QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.1

Tutorial on how to predict Skin sensitization potential taking into account alert performance

Outlook

- **Background**
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow

Background

- This is a step-by-step presentation designed to take the Toolbox user through the workflow of a data filling exercise accounting alert performance.

Outlook

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

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Define target endpoint;
- Relevancy of profiles and data availability;
- Calculation of alert performance (AP).

Outlook

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

Specific Aims

- To introduce to the Toolbox user the workflow of the defining target endpoint;
- To familiarize the user with the new interface of the Toolbox;
- To familiarize the user with the coloring of the profiles and databases;
- To familiarize the user with the calculation of alert performance;
- 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
Overview

- Alert performance (AP) is used to define how much relevant to a target endpoint an alert is;
- AP reflects the alerts usability for category formation;
- AP can be calculated for any endpoint and any profile; one should only have preliminary defined target endpoint;
- AP can be calculate for an alert with or without accounting metabolism;

Outlook

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

The Exercise

- In this exercise we will predict the skin sensitization potential for an untested compound, (*pyridaphenthion*) [CAS# 119-12-0], which will be the "target" chemical.
- We will preliminary define the target endpoint.
- This prediction will be accomplished by calculation of alert performance.
- The category will be defined by the mechanism of protein binding common to all the chemicals in the category.
- The prediction itself will be made by "read-across".

The Exercise
Sidebar On Sensitization

- Allergic contact dermatitis that results from skin sensitization is a significant health concern.
- Skin sensitization is a toxicological endpoint that is complex and conceptually difficult.
- However, there is growing agreement that most organic chemicals must react covalently with skin proteins in order to behave as skin sensitizers.
- Therefore, mechanisms by which organic chemicals bind with proteins are relevant to grouping chemicals that may be skin sensitizing agents.

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

Outlook

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

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.

Input
Ways of Entering a Chemical

User Alternatives for Chemical ID:

A. Single target chemical

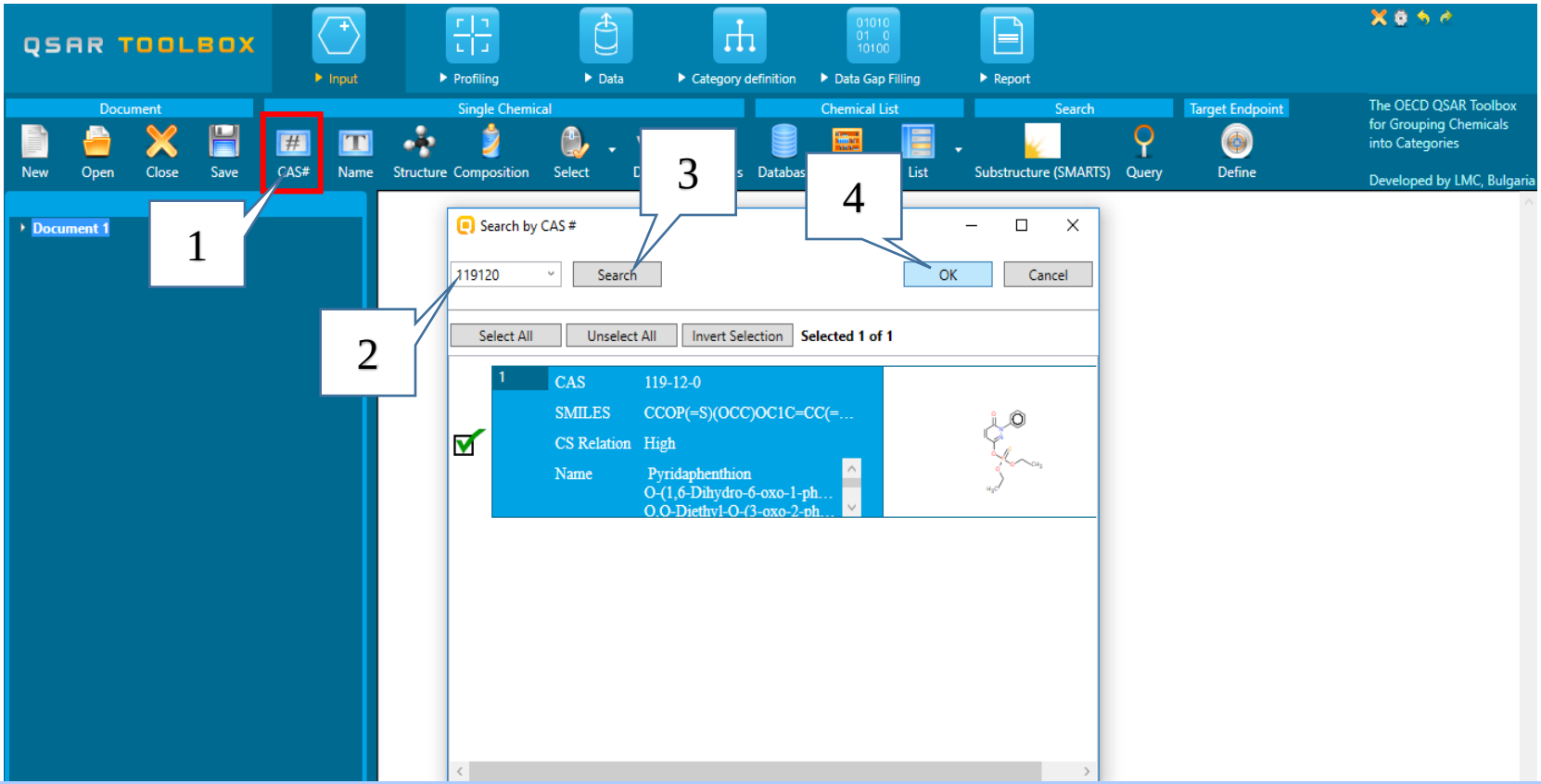
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Input Screen

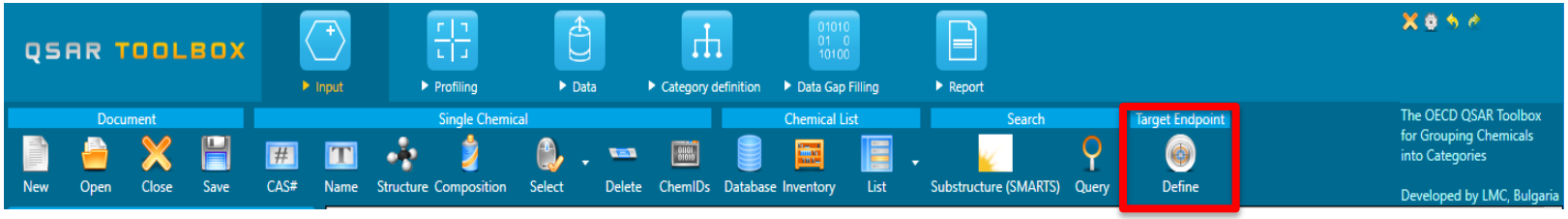
Input target chemical by CAS#



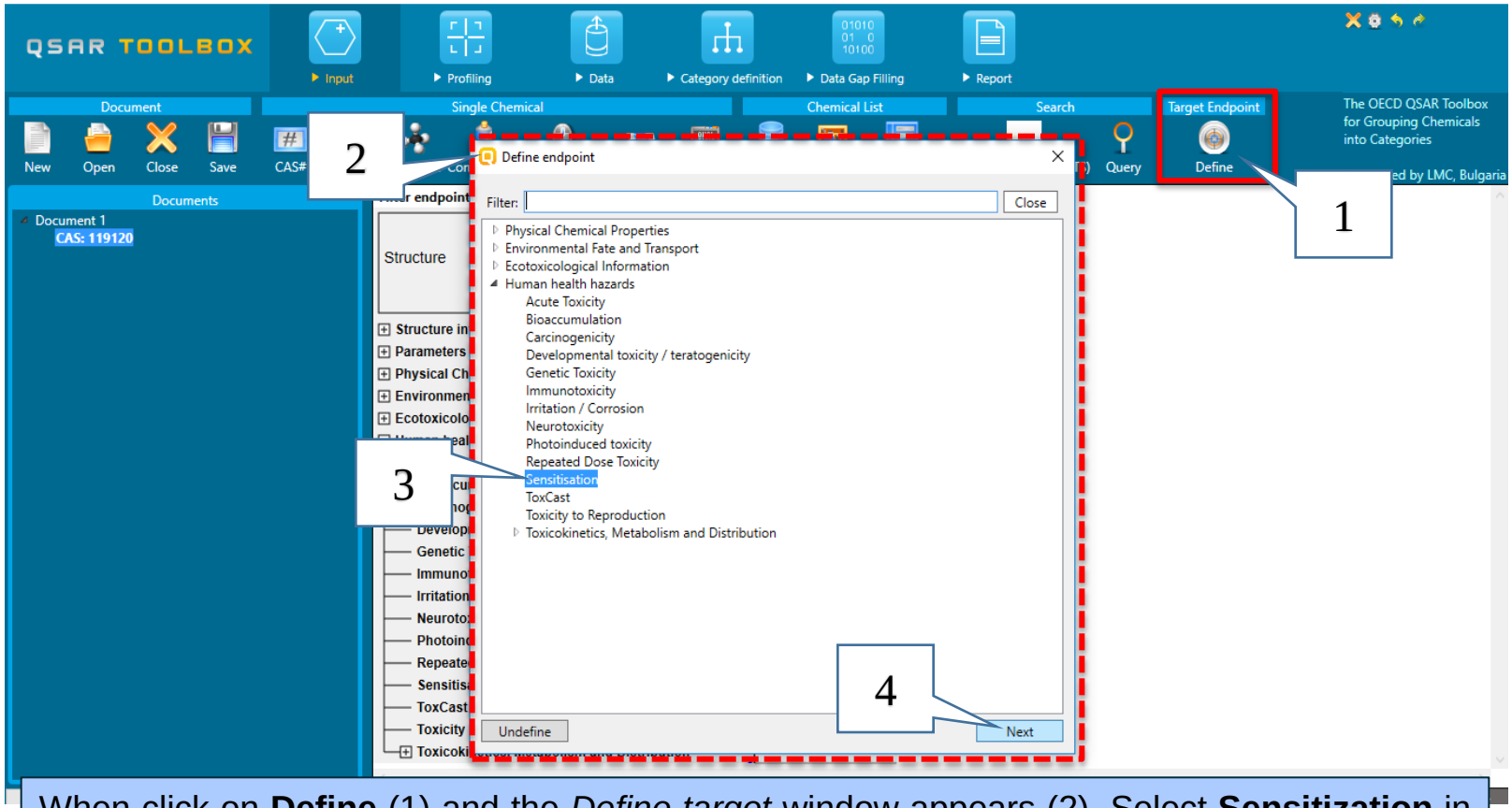
Click on **CAS#** (1); Insert CAS **119-12-0** in the blank field (2) and click on **Search** (3). When the structure appears, click on **OK** (4).

Input
Define target endpoint

Calculation of alert performance (AP) is only possible if the target endpoint is preliminary selected.
Defining of the endpoint allows entering the endpoint of interest e.g. EC3, LC50, gene mutation etc., along with specific metadata information. Based on the metadata, different relevancy scores for profiles could be provided for same endpoint.



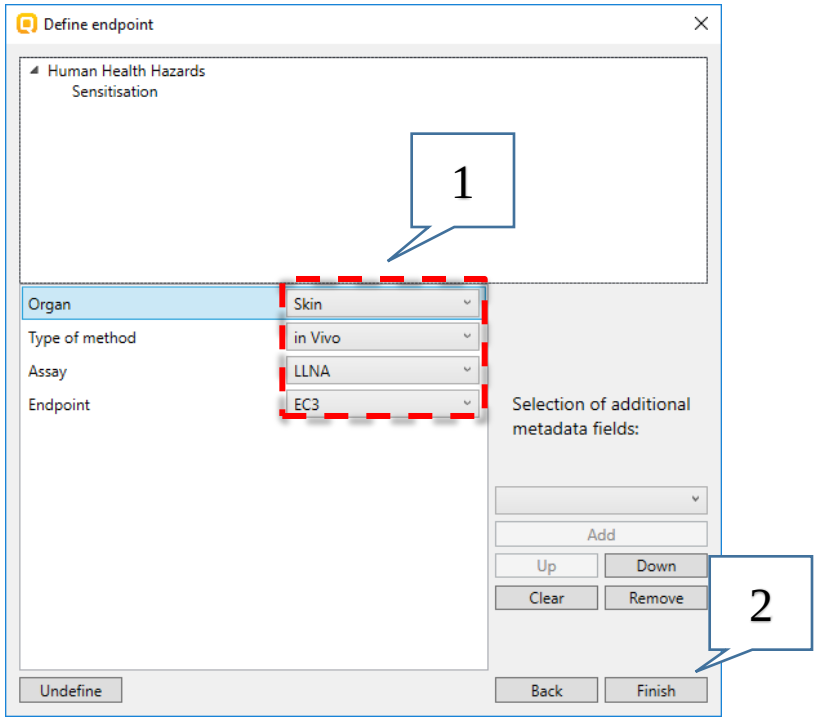
Input
Define target endpoint



When click on **Define** (1) and the *Define target* window appears (2). Select **Sensitization** in the *Human health hazards* category (3) and click on **Next** (4).

Input
Define target endpoint

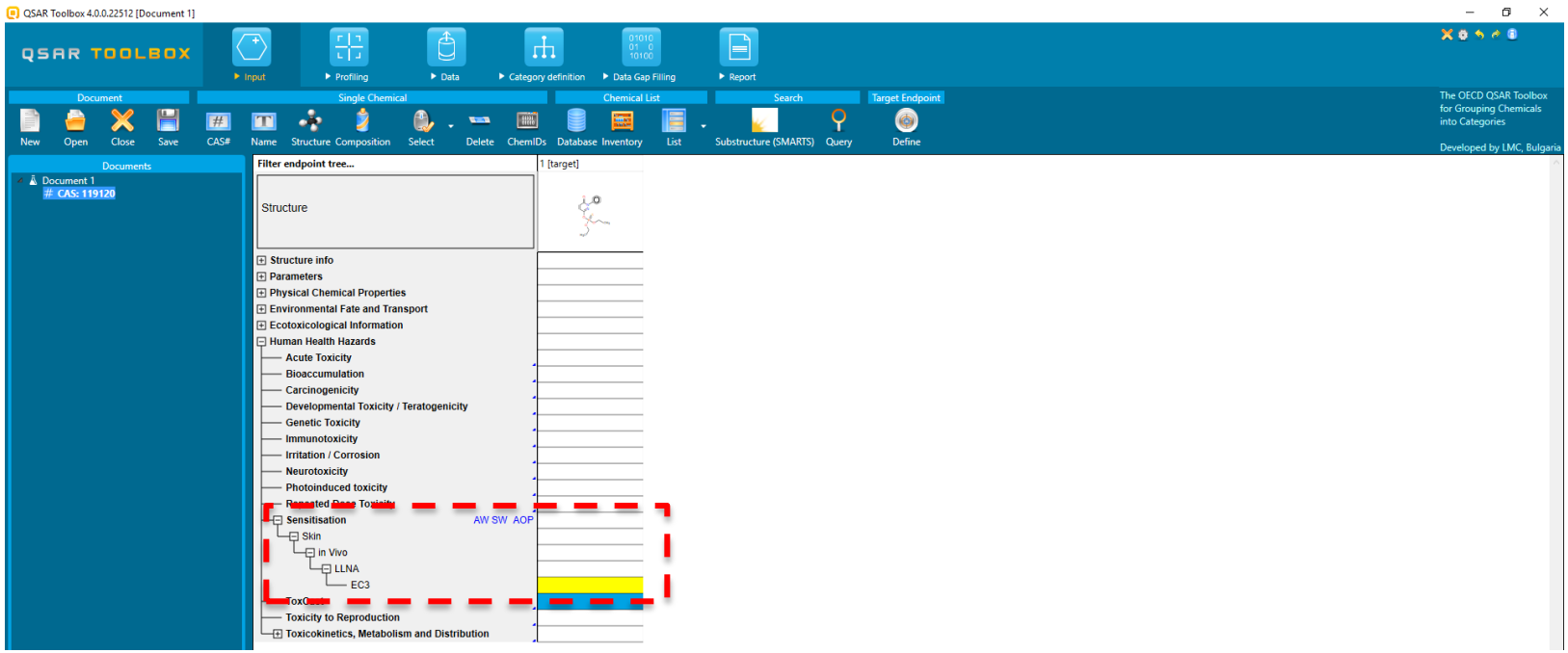
On the next step you have to select the endpoint of interest and additional metadata if needed.



1. Select *Organ*: **Skin**, *Type of method*: **In Vivo**, *Assay*: **LLNA**, *Endpoint*: **EC3**. Selection of additional metadata happens by click on the Add button and selection of the metadata from the drop-down menus; 2. Click on **Finish**

Input
Define target endpoint

Once the endpoint is defined along with its metadata, they appear in the endpoint tree and the corresponding row of the data matrix is highlighted.



Outlook

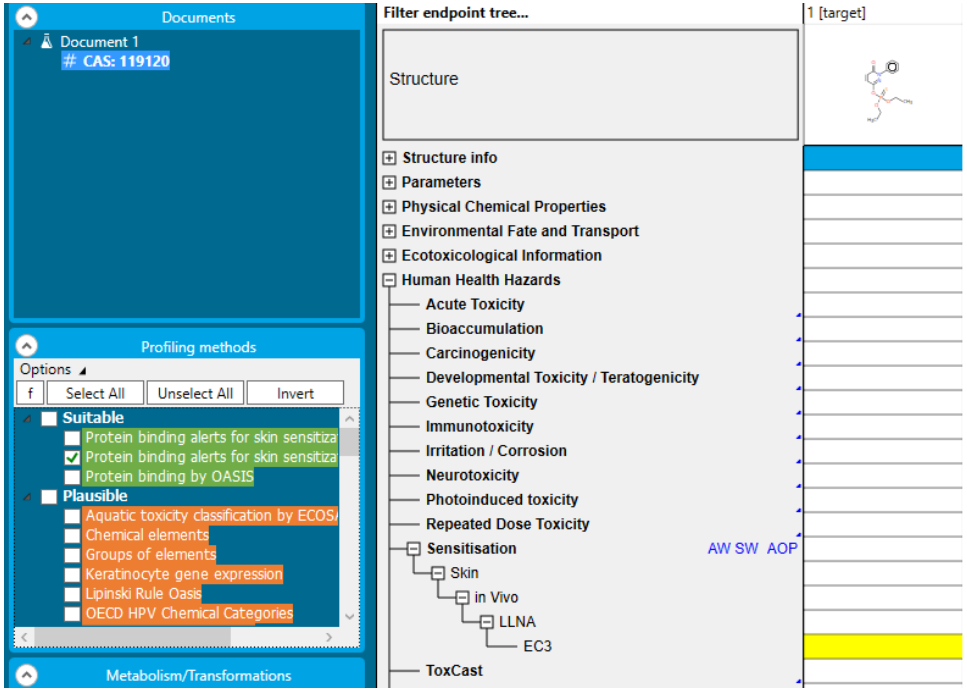
- Background
- Objectives
- Specific Aims
- Alert performance
- The exercise
- **Workflow**
  - Input
  - **Profiling**

Profiling
Overview

- "Profiling" refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- Available information includes likely mechanism(s) of action, as well as observed or simulated metabolites.

Profiling
Sidebar to profiles` relevancy

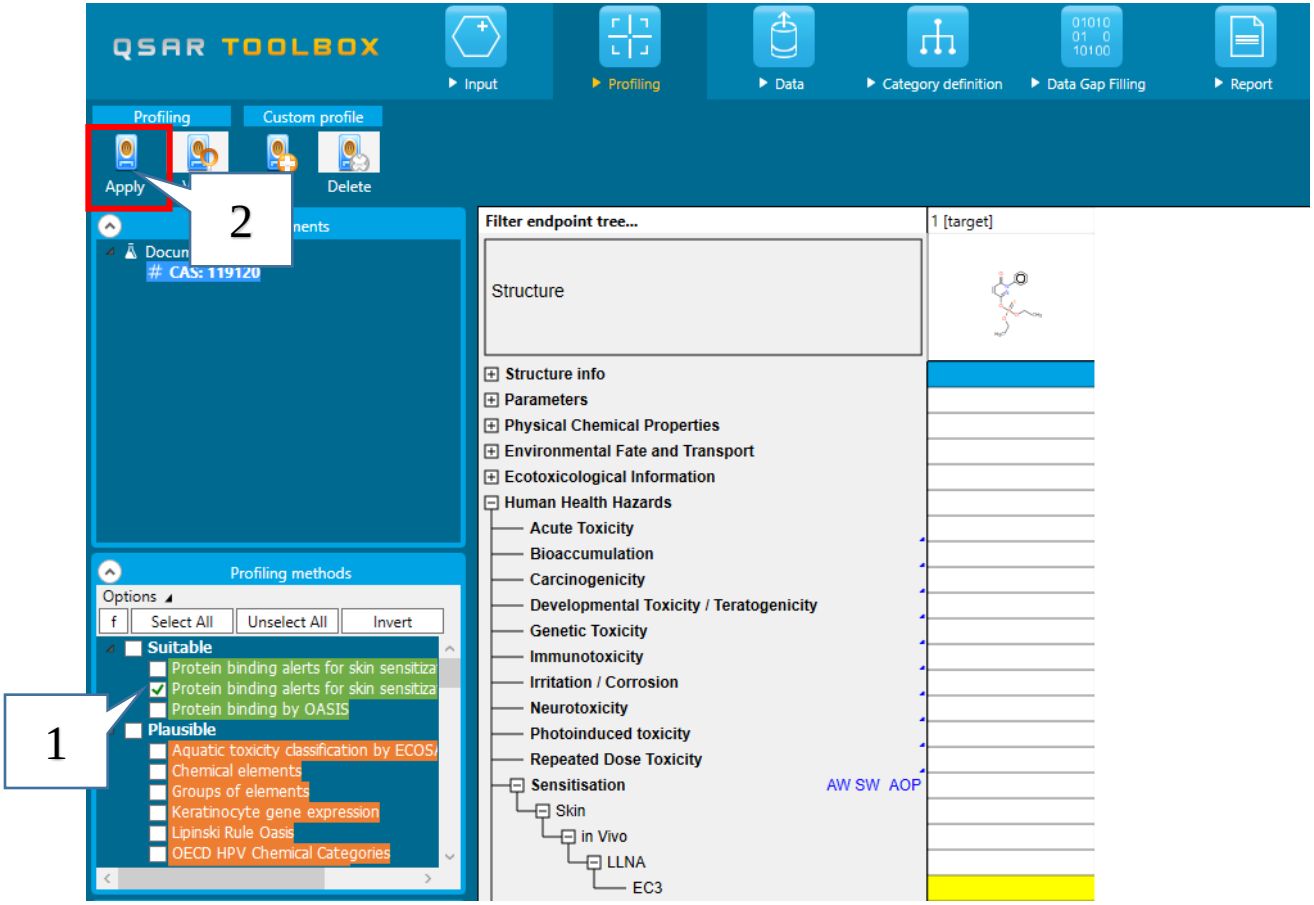
Once the endpoint is selected, the relevant profiles and metabolic transformations are highlighted.



- **Suitable** (in green) - developed using data/knowledge for the target endpoint;
- **Plausible** (in orange) - structure-based; form broader group of analogues;
- **Unclassified** (no color) – all profilers, which are not classified in any of the categories above.

Profiling

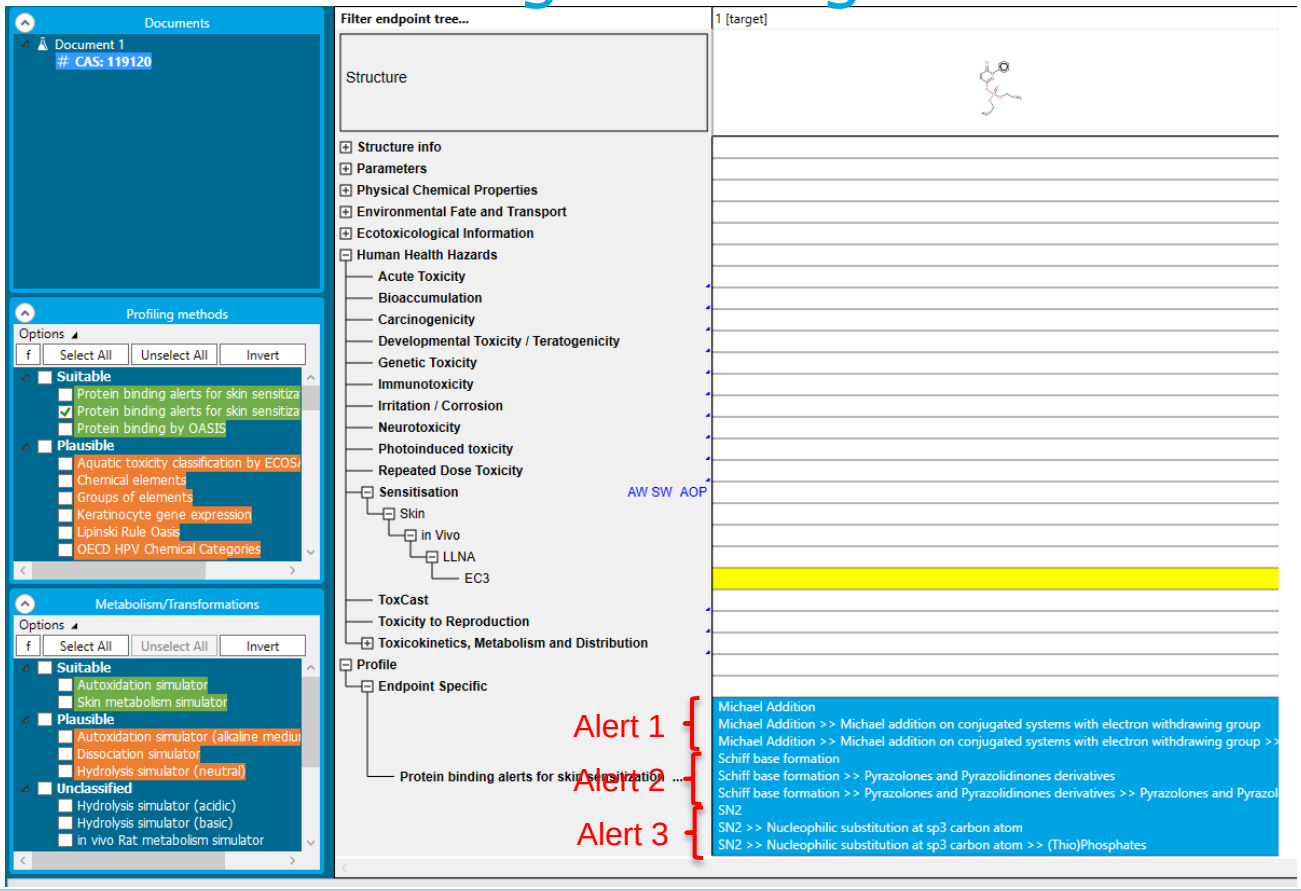
Profiling the target chemical



1. Select *Protein binding alerts for skin sensitization by OASIS*; 2. Click on **Apply**

Profiling

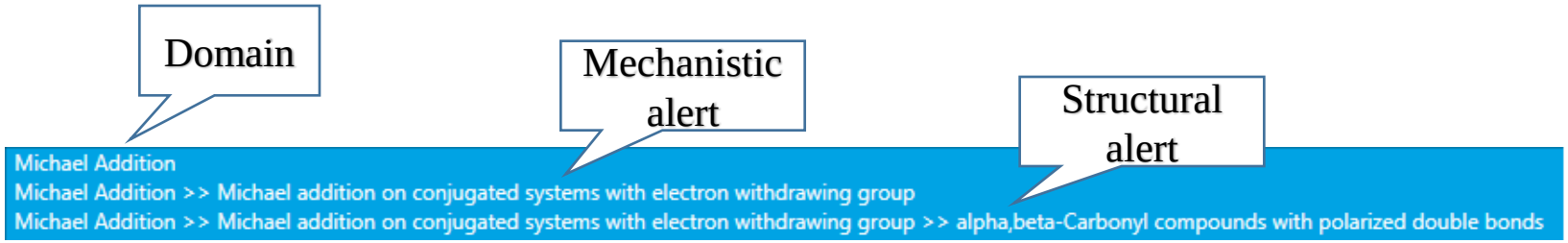
Profiling the target chemical



Three protein binding alerts for skin sensitization are found in the target chemical.

Profiling
Sidebar on the hierarchical type profiles

Protein binding alerts for skin sensitization by OASIS is a hierarchical profile. The organization of the hierarchical profiles includes three levels of information for each category – domain, mechanistic alert and structural alert.



Right click over the structural alert and select on *Explain* opens the profiling scheme, where the user can see more details about the current alert.

Outlook

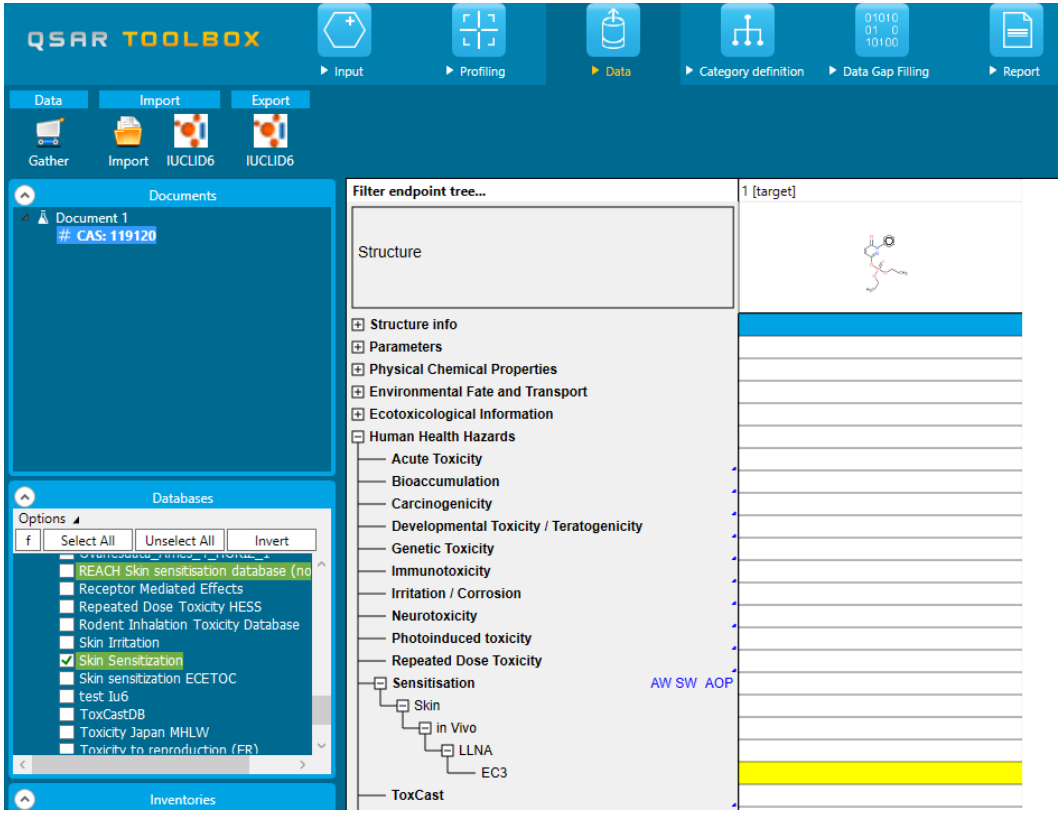
- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Input
  - Profiling
  - **Data**

Data
Overview

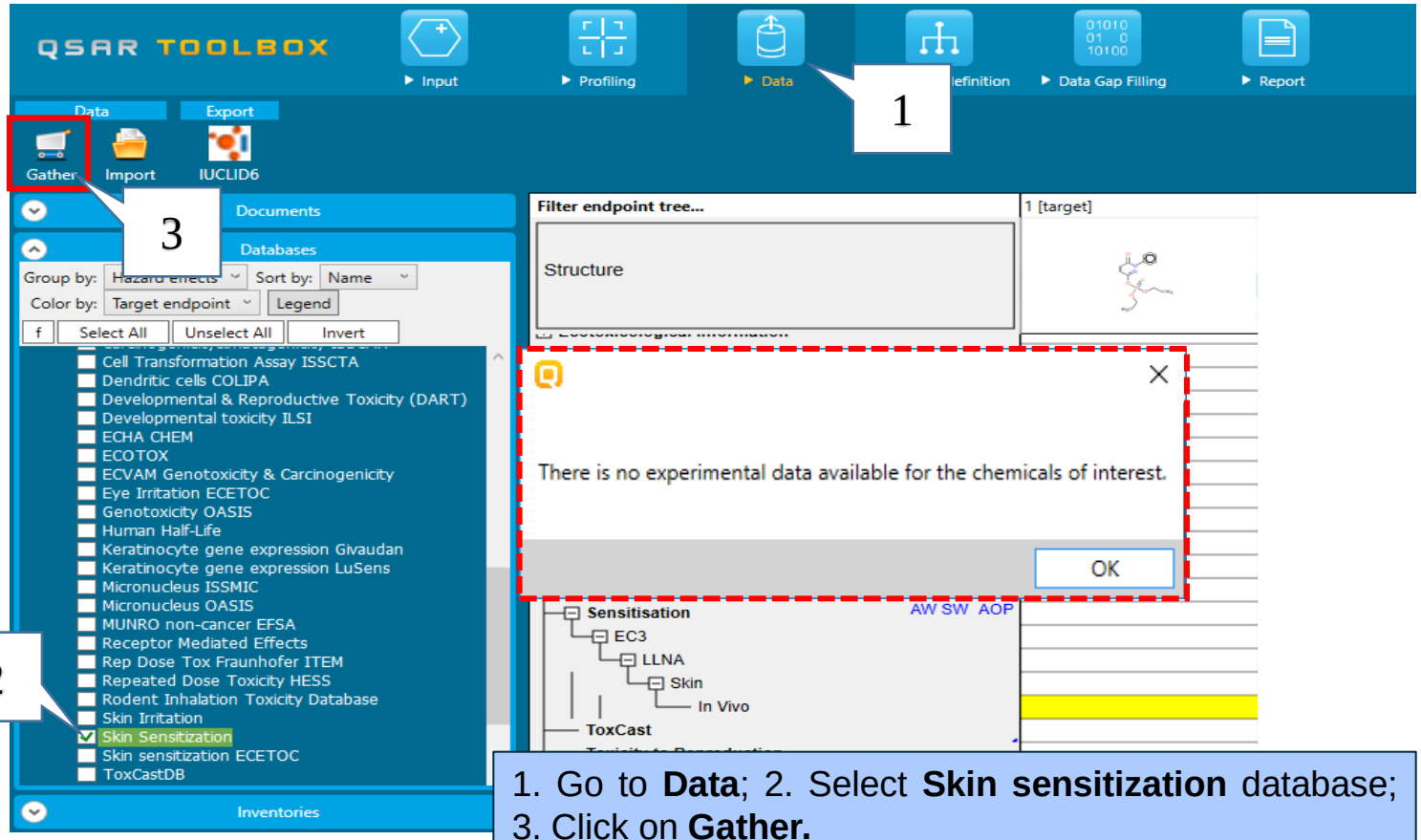
- "Data" refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.
- Data gathering can be executed in a global fashion (i.e., collecting all data for all endpoints) or on a more narrowly defined basis (e.g., collecting data for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

Data
Sidebar on Data availability

Once the endpoint is selected, the databases, which contain such type of data, are highlighted in green.



Data
Gather data



1. Go to **Data**; 2. Select **Skin sensitization** database; 3. Click on **Gather.**
A pop-up message informs that there is no experimental data for the target chemical.

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 data and analogues (and therefore calculation of AP) is performed only among the chemicals which are listed in the selected databases. In this example only the Skin sensitization database is selected.
- In this example, an insert window appears stating there was “no data found” for the target chemical.

Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - **Category definition**

Recap

- In module one, you have entered the target chemical and define the target endpoint.
- In the second module, you have profiled the target chemical with profiling scheme, which is suitable for the selected target endpoint.
- In the third module, you have seen the database corresponding to the defined target endpoint. You have found that no experimental data is currently available in the database for the structure.
- In other words, you have identified a data gap which you would like to fill.
- Click on “Category Definition” to move to the next module.

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 read-across.
- 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.

Category Definition

Protein binding by OASIS grouping method

- This is one of the best grouping methods in the Toolbox. It is built on conventional organic chemical reactions and as such is qualitative in character.
- This method is particularly relevant for respiratory and skin sensitization and acute aquatic toxicity, but also for chromosomal aberration and acute inhalation toxicity.

Category Definition
Sidebar to Protein binding by OASIS categorization

- This scheme includes 112 categories organized in three level of information:

  ✓ Level I: Mechanistic Domains (11 categories)
  ✓ Level II: Mechanistic alerts associated to each mechanistic domain are created on the basis of a common reactive centre being activated by a number of substituents (50 categories)
  ✓ Level III: A number of structural alerts specifying the substituents to a common reactive centre are made up each mechanistic alert (112 categories)

Category Definition

Sidebar to Protein binding by OASIS categorization

- Each category from level III is presented by defined 2-dimensional structural alerts that is responsible for the eliciting toxic effects, such as skin sensitization which are a result of protein binding.

- The associated chemical reactions are in accordance with existing knowledge on electrophilic interaction mechanisms of various structural functionalities.

Category Definition

Sidebar to Protein binding by OASIS categorization

- There is an agreement that most organic chemicals must react covalently with skin proteins in order to behave as skin sensitizers.
- Therefore, chemical reactions by which organic chemicals bind with proteins are relevant to grouping chemicals that may be skin sensitizing agents. So you have mechanistic plausibility for defining your category based on similar protein-binding mechanism.

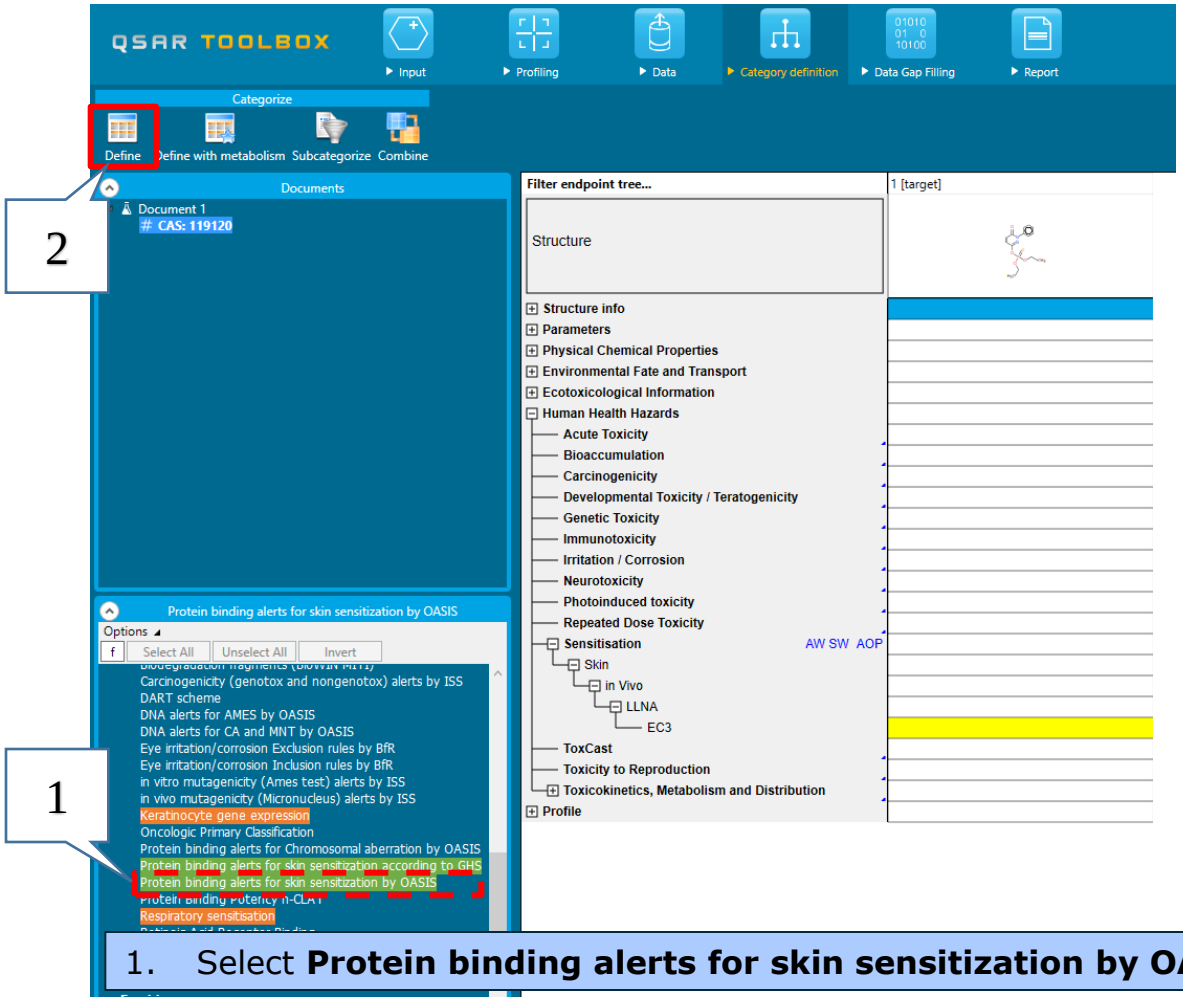
Category Definition

- When more than one alert is found in the target structure before or after metabolic activation, Alert performance could be used to define which of them is the most suitable for primary categorization

Alert performance
Overview

The performance of an alert represents the number of chemicals with data related to the predefined scale across all chemicals from the selected databases, which have the same alert. It provides and distribution of data according to a given effect (e.g. positive, negative) in percentages.

Category Definition
Calculation of Alert performance



1. Select **Protein binding alerts for skin sensitization by OASIS**; 2. Click on **Define**

Category Definition
Calculation of Alert performance

After the click on Define, the Categorization dialog appears. It consists of all protein binding alerts for SS found in the target structure.

The most suitable alert for category formation is determined by comparison of their alert performance.

Additional section for calculating "Alert performance" is designed in this dialogue, when the endpoint is preliminary defined. (see on the next slide).

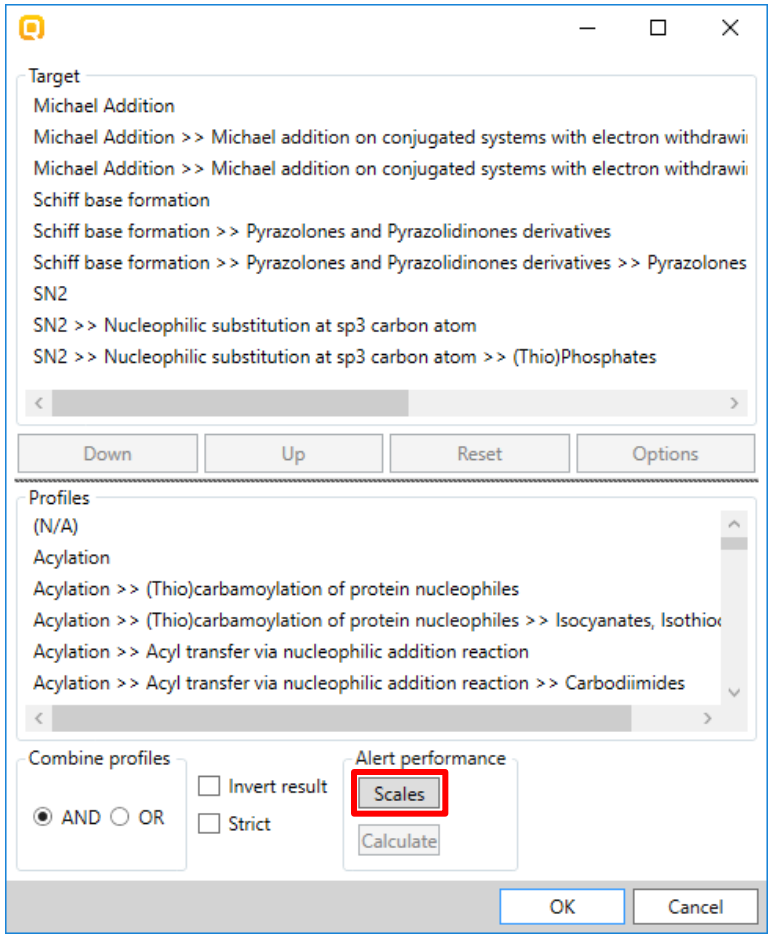
Alert performance can be calculate for only one, for combination of alerts or for all found alerts.

Category Definition
Calculation of Alert performance

alert 1

alert 2

alert 3



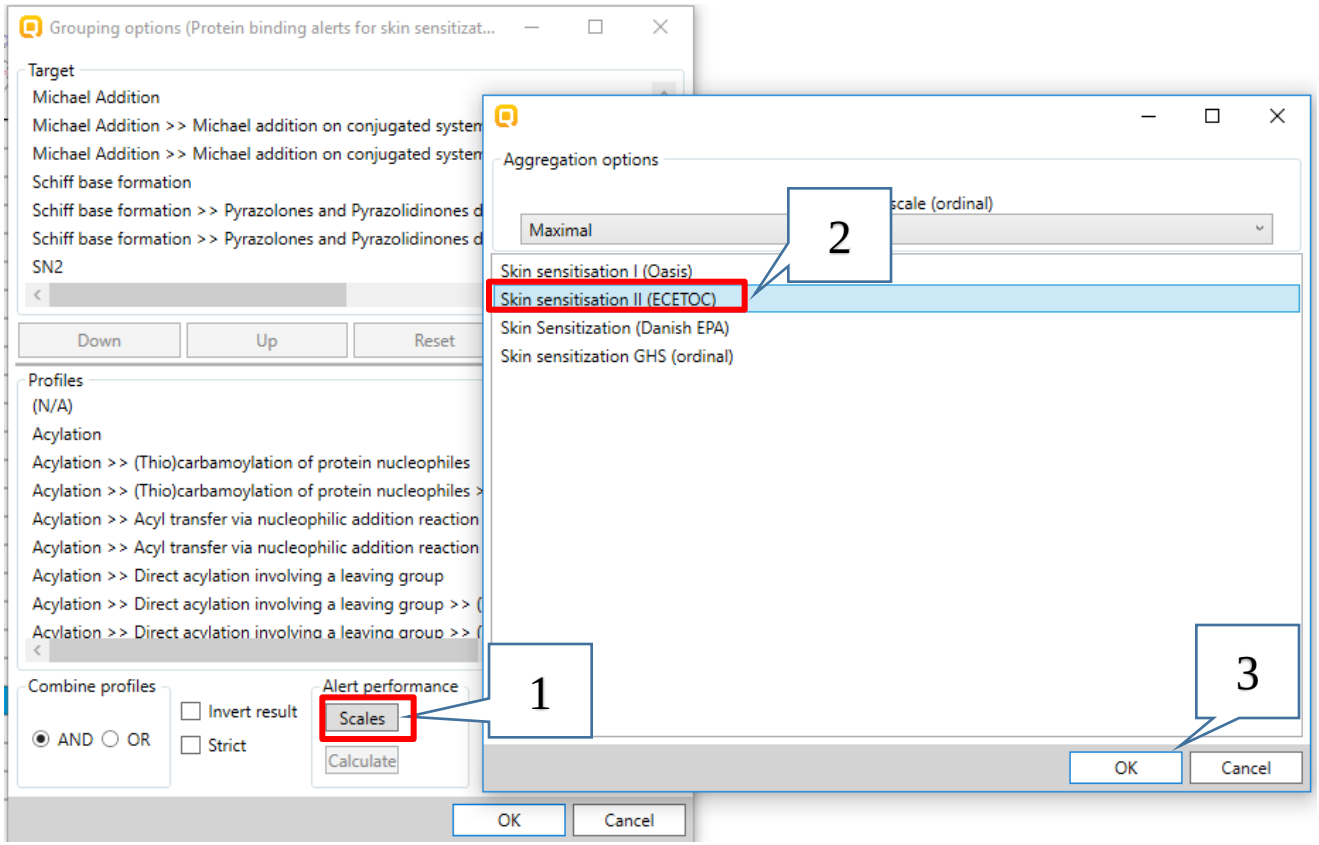
Category Definition
Calculation of Alert performance

In order to calculate the performance of an alert, first of all you have to click on Scales.

The main purpose of the scales is to unify all data available in the Toolbox databases for a certain endpoint. Therefore, the most appropriate scale is “Skin Sensitisation II (ECETOC)”. It is a dichotomous scale that converts all skin data into: Positive and Negative.

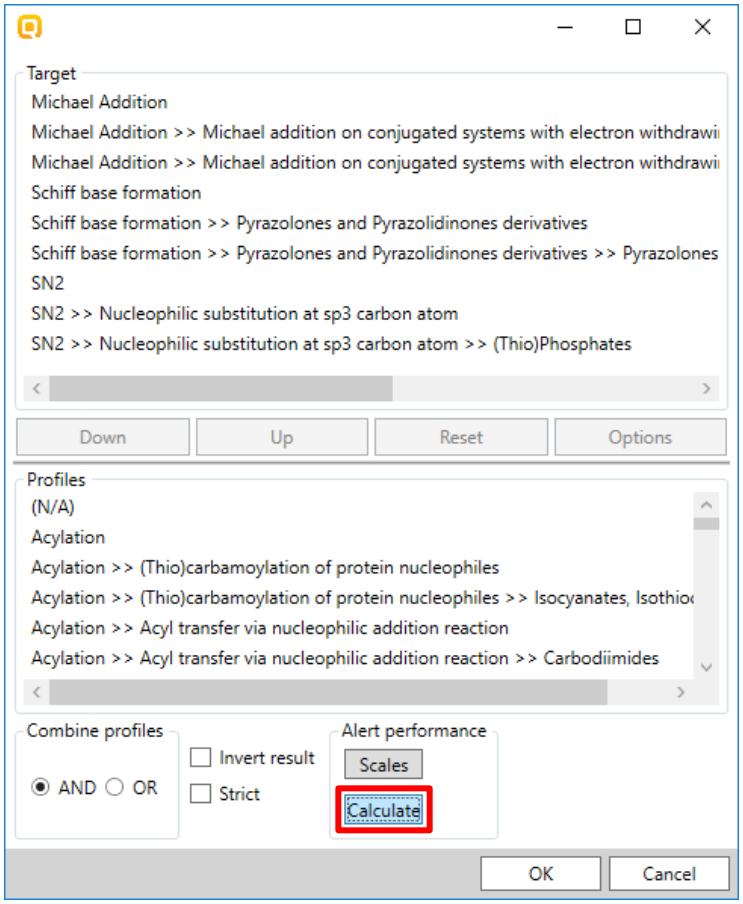
Additional option for applying different weight of the data that is available is also provided. Worst case scenario have been taken into account, i.e. “Maximal” data is set as default.

Category Definition
Calculation of Alert performance



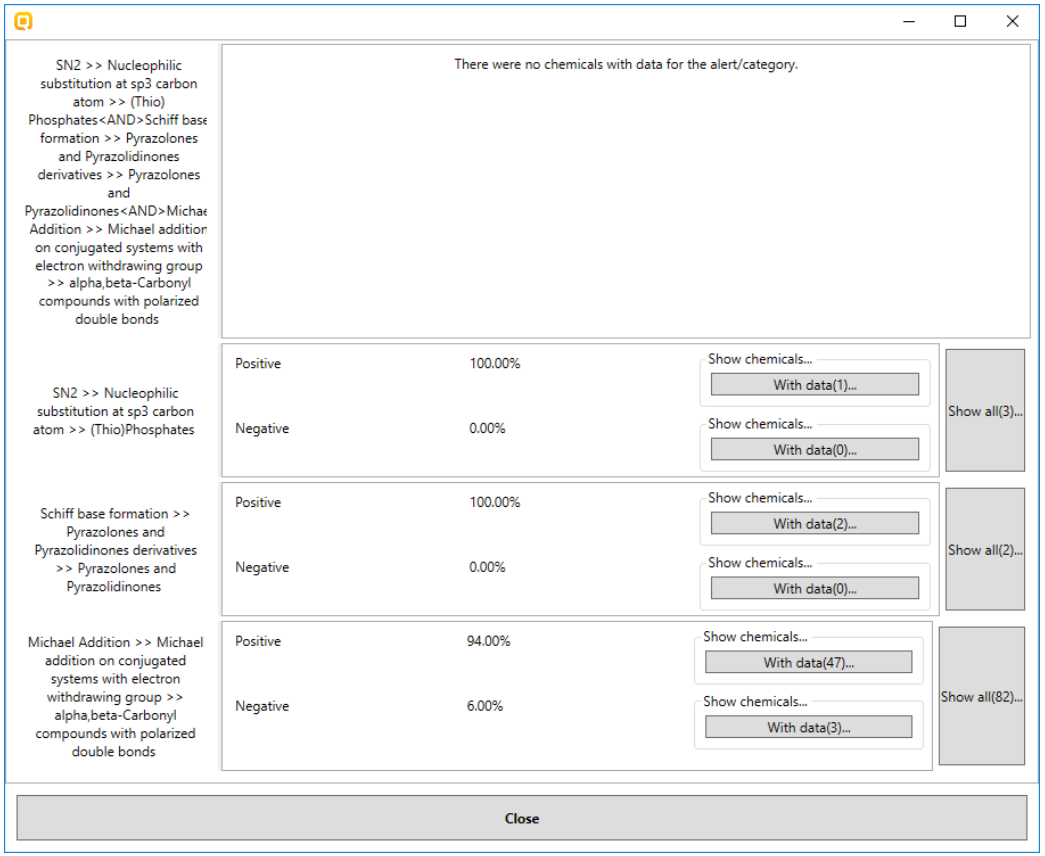
1. Click on **Scales**; 2. Select **Skin sensitization II (ECETOC)** scale; 3. Confirm with "**OK**"

Category Definition
Calculation of Alert performance



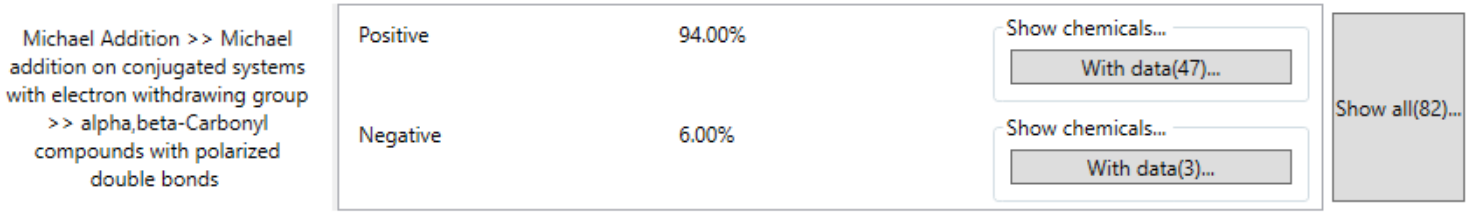
Category Definition
Calculation of Alert performance

Information for the calculated AP for each of the alerts appears in the following window:



Category Definition

Calculation of Alert performance for one alert



The system informs that 82 analogues with the searched alert have been found. Out of these 82 analogues data have been found for 50 substances. Of them:

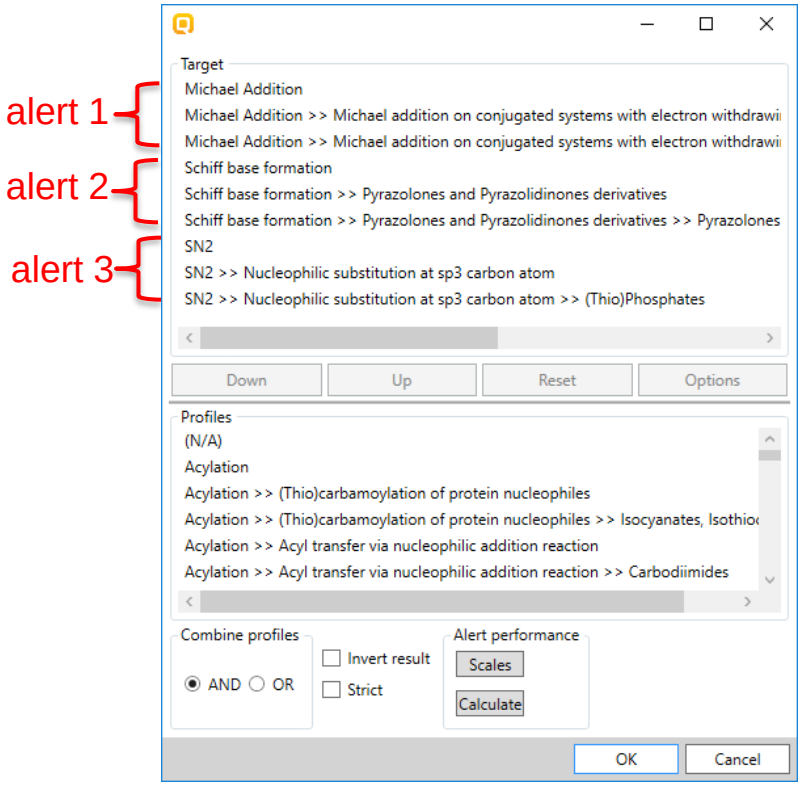
- 47 out of 50 chemicals have positive data (94%)
- 3 out of 50 chemicals have negative data (6%).

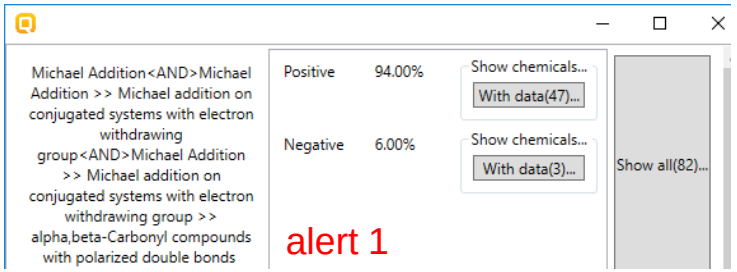
⚠ Keep in mind that the statistic is obtained from the chemicals and data, available in the selected databases

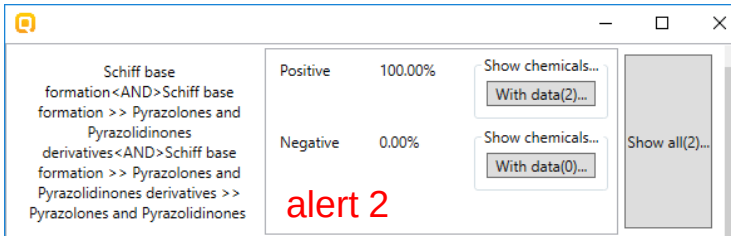
Category Definition

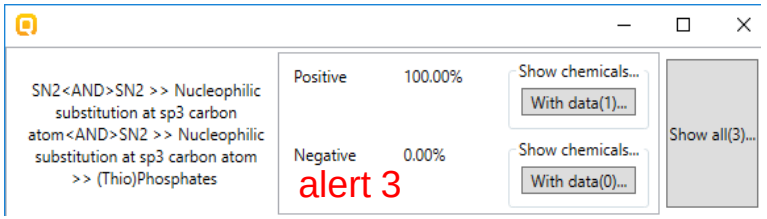
Calculation of Alert performance for one alert

In summary we see that the first alert is the most suitable to define a category.







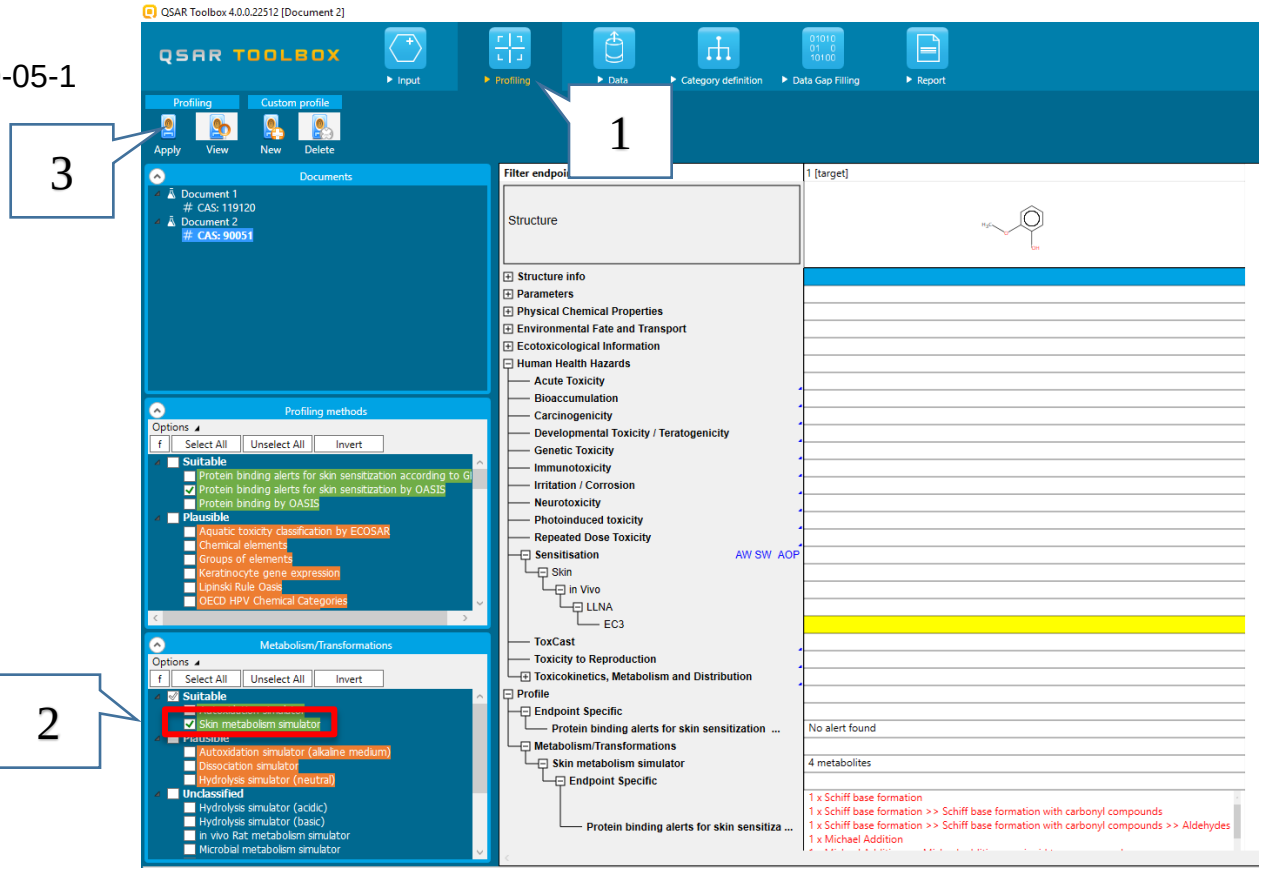


Category Definition

Sidebar on Alert performance accounting metabolism

AP can be also calculated for alert(s) identified after AU or SM activation.

Example: CAS 90-05-1

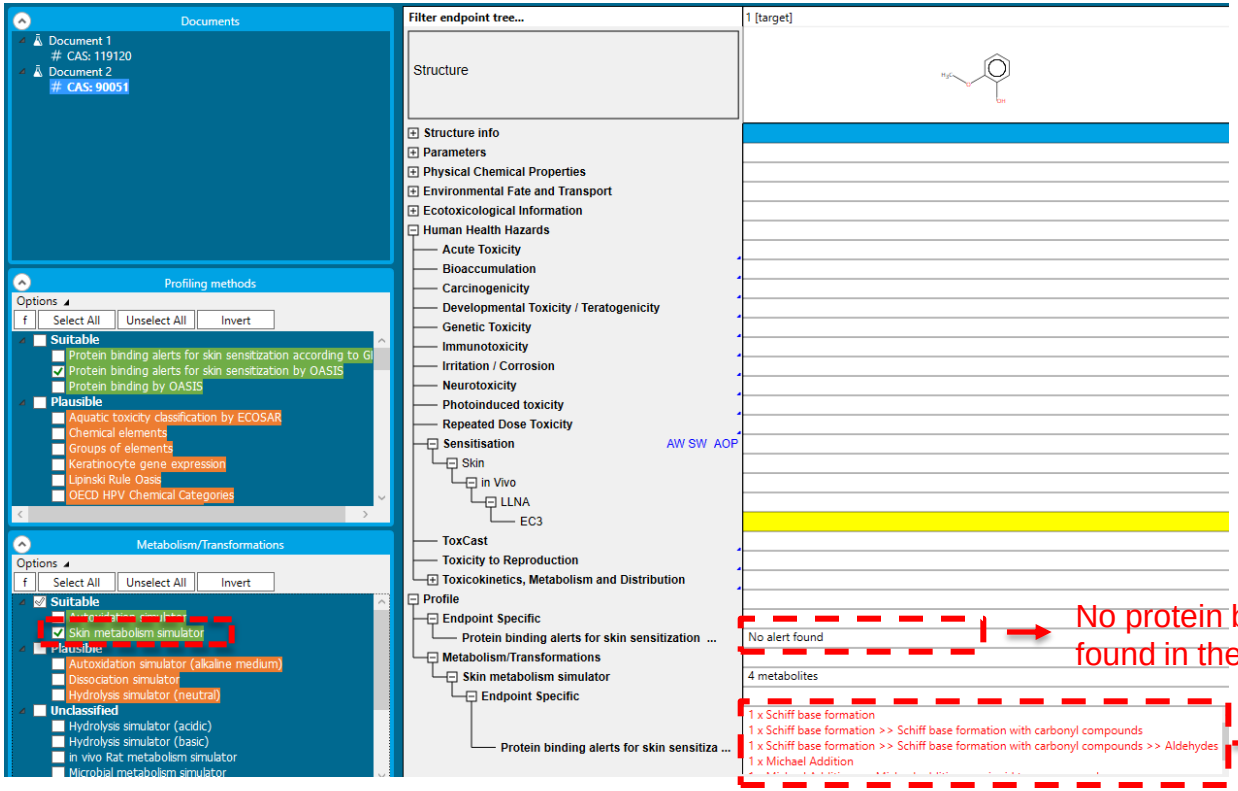


Category Definition

Sidebar on Alert performance accounting metabolism

AP can be also calculated for alert(s) identified after AU or SM activation.

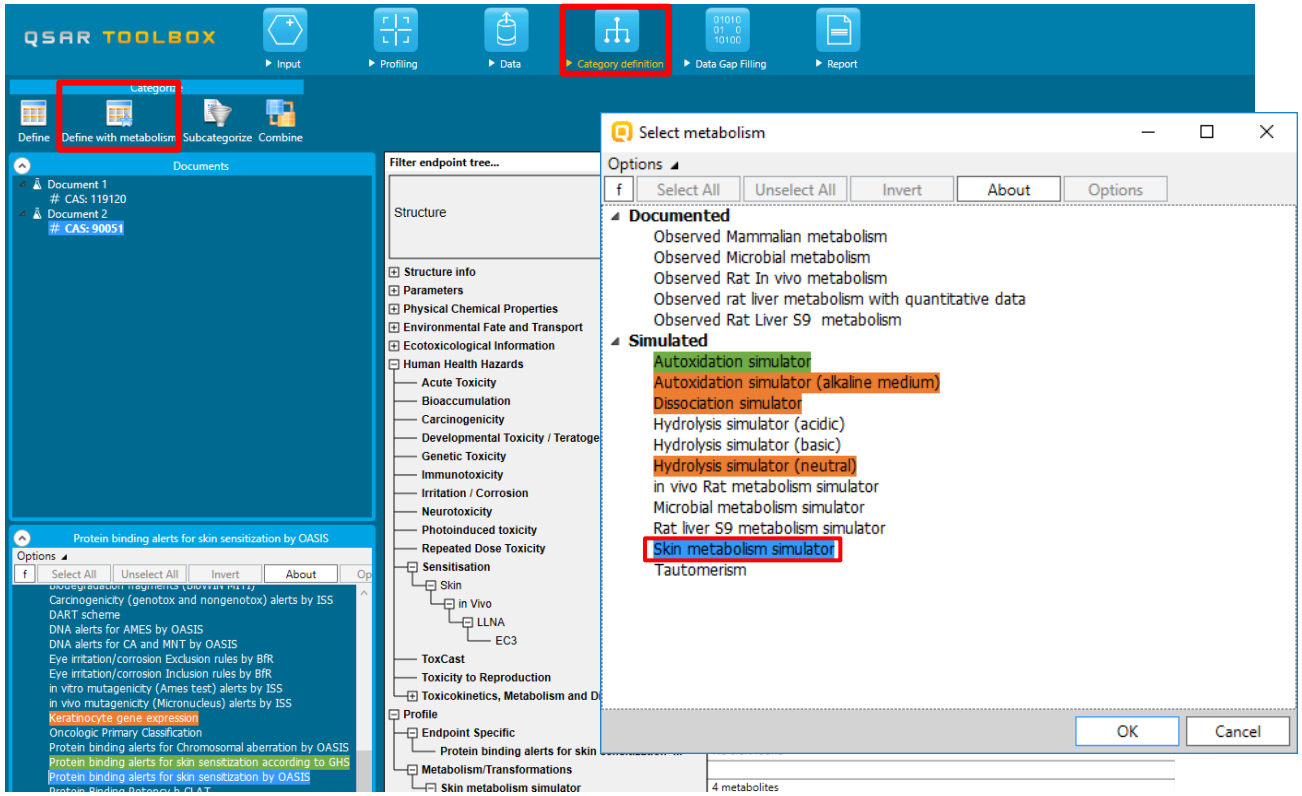
Example: CAS 90-05-1



Category Definition

Sidebar on Alert performance accounting metabolism

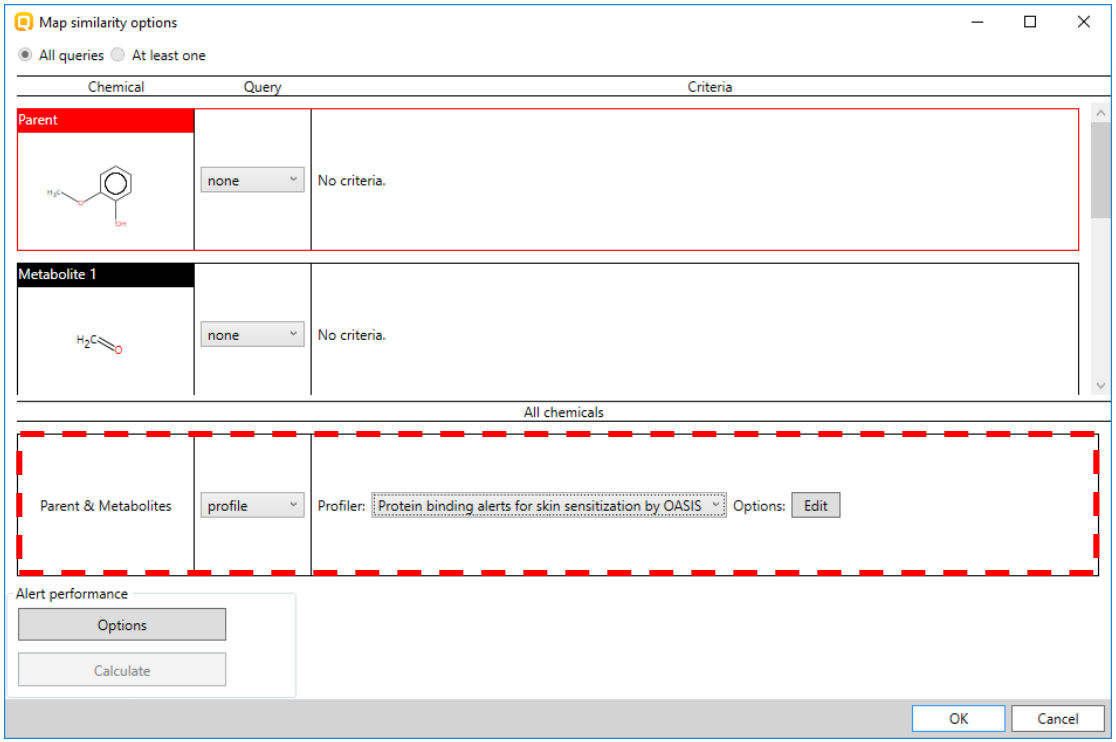
The target chemical has no alert for protein binding as parent but it is activated as a result of skin metabolism. In this respect, the primary category will be defined with accounting for the metabolic activation.



Category Definition

Sidebar on Alert performance accounting metabolism

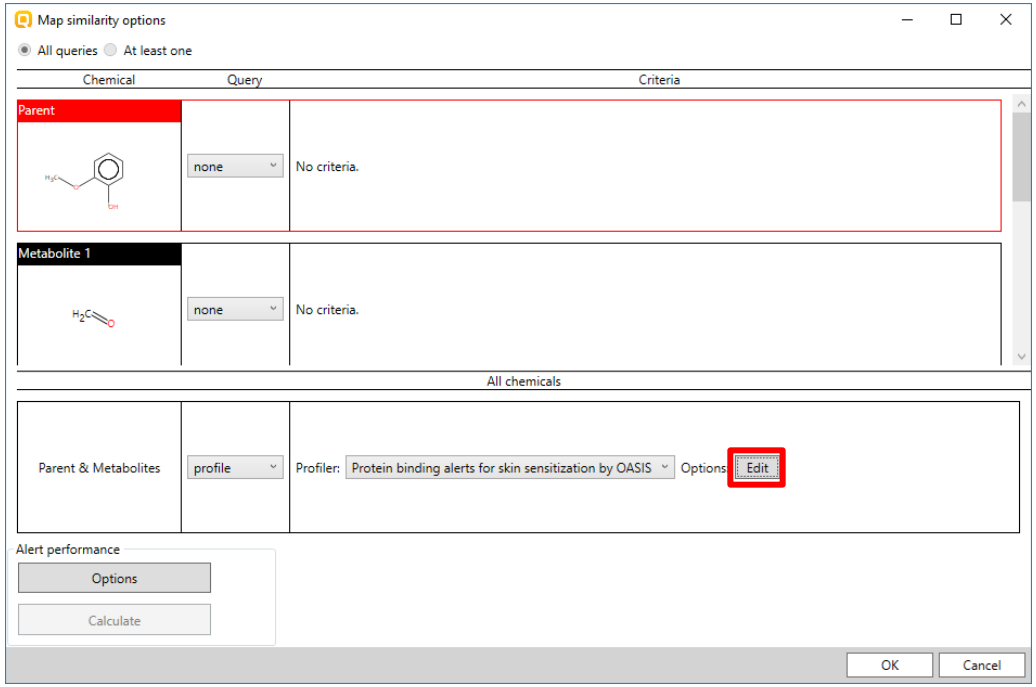
The system will search for chemicals which have similar distribution of the protein binding alerts as identified in the generated package parent and metabolites, accounting for the skin metabolism activation for the analogues.

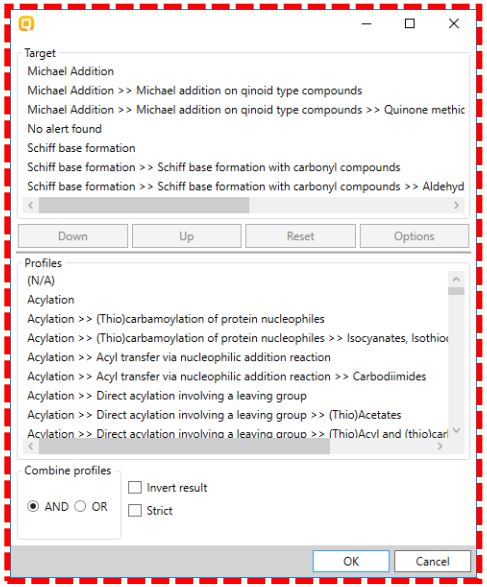


Category Definition

Sidebar on Alert performance accounting metabolism

The found alerts could be seen by click on the **Edit** button. To calculate AP for one alert – remove all alerts except the alert for which AP will be calculated (three levels of mechanistic information are required – domain, mechanistic and structural alert), select a scale in the **Options** and click on **Calculate**.

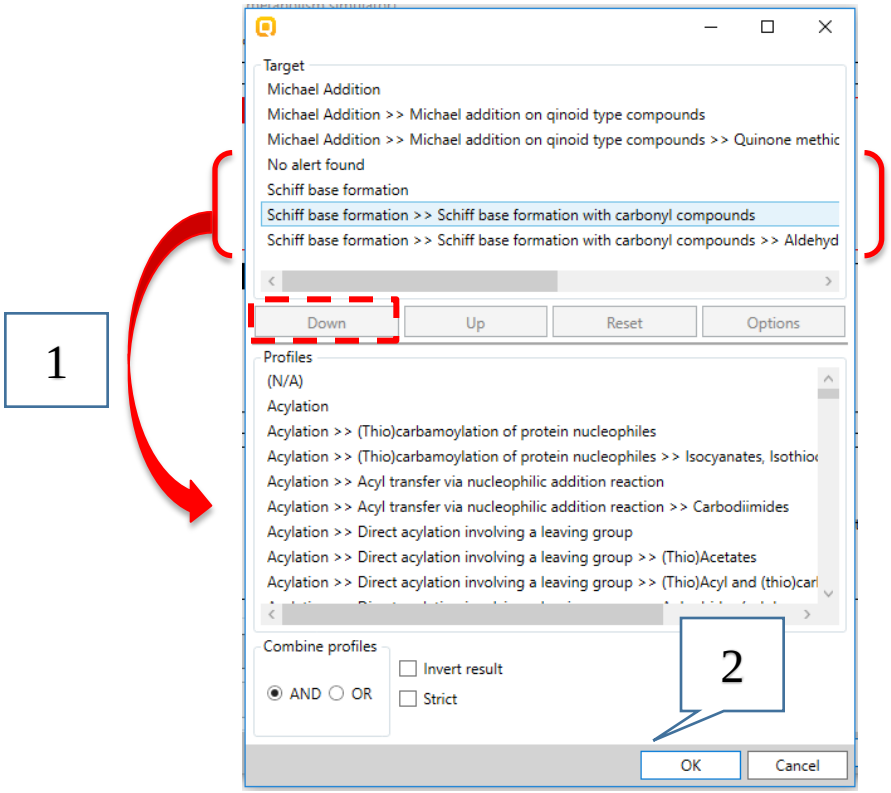




Category Definition

Calculation of Alert performance accounting metabolism for one alert

Now we will calculate AP for each of the alerts in order to see which of them is the most suitable for category formation.

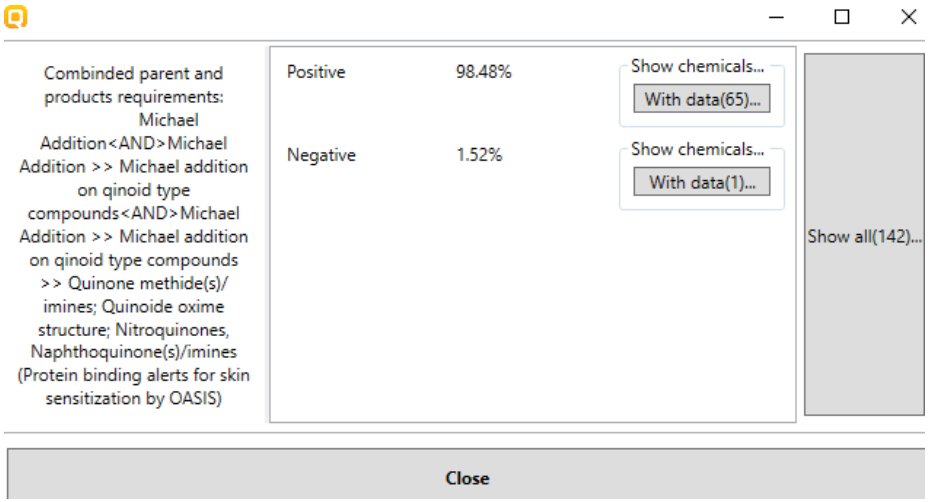


1. Remove No alert found and the second alert by double click or using “Down” button; 2. Click on OK button. Select again scale Skin sensitization (ECETOC)

Category Definition

Calculation of Alert performance accounting metabolism for one alert

Performance of the first alert appears in the following window:



The system informs that 142 analogues with the searched alert accounting for skin metabolism have been found. Out of these 142 analogues data have been found for 66 structures. Of them:

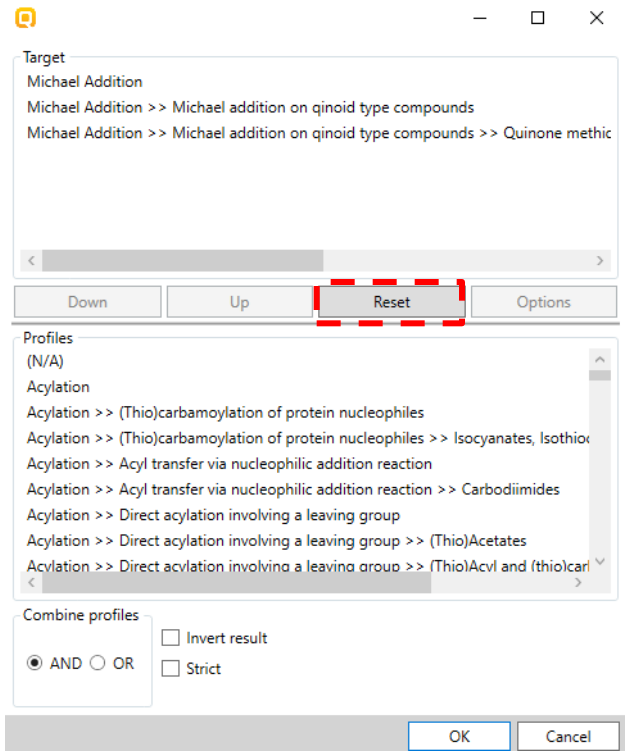
- 65 out of 66 chemicals have positive data (98.48%)
- 1 out of 66 chemicals have negative data (1.52%).

⚠ Keep in mind that the statistic is obtained from the chemicals and data, available in the selected databases

Category Definition

Calculation of Alert performance accounting metabolism for one alert

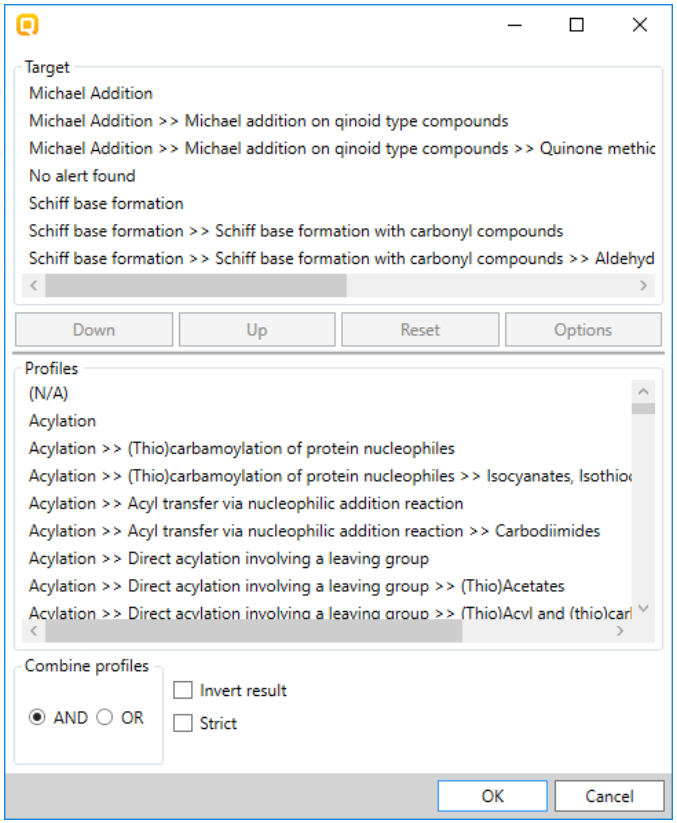
Click on the **Reset** button and repeat the alert performance calculation steps for the second alert.

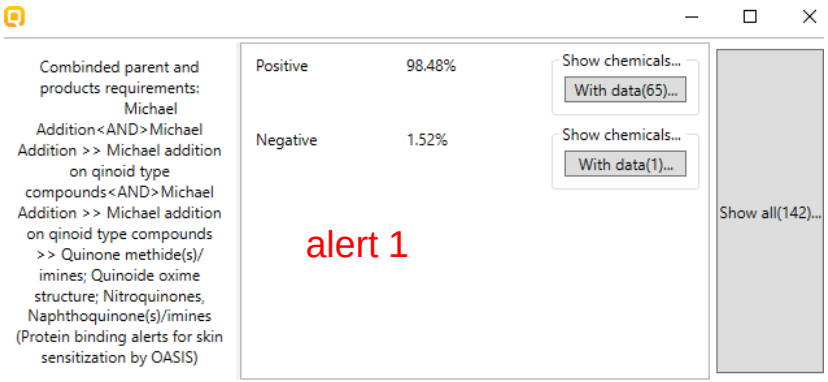


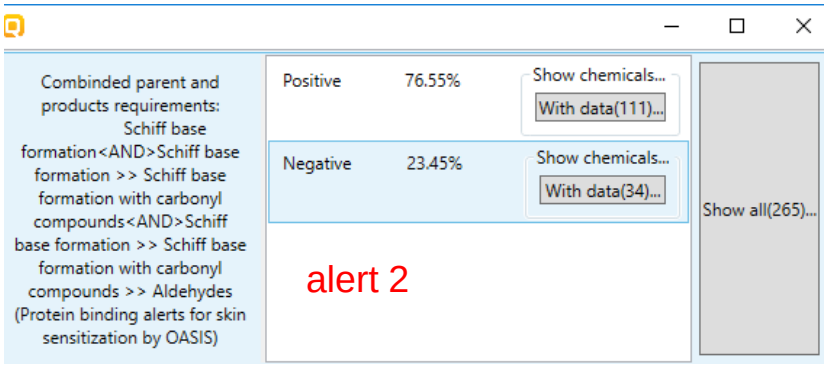
Category Definition

Calculation of Alert performance accounting metabolism for one alert

In summary we see that the first alert is the most suitable to define a category.

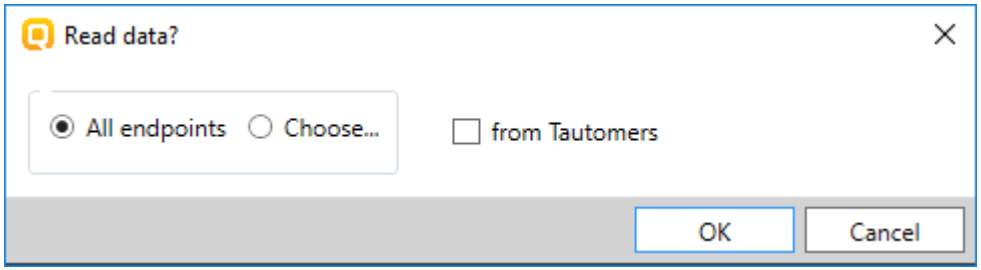






Category Definition
Analogues

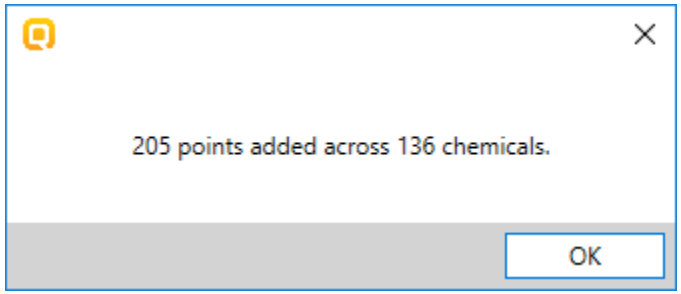
- Based on the defined category (*Michael Addition<<Michael addition on conjugated systems with electron withdrawing group<<α,β-Carbonyl compounds with polarized double bonds*) 136 analogues have been identified (including the target chemical CAS: 90-05-1).
- The Toolbox automatically requests the user to select the endpoint that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).



Only Skin sensitization database is selected now and we click on **OK**.

Category Definition
Read data for Analogues

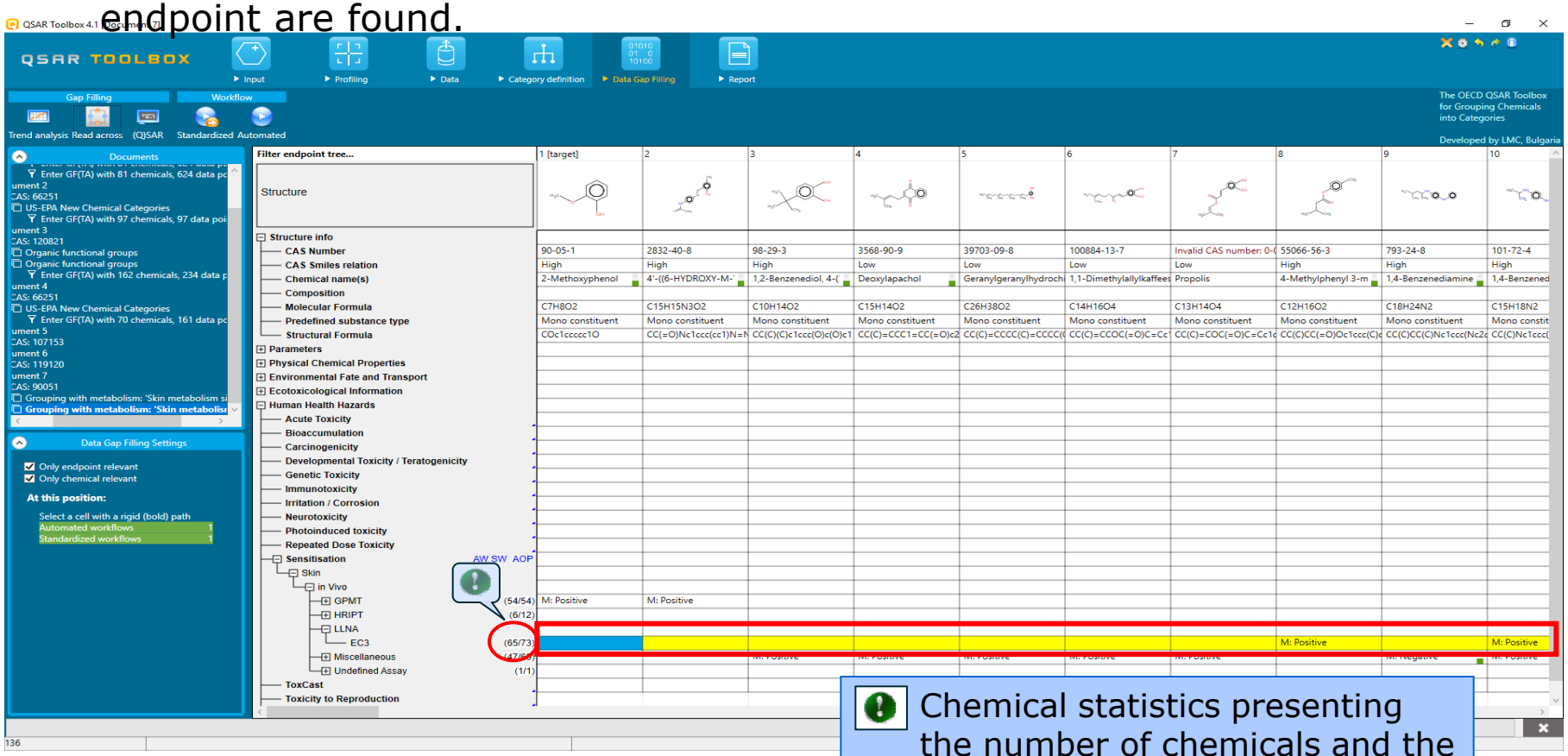
- The Toolbox automatically informs the user for the number of gathered data points across the chemicals in the category



- Click **OK** to confirm the appeared message

Category Definition
Summary information for Analogues

66 chemicals with 74 experimental results related to the defined target endpoint are found.



Chemical statistics presenting the number of chemicals and the available experimental data.

Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Category definition
  - **Data Gap Filling**

Recap

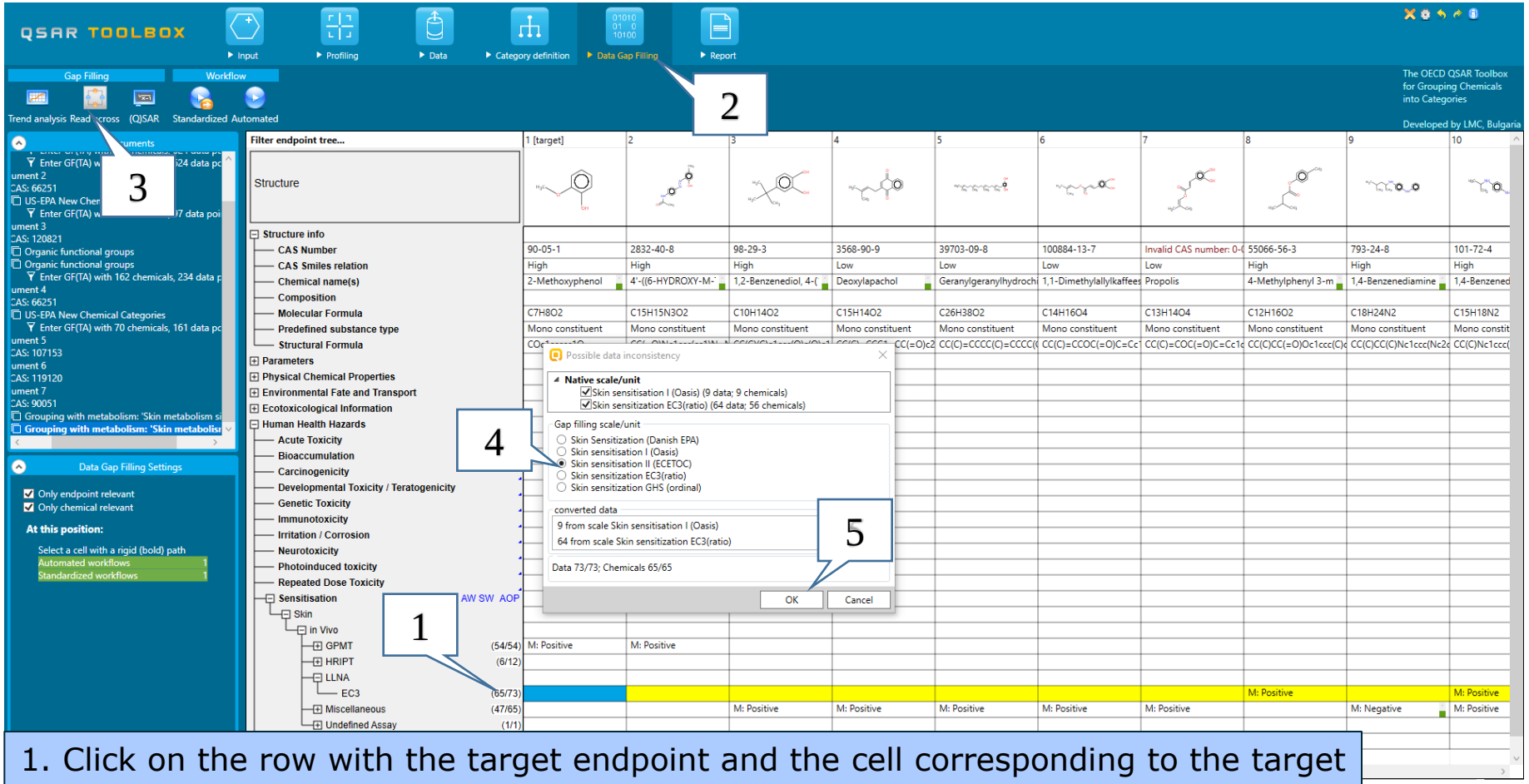
- You have identified three protein binding alerts for the target chemical (pyridaphenthion).
- You have calculated and compare the alert performance for each of the alerts.
- You have now retrieved in the available experimental results on skin sensitisation (EC3) values for 65 chemicals with the same mechanism of protein binding as the target compound, which were found in the "Skin Sensitisation" database.
- The user can now proceed to the next module; click on *"Data Gap Filling"*.

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

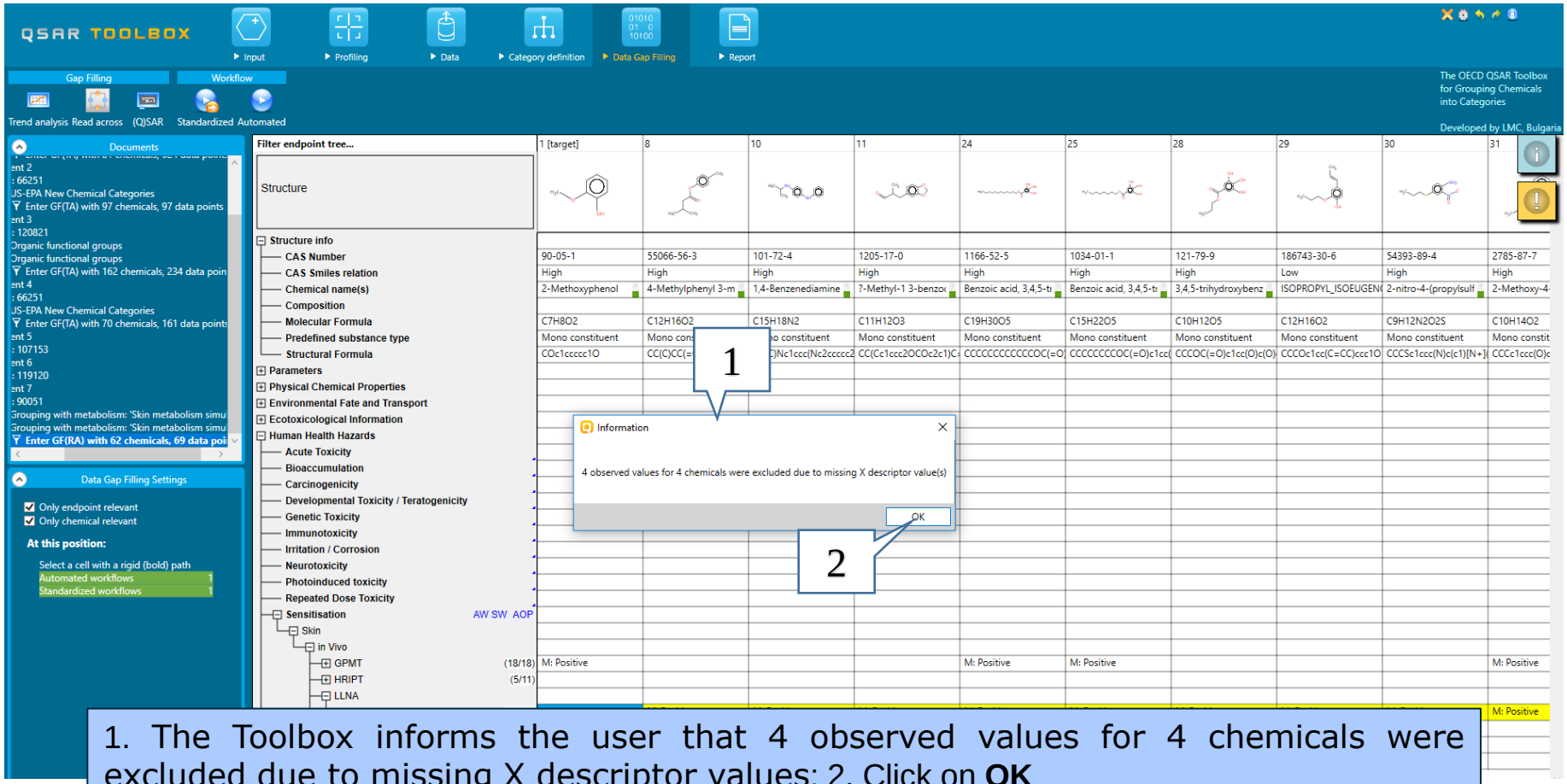
In this example we will use the manual read-across approach.

Data Gap Filling
Apply Read across



1. Click on the row with the target endpoint and the cell corresponding to the target chemical; 2. Go to **Data gap Filling**; 3. Select **Read-across**; 4. Select *Skin sensitisation II (ECETOC);* 5. Click on **OK**

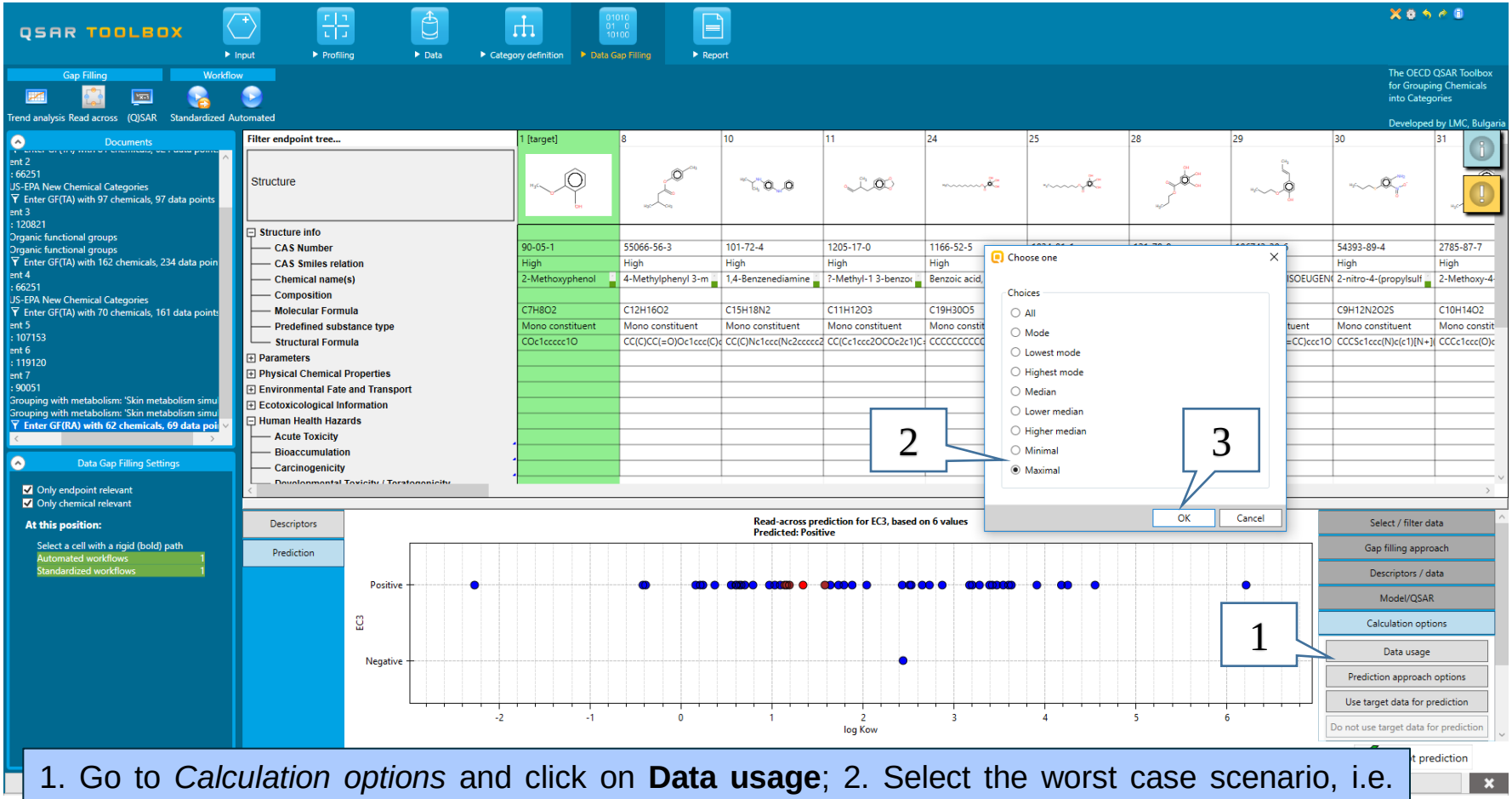
Data Gap Filling
Apply Read across



1. The Toolbox informs the user that 4 observed values for 4 chemicals were excluded due to missing X descriptor values; 2. Click on **OK**

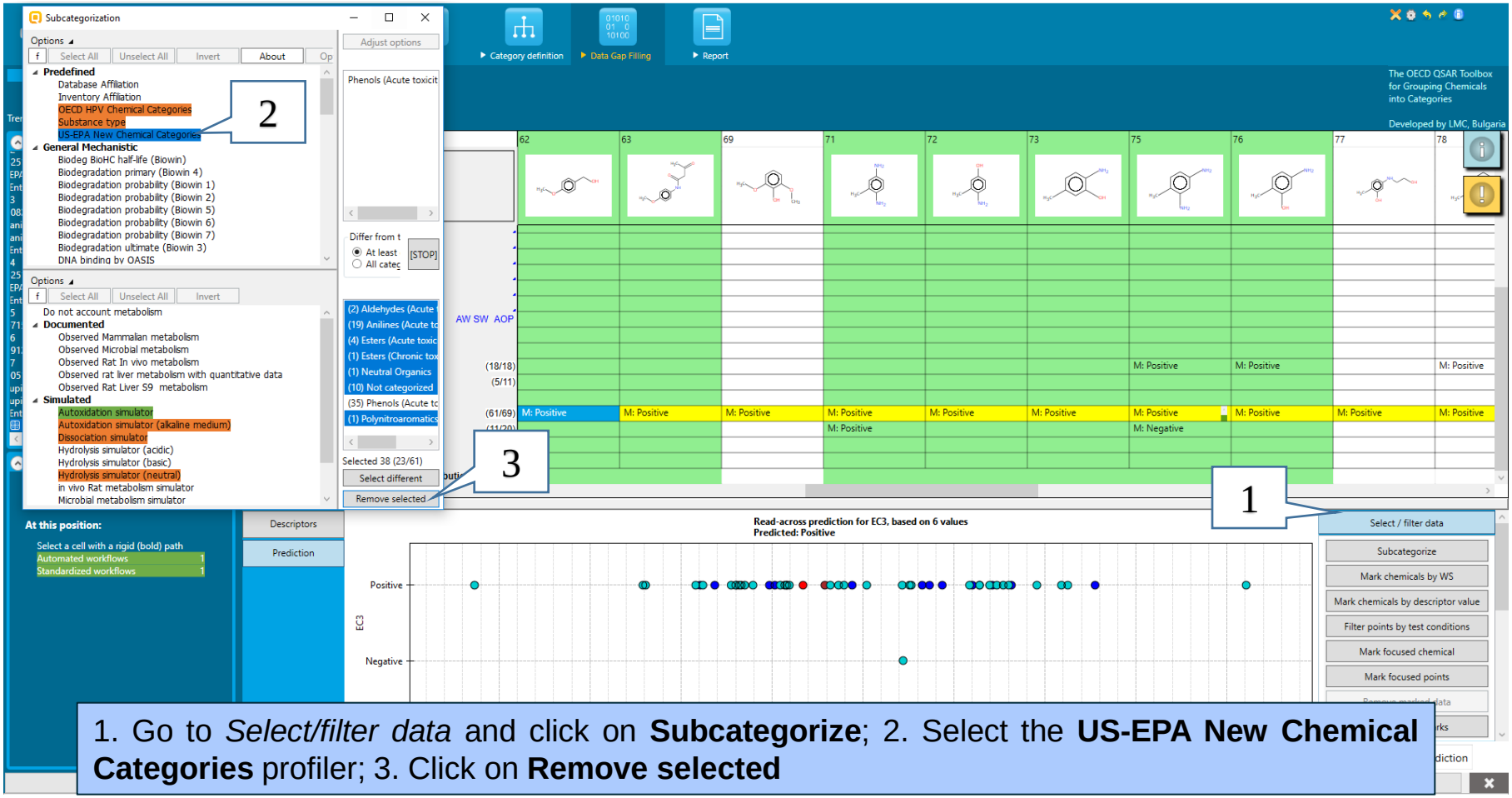
Data Gap Filling

Apply Read across



1. Go to *Calculation options* and click on **Data usage**; 2. Select the worst case scenario, i.e. "**Maximal**" data; 3. Click on **OK**

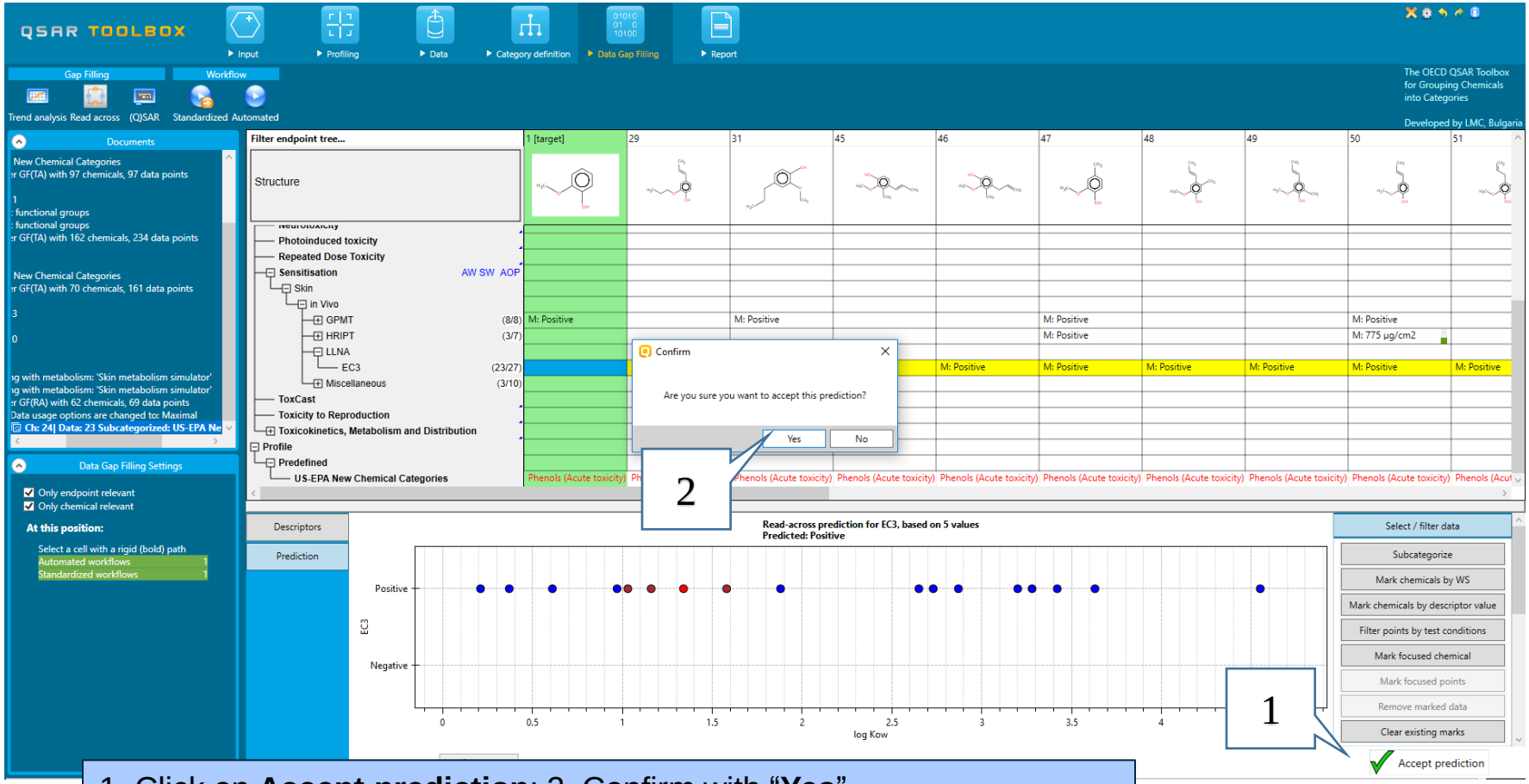
Data Gap Filling
Apply Read across



1. Go to *Select/filter data* and click on **Subcategorize**; 2. Select the **US-EPA New Chemical Categories** profiler; 3. Click on **Remove selected**

Data Gap Filling

Accepting the predicted result



1. Click on **Accept prediction**; 2. Confirm with "**Yes**"

Recap

- Read-across is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitisation. Since the most of the analogues and all five neighbouring tested chemicals in the category were positive, it was easy to accepting the prediction of positive for the target chemical.
- You are now ready to complete the final module and to create the report.
- Click on "Report" to proceed to the last module.

Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - Category definition
  - Data Gap Filling
  - **Report**

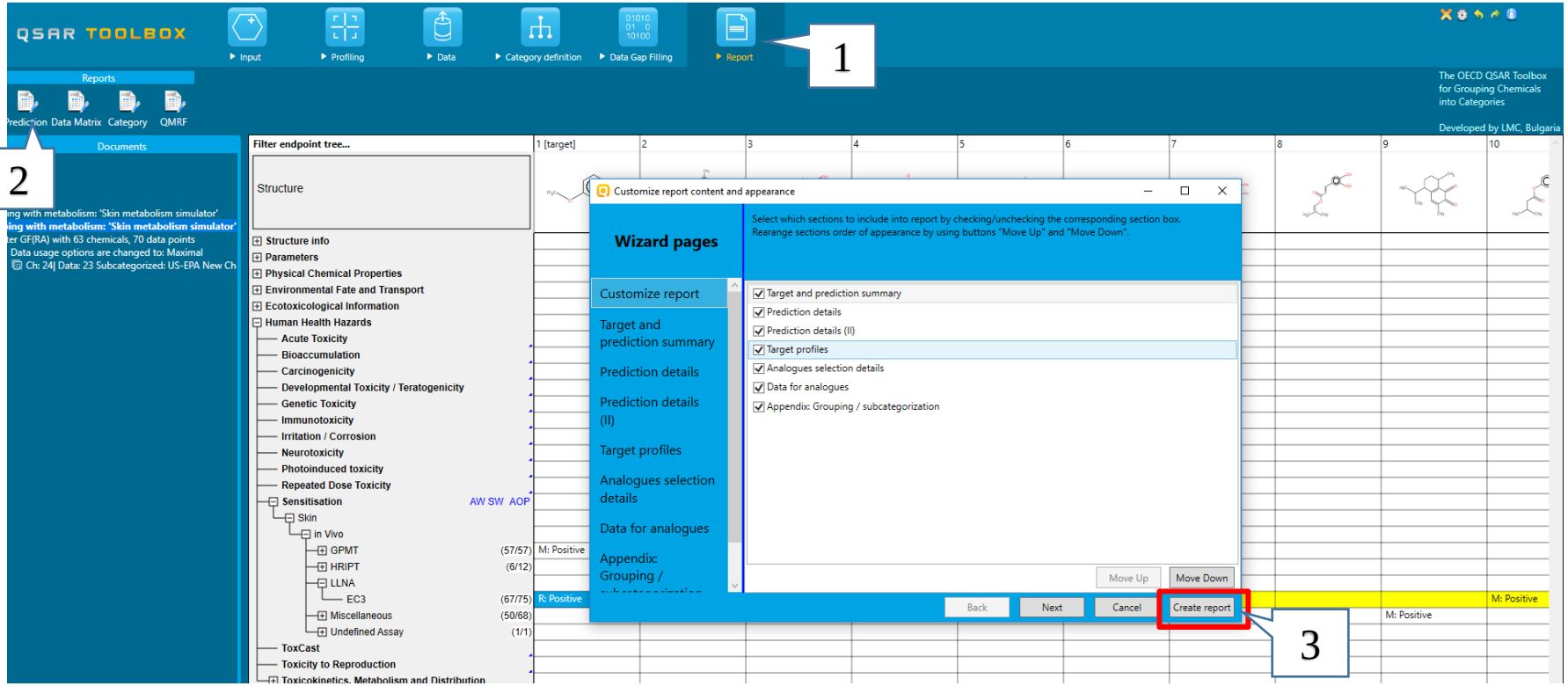
Report
Overview

- Report module could generate report on any of predictions performed with the Toolbox.
- Report module contains predefined report template which users can customize.
- Two type of report files are generated:

  - *Prediction report* – containing information for the target

  - *Data matrix* – containing information for the analogues used for the prediction.

Report
Generation report

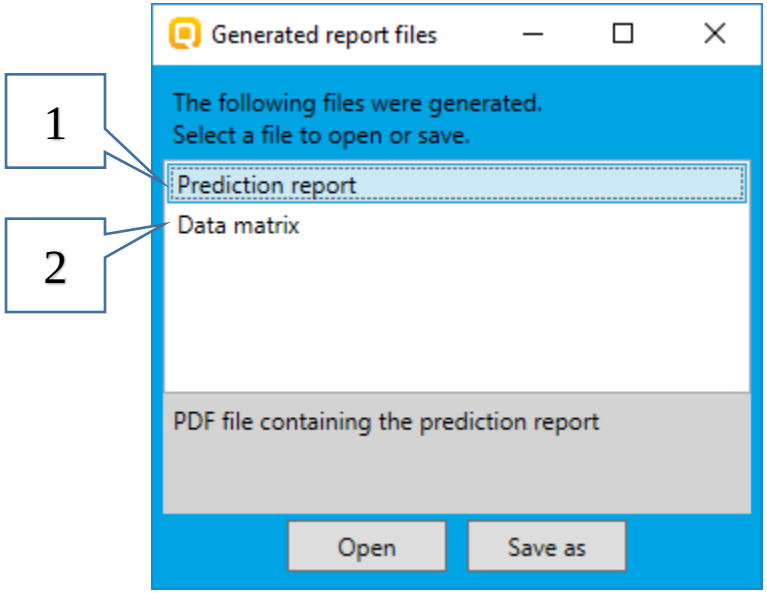


1. Go to the **Report** module; 2. Click on **Prediction**; 3**.** Click on **Create report**

Report
Generation report

After the click on the Create report button, *Generated report files* window appears. It contains two type of files:

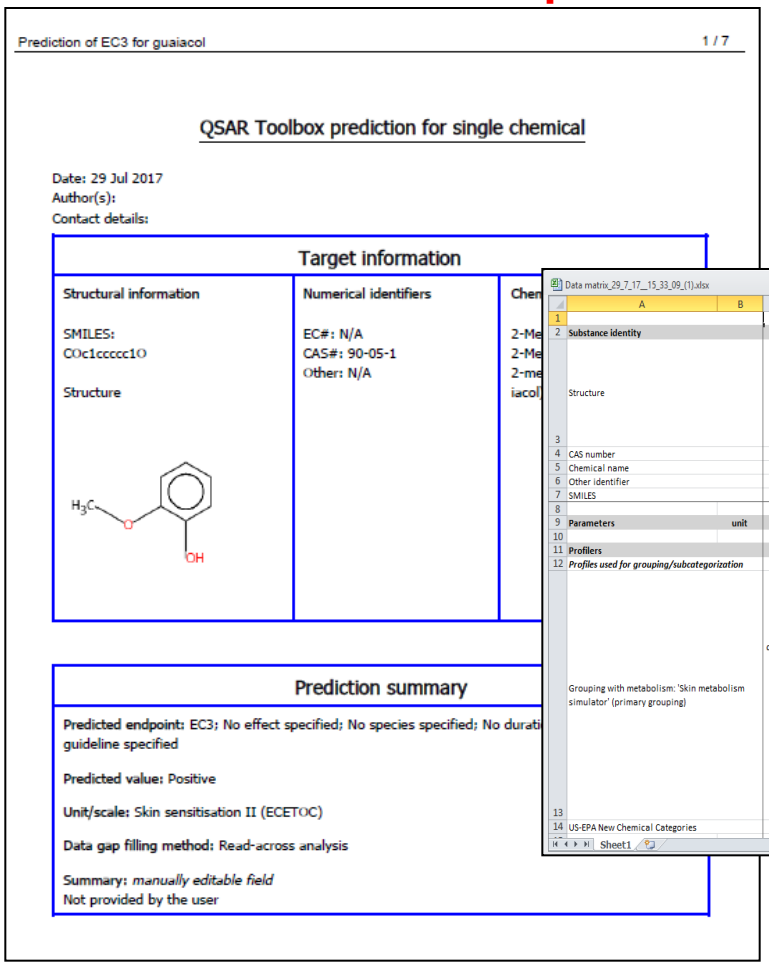
1) **Prediction report** - a PDF file containing the prediction information related to the target.
2) **Data matrix** - a MS Excel file containing chemicals used for prediction along with their data for selected parameters, profiles and endpoint tree positions.



Report

Generated report files

Prediction report



Prediction of EC3 for guaiacol 1 / 7

QSAR Toolbox prediction for single chemical

Date: 29 Jul 2017
Author(s):
Contact details:

Target information

Structural information

SMILES:
COc1ccccc1O

Structure

Numerical identifiers

EC#: N/A
CAS#: 90-05-1
Other: N/A

Prediction summary

Predicted endpoint: EC3; No effect specified; No species specified; No durati
guideline specified

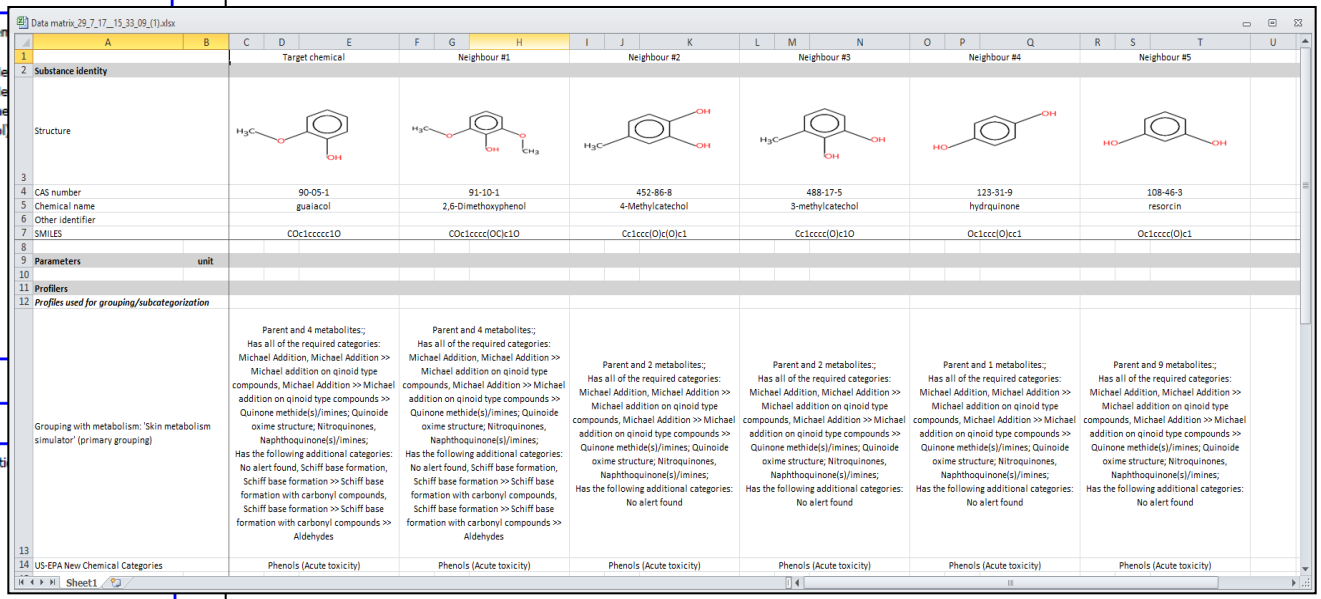
Predicted value: Positive

Unit/scale: Skin sensitisation II (ECETOC)

Data gap filling method: Read-across analysis

Summary: *manually editable field*
Not provided by the user

Data matrix report



Data matrix_29_7_17__15_33_09_(1).xlsx

		Target chemical	Neighbour #1	Neighbour #2	Neighbour #3	Neighbour #4	Neighbour #5
2	**Substance identity**						
	Structure						
4	CAS number	90-05-1	91-10-1	452-86-8	488-17-5	123-31-9	108-46-3
5	Chemical name	guaiacol	2,6-Dimethoxyphenol	4-Methylcatechol	3-methylcatechol	hydrquinone	resorcin
6	Other identifier						
7	SMILES	COc1ccccc1O	COc1cccc(OC)c1O	Cc1ccc(O)c(O)c1	Cc1cccc(O)c1O	Oc1ccc(O)cc1	Oc1cccc(O)c1
8							
9	**Parameters** **unit**						
10							
11	**Profilers**						
12	***Profiles used for grouping/subcategorization***						
13	Grouping with metabolism: 'Skin metabolism simulator' (primary grouping)	Parent and 4 metabolites:; Has all of the required categories: Michael Addition, Michael Addition >> Michael addition on qinoid type compounds, Michael Addition >> Michael addition on qinoid type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines; Has the following additional categories: No alert found, Schiff base formation, Schiff base formation >> Schiff base formation with carbonyl compounds, Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Parent and 4 metabolites:; Has all of the required categories: Michael Addition, Michael Addition >> Michael addition on qinoid type compounds, Michael Addition >> Michael addition on qinoid type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines; Has the following additional categories: No alert found, Schiff base formation, Schiff base formation >> Schiff base formation with carbonyl compounds, Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Parent and 2 metabolites:; Has all of the required categories: Michael Addition, Michael Addition >> Michael addition on qinoid type compounds, Michael Addition >> Michael addition on qinoid type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines; Has the following additional categories: No alert found	Parent and 2 metabolites:; Has all of the required categories: Michael Addition, Michael Addition >> Michael addition on qinoid type compounds, Michael Addition >> Michael addition on qinoid type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines; Has the following additional categories: No alert found	Parent and 1 metabolites:; Has all of the required categories: Michael Addition, Michael Addition >> Michael addition on qinoid type compounds, Michael Addition >> Michael addition on qinoid type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines; Has the following additional categories: No alert found	Parent and 9 metabolites:; Has all of the required categories: Michael Addition, Michael Addition >> Michael addition on qinoid type compounds, Michael Addition >> Michael addition on qinoid type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines; Has the following additional categories: No alert found
14	US-EPA New Chemical Categories	Phenols (Acute toxicity)	Phenols (Acute toxicity)	Phenols (Acute toxicity)	Phenols (Acute toxicity)	Phenols (Acute toxicity)	Phenols (Acute toxicity)

Sheet1

Congratulation

- You have now been introduced to the defining of target endpoint;
- You have now been introduced to the coloring of profilers and databases.
- You have now been introduced to the consecutive steps of the calculation of alert performance.
- Note proficiency comes with practice.