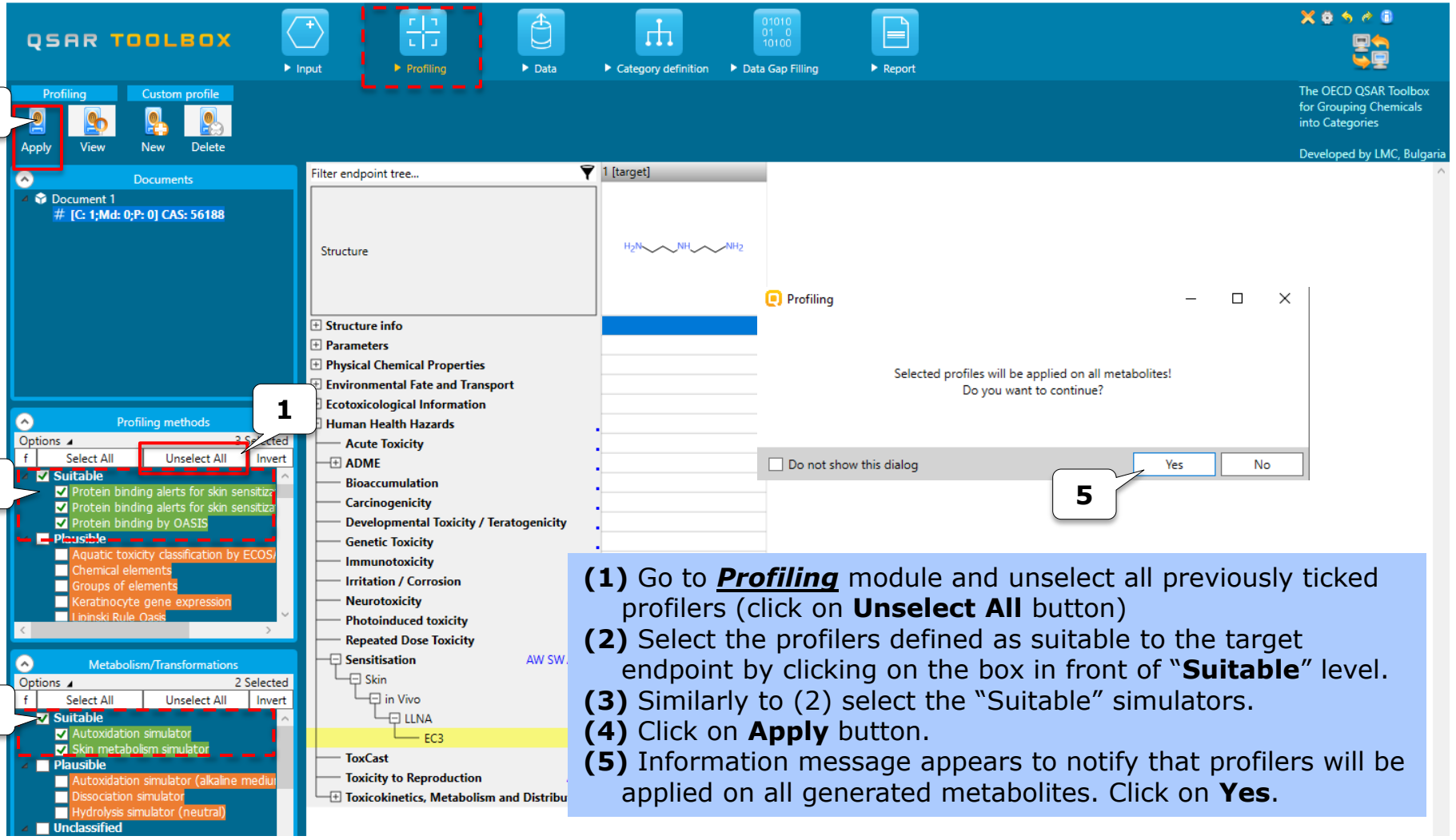
Profiling
Overview

- As you know, the "Profiling" refers to the electronic process of retrieving relevant information on a compound which is stored in the Toolbox, other than fate and (eco)toxicity data.
- The available information includes likely mechanism(s) of action, as well as observed and/or simulated metabolites.
- Based on the already defined target endpoint (slides #20-22) the profilers and metabolism simulators are automatically grouped by their relevancy to the endpoint*.

*More details regarding the grouping of profilers by relevancy could be seen in tutorial *"Example for predicting skin sensitisation taking into account alert performance".*

Profiling

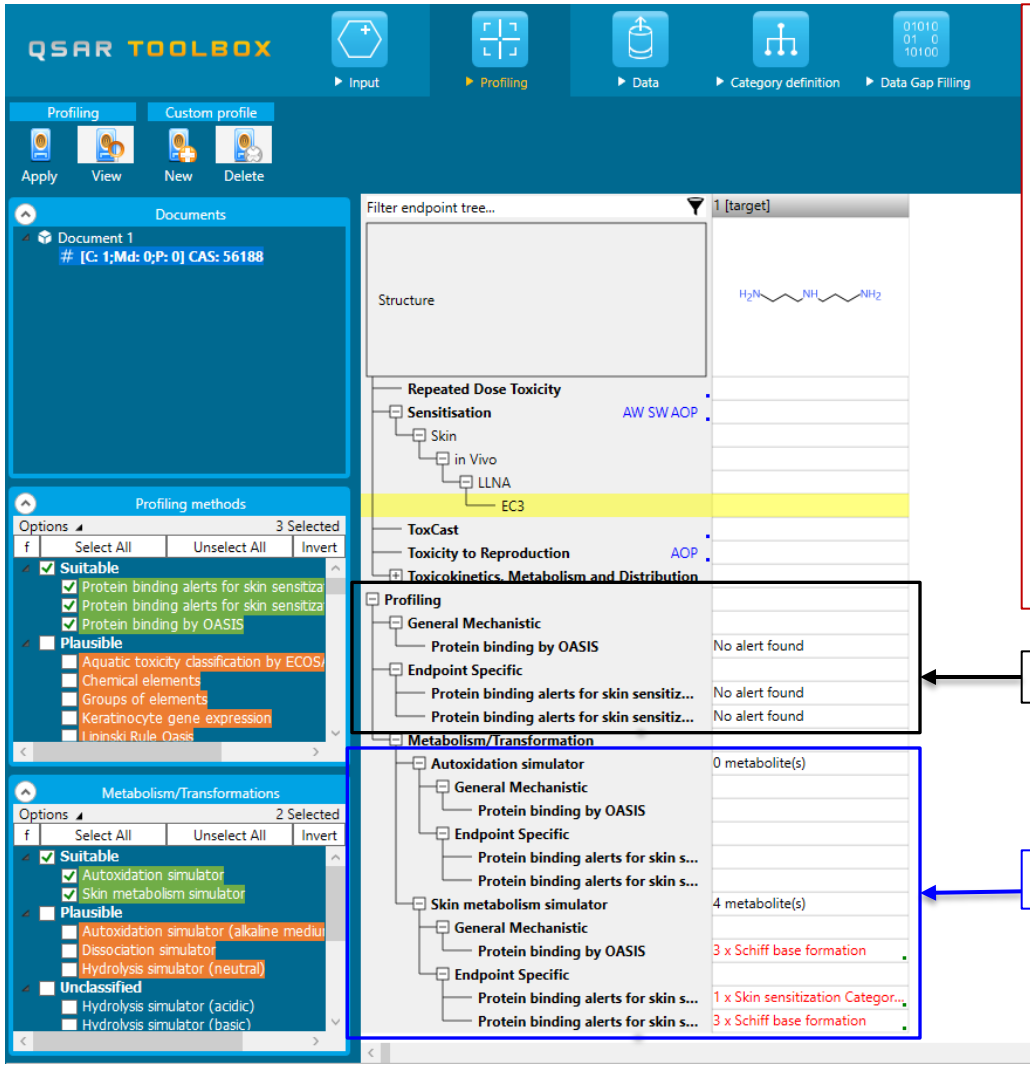
Profiling the target



(1) Go to ***Profiling*** module and unselect all previously ticked profilers (click on **Unselect All** button)
(2) Select the profilers defined as suitable to the target endpoint by clicking on the box in front of "**Suitable**" level.
(3) Similarly to (2) select the "Suitable" simulators.
(4) Click on **Apply** button.
(5) Information message appears to notify that profilers will be applied on all generated metabolites. Click on **Yes**.

Profiling

Profiling the target



As a result of the applying of the profiles and metabolic simulators suitable to the target endpoint:

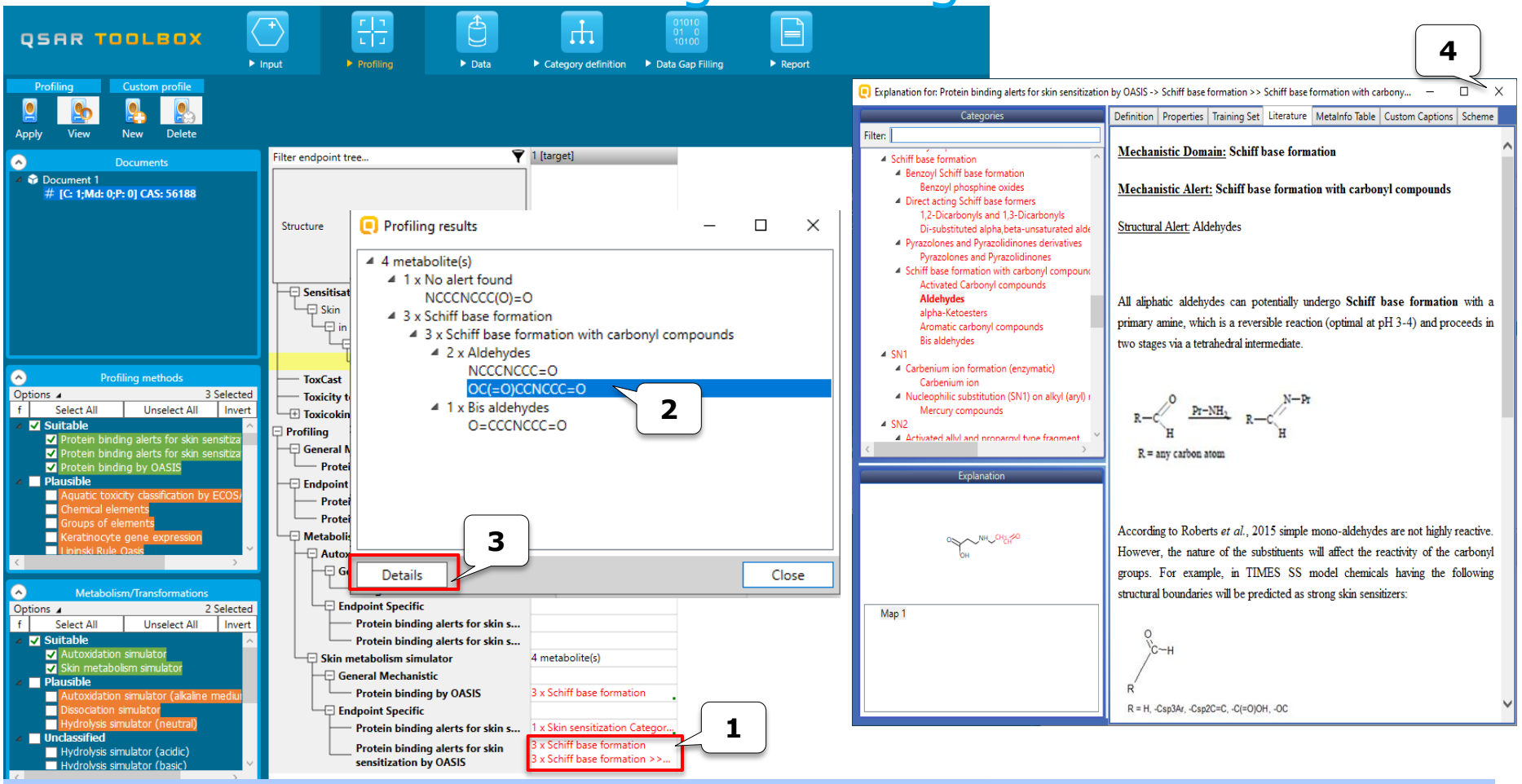
- **No alerts are found** in the target structure;
- The target chemical does not autoxidized.
- Four metabolites are simulated by the Skin metabolism simulator
- **Protein binding alerts are identified** in some of the generated skin metabolites.
- More details for the mechanism of interaction and some additional information is also provided (*see next page*)

Profiling result of the target structure

Profiling result of the generated metabolites

Profiling

Profiling the target



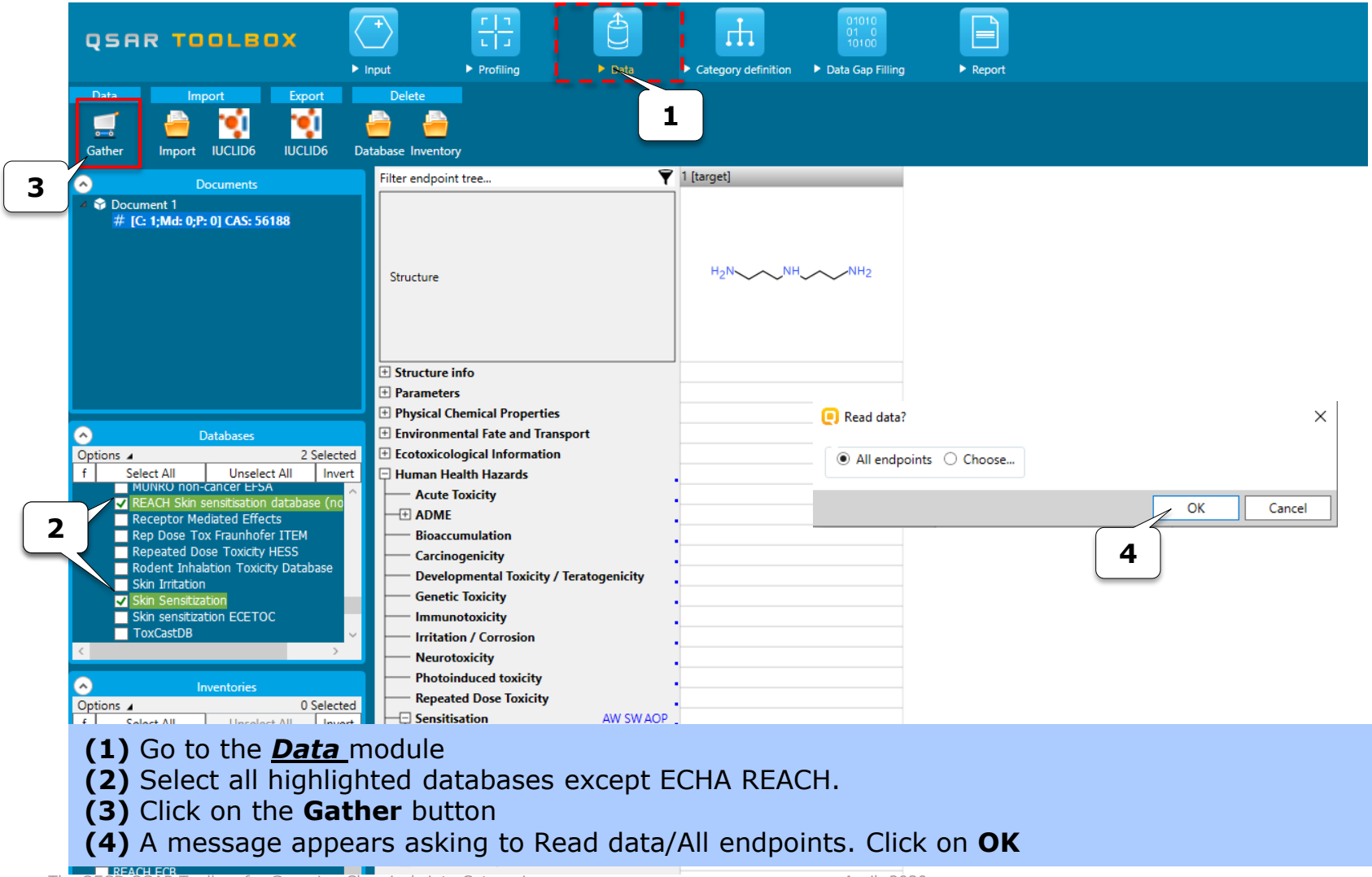
Double click on the cell with profiling a result **(1)** (or right click and select "Explain");
A new dialogue appears where the SMILES of the generated metabolites are provided along with the respective profiling result. Select a metabolite **(2)** and click on **Details (3)**. The explanation of the mechanism associated with the identified alert will be shown. Close the explanation window **(4)**.

Data
Overview

- "Data" module refers to the electronic process of retrieving environmental fate, ecotoxicity and toxicity data stored in the Toolbox databases.
- 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 single or 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

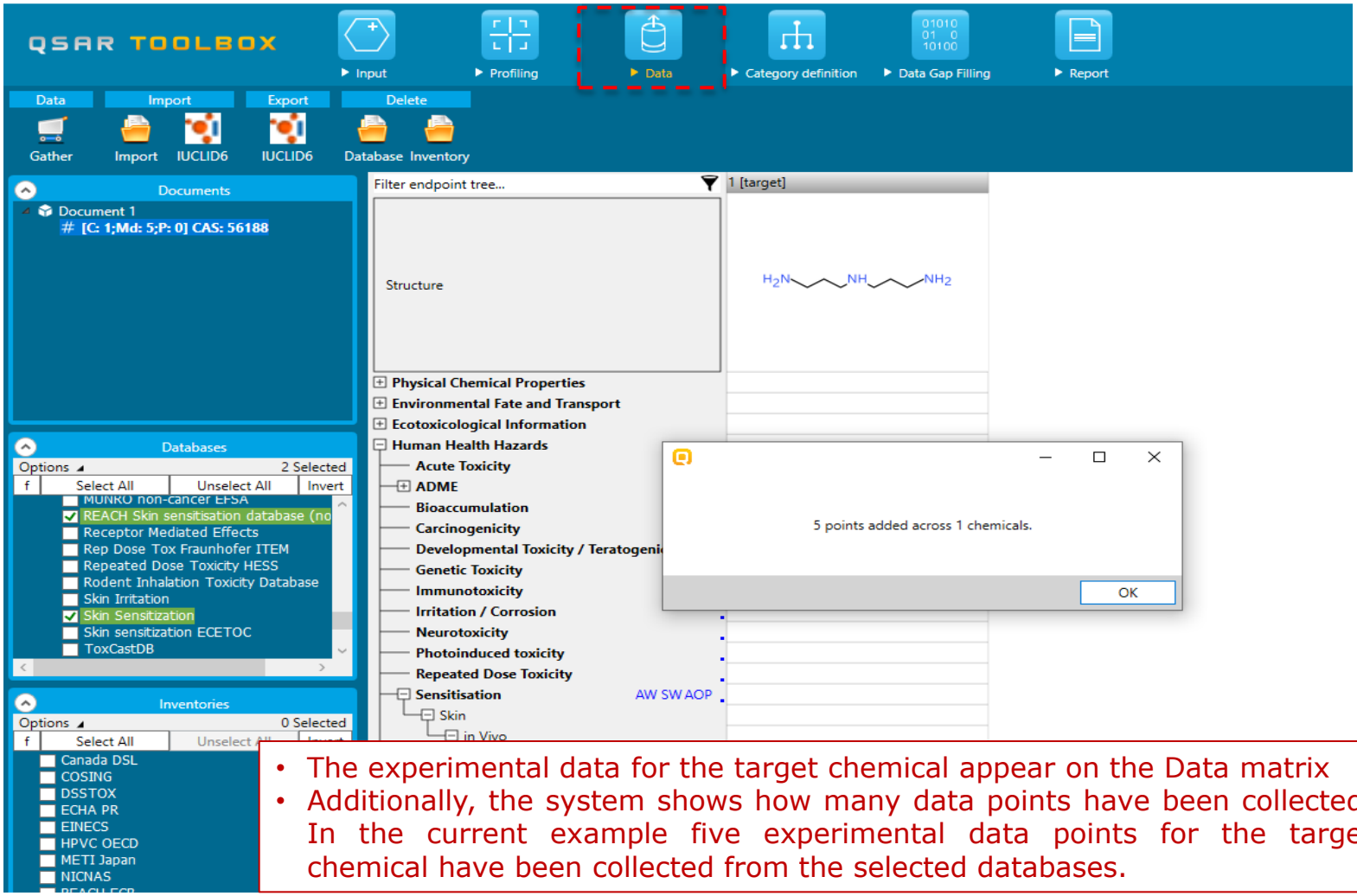
Collecting the experimental data



(1) Go to the ***Data*** module
(2) Select all highlighted databases except ECHA REACH.
(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 on the Data matrix
- Additionally, the system shows how many data points have been collected. In the current example five experimental data points for the target chemical have been collected from the selected databases.

Category Definition
Overview

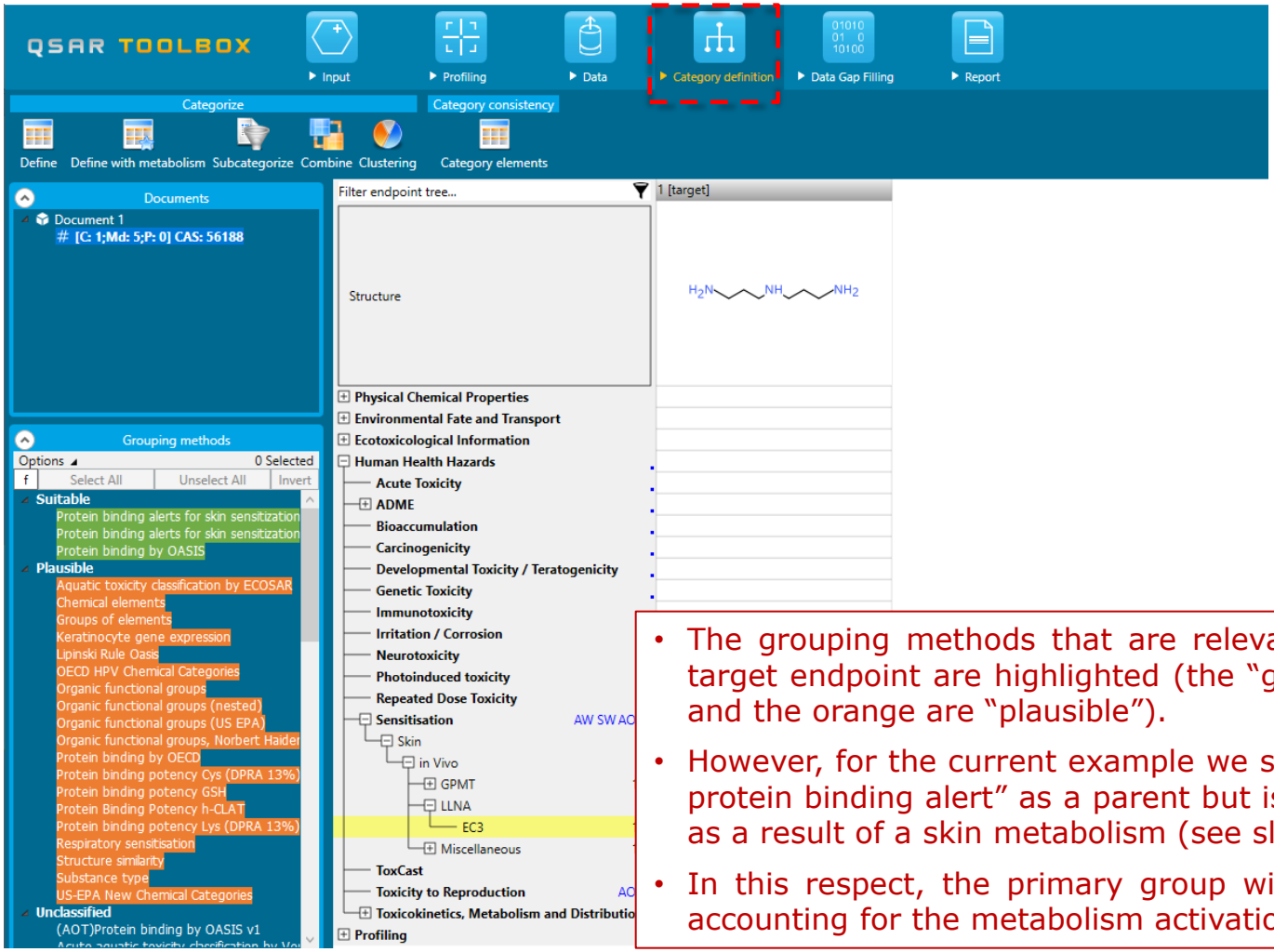
- The next step of the TB workflow is to collect analogues of the target chemical. This is the critical step in the workflow.
- The analogues search happens in the Category definition module.
- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- 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 in the form of profilers appear here as grouping methods.
- For example, if a specific protein binding mechanism is identified in the target chemical, it could be used for searching analogues acting by the same mechanism of interaction with biomacromolecules.
- The profilers, relevant to the defined target endpoint, are highlighted and could be used to define the category.
- The group could be defined accounting for the metabolic activation of the chemicals in the cases where an alert is identified in some of the target` metabolites.

Category Definition

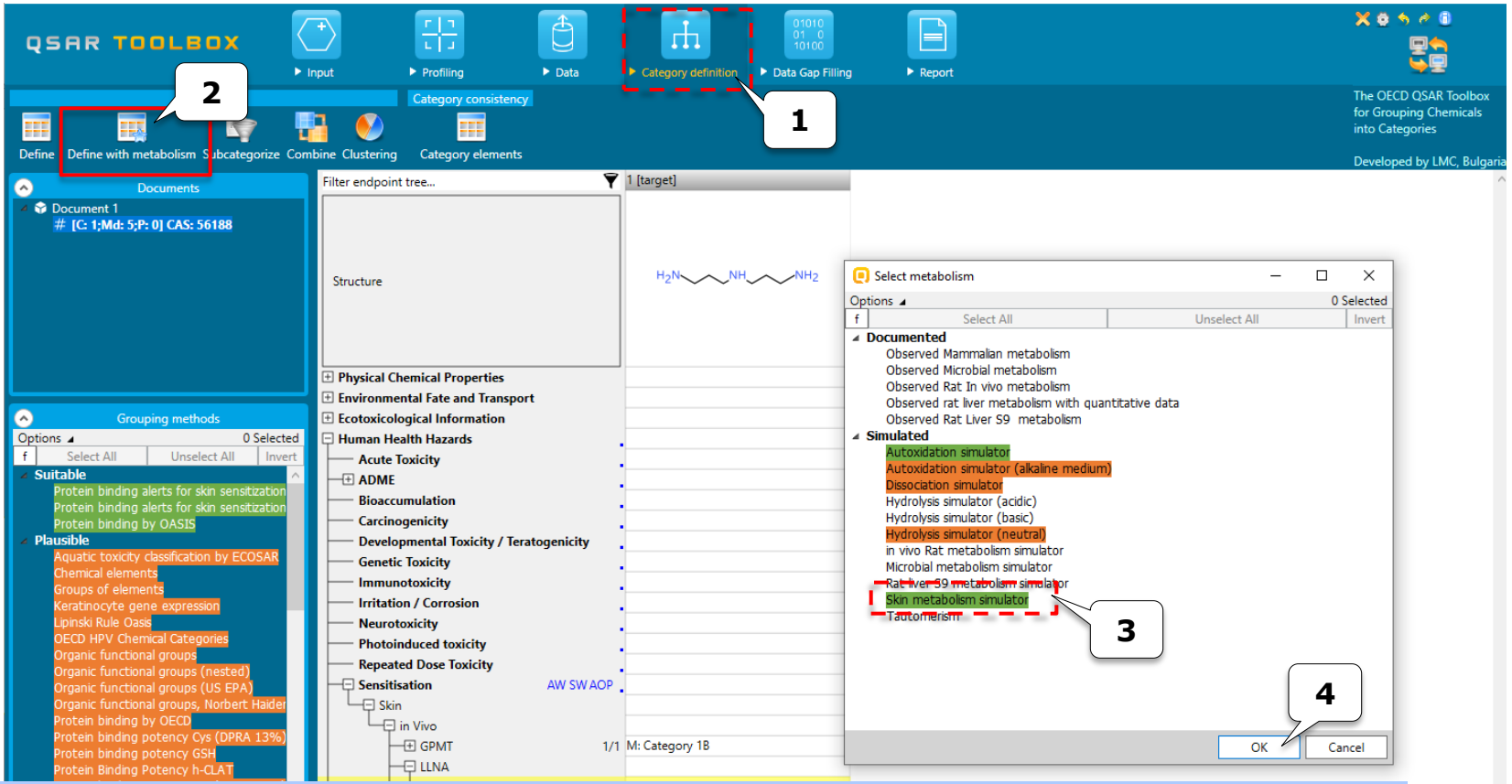
Define the category accounting for a metabolism



- 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 slides 24-26).
- In this respect, the primary group will be defined with accounting for the metabolism activation of the target.

Category Definition

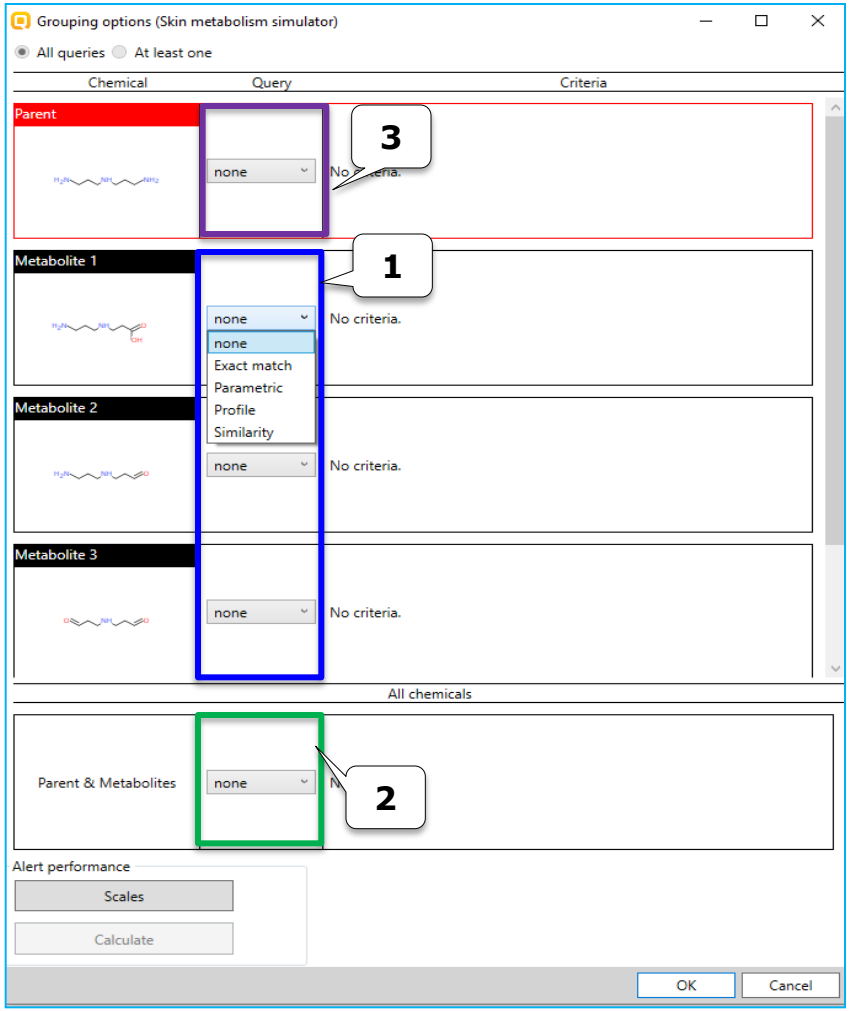
Define the category accounting for a metabolism



(1) Move to the ***Category definition*** module;
(2) Click on the **Define with metabolism** button;
A new dialogue appears with all available documented and simulated metabolic simulators highlighted in the respective color (suitable – green, plausible-orange).
(3) Select the ***Skin metabolism simulator***;
(4) Click **OK** to confirm the selection.

Category Definition

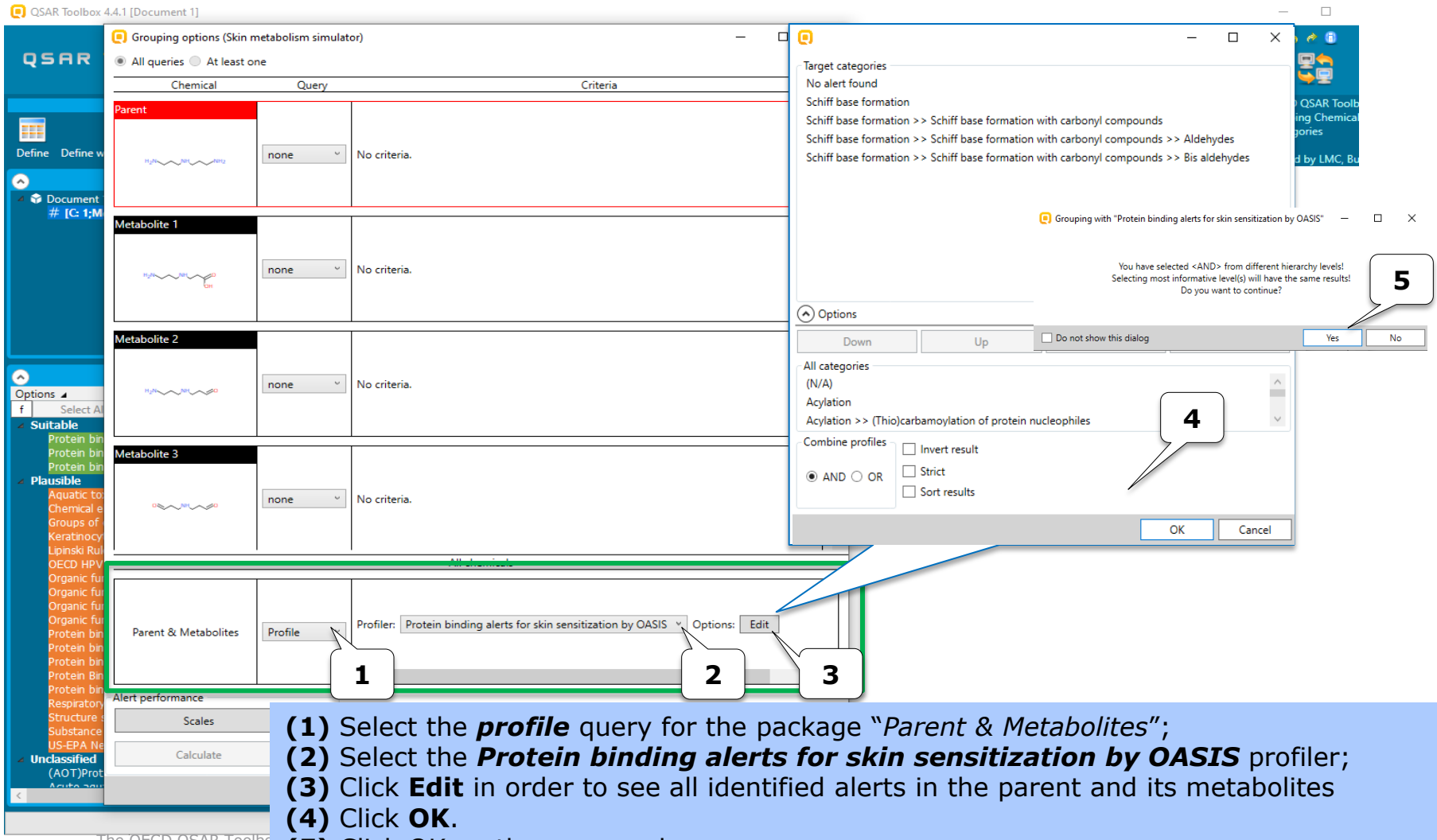
Define the group with accounting for a metabolism



- The newly appeared window shows the parent and all generated metabolites produced by the selected metabolic simulator (*Skin metabolism 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 (3) could be also defined as an addition. However, 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 chemicals having exactly the same metabolite;
  - ***Parametric*** – the searched chemical (parent or metabolite) to have specific value or a range of variation of a defined parameter (a list with all parameters currently available in the Toolbox is provided);
  - ***Profile*** – the searched chemical (parent or metabolite) to have a specific category by the selected profiler (a list with all profilers is provided);
  - ***Structural*** – the searched chemical (parent or metabolite) to be structurally similar to the current chemical above a defined threshold.
- Calculation of the AP will take into account all defined criteria.

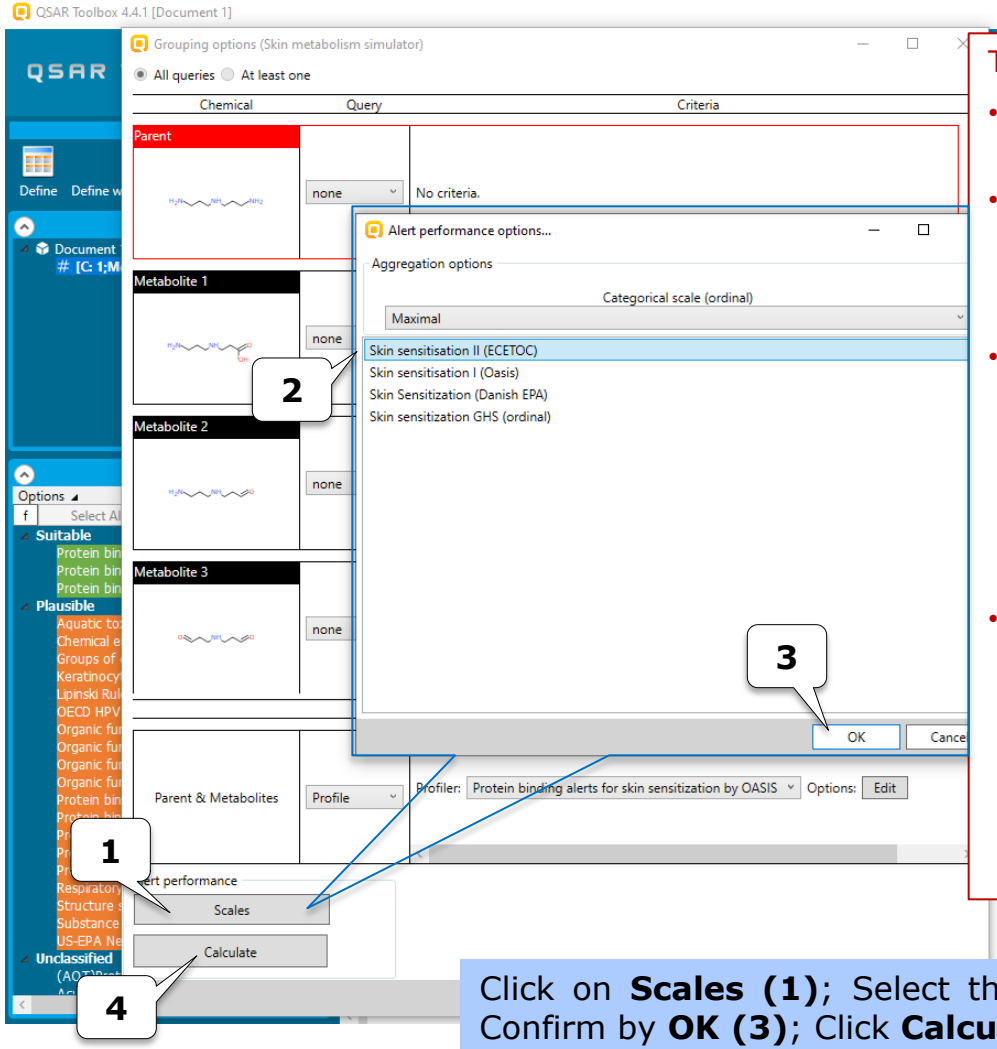
Category Definition

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
(4) Click **OK**.
(5) Click OK on the appeared message

Category Definition
Calculation of the alert performance

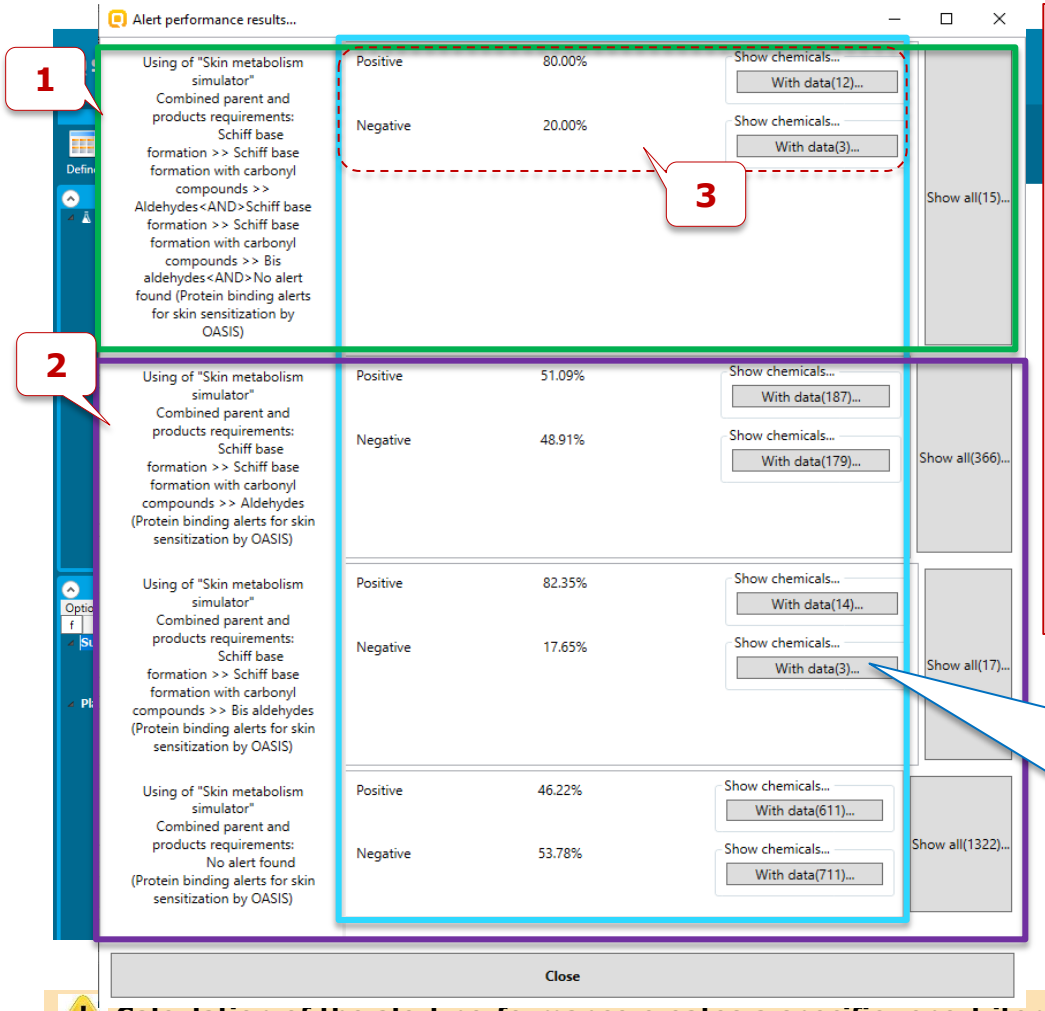


The alert performance results depend on:

- **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 in our case.
- **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.
- **Aggregation options** – this options 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.

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

Category Definition
Calculation of the alert performance



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

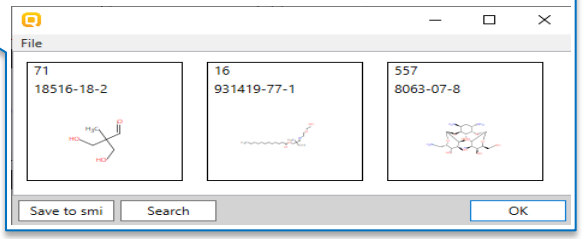
1) The AP statistic accounting for all set criteria and all identified alerts in case of a selected *profile* query.

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 used for the statistic. The user is also able to see the corresponding chemicals by single click (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.

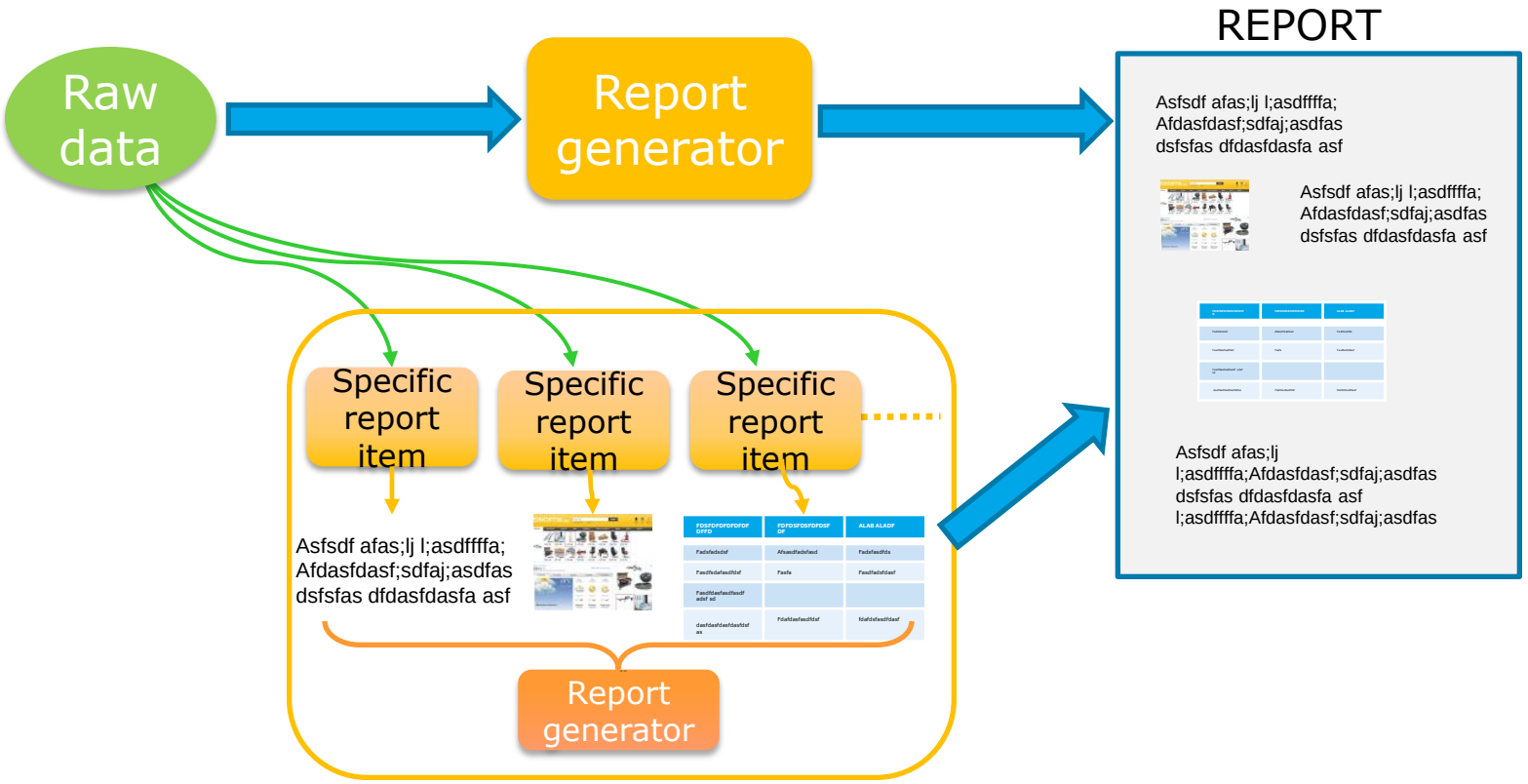
You can see the chemicals used for the AP calculation by single click



Calculation of the alert performance creates a specific report item stored in the so-called *Report basket (see the next slide)*.

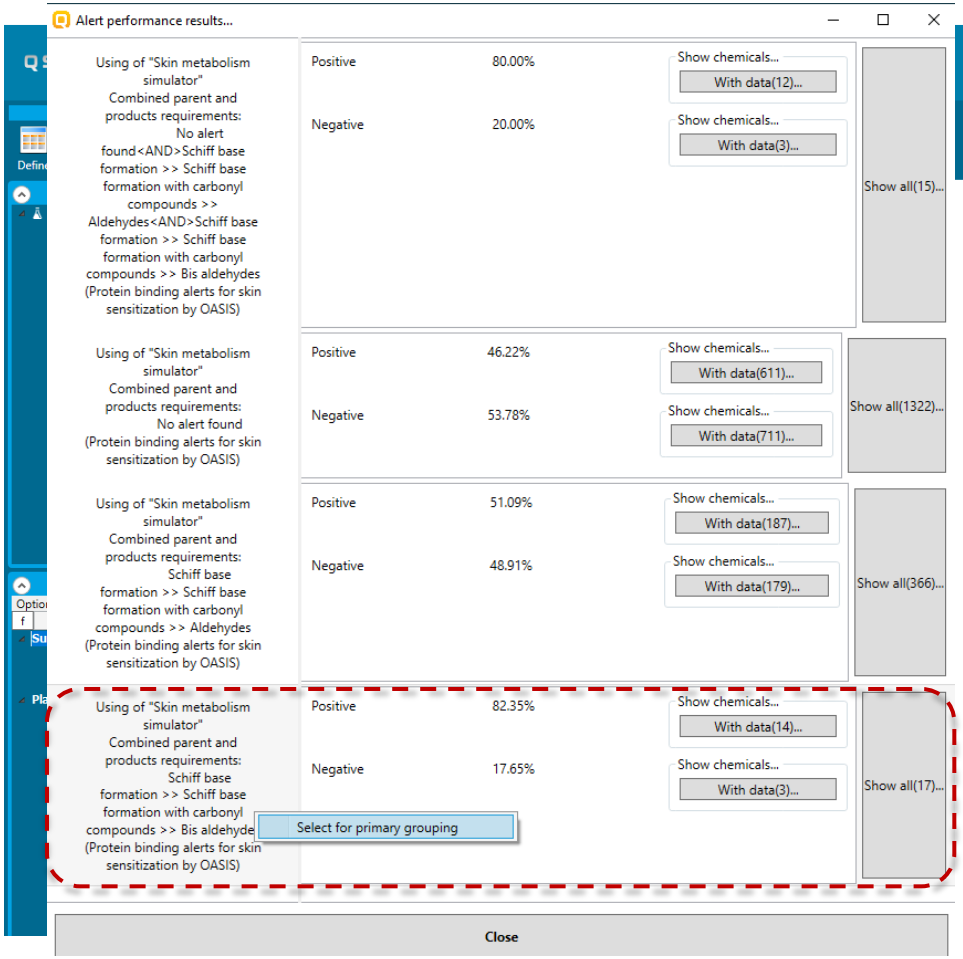
Sidebar on the Report basket

- Specific report items are collected during the workflow (e.g. calculation of AP). All items are stored in the "*Report basket*" and can be used in the report to support or justify the consistency of a category.
- Items with external content - text and/or image (e.g. schemes of mechanisms of interactions, metabolic maps, snapshots from external modeling sources, etc.) could be also created within the Report basket and further included to the report.



Category Definition

Selection of an alert for category definition



*By analyzing the results, we can see that the **Bis aldehydes** alert shows the best predictability (82%) with respect to the defined endpoint and selected databases (i.e. 14 out of 17 chemicals having the same alert are positive by experimental data).*

So, we can use only this alert for category definition. The category will consists of chemicals having Bis aldehyde alert as parents or as a result of skin metabolism.

(1) Apply right click on the row for Bis aldehydes alert and click ***Select for primary grouping***.

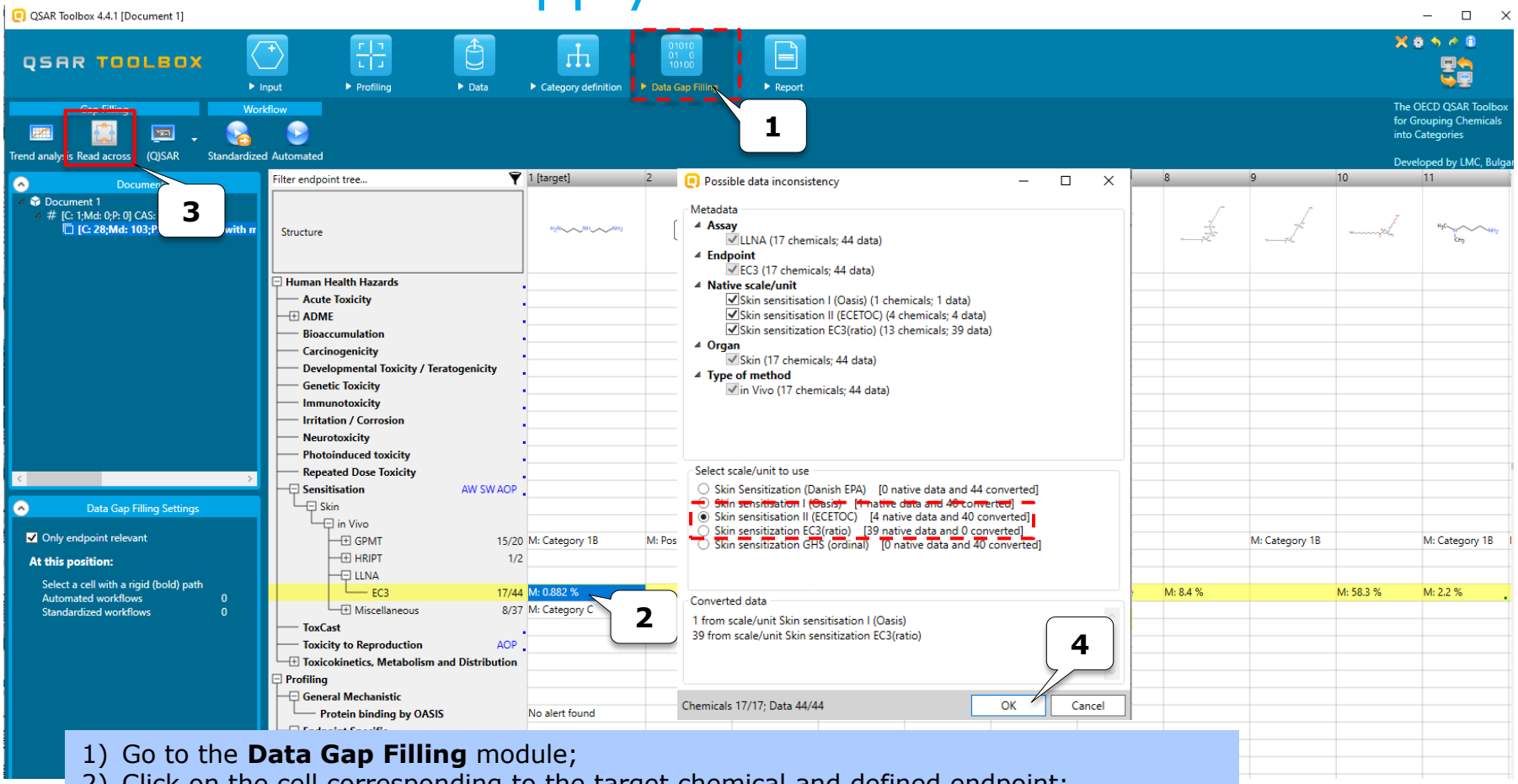
Data Gap Filling
Overview

- "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 the prediction.

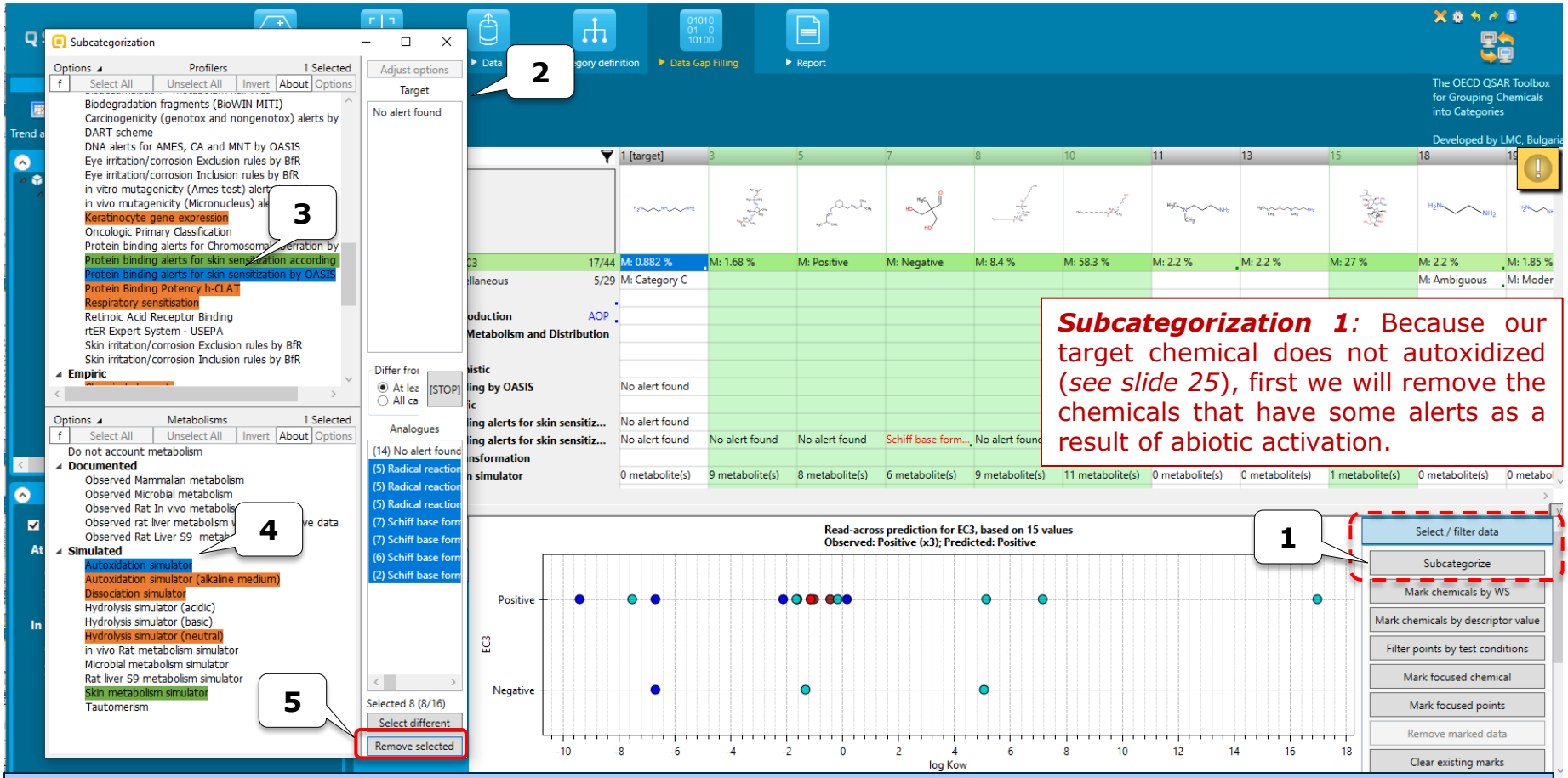
In this example we will use the read-across approach.

Data Gap Filling
Apply Read across



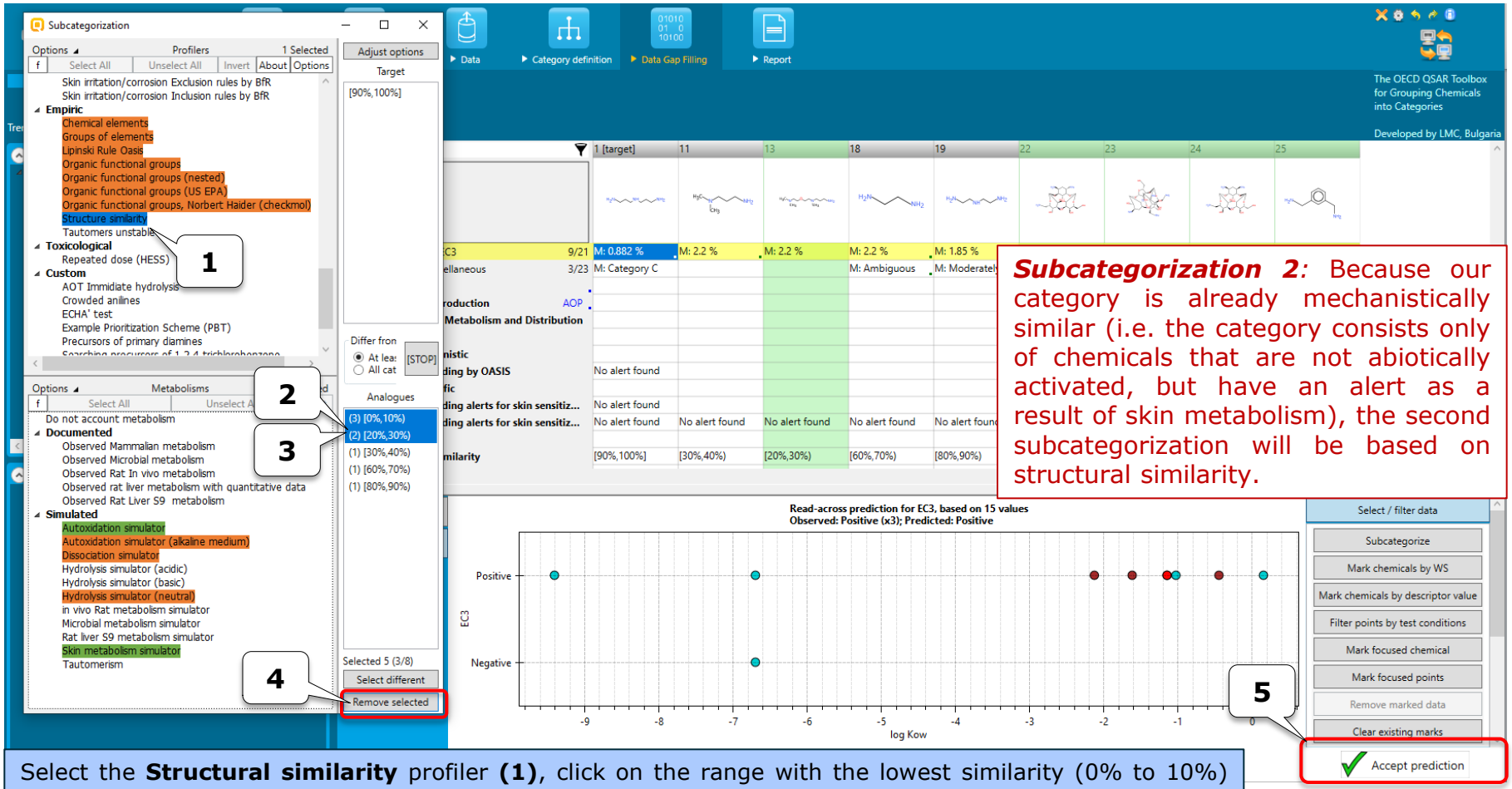
1) Go to the **Data Gap Filling** module;
2) Click on the cell corresponding to the target chemical and defined endpoint;
3) Click the **Read across** button;
4) *Skin sensitisation II (ECETOC)* scale is selected by default. Confirm by **OK**.

Data Gap Filling
Subcategorization



In order to subcategorize your category (i.e. to keep only the most similar analogues to the target chemical) Go to ***Select / filter data >> Subcategorize*** **(1)**. The Subcategorization window with all available profilers will appear **(2)**. Select the ***Protein binding alerts for skin sensitization by OASIS*** profiler **(3)** and then the ***Autoxidation simulator*** **(4)**. The chemicals different to the target will be highlighted. Remove them by click on **Remove selected** button **(5)**.

Data Gap Filling
Subcategorization



Subcategorization 2: Because our category is already mechanistically similar (i.e. the category consists only of chemicals that are not abiotically activated, but have an alert as a result of skin metabolism), the second subcategorization will be based on structural similarity.

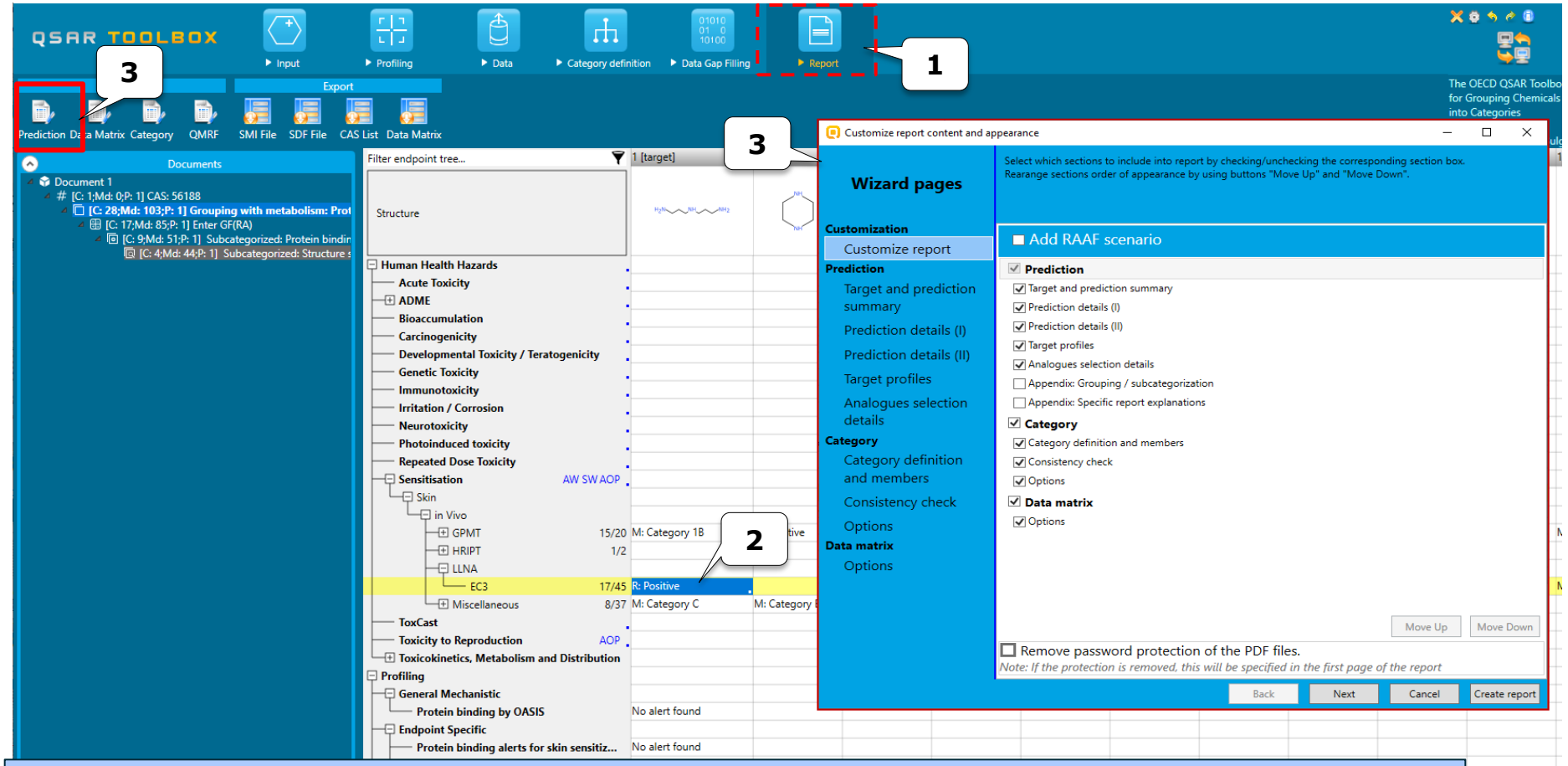
Select the **Structural similarity** profiler **(1)**, click on the range with the lowest similarity (0% to 10%) **(2)**, press the *Ctrl* button of the keyboard and click on the second similarity range (20% to 30%) **(3)**. Remove the selected dissimilar chemicals by click on **Remove selected** button **(4)**. Finally, click on **Accept prediction** button **(5)**.

Report
Overview

- The report module can generate a report on predictions performed within 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 selected. 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.

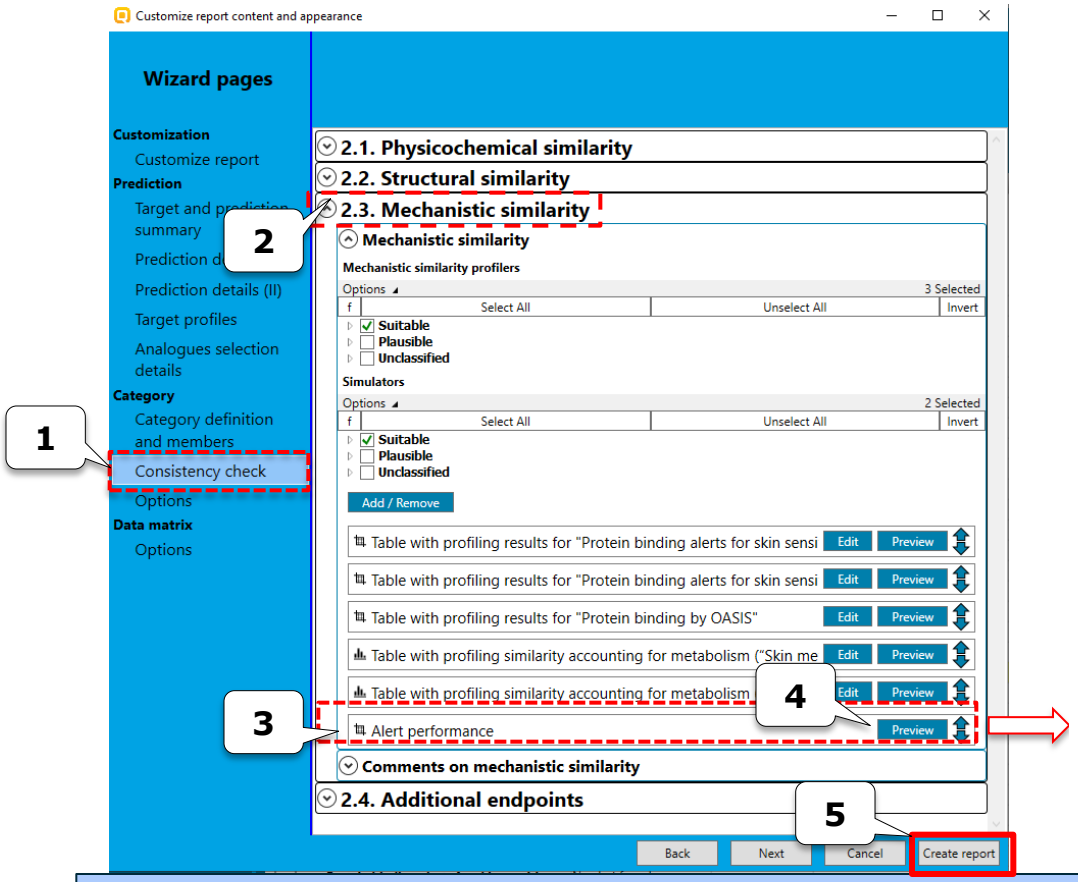
Report

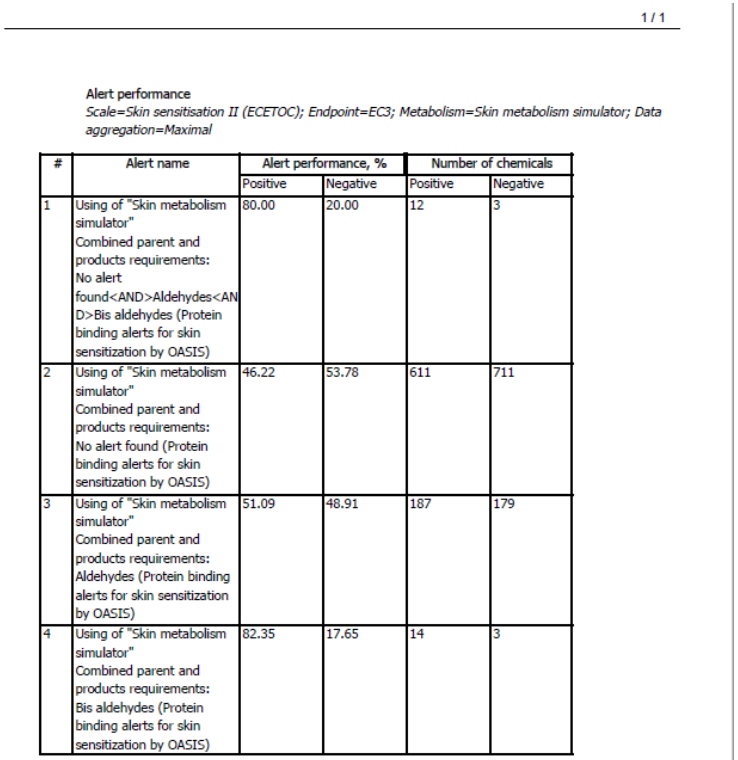
Generating a prediction report



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

Report
Generating a prediction report





1 / 1

Alert performance

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

#	Alert name	Alert performance, %		Number of chemicals	
		Positive	Negative	Positive	Negative
1	Using of "Skin metabolism simulator" Combined parent and products requirements: No alert found<AND>Aldehydes<AND>Bis aldehydes (Protein binding alerts for skin sensitization by OASIS)	80.00	20.00	12	3
2	Using of "Skin metabolism simulator" Combined parent and products requirements: No alert found (Protein binding alerts for skin sensitization by OASIS)	46.22	53.78	611	711
3	Using of "Skin metabolism simulator" Combined parent and products requirements: Aldehydes (Protein binding alerts for skin sensitization by OASIS)	51.09	48.91	187	179
4	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

1) Go to the ***Consistency check*** section of the report; **2)** Open section ***2.3. Mechanistic similarity;*** **3)** The ***Alert performance*** item appears below the other automatically included items; **4)** Click **Preview** button to see the information of this item will look like in the generated report; **5)** Finally, click the **Create report** button to generate the report files. The AP item will be included in the Category report file.

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.

Congratulation

- You now know how to define a target endpoint;
- You now know how to calculate the alert performance accounting for the metabolic activation of the chemicals;
- You now know where the Alert performance item appears in the report;
- Note, proficiency comes with practice!