QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.1

Step-by-step example for predicting skin sensitization accounting for abiotic activation of chemicals

Outlook

- **Background**
- Objectives
- The exercise
- Workflow

Background

- This is a step-by-step presentation designed to take the user through the Toolbox workflow for predicting skin sensitization potential of eugenol using a newly implemented categorization tool taking into account its abiotic activation.

Outlook

- Background
- **Objectives**
- The exercise
- Workflow

Objectives

This presentation demonstrates a number of functionalities of Toolbox 4.1:

- Profiling the target chemical.
- Identifying analogues of the target chemical.
- Filling in data gaps of target chemical by means of read-across.
- Profiling target chemical by taking into account its (a)biotic activation.
- Identifying analogues of the target by using a new categorization functionality allowing (a)biotic activation to be taken into account.
- Filling in data gaps by read-across when (a)biotic activation is taken into account.

Outlook

- Background
- Objectives
- **The exercise**
- Workflow

The Exercise

- In this exercise we will predict the skin sensitization potential of target chemical **Eugenol [CAS# 97-53-0].**
- Profile and gather data for the target chemical.
- Two types categorizations are applied:
  - Identifying analogues by using well-known categorization group
  - Identifying analogues based on autoxidation activation of the target illustrating new categorization functionality
- Filling in data gaps by read-across.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**

Workflow

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

Outlook

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

Chemical Input
Overview

- This module provides the user with several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.

Chemical Input
Ways of Entering a Chemicals

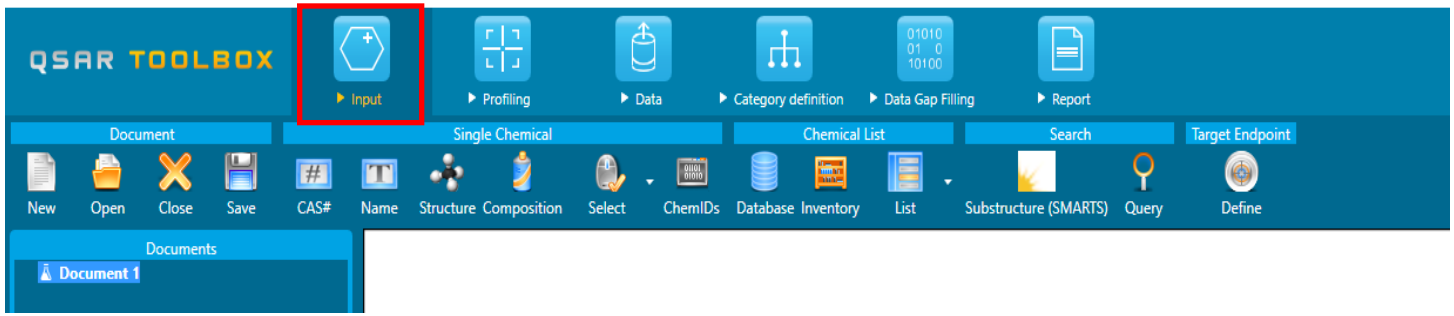
Alternative ways for input of Chemical:

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

2. Group of chemicals
   - User's List/Inventory
   - Customized Databases

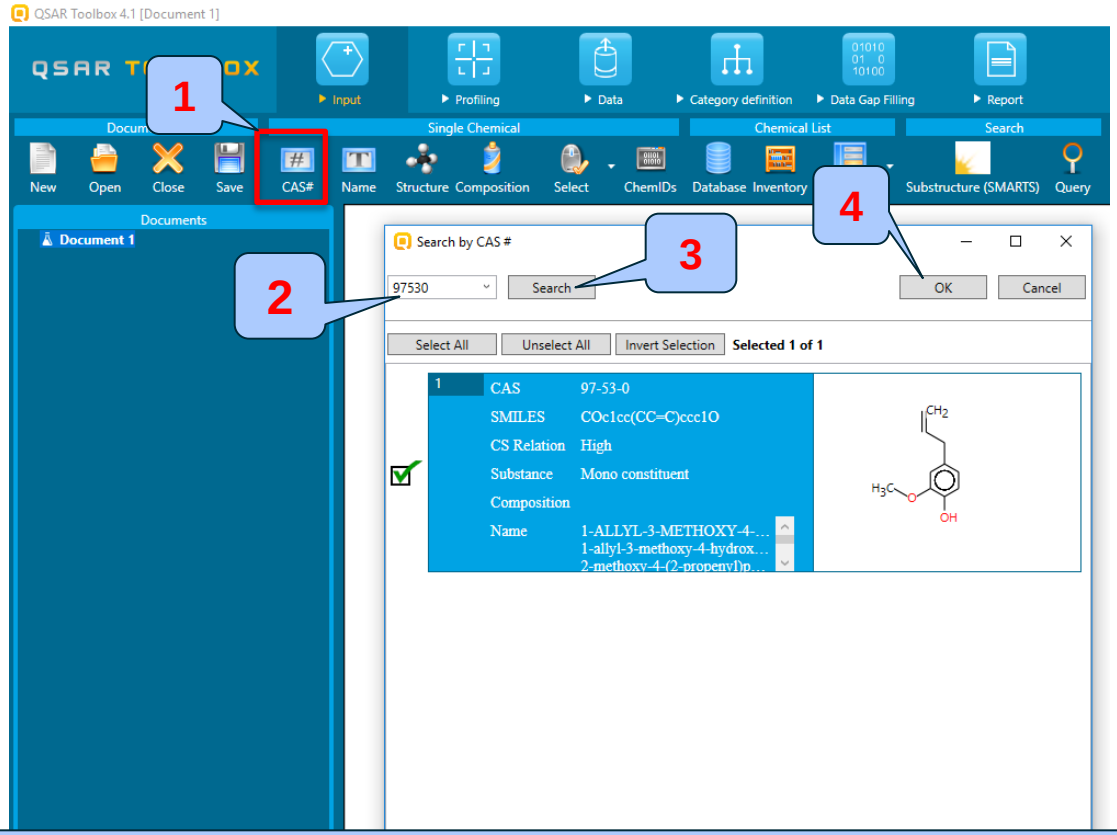
Chemical Input
Input Screen

- Open the Toolbox.
- The six modules in the workflow are seen listed next to "QSAR TOOLBOX" title.
- Click "Input"



Chemical Input

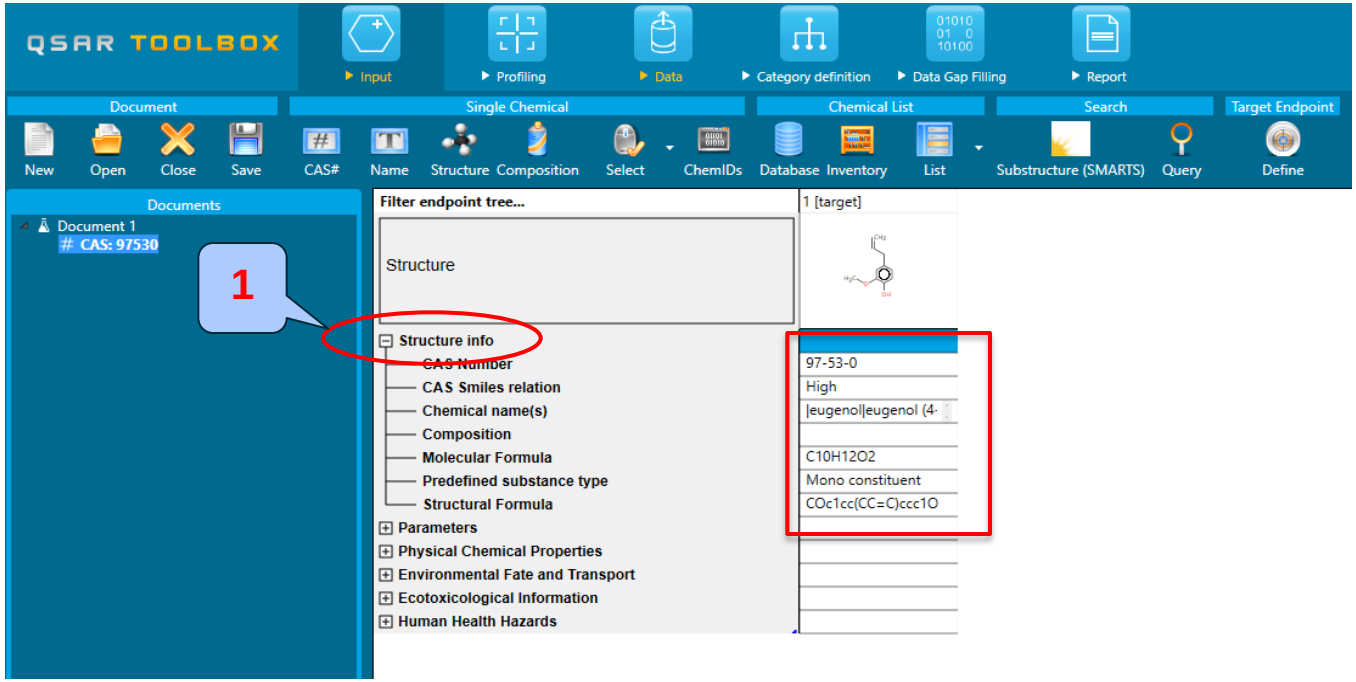
Input target chemical by CAS#



1. Click **CAS#**; 2. Enter the CAS#97-53-0 in the blank field; 3. Click ***Search*** : Eugenol chemical is found in TB databases; 4. Click ***OK*.**

Chemical Input
Target chemical identity

Expanding ***Structure Info*** (1) displays the chemical identification information.



Chemical Input
Target chemical identity

The labels indicated different reliability of the CAS-SMIES relationship are as follows:

High:

This reliability corresponds to high reliability of CAS-SMILES relation. This label is assigned if the chemical belongs to at least one high quality data source (database or inventory)

Moderate:

This reliability corresponds to moderate reliability of CAS-SMILES relation. The moderate label is assigned if the chemical belongs to three “Distribute to QA” data sources.

Low:

This reliability corresponds to poor reliability of CAS-SMILES relation. This label is assigned if the chemical belongs to less than three, but at least one “Distribute to QA” data sources.

Outlook

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

Profiling
Overview

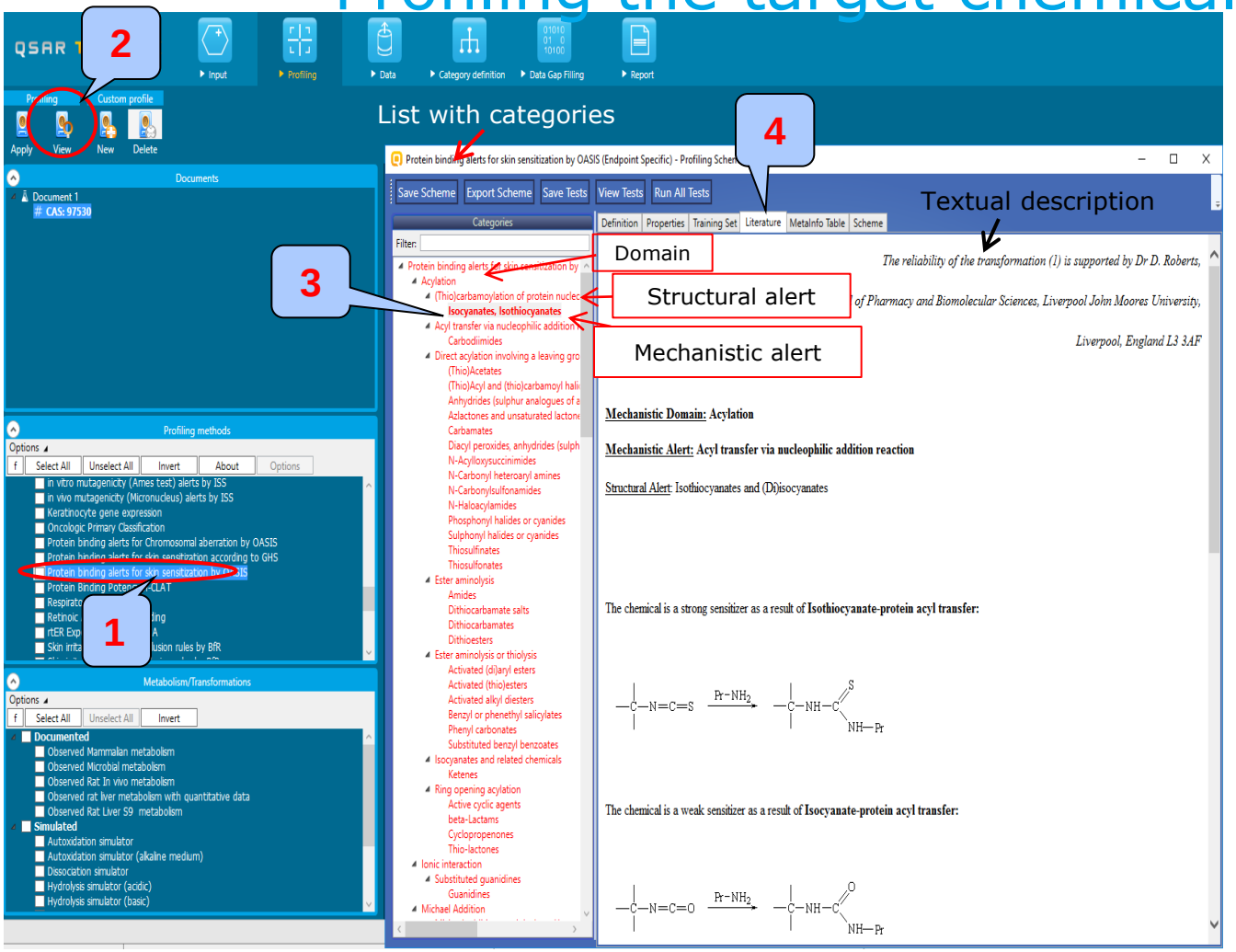
- "Profiling" refers to the electronic process of retrieving relevant information for the target compound, other than its environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- Available information includes the probable mechanism(s) of action, as well as observed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

Profiling

- For most of the profiles, background information can be retrieved by highlighting one of the profiles (for example, Protein binding alerts for skin sensitization by OASIS and clicking “View” (see next screen shot).

Profiling

Profiling the target chemical



1. Highlight the profiler (1); 2. Click ***View*** (2); 3. Click "Isocyanates, Isothiocyanates" 4. Click ***Literature*** to see textual description associated with the category.

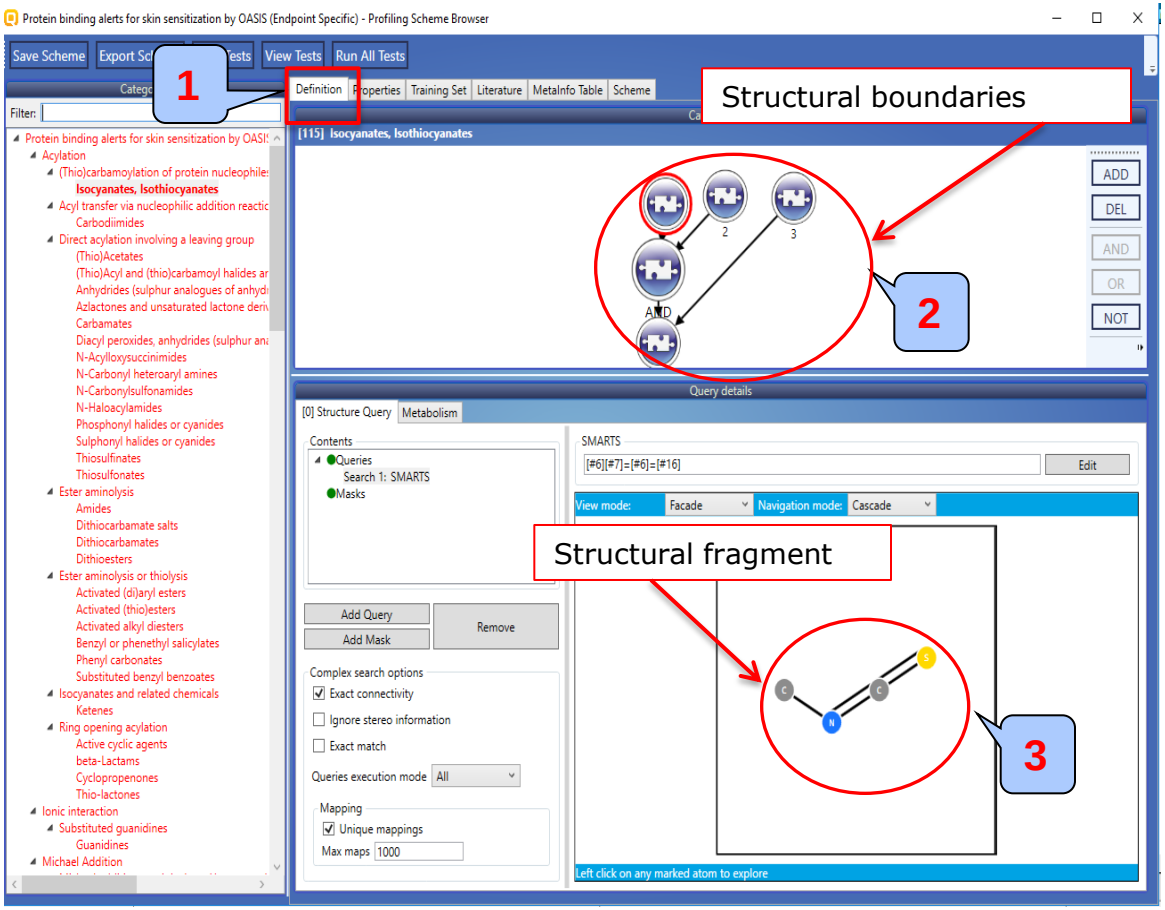
Profiling

Profiling the target chemical

1. Click ***Definition*** in order to see more details about the structural boundaries of *Isocyanates, Isothiocyanates;*

2. The structural boundaries implemented in the category are shown in (2);

3. The structural fragment in the first structural boundary (circled in red) is shown in (3).



Profiling

Profiling the target chemical

- The outcome of the profiling determines the most appropriate way to search for analogues.
- The following profiling schemes are relevant to the **Skin sensitization**:
  - Protein binding by OECD – general mechanistic
  - Protein Binding Potency – general mechanistic
  - Protein binding alerts for skin sensitization by OASIS – endpoint specific

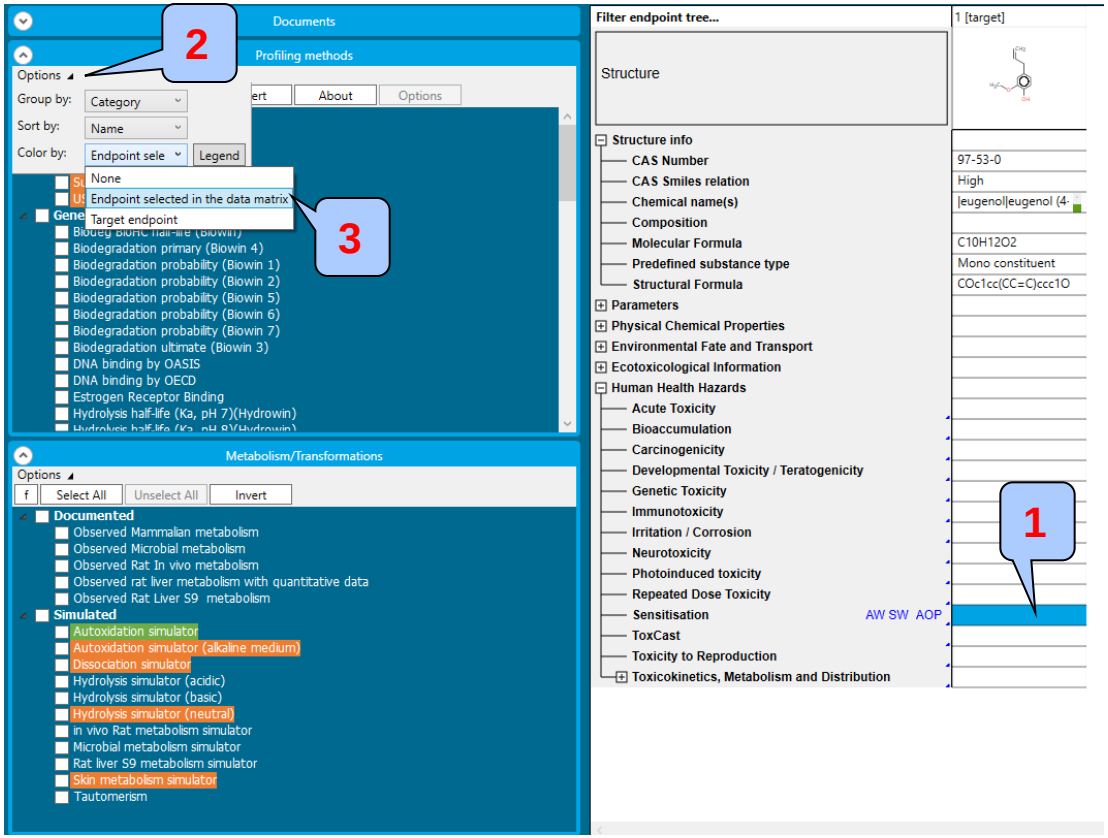
Profiling

Profiling the target chemical

- Click in the box next to the name of the profiling methods related to the target endpoint.
- This selects (a green check mark appears) or deselects (the green check mark disappears) profilers.
- An option color by ***Endpoint selected in the data matrix*** is implemented, which highlights all relevant profiles (see next slide).

Profiling

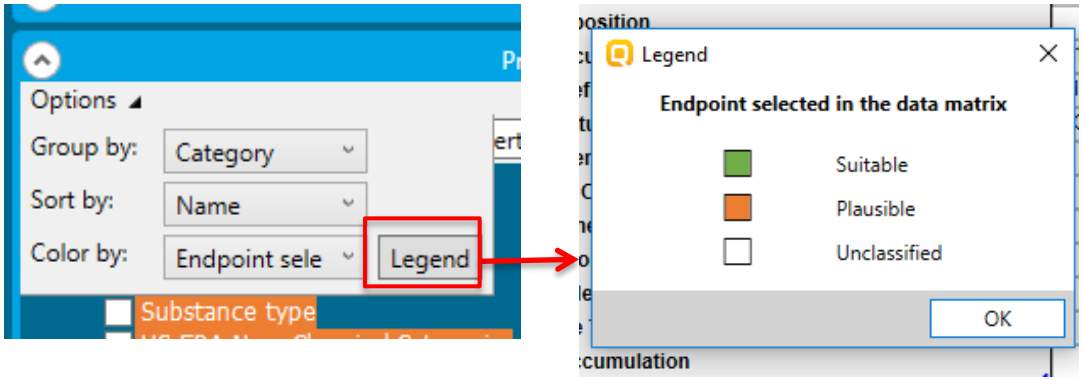
Profiles' relevancy- new functionality



1. Open the endpoint tree and select the cell at the **Sensitization** level (1)
2. Expand ***Options*** (2);
3. Select ***Color by Endpoint selected in the data matrix*** (3);
4. All profiles and metabolic simulators relevant to the selected endpoint will be colored (see next slide for color legend).

Profiling

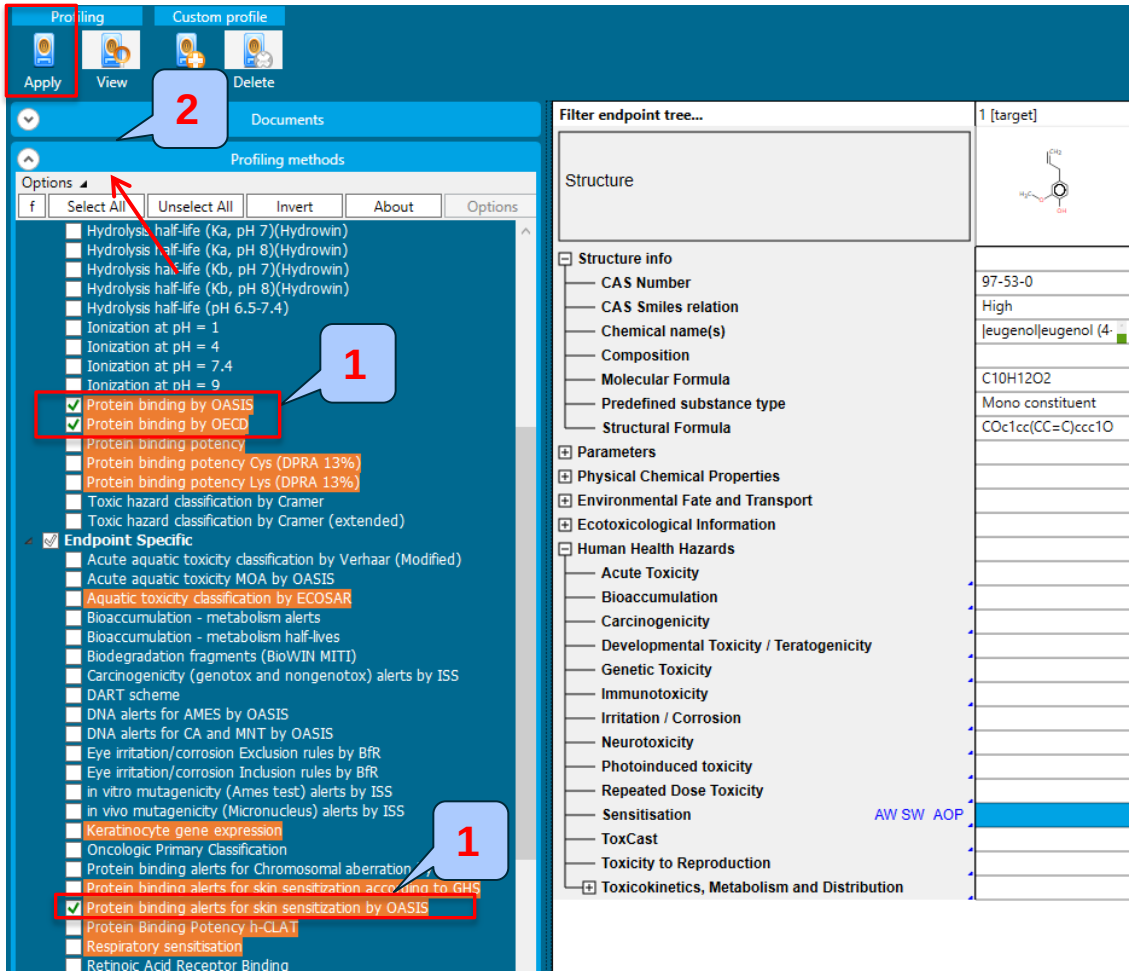
Profiles' relevancy- new functionality



- **Suitable** - developed using data/knowledge for the target endpoint;
- **Plausible** – not endpoint specific; structure-based; form broader group of analogues;
- **Unclassified** – all profiles, which are not classified in any of the categories above.

Profiling

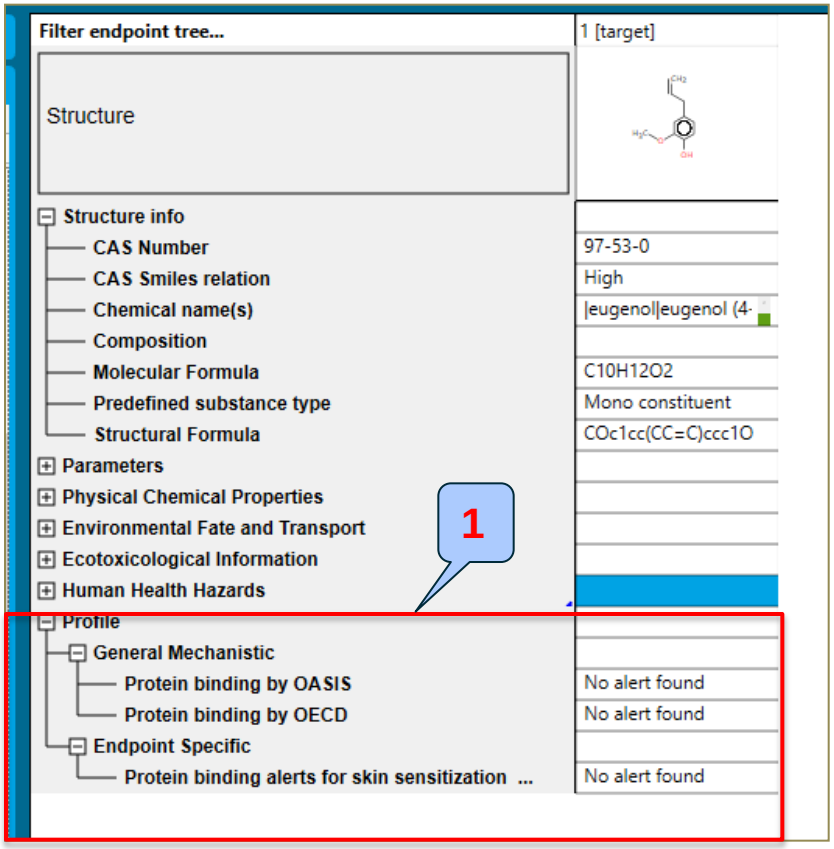
Profiling the target chemical



1. Tick the checkboxes of Protein binding alert by OASIS, Protein binding by OECD and protein binding alerts for skin sensitization by OASIS.
2. Click ***Apply***;

Profiling

Profiling the target chemical



- The actual profiling takes up to several seconds depending on the number and type of profiles selected.
- The results of profiling automatically appear at the bottom of the endpoint tree (1).
- No protein binding alert has been found for the target compound (eugenol) based on three protein binding profilers.

Outlook

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

Data
Overview

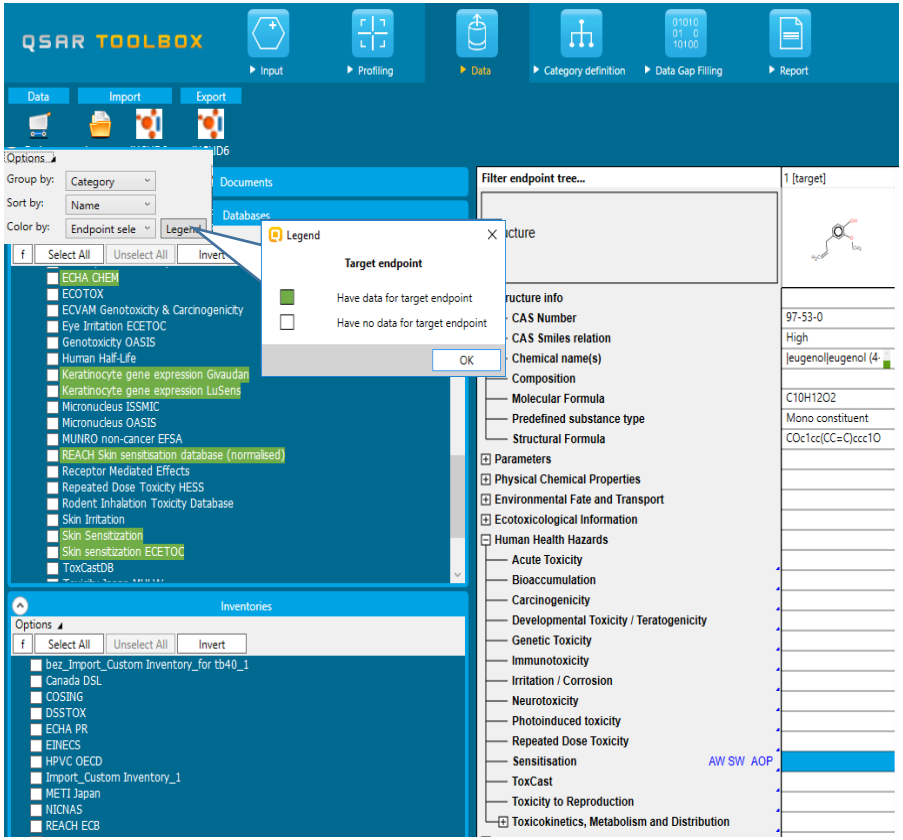
- "Data" refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.
- Data gathering can be executed in a global fashion (i.e., collecting all data for all endpoints) or on a more narrowly defined basis (e.g., collecting data for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

Data
Case study

- In this example, we limit our data gathering to a single toxicity endpoint (skin sensitization).
- In this example, we collect data from the databases containing experimental results for Skin sensitisation (Skin sensitisation and Skin sensitisation ECETOC).

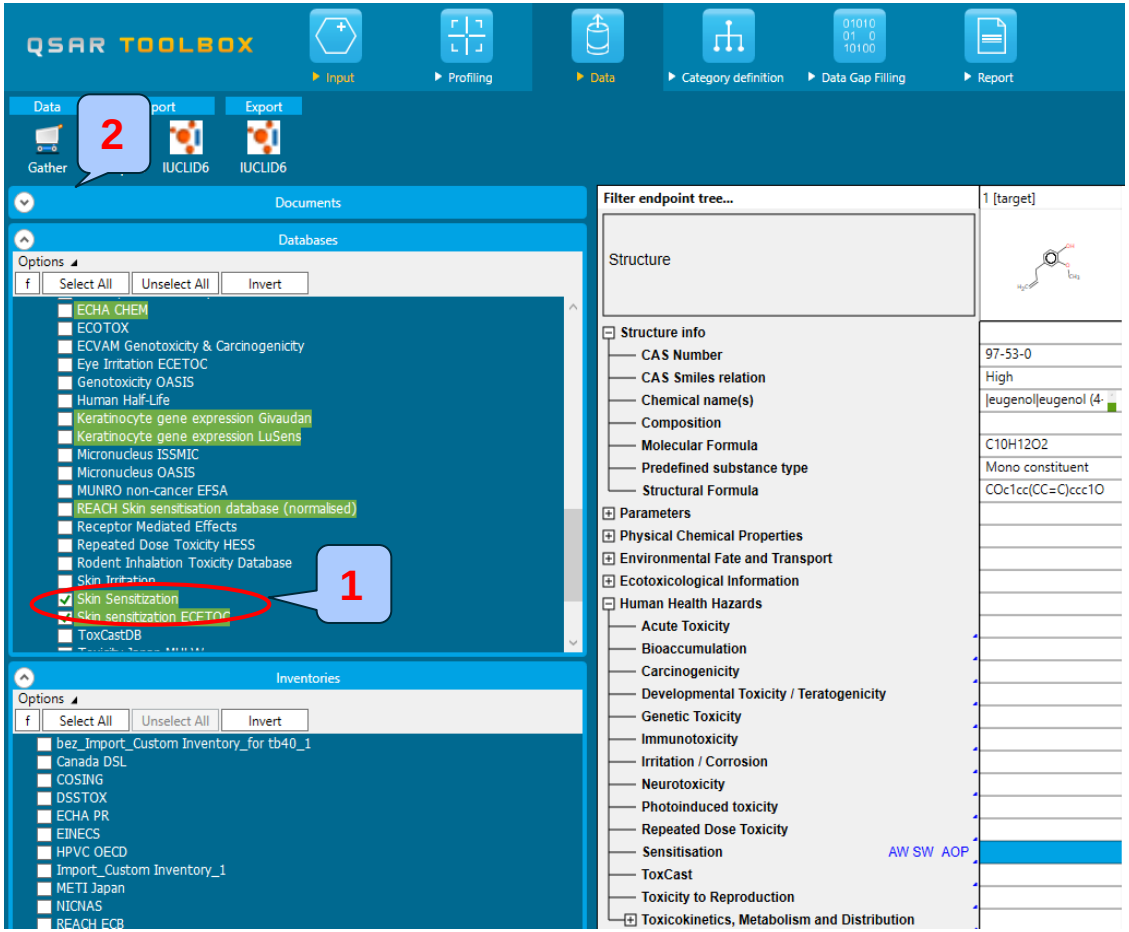
Data

Database` relevancy-new functionality



1.Click the cell at the ***Sensitization*** level (1);
2. Expand ***Options*** (2);
3. Select ***Colour by Endpoint selected in the data matrix*** (3);
4. All databases that have data for the target endpoint are coloured in green.
5 . Click ***Legend*** to open the colour legend (5).

Data
Gather data



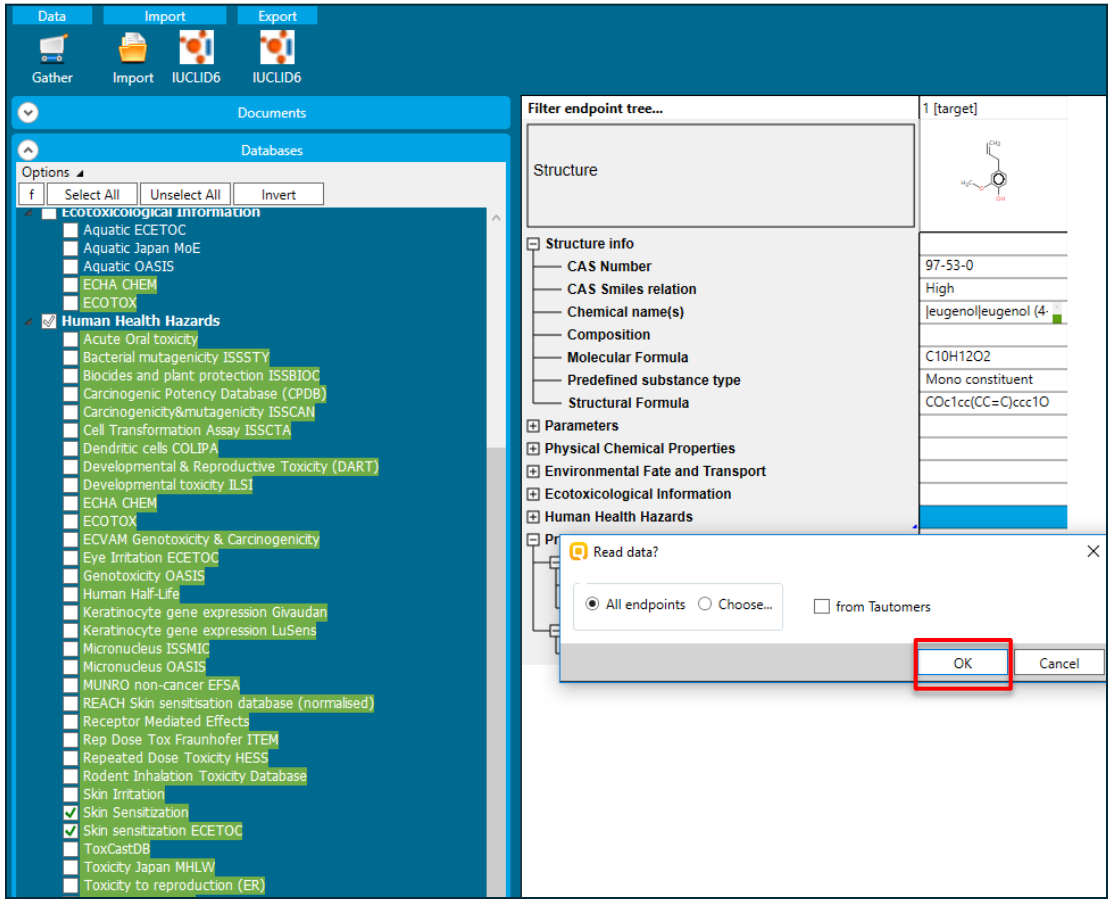
1. Tick *Skin Sensitization* and *Skin Sensitization ECETOC* (1);
2. Click ***Gather*** (2);

Data
Gather data

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

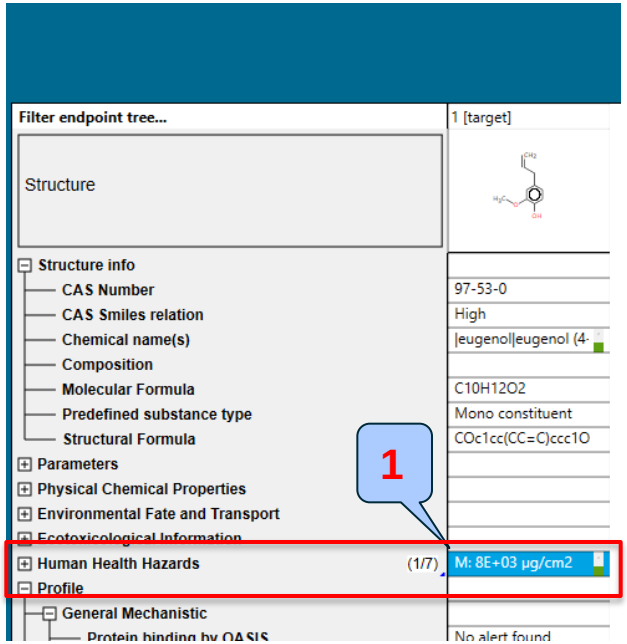
- It should be kept in mind that the search for data and analogues is performed only among the chemicals which are listed in the selected databases, which in this example are **Skin sensitization** and **Skin sensitization ECETOC .**

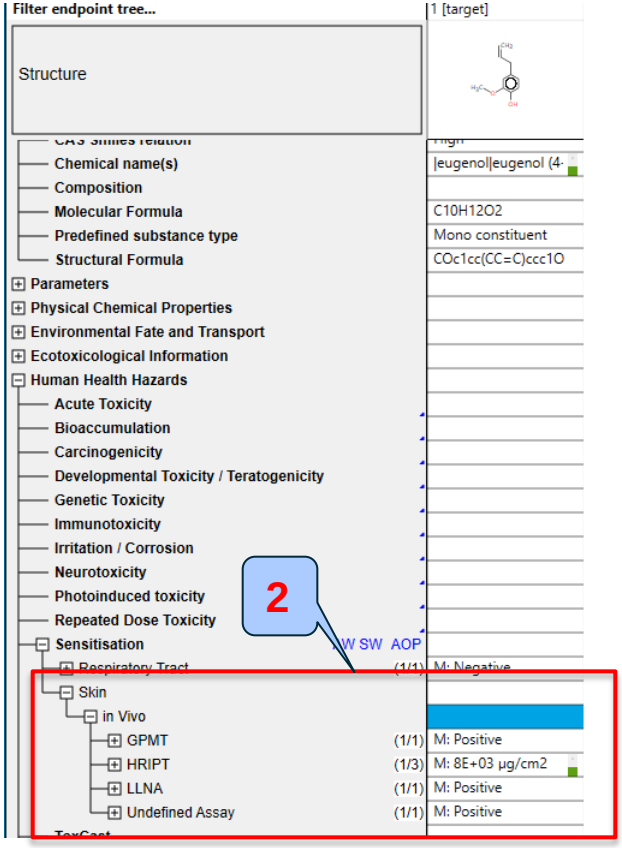
Data
Gather data



1. A "Read data?" window appears.
2. You could choose to gather "all" or "endpoint specific" data.
3. Click ***OK*** to read all available data;

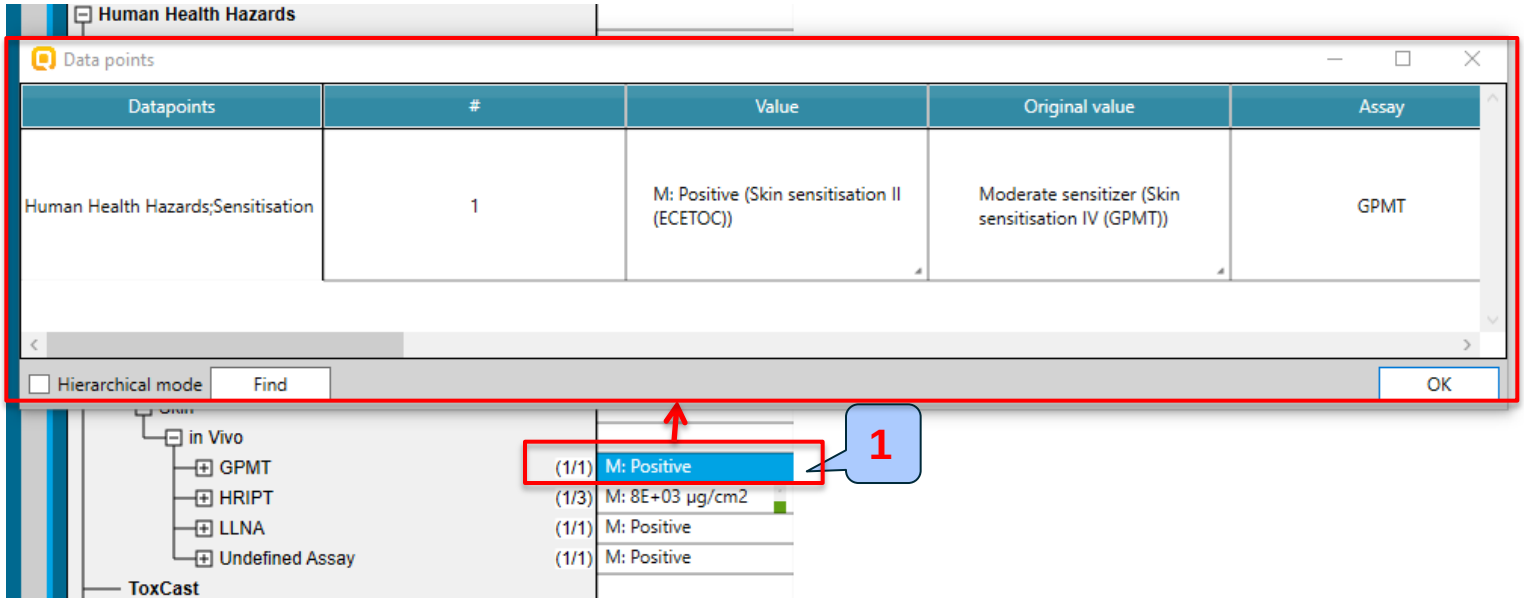
Data
Gather data





1. The extracted data is displayed on the data matrix;
2. Expand the Human Health node to display all skin experimental data;
3. Positive experimental data is found.

Endpoint
Gather data



1. **Double-click** on the cell displays metadata information for the observed data.

Recap

- The first module, introduces the target chemical, ensure for correctness of the structure.
- The second module shows that there is no protein binding alert for the target chemical.
- In the third module, you have found that the target chemical has positive skin sensitization data.
- In the further read-across analysis we will try to reproduce positive skin sensitization data.
- The study continues with identifying analogues and applying read-across.

Outlook

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

Category Definition
Grouping methods

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

Category Definition
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.

Basic guidance for category formation and assessment

Suitable categorization phases:

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

Performing categorization:

1. The categorization phases should be applied successively.
2. The application order of the phases depend on the specificity of the data gap filling.
3. More categories of same Phase could be used in forming categories.
4. Some of the phases could be skipped if consistency of category members is reached.

Graphical illustration of suitable categorization phases is shown on next slide

Suitable Categorization/Assessment Phases

Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Broad grouping
Endpoint Non-specific

Repeating Phase I due to Multifunctionality of chemicals

Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- Genotoxicity/carcinogenicity
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules

Subcategorization
Endpoint Specific

Metabolism accounted for

Phase III. Eliminating dissimilar chemicals

Apply Phase I – for structural dissimilarity
Filter by test conditions – for Biological dissimilarity

Subcategorization
Endpoint Specific

Category Definition

Grouping methods – phase I

Suitable Categorization/Assessment Phases

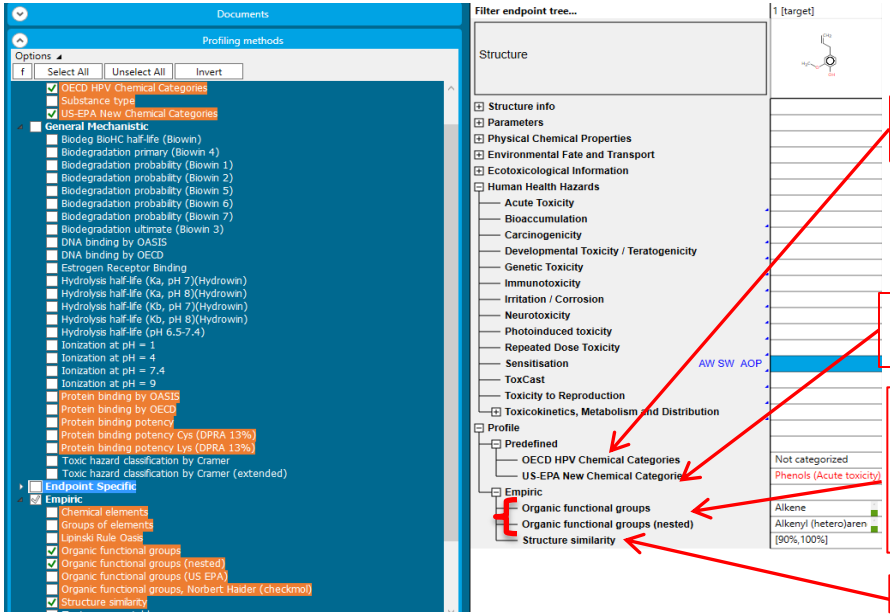
Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity

Broad grouping
Endpoint Non-specific

Each of the above grouping method is applied to the target chemical and number of the identified analogues are provided below. In order to preserve the basic functional groups available within the molecule: Allyl, Ether and Phenol, OFG are used for categorization purposes. US-EPA and OECD are not used because they omit the other two important functionalities: Allyl and Ether. Str. similarity identifies small set of analogues and apparently could not be used for categorization.

Phase I categorization in Toolbox



1007 analogues are identified.

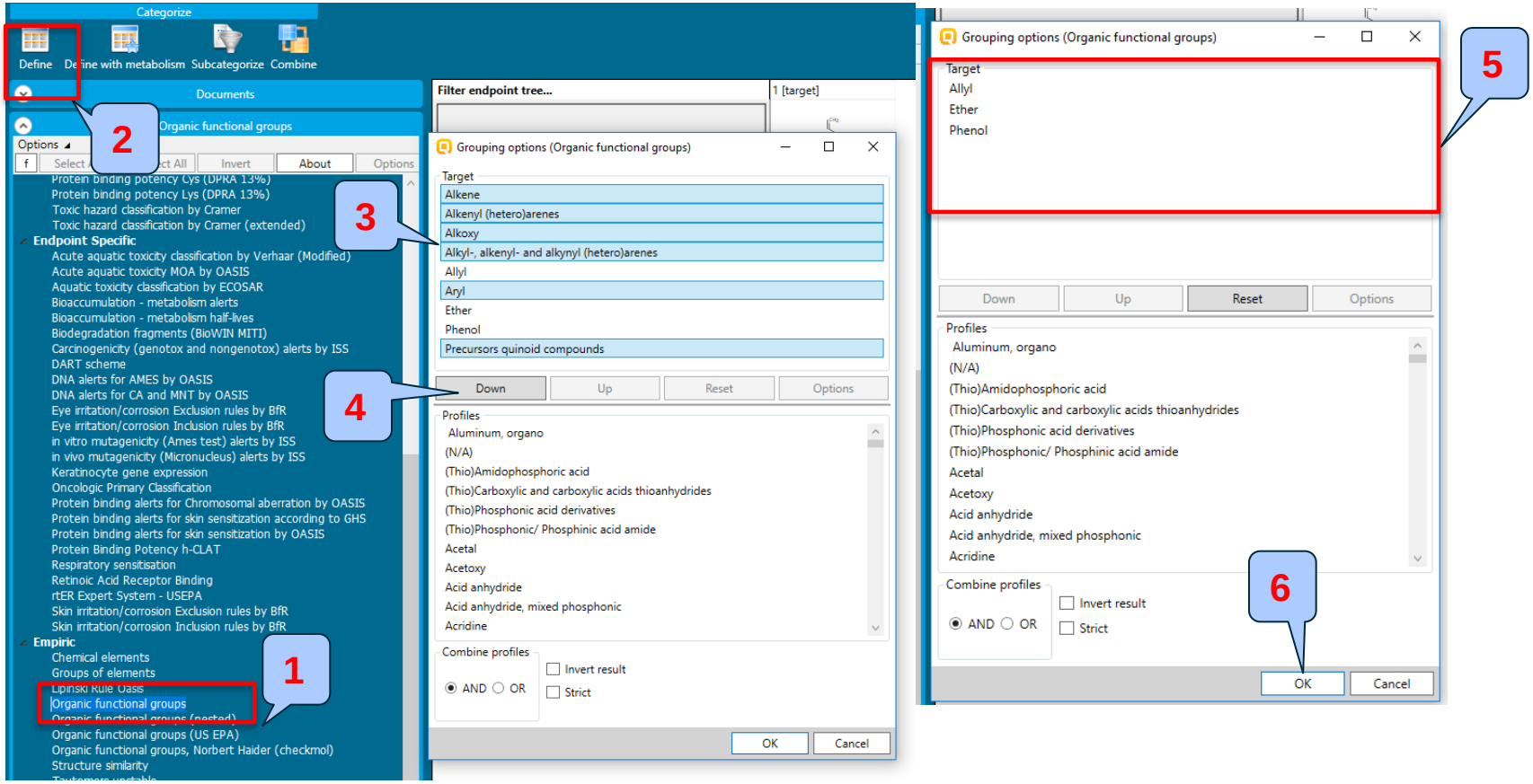
127 analogues are identified

2 analogues are identified (in case all categories are preserved)
10 analogues are identified (in case Allyl, Ether and Phenol are preserved)

38 analogues are identified

Category Definition

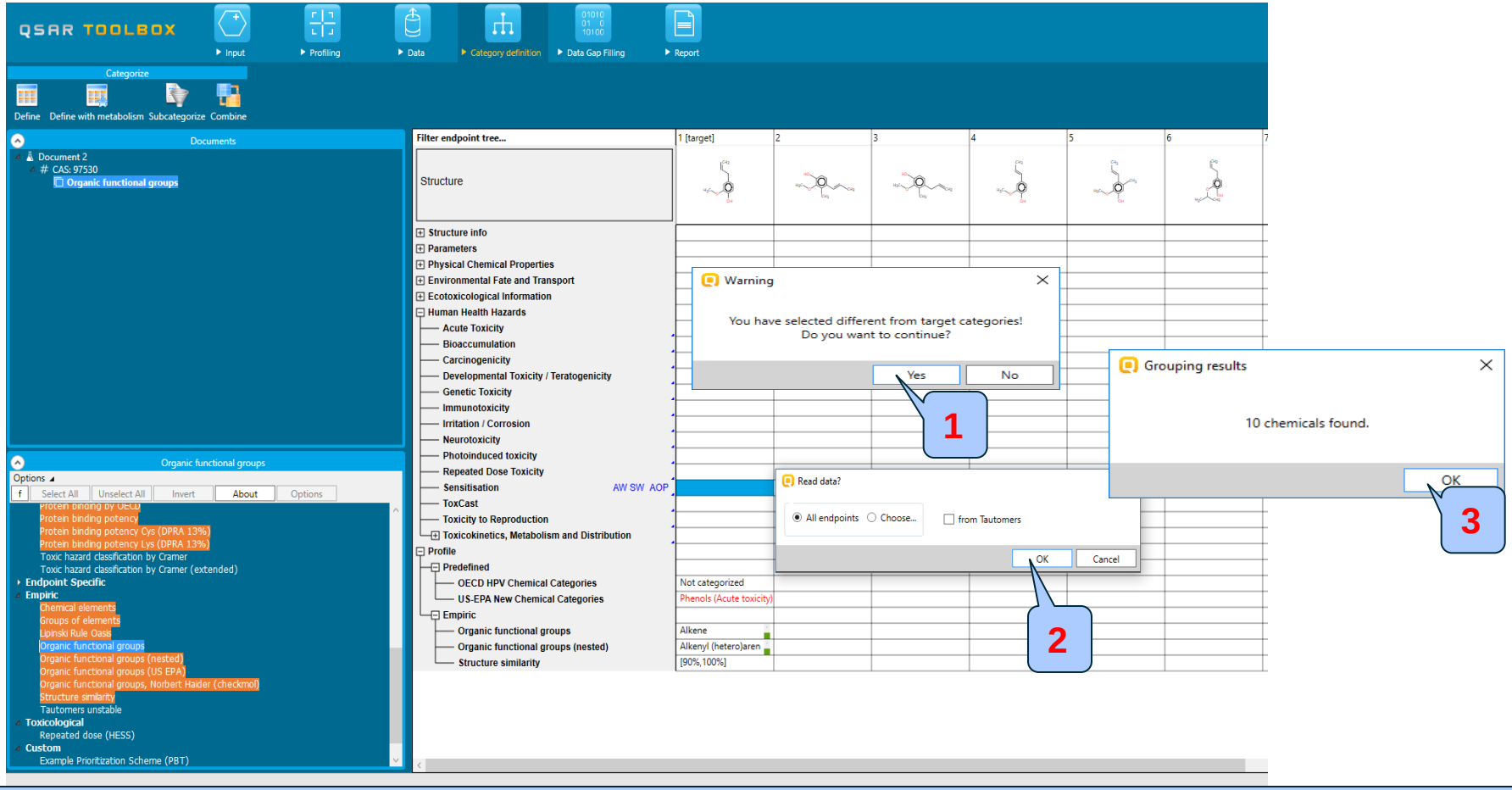
Define category by OFG



1. Select ***Organic functional groups***; 2. Click ***Define*** 3. Click all groups (highlighted in blue) but Allyl, Ether and Phenol by also holding the Ctrl button. 4. Click ***Down*** to remove them. They are moved in the panel down; 5. Only Allyl, Ether and Phenol should remain in the upper panel; 6.Click *OK*.

Category Definition

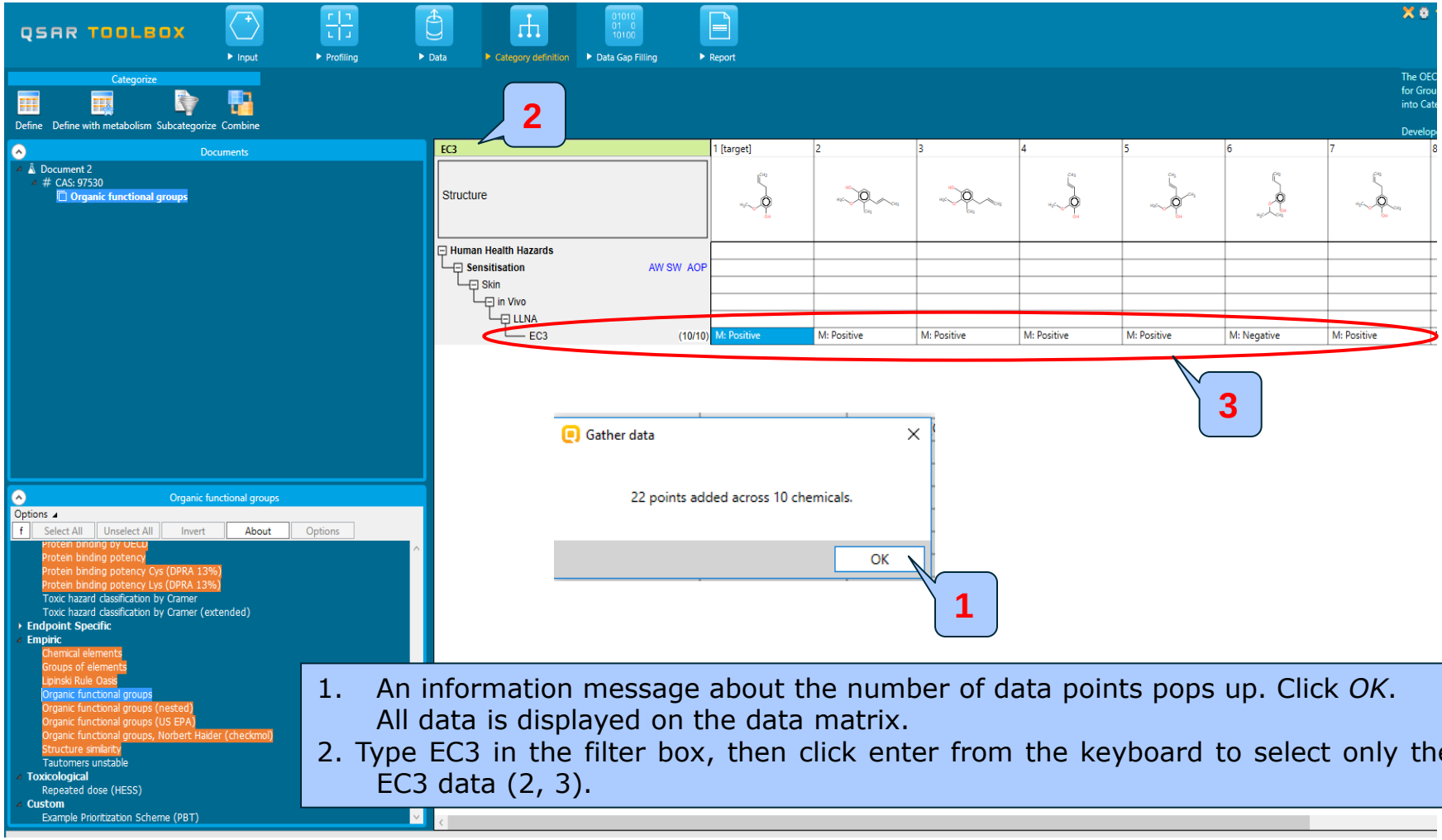
Define category by OFG



1. A warning message is displayed informing that the selected category differ from the target ones. Click *Yes*.
2. Click *OK* to confirm the grouping results;
3. Click *OK* to read all available data;

Category Definition

Gather data for analogues chemicals



1. An information message about the number of data points pops up. Click *OK*. All data is displayed on the data matrix.
2. Type EC3 in the filter box, then click enter from the keyboard to select only the EC3 data (2, 3).

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - **Data gap filling without accounting metabolism**

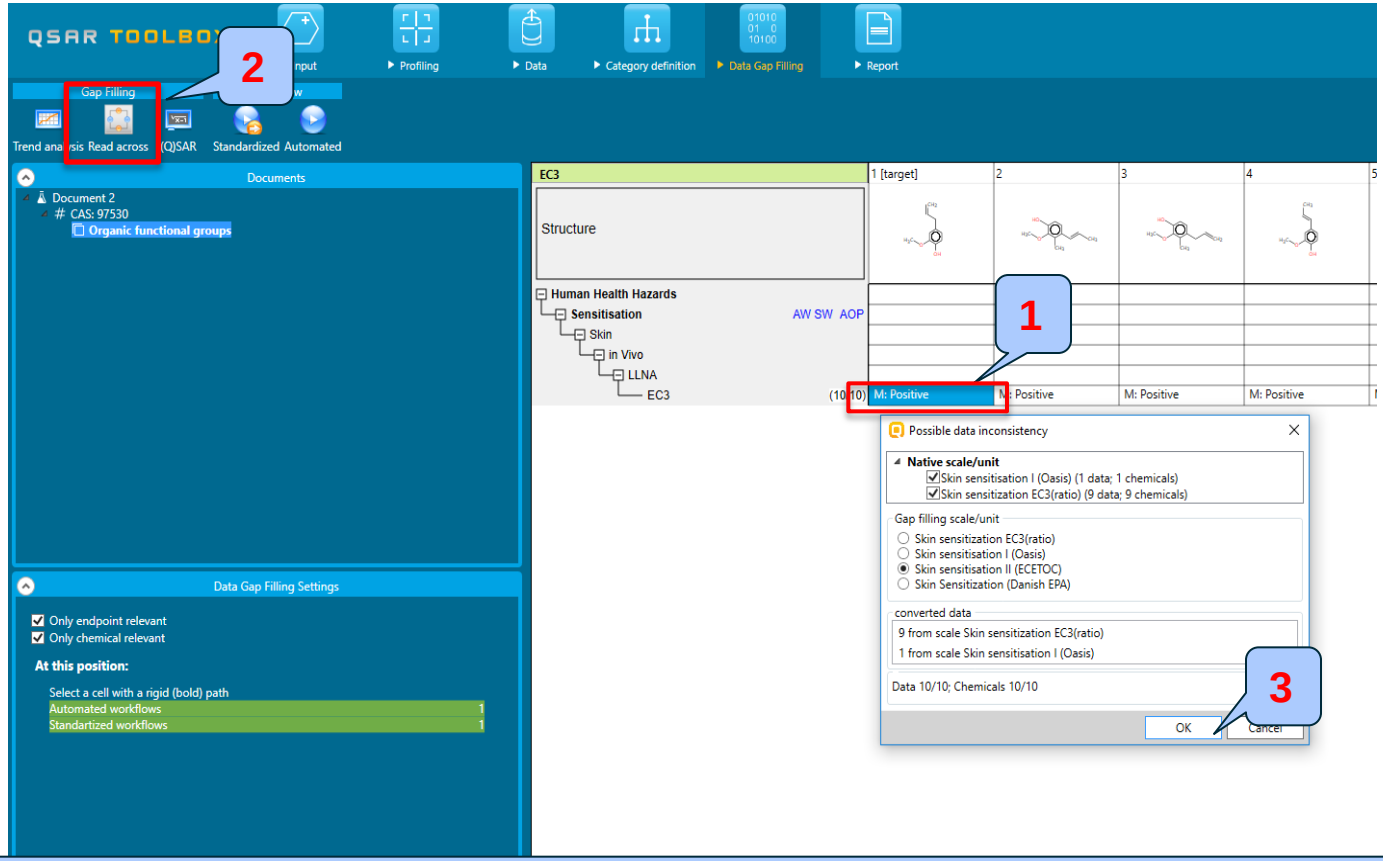
Data Gap Filling
Overview

- "Data Gap Filling" module provides three different data gap filling tools:
  - Read-across
  - Trend analysis
  - (Q)SAR models

- The most relevant data gap mechanism should be used by 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.

- In this example, we use read-across.

Data gap filling
Apply Read-across



1. Click the cell corresponding to *Sensitization>>Skin>>In Vivo>>LLNA>>EC3;*
2. Select ***Read-across****;*
3. A pop-up window informing about possible data inconsistency appears.

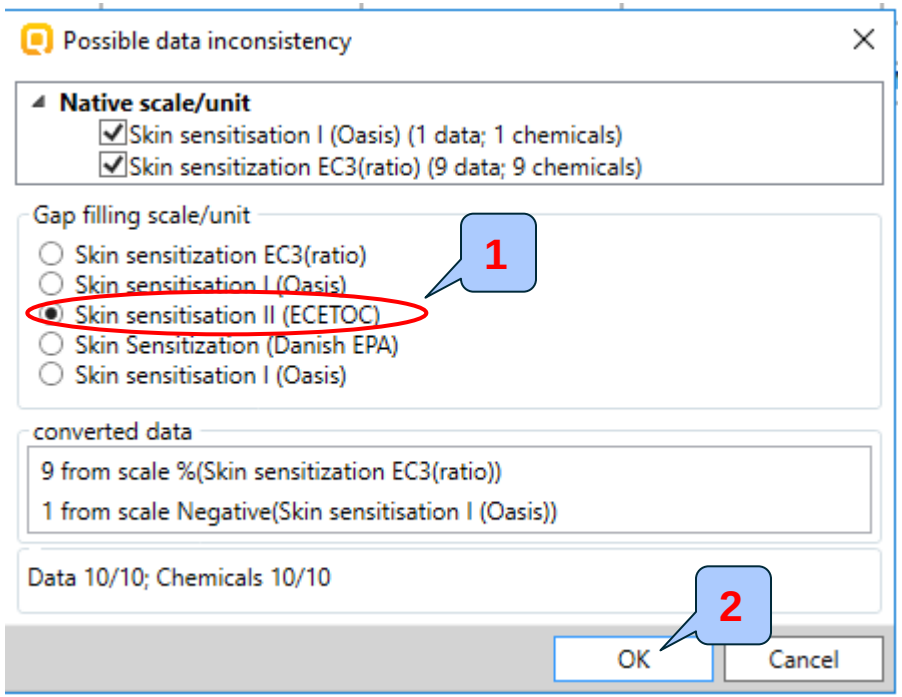
More details about scale definitions is provided on next slide.

Data gap filling
Scale definition

- Skin sensitisation is a "qualitative" endpoint for which the results are presented with categorical data (for example: positive; negative; weak sensitizer; strong sensitizer ,etc).
- The skin sensitisation potential of chemicals is coded in different scales depending on their source (for example: data from John Moores University of Liverpool is classified as: *Strongly sensitizing, Moderately sensitizing etc*.; data from European centre for Ecotoxicology and Toxicology of chemicals is classified as: *Positive, Negative, and Equivocal*).
- The main purpose of the scales is to unify all data available in the Toolbox databases for a certain endpoint.
- The default scale for Skin Sensitisation is "Skin Sensitisation ECETOC". It converts all skin data into: Positive, Negative.

Data gap filling
Scale definition

Back to our example:

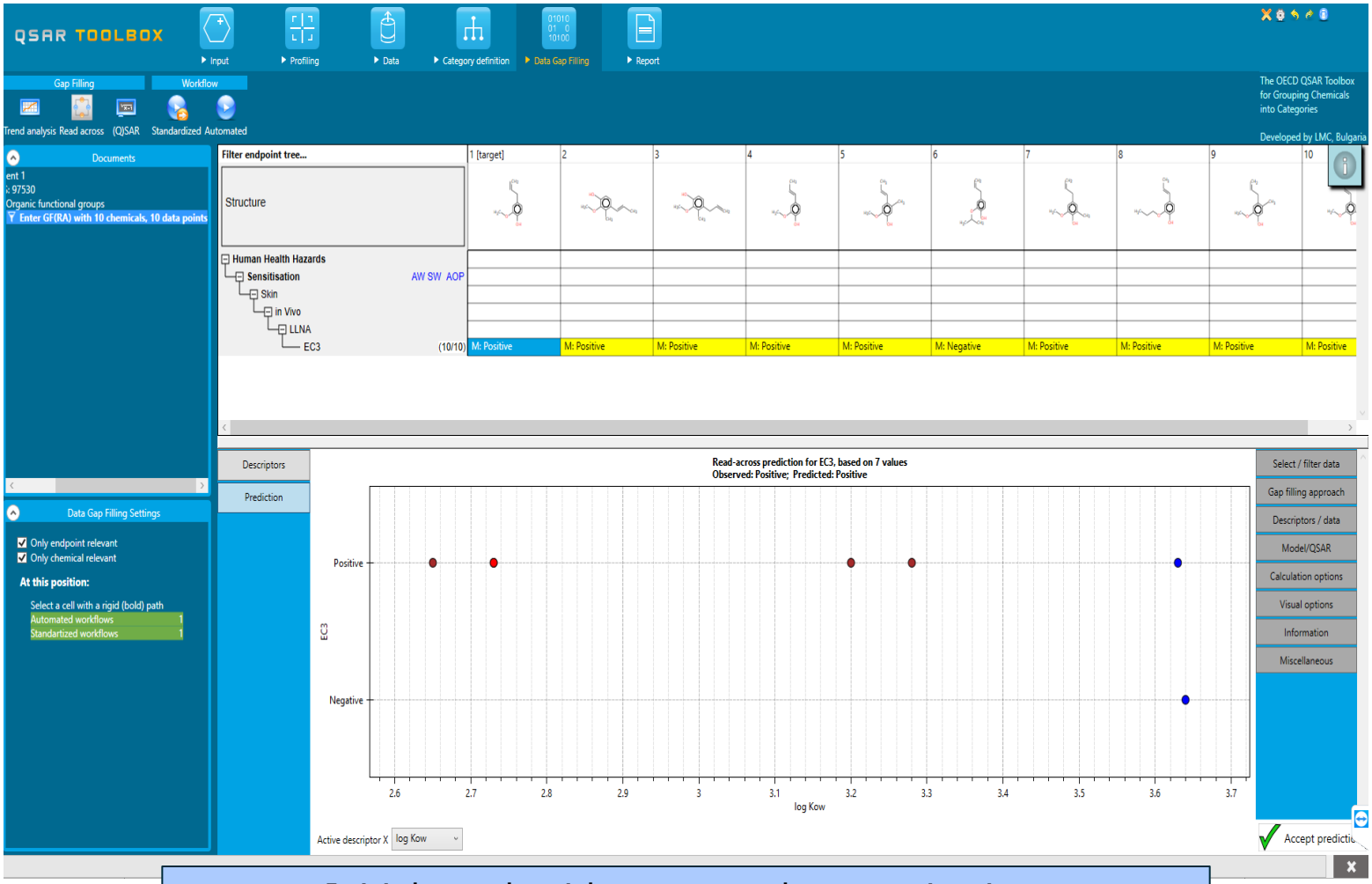


1. Verify that the default scale "Skin sensitisation II (ECETOC)" is selected;
2. Click ***OK***;

Data gap filling
Read-across

- The resulting plot places the experimental results of all analogues (Y axis) against the descriptor (X axis) which by default is log*Kow* (see next screen shot).
- The **RED** point represents predicted results for the target chemical.
- The **BROWN** points represent the experimental data of the analogues that are used for the read-across (by default are the five nearest neighbours to the target)
- The **BLUE** points represent the experimental data of the analogues not used in the read-across.

Data gap filling
Read-across



Initial graph without any subcategorization.

Data Gap Filling
Subcategorizations

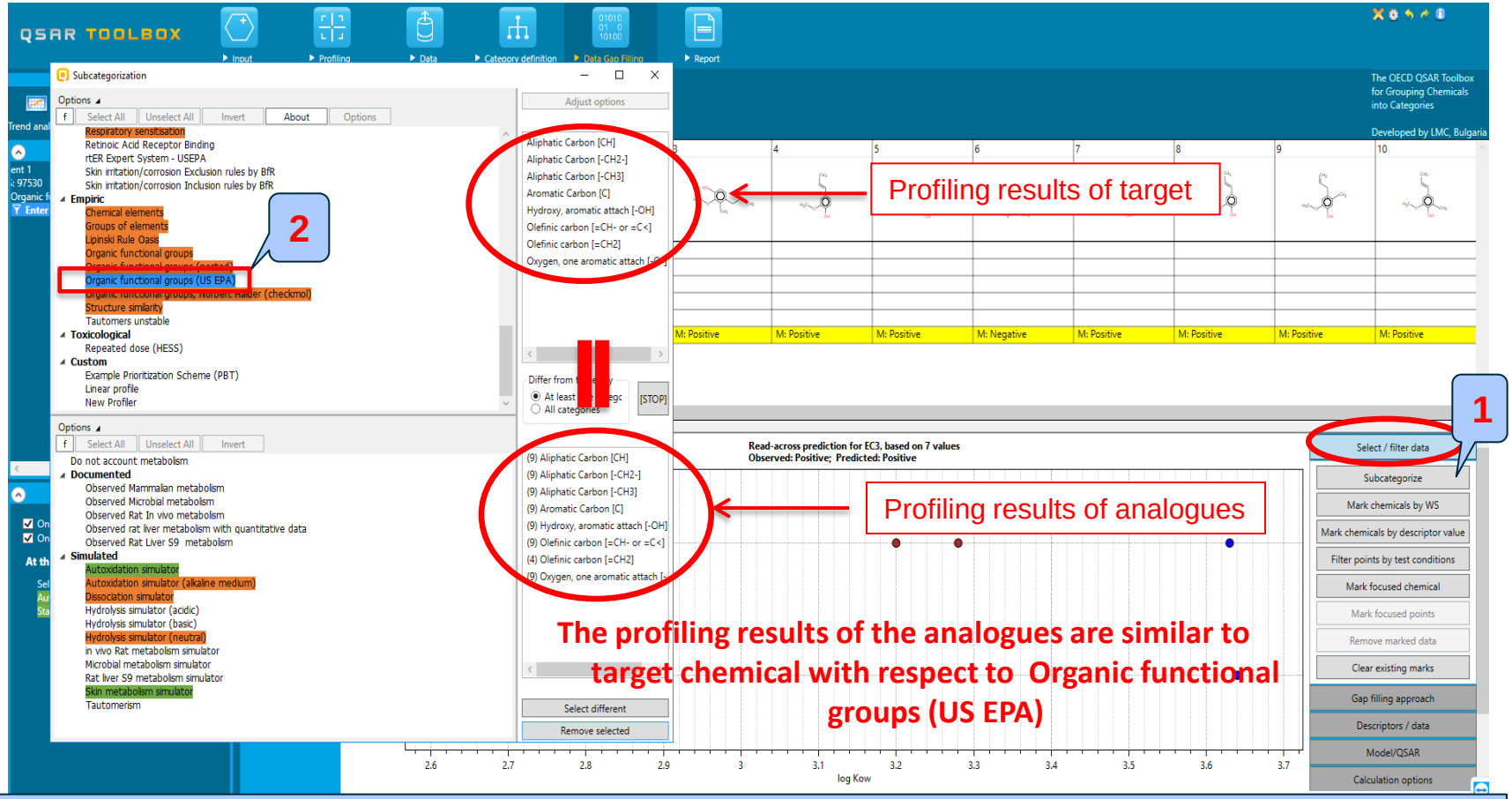
In this example two subcategorizations are applied in order to eliminate dissimilar analogues:

- Organic functional group (US-EPA) – phase I is repeated in order to eliminate multifunctional analogues (subcategorization 1)
- Protein binding alerts for skin sensitization by OASIS (subcategorization 2)

See next slides.

Data gap filling

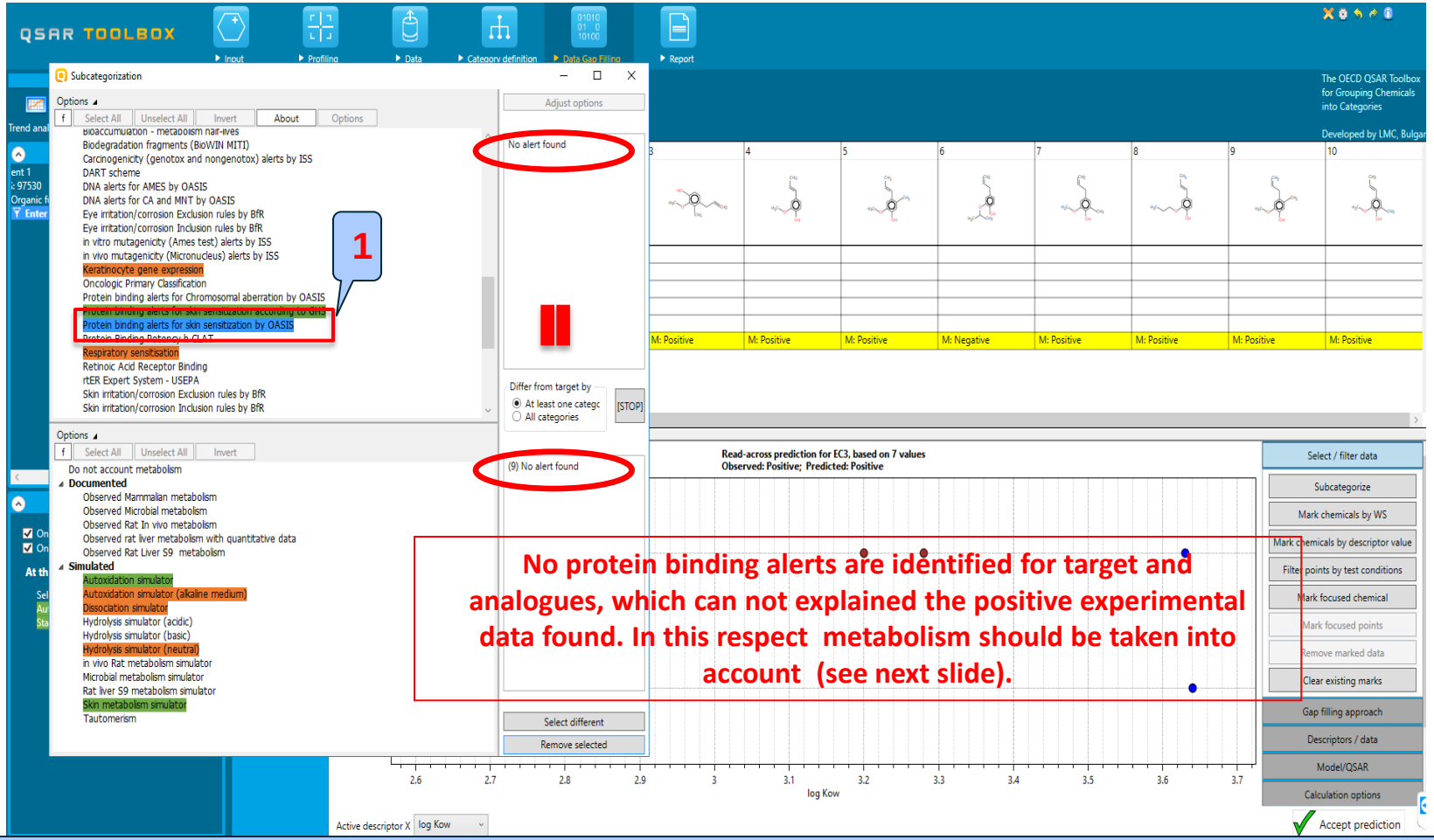
Subcategorization 1: Organic functional groups (US EPA)



1. Open ***Select/filter data/Subcategorize;*** 2. Select ***Organic functional groups (US EPA).***

Data gap filling

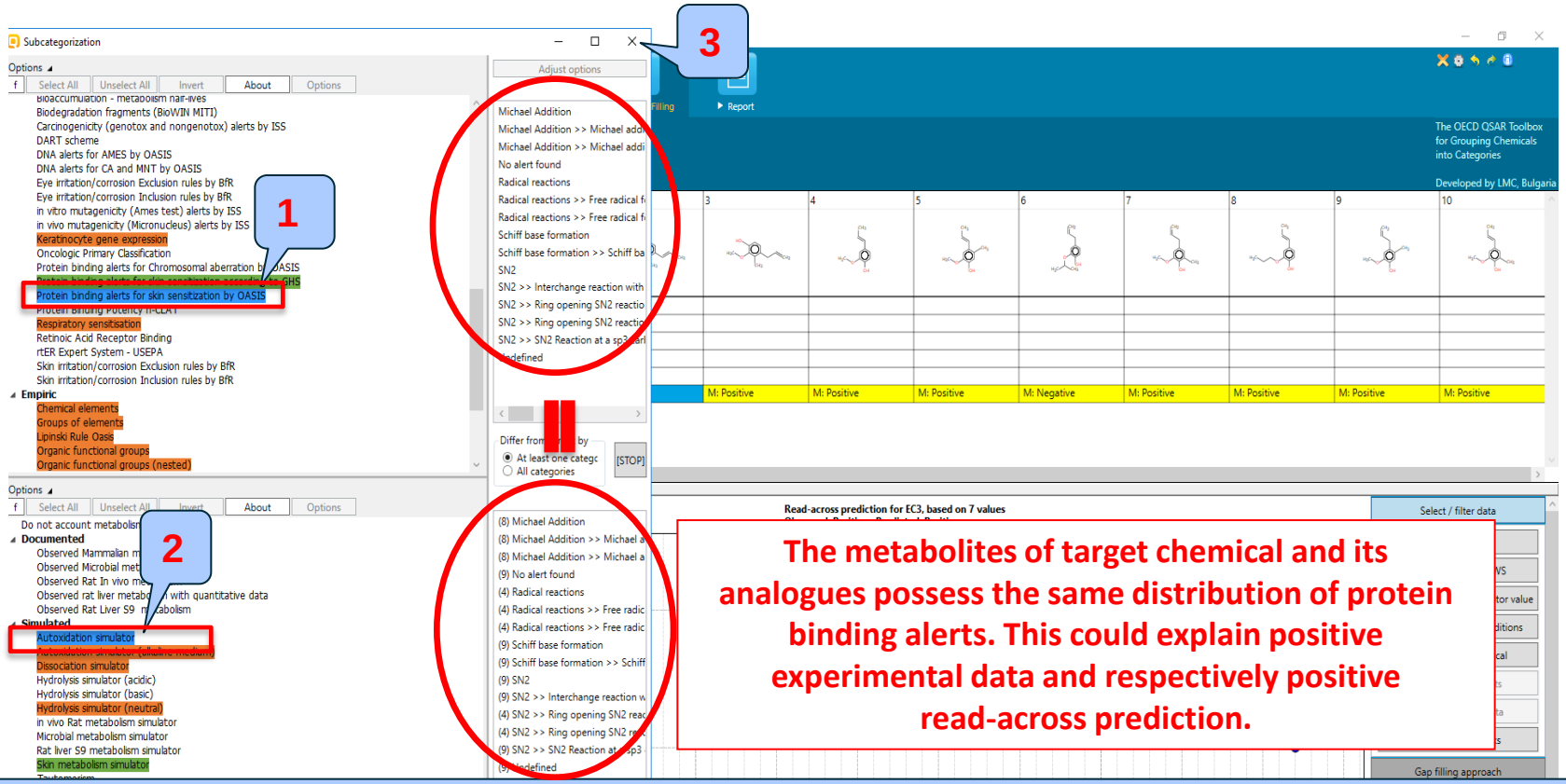
Subcategorization 2: Protein binding alerts for skin sensitization by OASIS



1.Select *Protein binding alerts for skin sensitization by OASIS.*

Data gap filling

Subcategorization when metabolism is taken into account



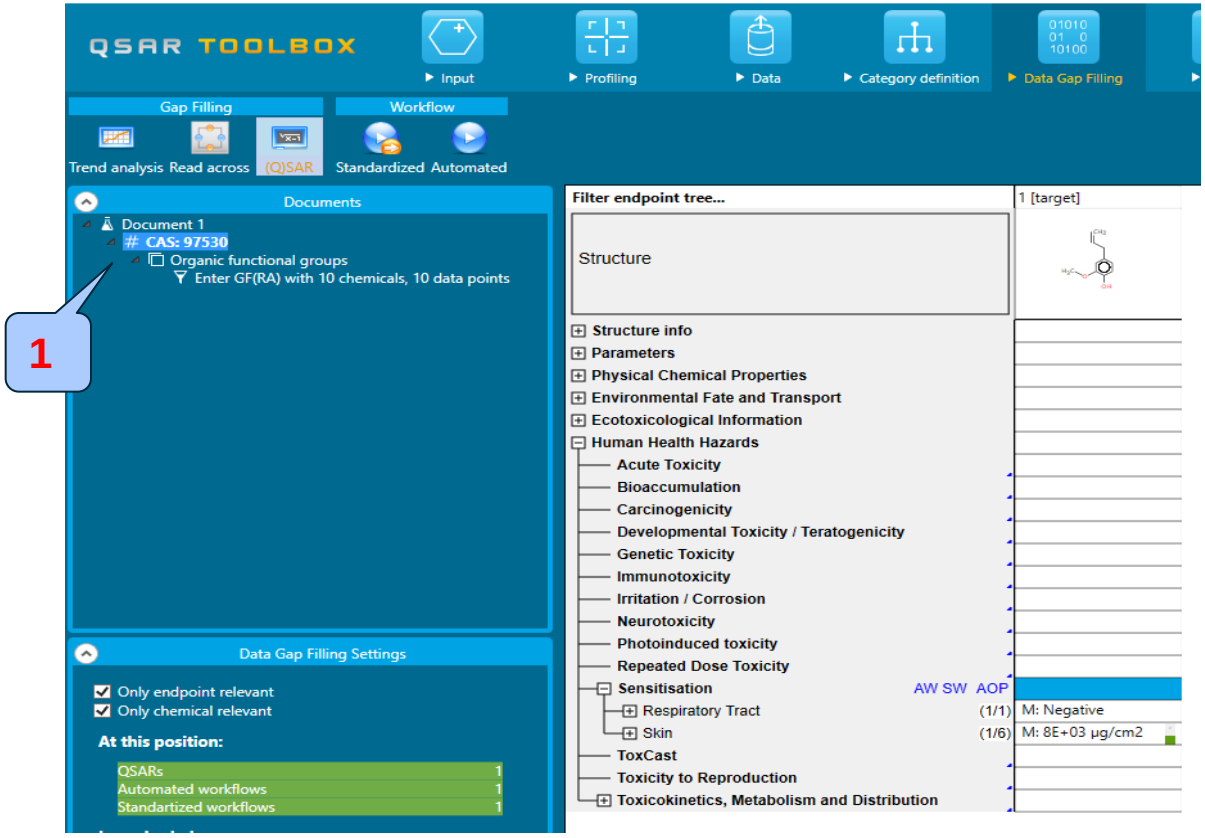
Subcategorization is applied by taking into account autoxidation simulation in combination with *Protein binding alerts for skin sensitization by OASIS:*

1. Select *Protein binding alerts for skin sensitization by OASIS;*
2. Select *Autoxidation simulator*; 3. Close the *Subcategorization* window;

Data gap filling
Interpreting Read-across

- In this example the target and all analogues have no protein binding alerts.
- All analogues along with the target possess the same distribution of positive protein binding alerts when autoxidation is taken into account.
- The latter could explain the positive experimental data of the target compound.

Data gap filling
Return to data matrix



To return to data matrix go to the document tree and click the node CAS: 97530 (1).

Data gap filling
Next actions

- The study continues with a second data gap filling where a category of analogues is defined by using a new categorization functionality allowing to define category accounting for (a)biotic activation of the target.

- Before proceeding with Data gap filling the following two procedures will be illustrated intended to explain and support the analysis. Following the steps is not necessary.

  - Multiplication of the target chemical

  - Profiling the parent and metabolites based on (a)biotic activation

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - **Multiplication of the target chemical**

Multiplication of the target chemical

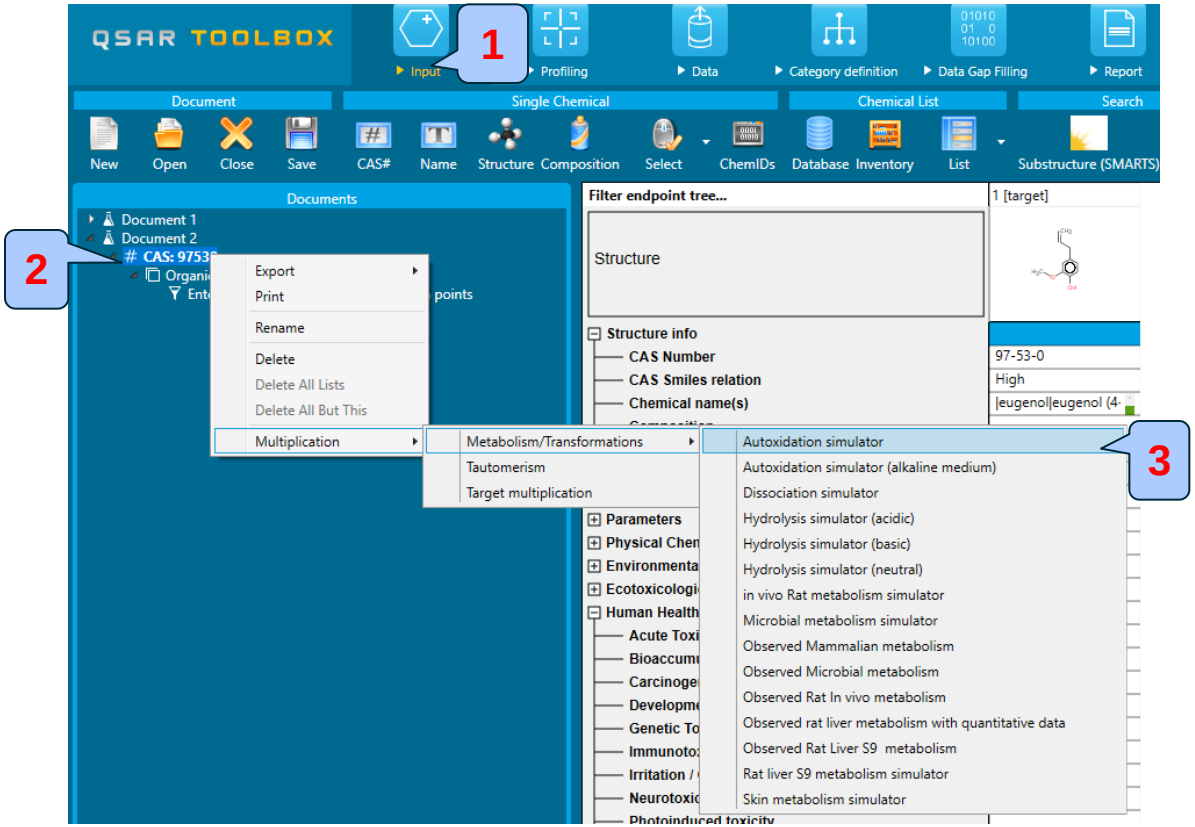
Multiplication of the target chemical could be accomplished in two ways:

- Scenario 1: In the **Input** section outside data gap filling module (slide 65)
- Scenario 2: In the **Profiling** section (slide 73)

Outlook

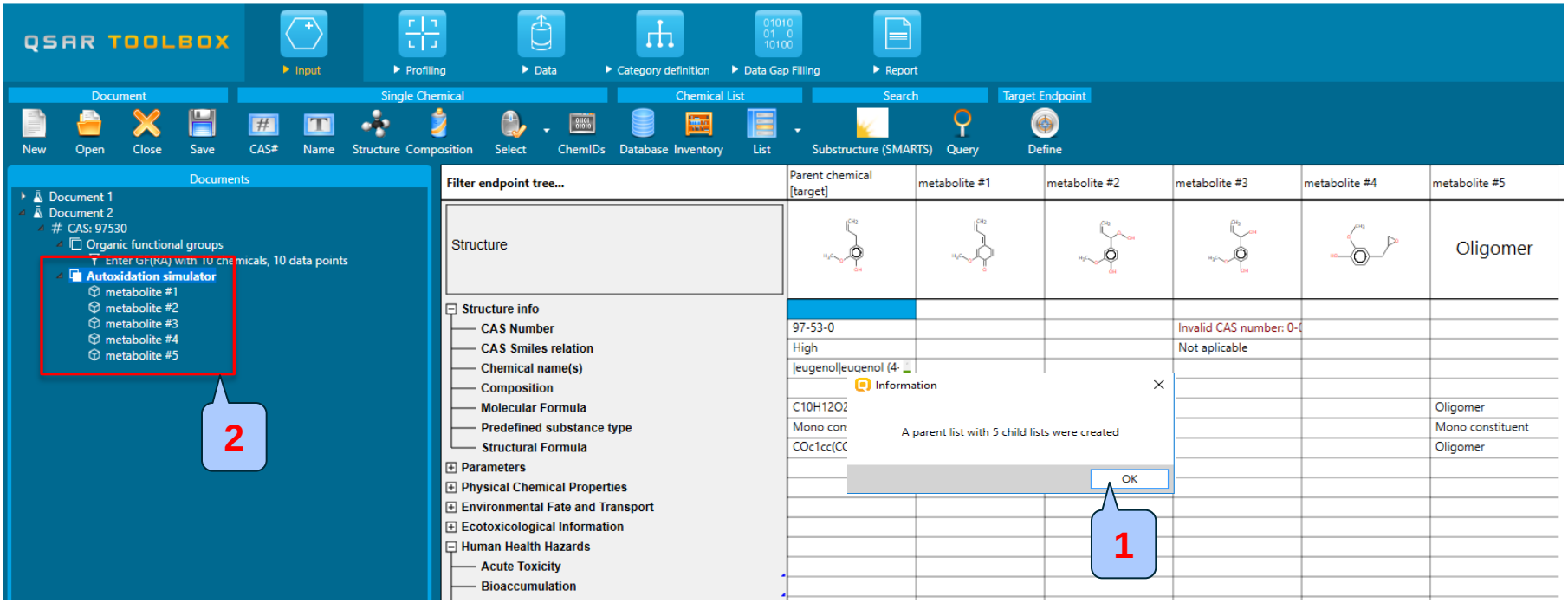
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - **Multiplication of the target chemical**
    - **In the Input section (scenario 1)**

Multiplication of target chemical in the Input section (scenario 1)



1. Go to *Input*
2. Select the CAS of the target chemical and right-click on it .
3. Select *Multiplication-Metabolism/Transformations/Autoxidation simulator*

Multiplication of target chemical in the Input section (scenario 1)

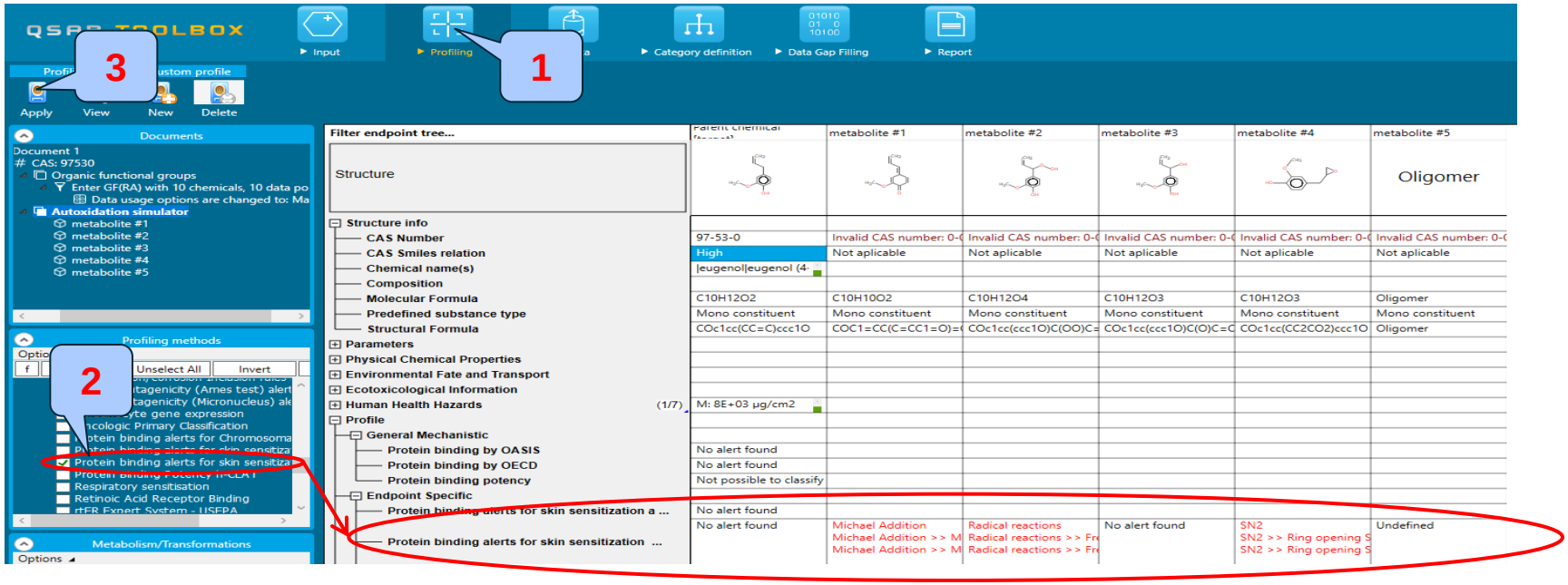


1. Select ***OK*** and then parent and product(s) are visualized on data matrix.
2. The generated metabolites are listed in the Documents tree ;

Protein binding result for parent and metabolites multiplied in the Input section

The profiling result indicates no protein binding alerts for target chemical. However, three of simulated autoxidation metabolites exhibit interaction with proteins via three different protein binding mechanisms (Michael Addition, Radical reactions, and SN2).

Autoxidation simulator

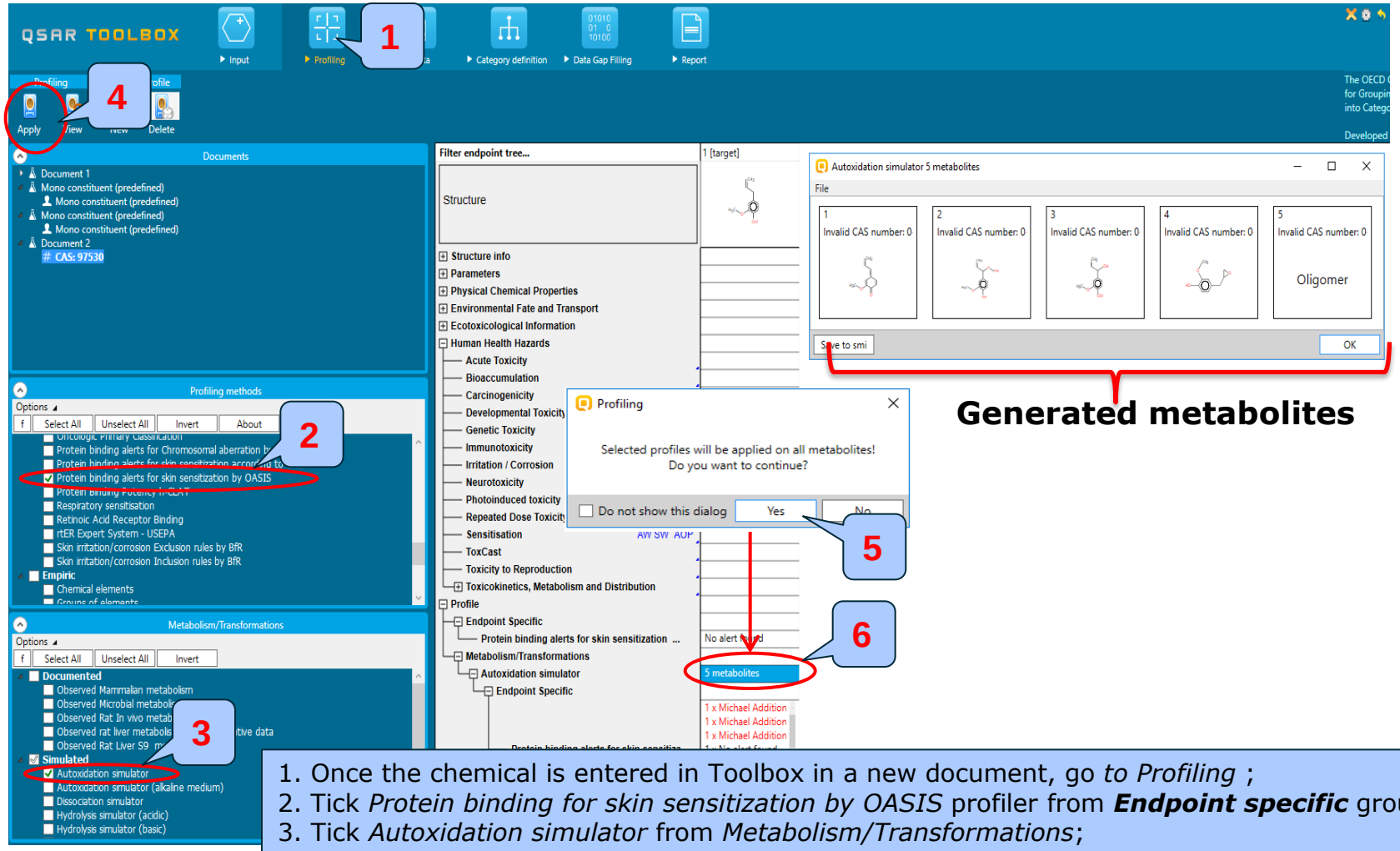


Once the chemical is multiplied in the Input section and metabolites are visualized (distributed on data matrix); 1. Go to *Profiling*; 2. Check *Protein binding alerts for skin sensitization by OASIS*; 3. Click *Apply*;

Outlook

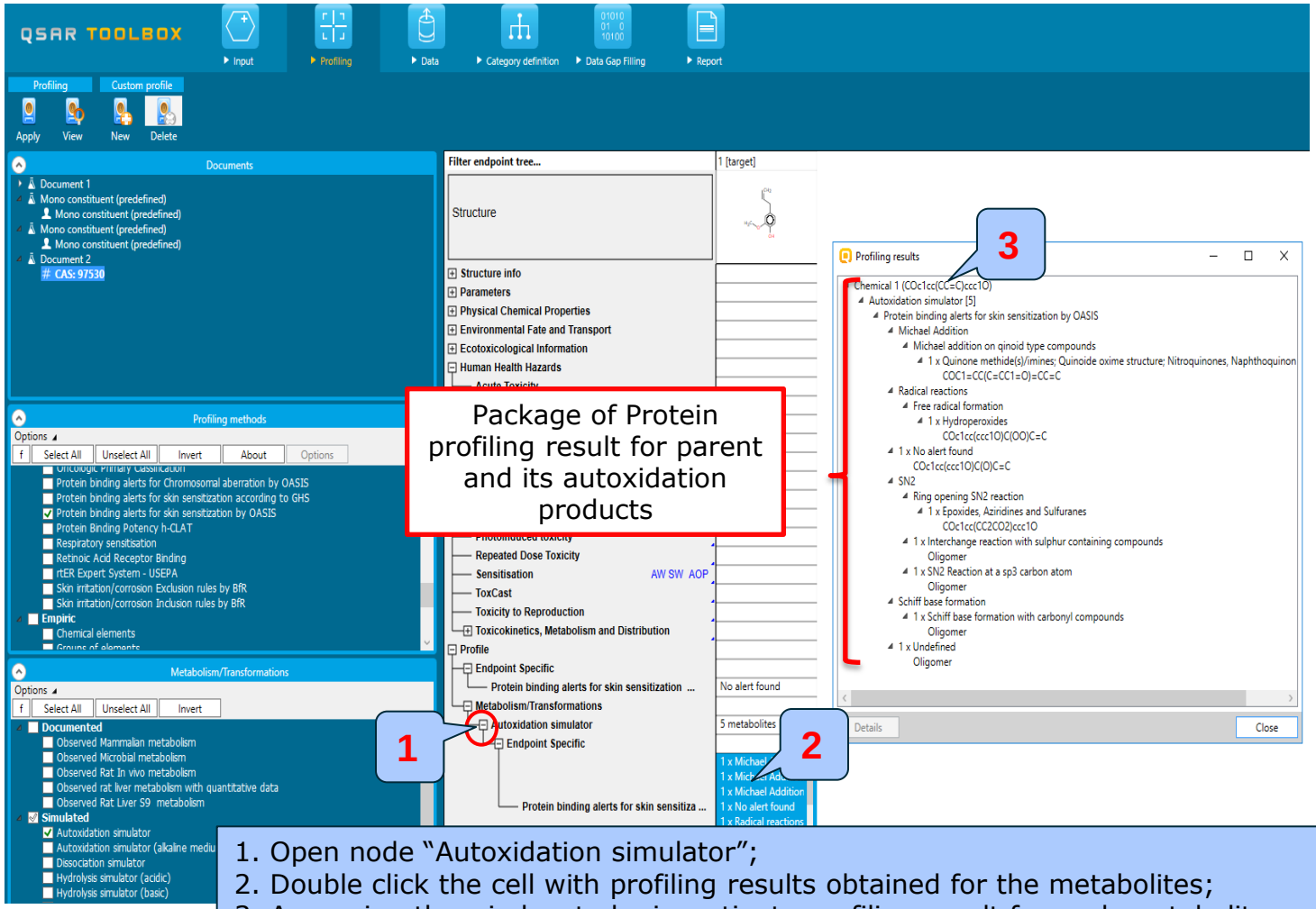
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - **Multiplication of the target chemical**
    - In the Input section (scenario 1)
    - **In the Profiling section (scenario 2)**

Multiplication of target chemical in the Profiling section (scenario 2)



1. Once the chemical is entered in Toolbox in a new document, go *to Profiling* ;
2. Tick *Protein binding for skin sensitization by OASIS* profiler from ***Endpoint specific*** group;
3. Tick *Autoxidation simulator* from *Metabolism/Transformations*;
4. Click ***Apply***; 5. Click ***Yes***;.
6. Double click the cell containing the 5 metabolites in order to visualized them.

Protein binding result for parent and metabolites multiplied in the Profiling section

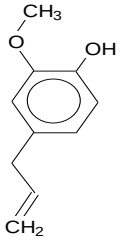


1. Open node "Autoxidation simulator";
2. Double click the cell with profiling results obtained for the metabolites;
3. Appearing the window to be investigate profiling result for each metabolite;

Recap

Parent

Eugenol
CAS# 97-53-0



Protein binding alert
(Protein binding by OASIS)

No alert found

Experimental data

LLNA
Strong sensitizer

Data gap filling

No protein binding alerts

Positive RA prediction

How to explain positive RA ?

Autoxidation Simulator/Skin metabolism

Protein binding alert found for package of metabolites

next

Categorization with metabolism

Identifying protein binding analogues of the target taking into account its metabolic activation

Outlook

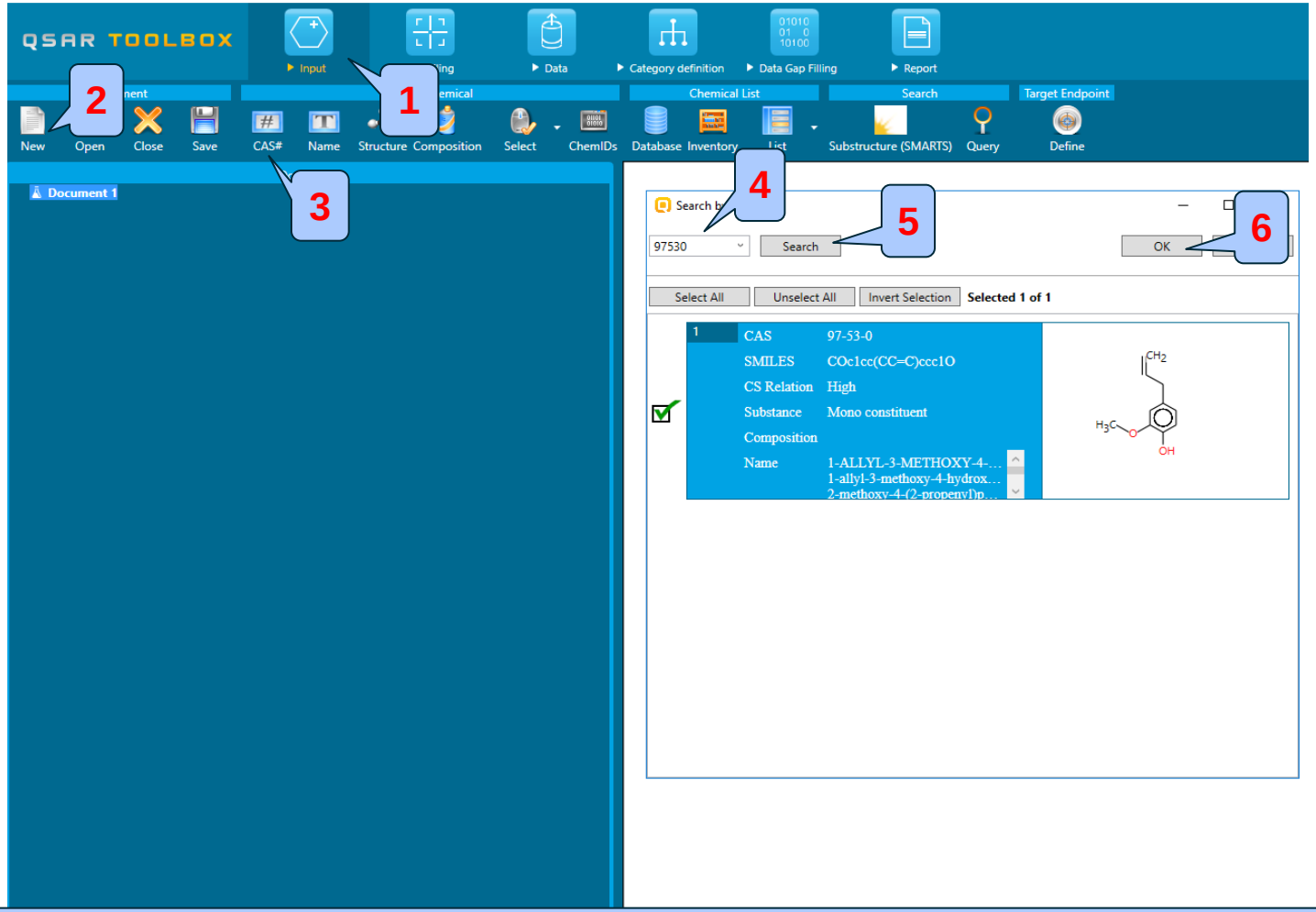
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - Multiplication of the target chemical
  - **Category definition with applying metabolism**

Category definition with applying metabolism

The advantages of the new functionality are:

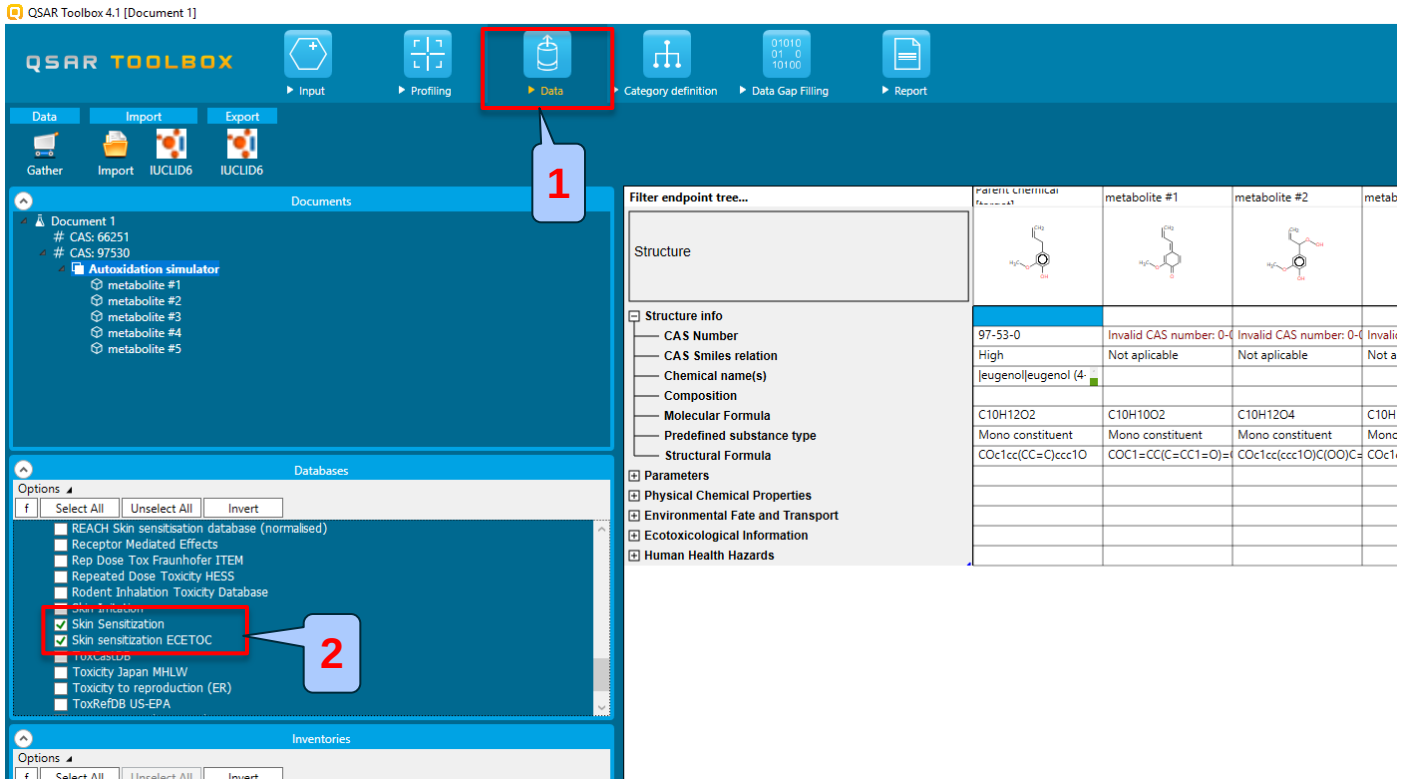
- Application of metabolism for analogues identification during process of categorization.
- A category can be defined with and without metabolism.
- This is a critical step in the workflow.
- Possibility to expand the chemical domain of the category and to identify analogues based on metabolism approach.
- Before proceeding with categorization accounting for (a)biotic activation of the target input the target in a new document (see next slide).

Category definition with applying metabolism



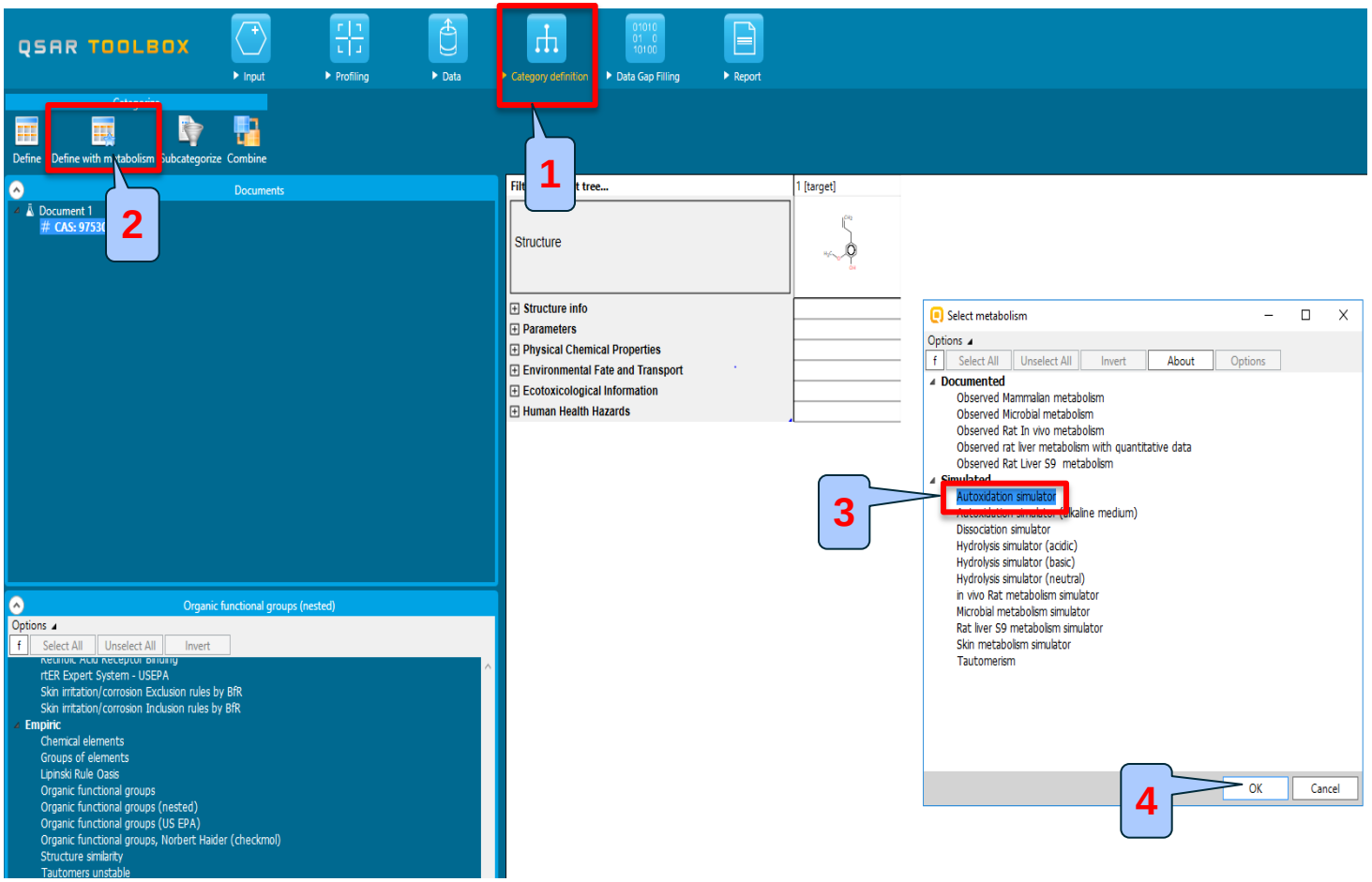
1. Go to ***Input***; 2. Click ***New***; 3. Click ***CAS#***; 4. Enter the CAS number of the target; 5. Click ***Search*** 5. Click ***OK***.

Category definition with applying metabolism



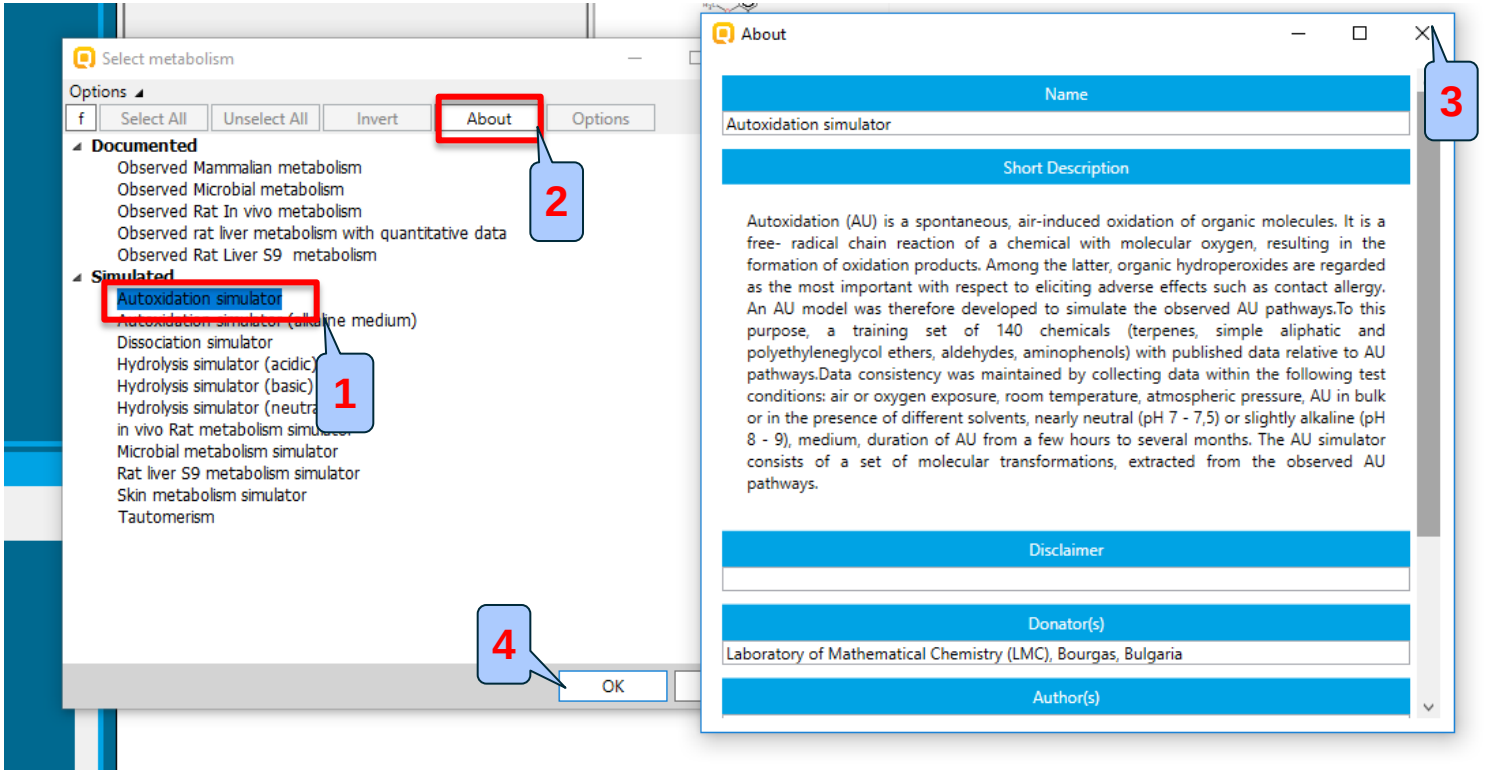
1. Go to ***Data***; 2. Tick *Skin Sensitization* and *Skin Sensitization ECETOC* databases;

Category definition with applied metabolism



1. Go to ***Category definition***; 2. Click ***Define with metabolism***; 3. Select ***Autoxidation simulator***; 4. Click ***OK.***

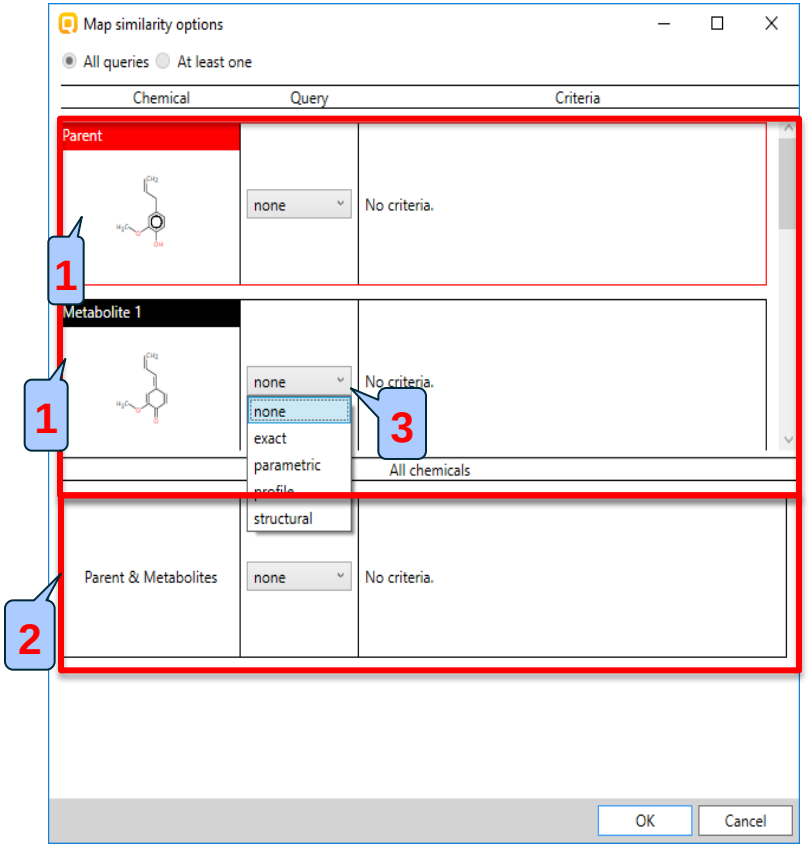
Category definition with applied metabolism



All available transformation maps – documented and simulated in Toolbox can be used in the primary grouping.

1. Click *Autoxidation simulator*; 2. Click *About* to read the short description of the simulator; 3. Close the About window; 4. Click *OK*.

Category definition with applied metabolism



Map similarity options dialogue appears. It shows all the generated metabolites of the target chemical (use the scroll bar to see them). It has two subsections:

- (1) shows the parent and each of the generated metabolites. This allows defining different criteria for each structure when looking for analogues.
- (2) treats the parent and its metabolites as a whole. i.e. the criteria is provided for the whole package but not for separate metabolites.

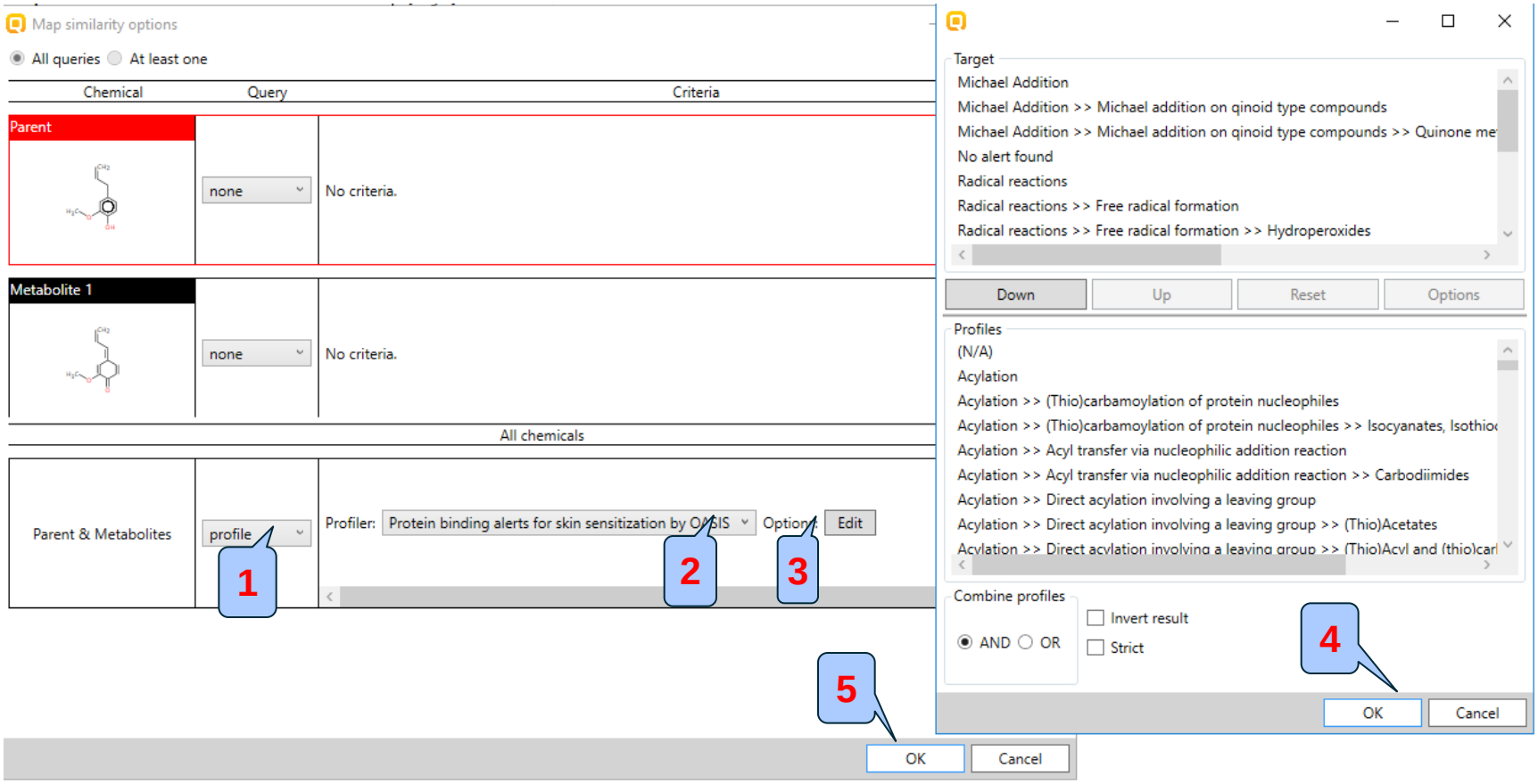
A drop down menu (3) is available for each of the structures (in the column "Query") which allow setting the type of criteria for looking for analogues.

Category definition with applying metabolism

Explanation of different options from the drop down menu:

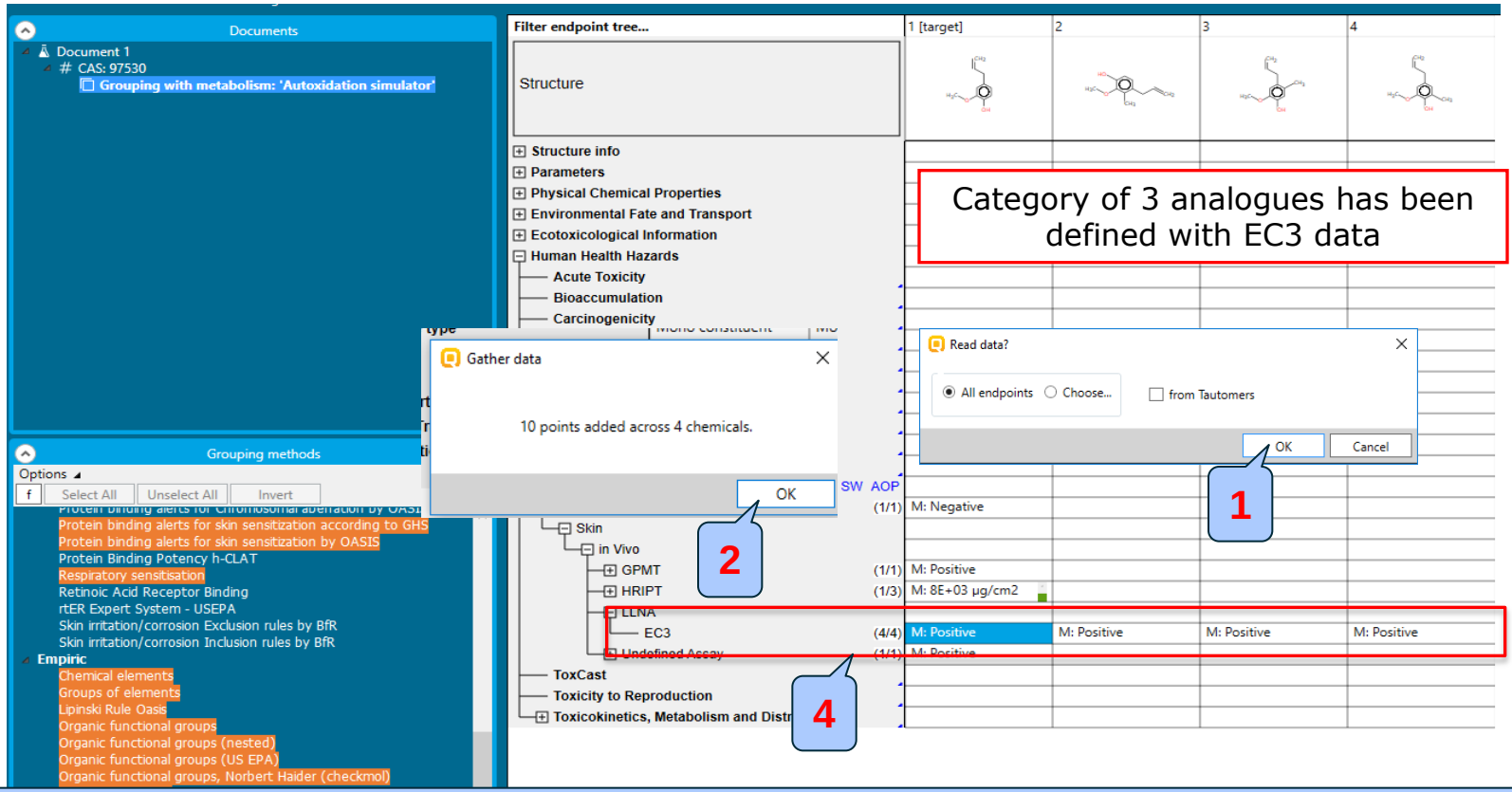
- **None** – default options; no criteria is set;
- **Exact** – search for analogues which metabolites have the exact structure of the target metabolite; only available for the metabolites and the package "parent + metabolites" but not for the parent chemical;
- **Parametric** – to have specific value or range of variation of defined parameter (a list with all parameters currently available in the Toolbox is provided);
- **Profile** – to have specific category by selected profiler (a list with all profilers is provided);
- **Structural** –allows structural similarity assessment based on the atom centered fragments.
- The user can select a profiling, parametric or structural query for both target and its metabolites.

Category definition with applied metabolism



1. Select a profile option for the package "parent & metabolites";
2. Select "Protein binding alerts for SS by OASIS" profile;
3. Click ***Edit*** . The profiling results of the parent structure and its metabolites are shown.
4. Click **OK** to confirm the defined search criteria.
5. Click ***OK*** in Map similarity options window to execute the search.

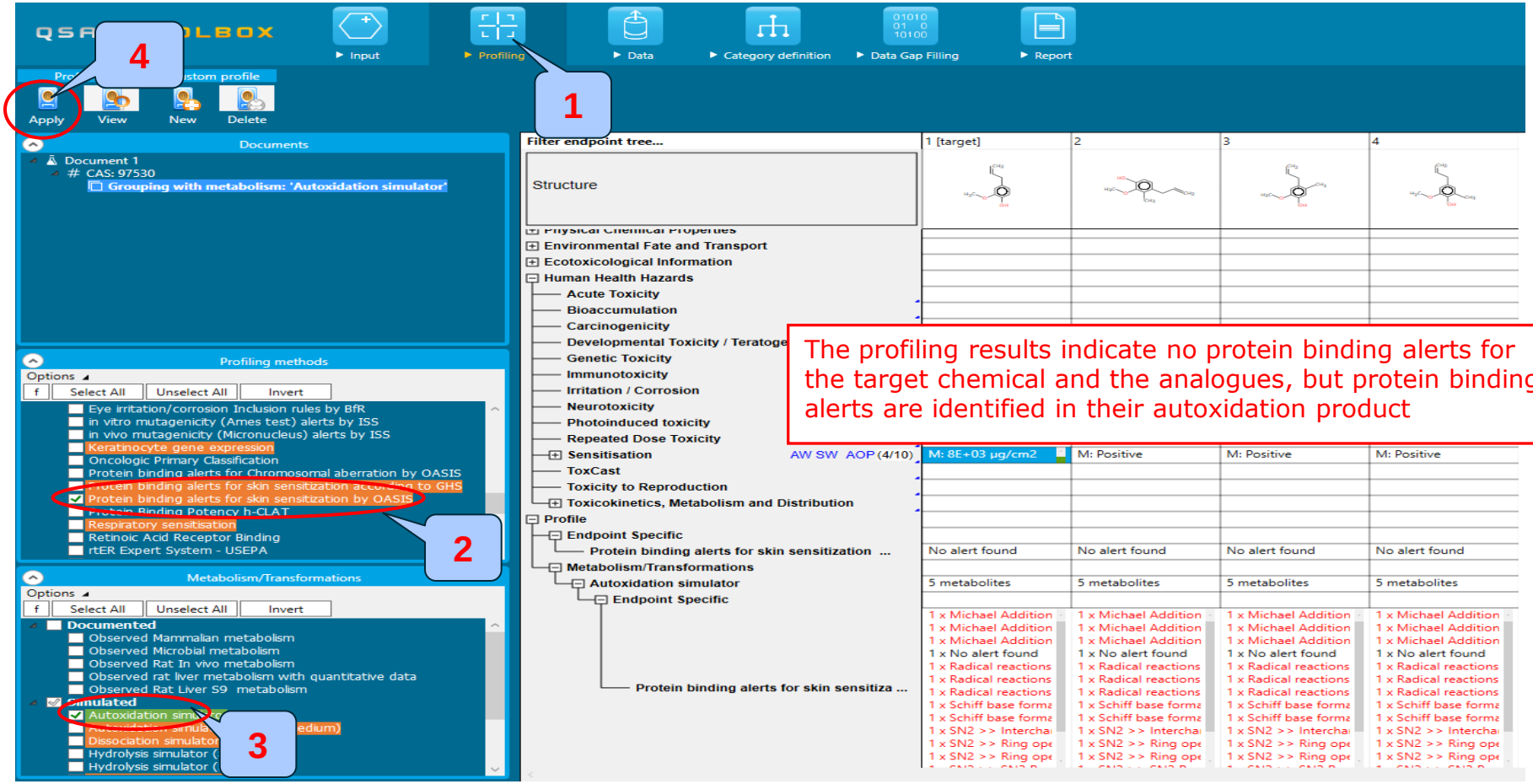
Category definition with applied metabolism



1. Click ***OK*** to read all data; 2. An information window appears informing about the number of experimental data collected and the number of chemicals in the category (1 target and 3 analogues), click ***OK***. 3. The analogues and their experimental data displayed on the matrix. Target and analogues have one experimental EC3 data each.
The forthcoming two slides illustrates how consistent is the identified category with respect to protein binding alerts when metabolism is taken into account

Categorization with applied metabolism

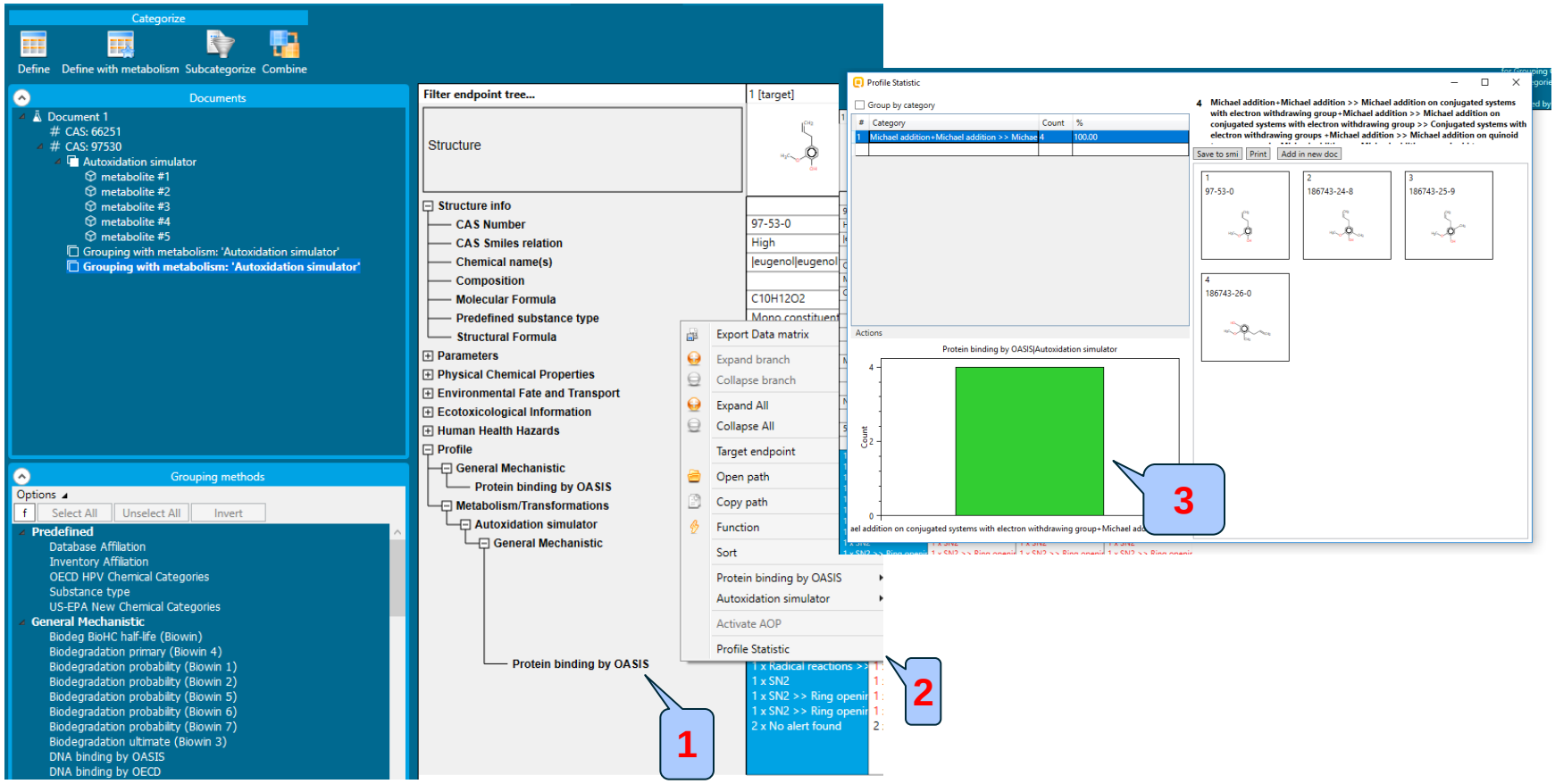
Profiling results for parent and metabolites



1. Go to ***Profiling***; 2. Check ***Protein binding alerts for skin sensitization by OASIS***
3. Check ***Autoxidation simulator***; 4.Click ***Apply***;

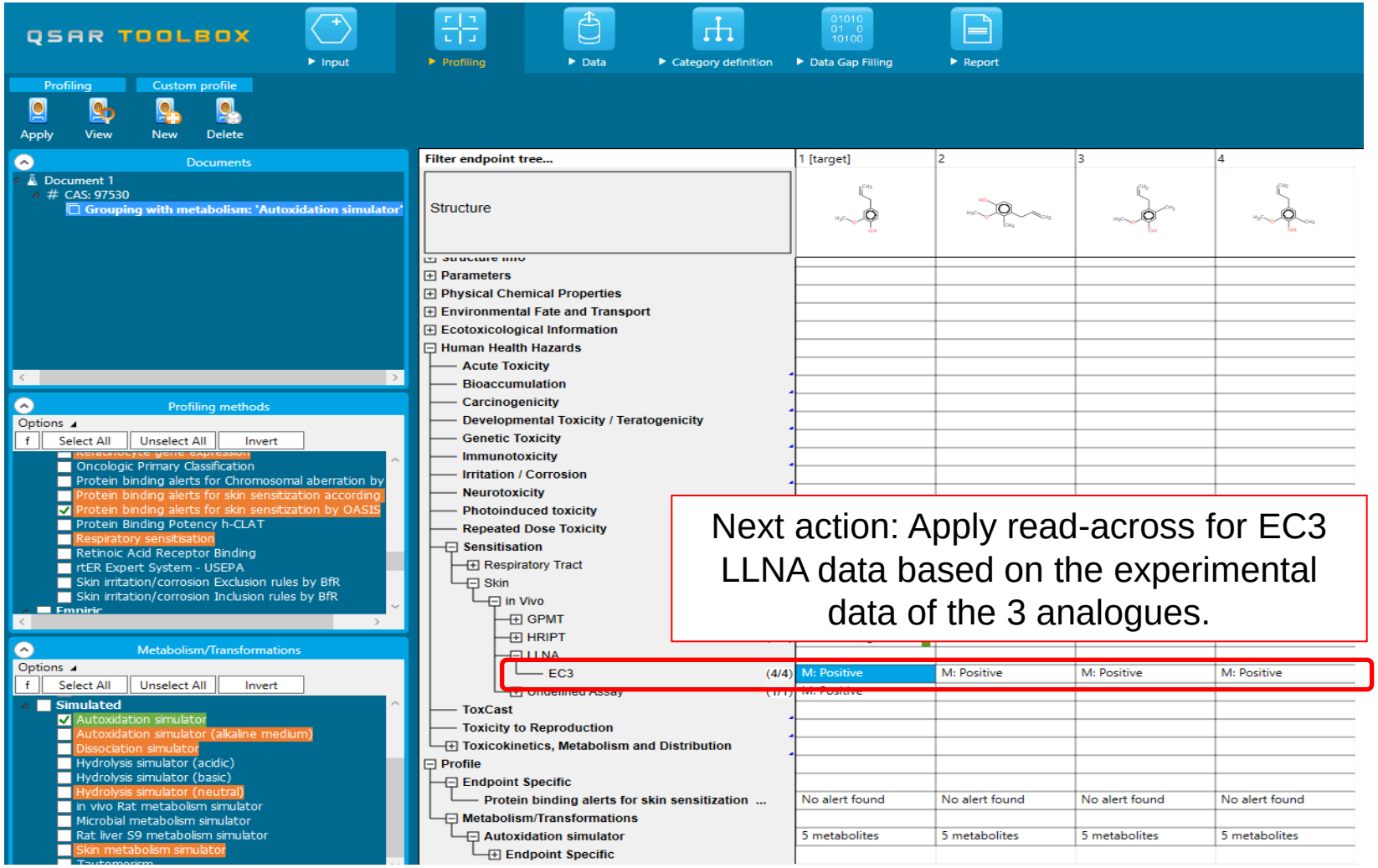
Categorization with applied metabolism

Profile statistic



1. Right-click next to Metabolism/Transformations/Autoxidation simulator/General mechanistic/Protein binding by OASIS Protein binding by OASIS in the grey field; 2. Click ***Profile statistics***. 3. All autoxidation products have the same distribution of the protein binding alerts.

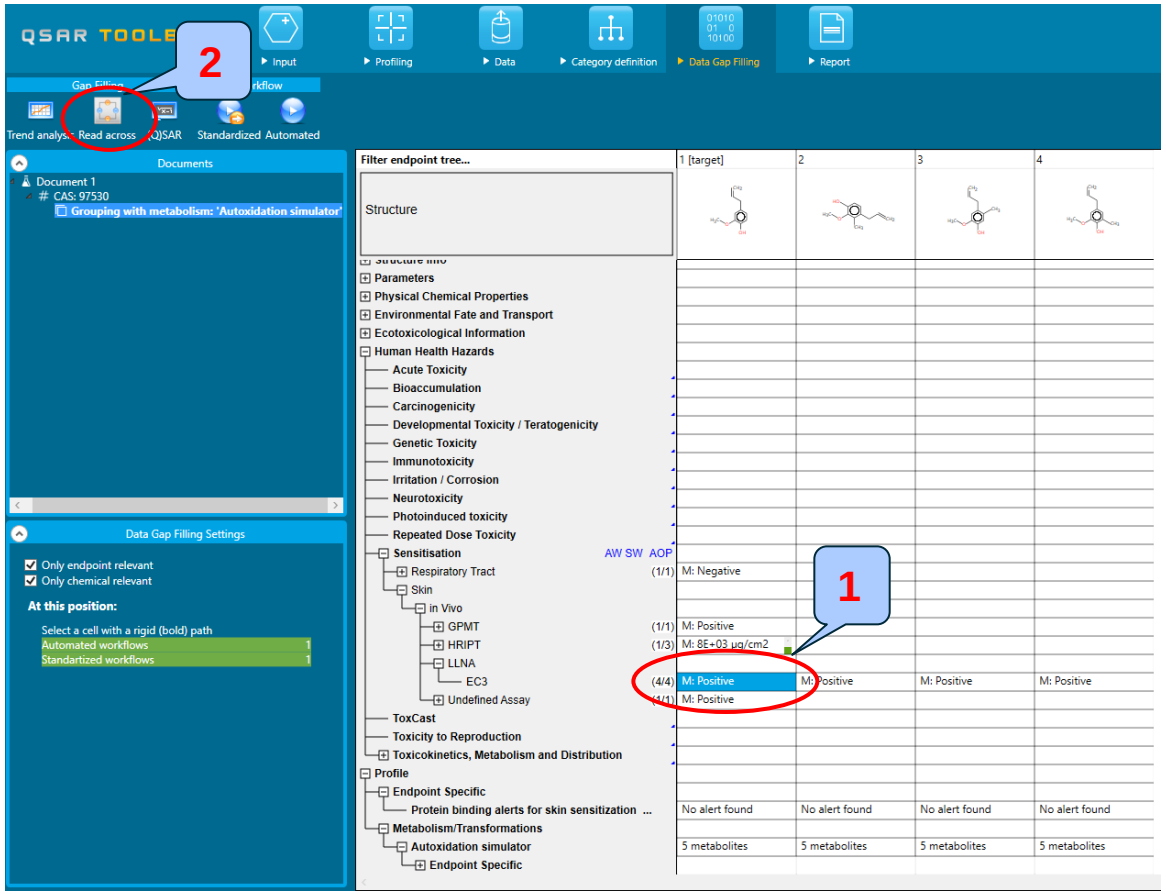
Categorization with applied metabolism



Outlook

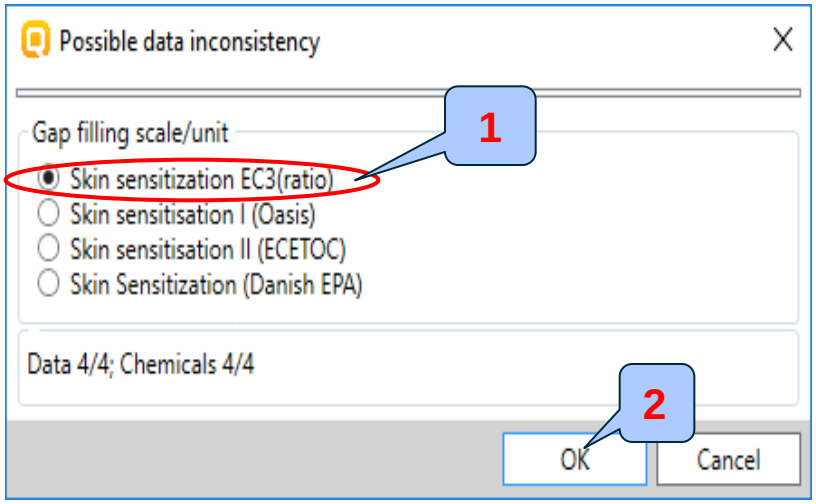
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - Categorization applying metabolism
  - **Data gap filling by taking into account metabolism**

Data gap filling
Apply Read across



1. Click the cell corresponding to *Sensitization>>Skin>>In Vivo>>LLNA>>EC3* for the target chemical; 2. Click *Read-across* .

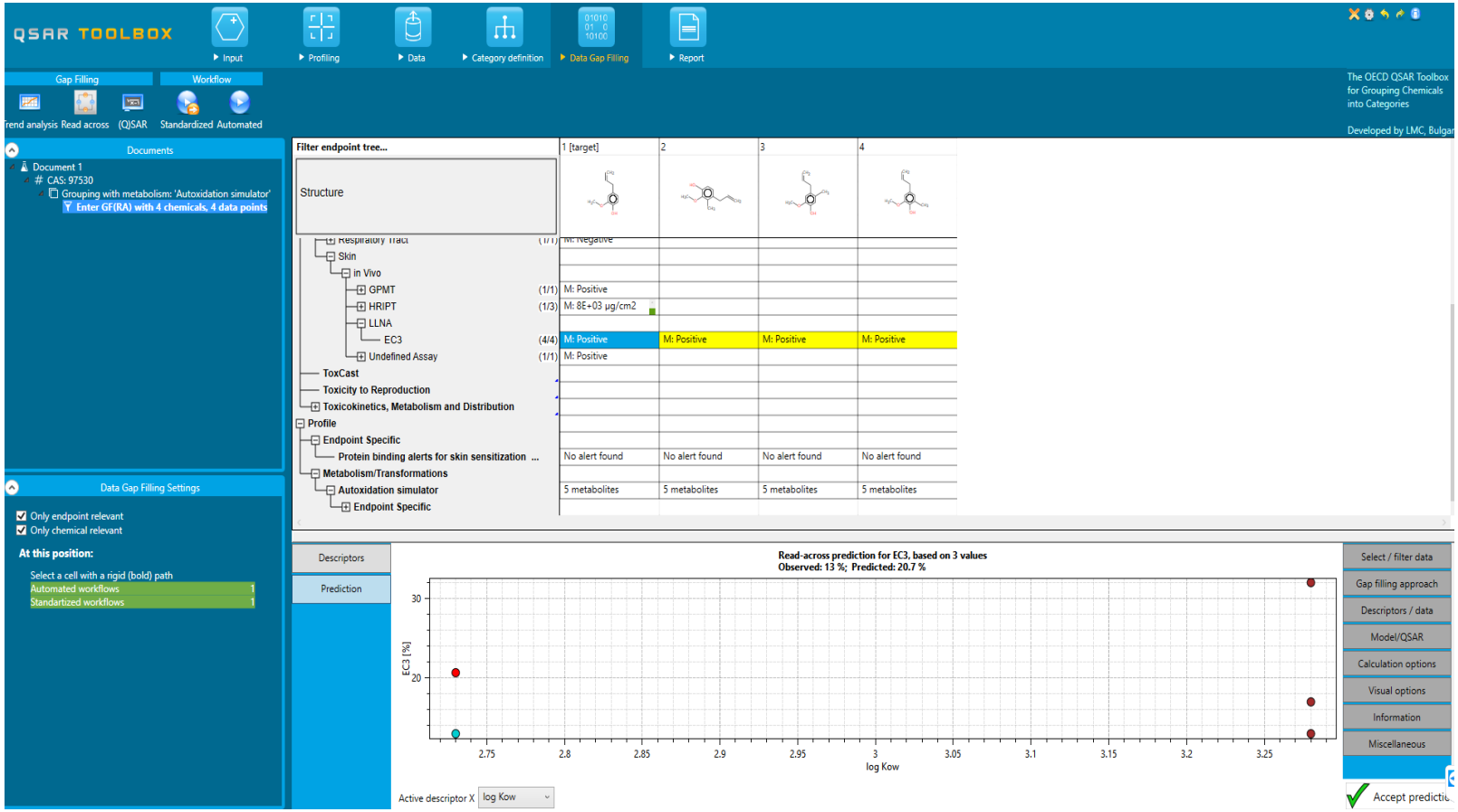
Data gap filling
Scale definition



1. Select scale *Skin sensitisation EC3 (ratio)*;
2. Click ***OK***.

Data gap filling

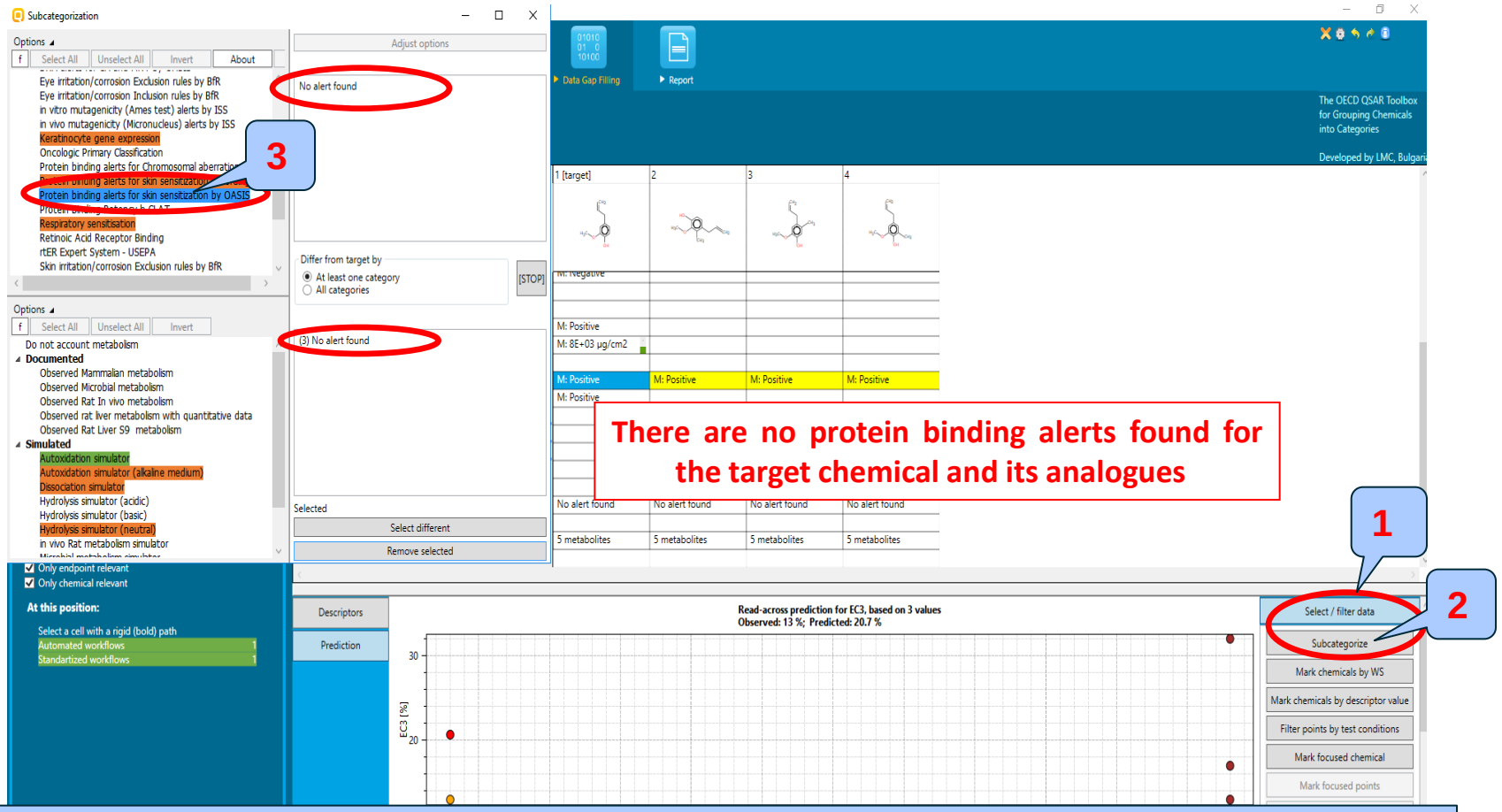
Read-across



Initial graph without any subcategorizations.

Data gap filling

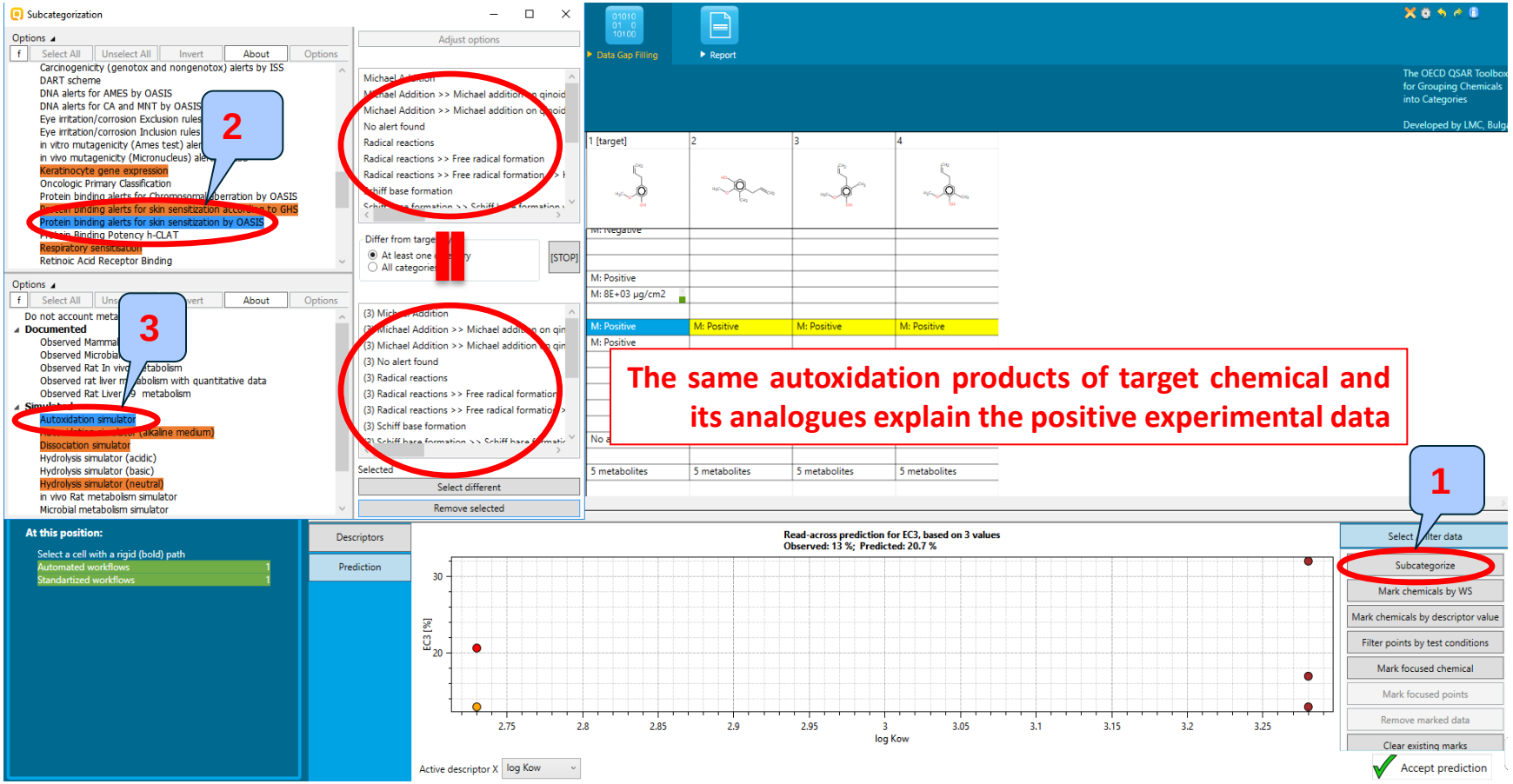
Subcategorization 1: Protein binding alerts for skin sensitization by OASIS



1.Open *Select filter data/subcategorize*; 2. Click ***Subcategorize***; 3.Select Protein binding alerts for skin sensitization by OASIS

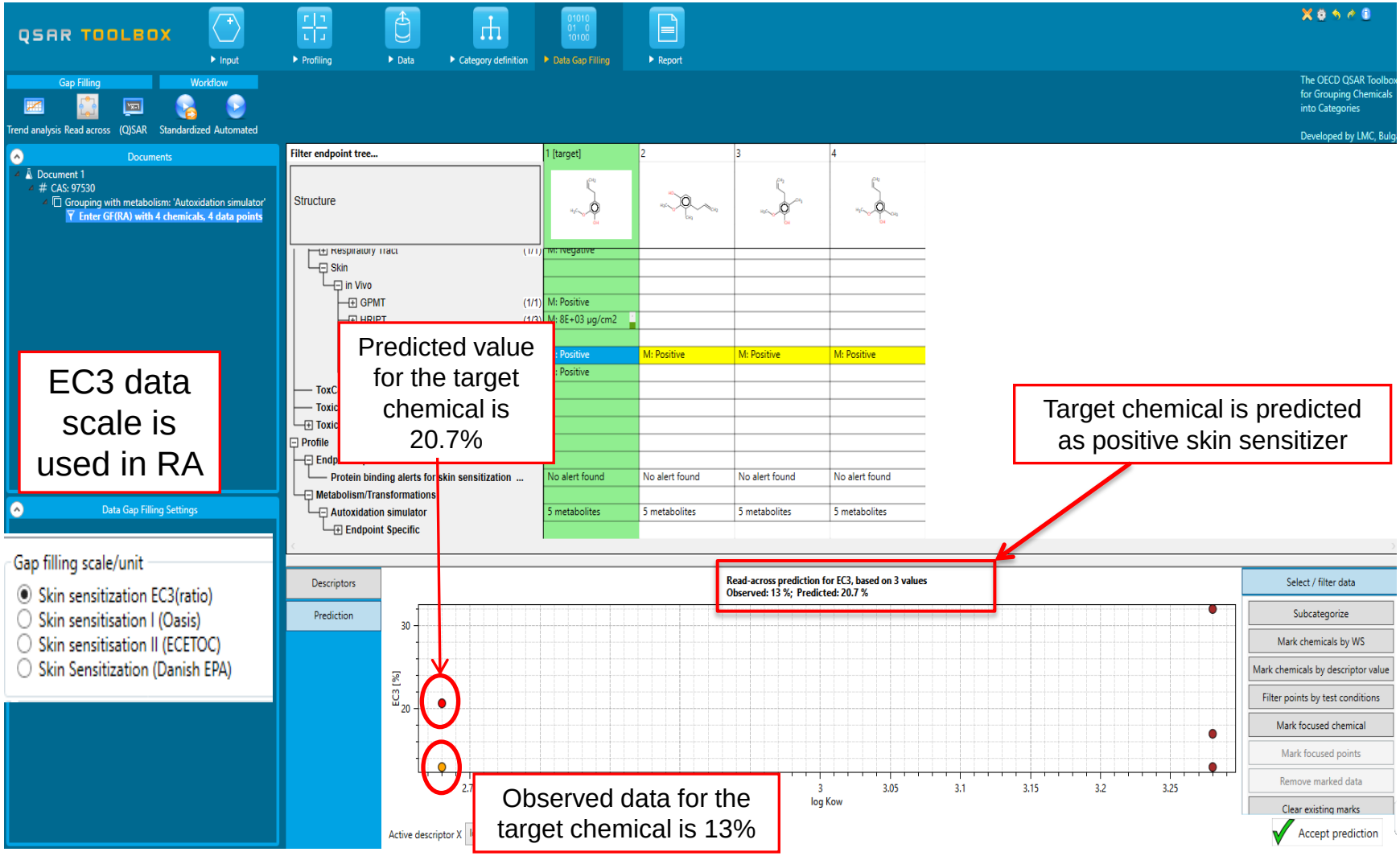
Data gap filling

Subcategorization 2: Protein binding alerts for SS when AO is taken into account

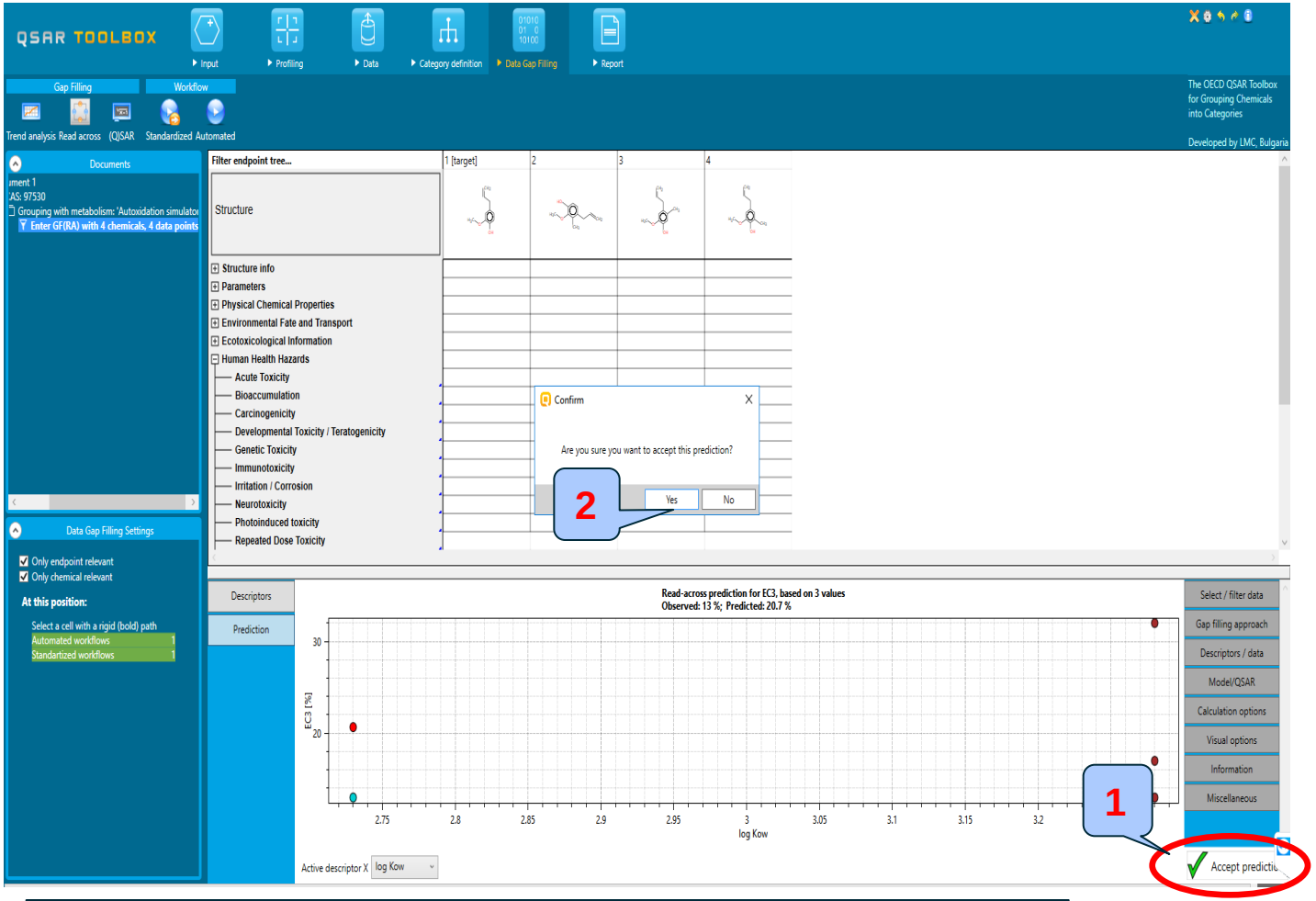


1. Click ***Select/filter data***; *2.* Click ***Subcategorize***; 2. Select *Protein binding alerts for skin sensitization by OASIS*; 3. Select *Autoxidation simulator.*

Data gap filling prediction



Data gap filling prediction

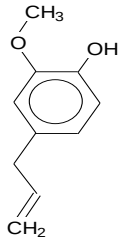


1.Click *Accept prediction*; 2. Click ***Yes*** to confirm the prediction.

Recap

Parent

Eugenol
CAS# 97-53-0



Protein binding alert
(Protein binding by OASIS)

No alert found

Experimental data

LLNA Strong sensitizer

Data gap filling

No protein binding alerts

Positive RA prediction

How to explain positive RA ?

Autoxidation Simulator/Skin metabolism

Protein binding alert found for package of metabolites

Categorization with metabolism

Identifying protein binding analogues of the target taking into account its metabolic activation

Positive Protein binding alerts

Positive RA prediction

Outlook

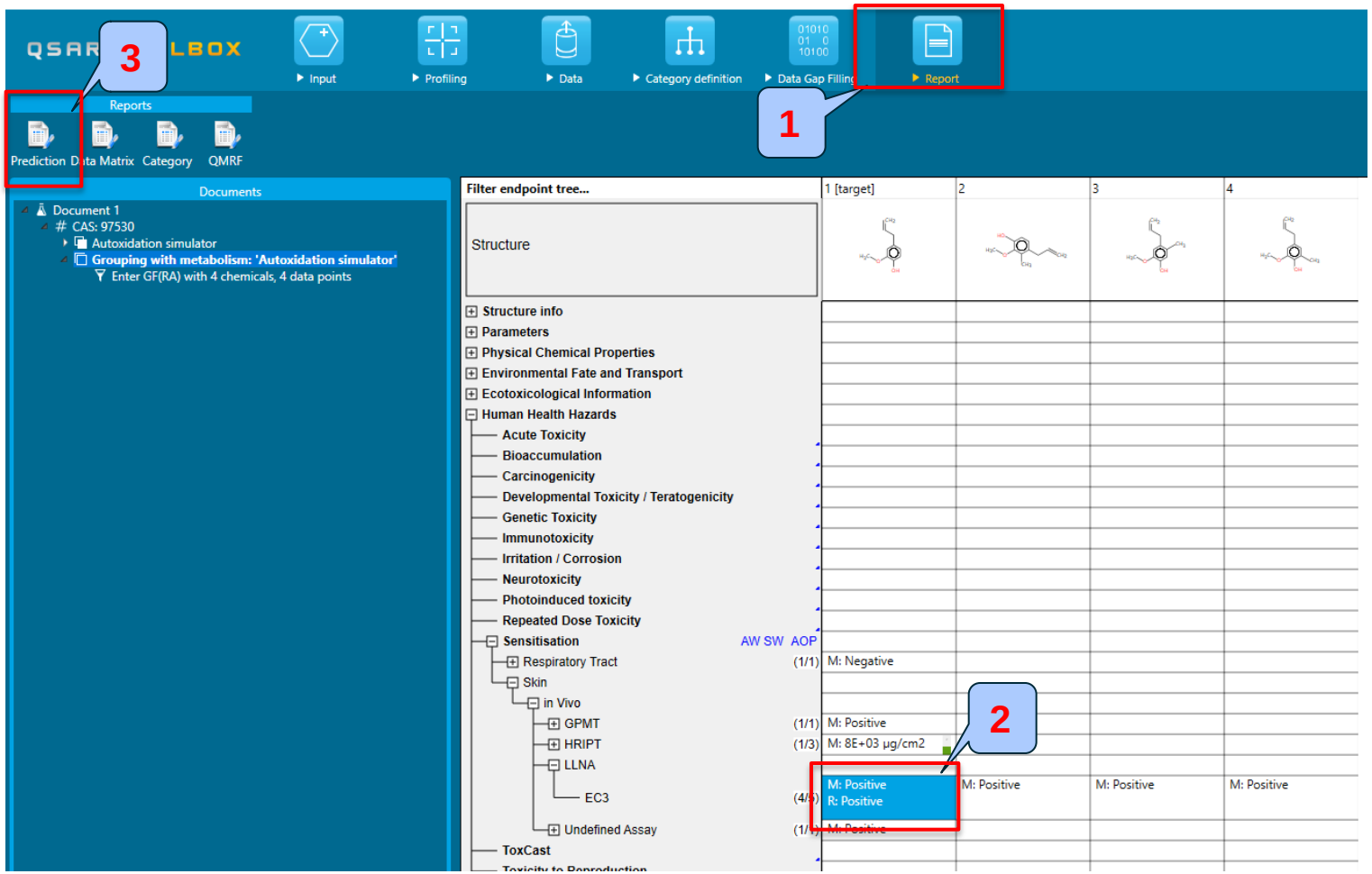
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - Categorization applying metabolism
  - Data gap filling by taking into account metabolism
  - **Report**

Report

- The report module allows you to generate a report on the predictions obtained with the Toolbox.
- This module contains predefined report templates .
- The report can then be printed or saved in different formats.

The Generating of a report is shown on next slides.

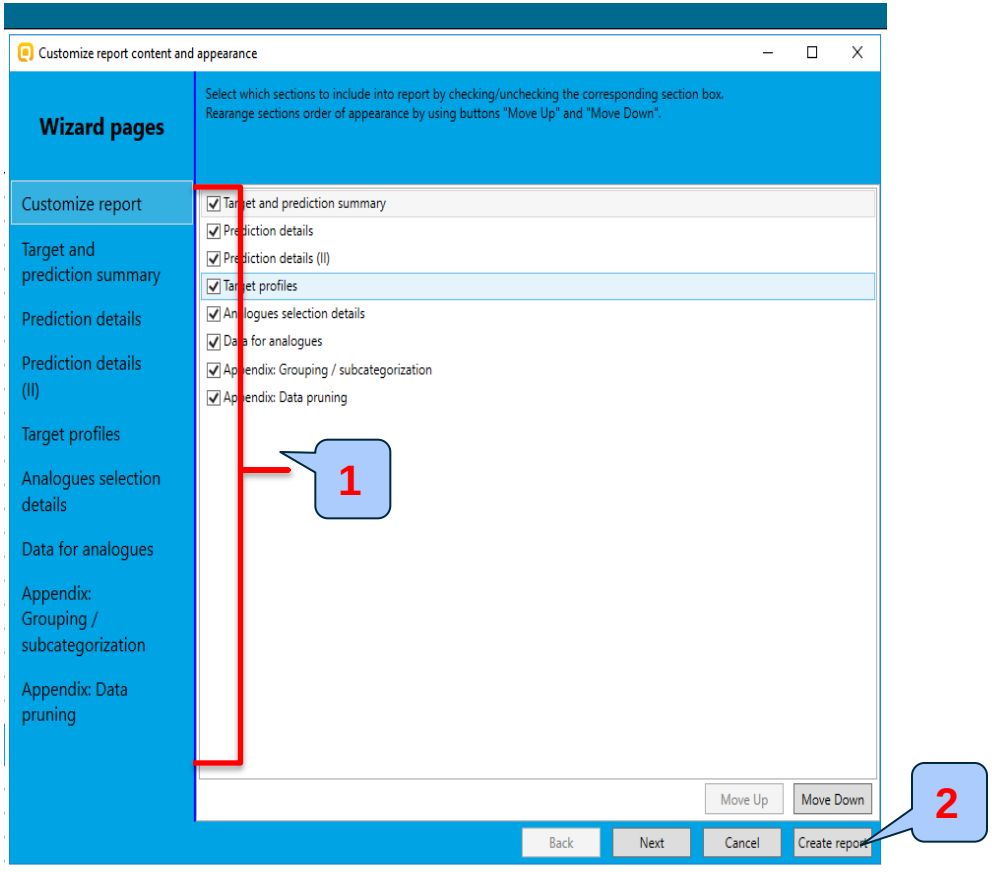
Report



1. Go to ***Report*** section; 2. Highlight the prediction result – "Positive"; 3.Click ***Prediction***;

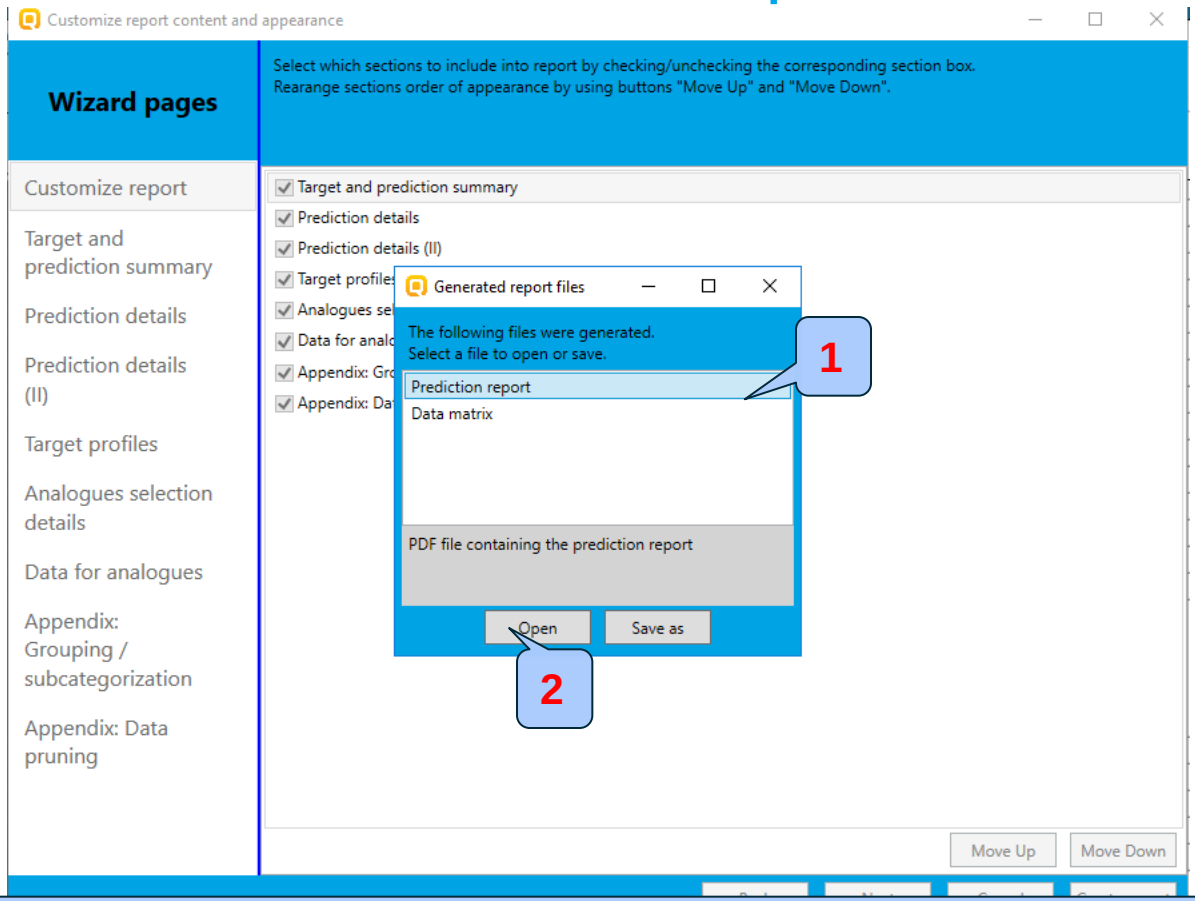
Report

Inserting additional information



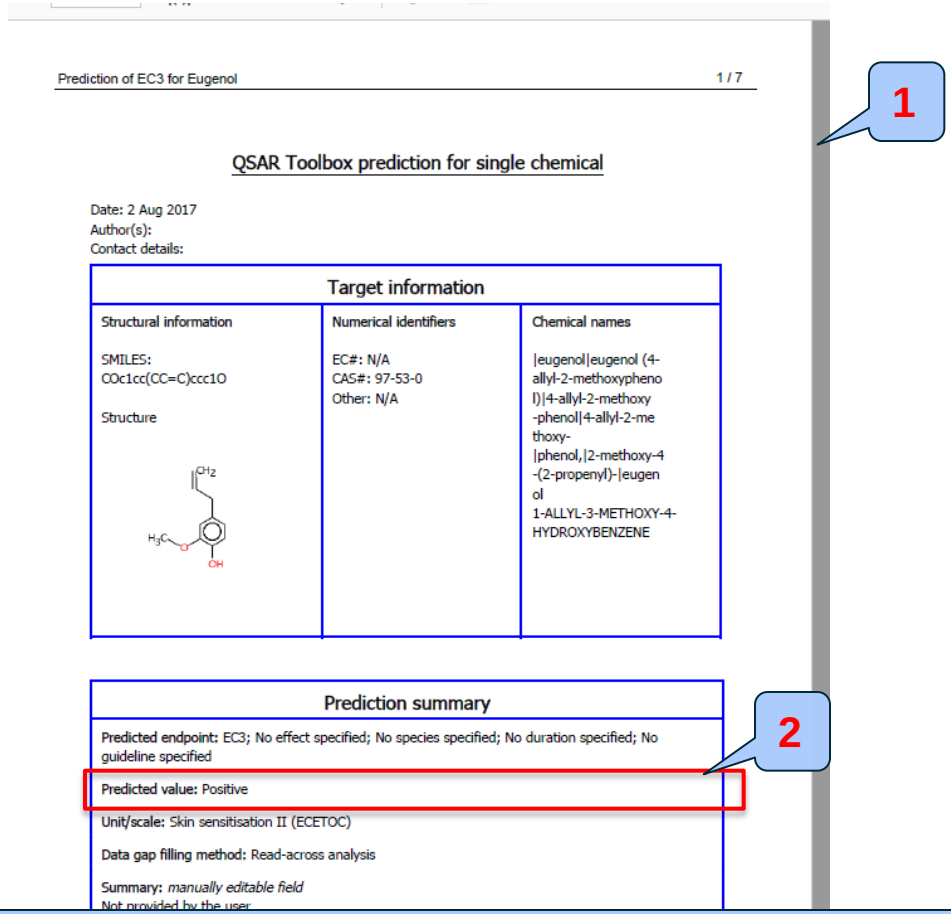
1. The user could fill in additional information in some some of the fields;
2. Select ***Create report***;

Report
Generation report



1. Two files are generated, which can be selected from the Generated report files window;
2. Select the file and then click ***Open***.

Report
Overview



Prediction of EC3 for Eugenol 1 / 7

QSAR Toolbox prediction for single chemical

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

Target information

Structural information	Numerical identifiers	Chemical names
SMILES: COc1cc(CC=C)ccc1O Structure	EC#: N/A CAS#: 97-53-0 Other: N/A	|eugenol|eugenol (4-allyl-2-methoxyphenol)|4-allyl-2-methoxy-phenol|4-allyl-2-methoxy-|phenol,|2-methoxy-4-(2-propenyl)-|eugenol 1-ALLYL-3-METHOXY-4-HYDROXYBENZENE

Prediction summary

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

Predicted value: Positive

Unit/scale: Skin sensitisation II (ECETOC)

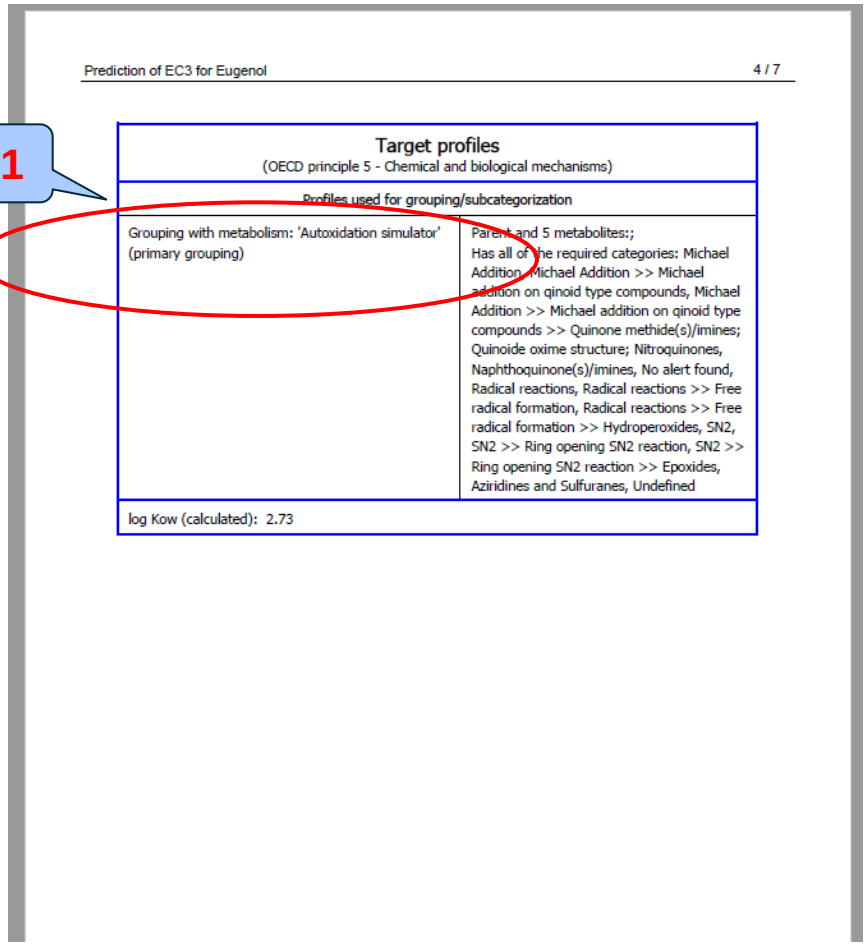
Data gap filling method: Read-across analysis

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

1
2

1. The prediction report is a file in PDF format
2. Predicted value is included in the Prediction summary

Report



Prediction of EC3 for Eugenol 4 / 7

1

Target profiles
(OECD principle 5 - Chemical and biological mechanisms)

Profiles used for grouping/subcategorization

Grouping with metabolism: 'Autoxidation simulator' (primary grouping)	Parent and 5 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, No alert found, Radical reactions, Radical reactions >> Free radical formation, Radical reactions >> Free radical formation >> Hydroperoxides, SN2, SN2 >> Ring opening SN2 reaction, SN2 >> Ring opening SN2 reaction >> Epoxides, Aziridines and Sulfuranes, Undefined

log Kow (calculated): 2.73

Information that primary grouping with metabolism (Autoxidation simulator) was taken into account when predicting skin sensitization is included (1).

Report

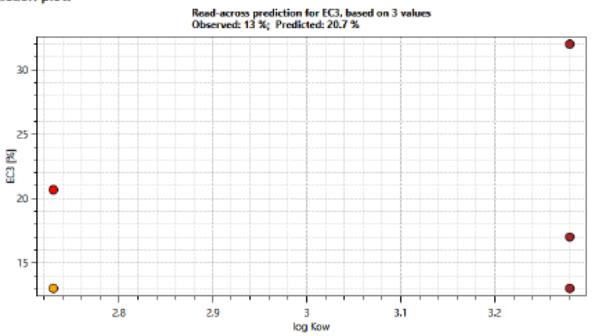
1

Prediction details (I)

Predicted value: 20.7 %, conf.range: (-29.1 ; 70.4) at 95.0%

Predicted endpoint (OECD Principle 1 - Defined endpoint): Human Health Hazards -> Sensitisation -> EC3 -> LLNA -> in Vivo -> Skin

Prediction plot:



Calculation approach (OECD principle 2 - Unambiguous algorithm): takes the arithmetic mean (average) value from the nearest 3 neighbours

Active descriptor: log Kow (calculated)

Data usage: Arithmetic mean (average) value*

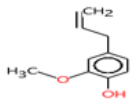
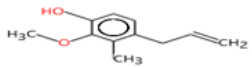
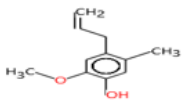
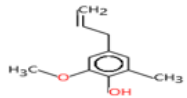
*When multiple values are available for the same chemical, their arithmetic mean (average) value is taken in prediction calculations

1. Predicted value

Data matrix

Overview

1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1				Target chemical			Neighbour #1			Neighbour #2			Neighbour #3	
2	**Substance identity**													
3	Structure													
4	CAS number			97-53-0			186743-26-0			186743-25-9			186743-24-8	
5	Chemical name			Eugenol			3-METHYL_EUGENOL			5-METHYL_EUGENOL			6-METHYL_EUGENOL	
6	Other identifier													
7	SMILES			COc1cc(CC=C)ccc1O			COc1c(O)ccc(CC=C)c1C			COc1cc(CC=C)c(C)cc1O			COc1cc(CC=C)cc(C)c1O	
15	**Measured and predicted data**													
16	***Data used for prediction***													
17	*environment*	*endpoint*	*value*	*unit*	*species, duration, test type, type of method, assay, strain, test guideline, year, referen*	*value*	*unit*	*species, duration, test type, type of method, assay, strain, test guideline, year, referen*	*value*	*unit*	*species, duration, test type, type of method, assay, strain, test guideline, year, referen*	*value*	*unit*	*species, duration, test type, type of method, assay, strain, test guideline, year, referen*
18	Sensitisation	EC3				32	%	mouse in Vivo LLNA 2005 Dermatitis, 16 (4): 1-46 Gerberick, G.F.,Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. Compilation of historical Local Lymph	13	%	mouse in Vivo LLNA 2005 Dermatitis, 16 (4): 1-46 Gerberick, G.F.,Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. Compilation of historical Local Lymph	17	%	mouse in Vivo LLNA 2005 Dermatitis, 16 (4): 1-46 Gerberick, G.F.,Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. Compilation of historical Local Lymph

1. The data matrix is an *Excel* file, which contains information about the analogues

Outlook

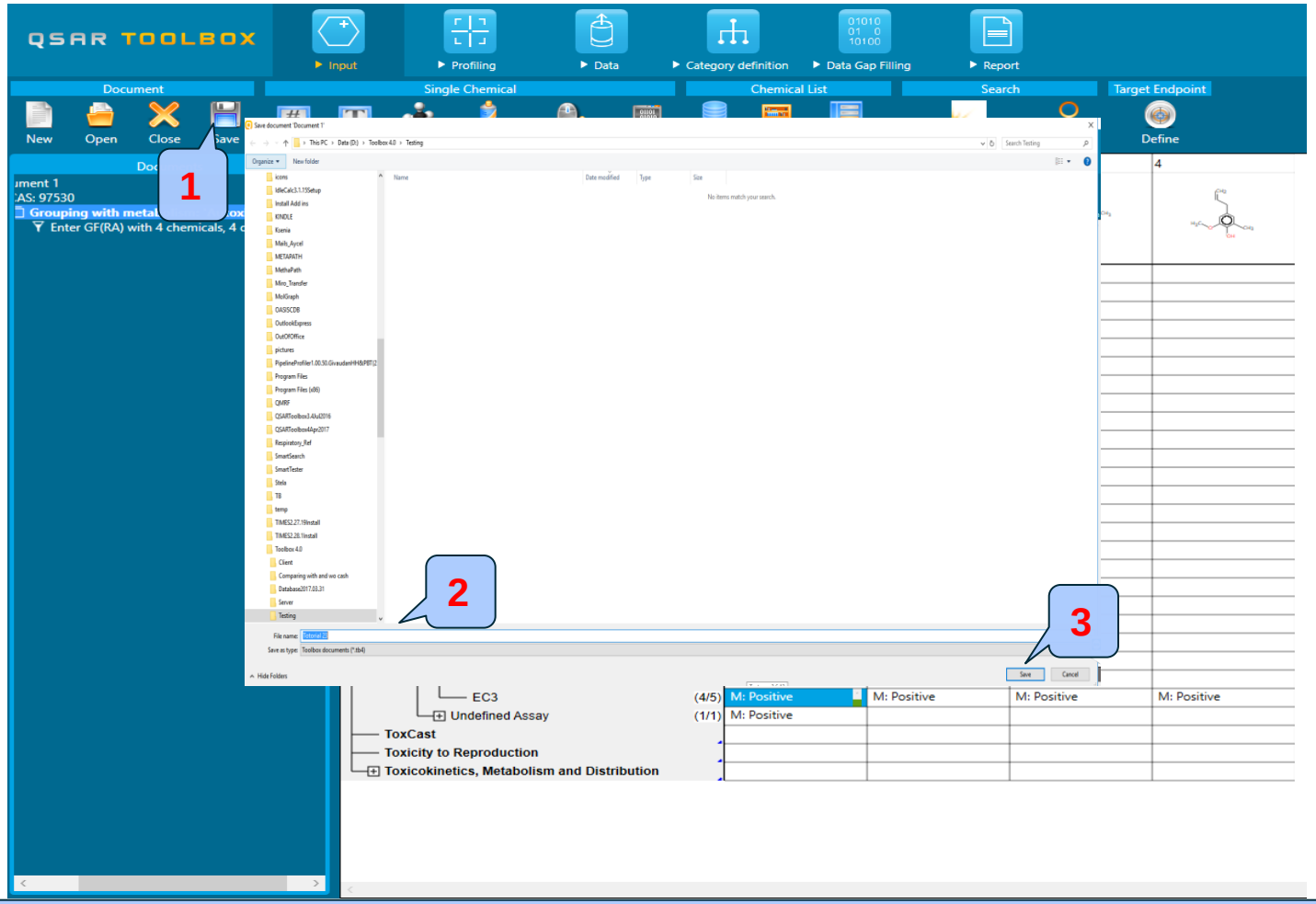
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - Categorization applying metabolism
  - Data gap filling by taking into account metabolism
  - Report
  - **Save the prediction**

Saving the prediction result

- This functionality allows storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc, on the same computer. The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document and later executing these actions in the same sequence in order to get the same result(s).

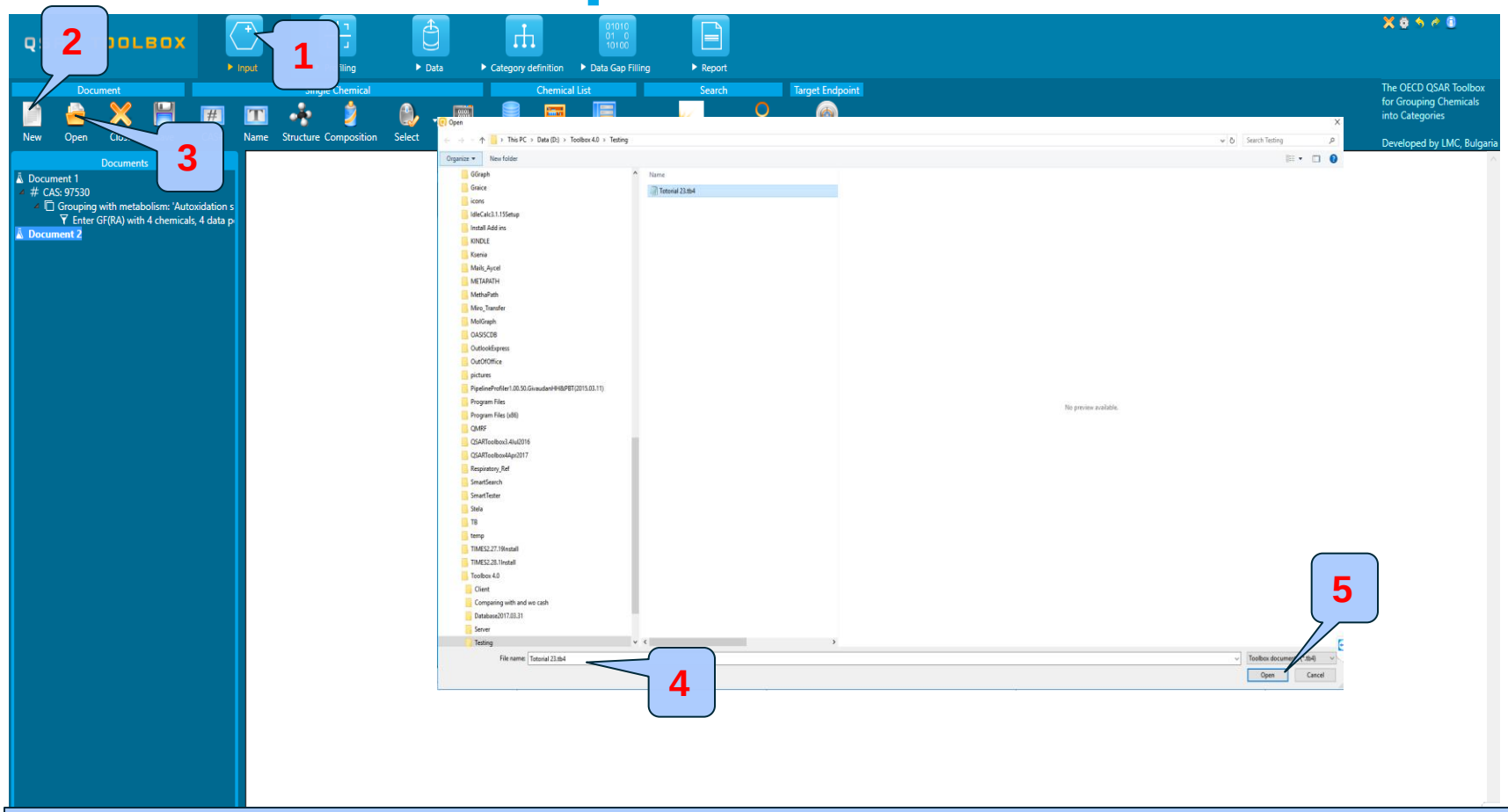
- Saving/Loading the file with TB prediction is shown on next screenshots.

Saving the prediction result



1. Click ***Save*** button; 2. Browse and type in the name of the file; 3. Click ***Save***.

Open saved file



Once the file has been saved 1. Go to ***Input***; 2.Create new document 3.Click ***Open***; 4. Browse and select the file; 5. Click ***Open***;