QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Predicting skin sensitisation potential of a chemical using skin sensitization data extracted from ECHA CHEM database

Outlook

- **Background**
- The exercise
- Workflow
- Save prediction

Background

- Read-across can be used to estimate missing data from a single or limited number of chemicals using an analogue approach. It is especially appropriate for "qualitative" endpoints for which a limited number of results are possible (e.g. positive, negative, equivocal).
- In the analogue approach, endpoint information for a single or small number of tested chemicals is used to predict the same endpoint for an untested chemical that is considered to be "similar".
- Analogous sets of chemicals are often selected based on the hypothesis that the toxicological effects of each member of the category will show a common behaviour.
- In this test case ECHA Chem skin data are used in the read-across analysis

Side-Bar 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
- **The exercise**
- Workflow
- Save prediction

The Exercise

- In this exercise we will predict the skin sensitization potential for an untested compound, (bis(tert-butyldioxyisopropyl)benzene) [CAS # 25155-25-3], which will be the "target" chemical.
- This prediction will be accomplished by collecting a small set of test data for chemicals considered to be in the same category as the target molecule.
- The category will be defined by the structural alerts common to all the chemicals in the category.
- The additional data extracted from ECHA CHEM database will be used in order to expand the group of analogues
- The prediction itself will be made by "read-across".
- The obtained prediction will be saved in a file

Outlook

- Background
- 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 Chemical

User Alternatives for Chemical ID:

A. Single target chemical

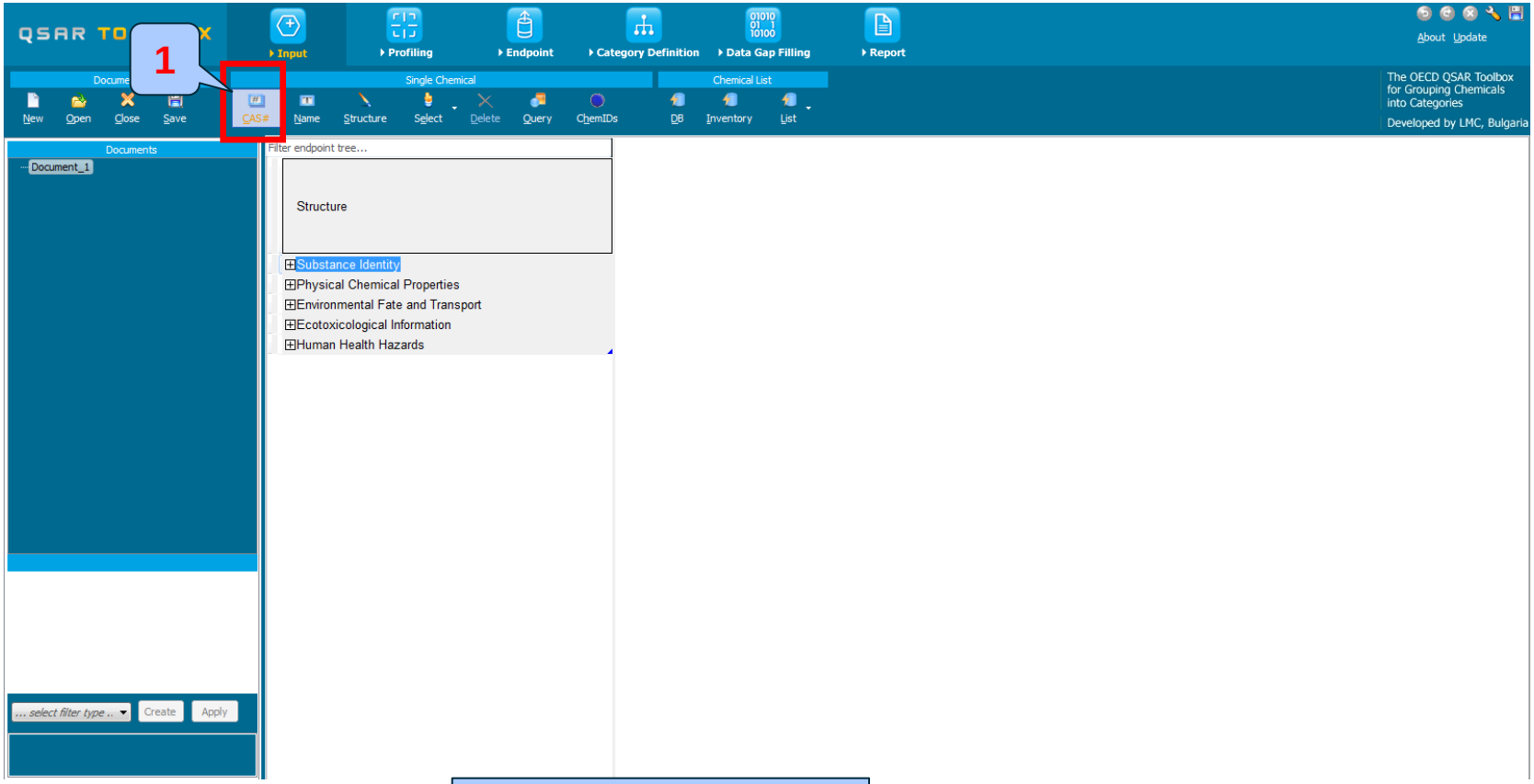
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases
- Chemical IDs such as EC number, Einecs number

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Chemical Input Screen

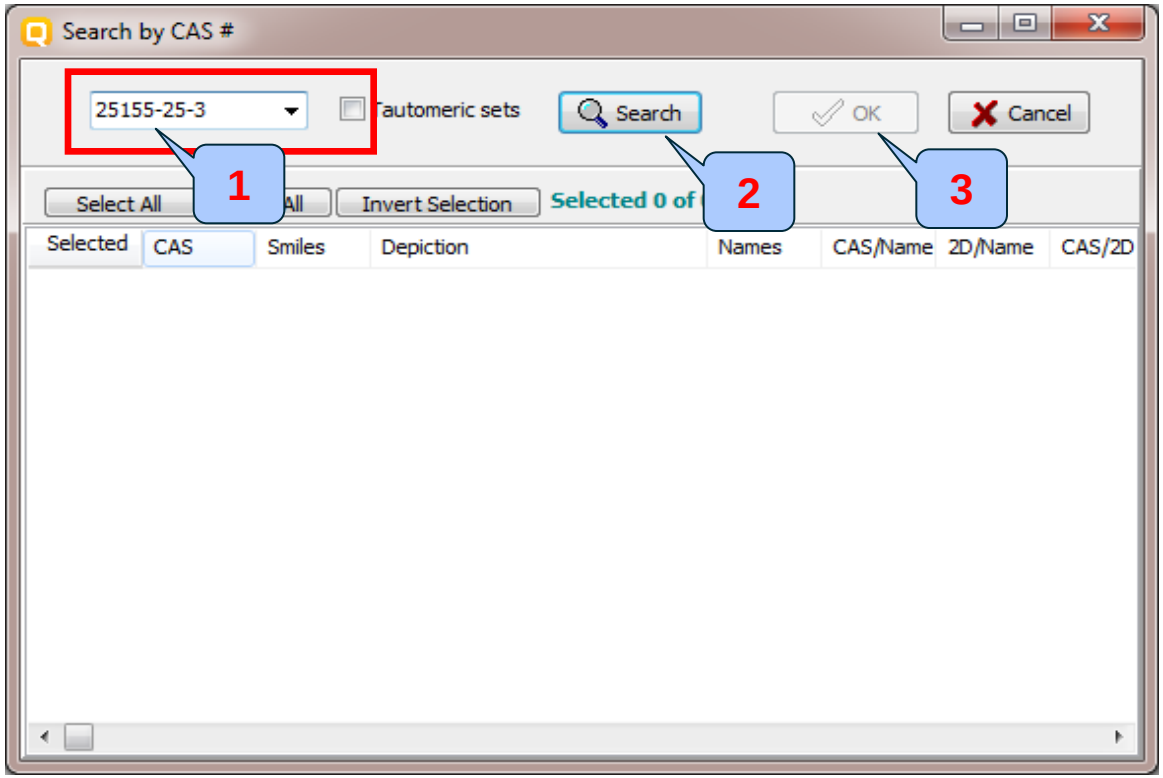
Input target chemical by CAS#



1. **Click** on CAS#

Chemical Input Screen

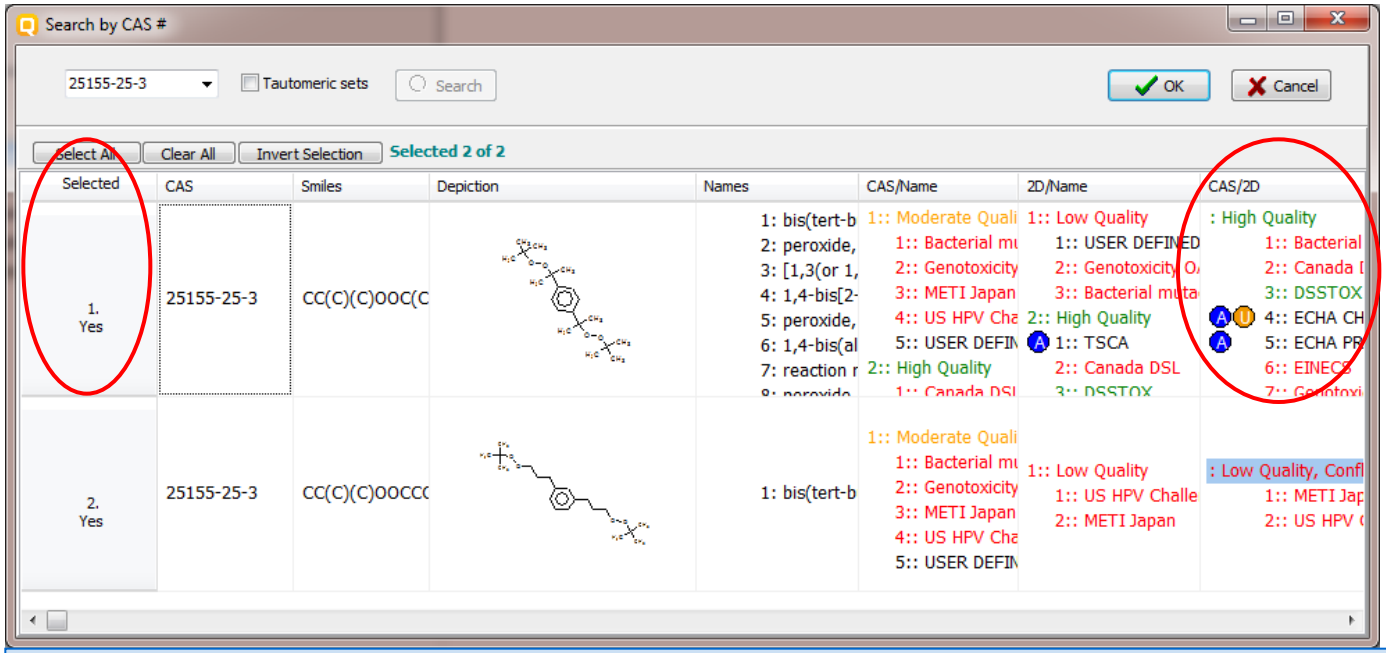
Enter CAS# 25155-25-3



1. **Enter** the CAS# In the field; 2. **Click** Search button; 3. **Press** OK

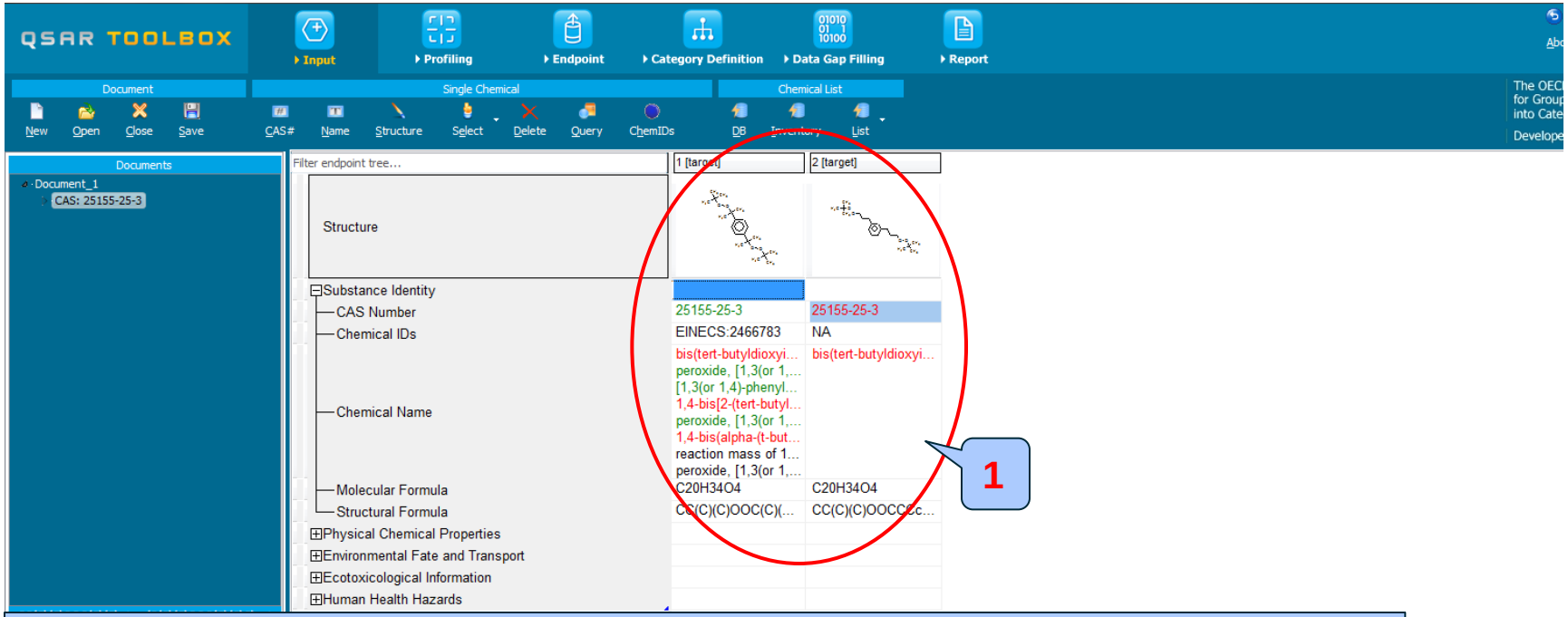
Chemical Input
Target chemical identity

The Toolbox now searches the databases to find out if the CAS# you entered is linked to a molecular structure stored in the Toolbox. It is displayed as a 2-demensional depiction.



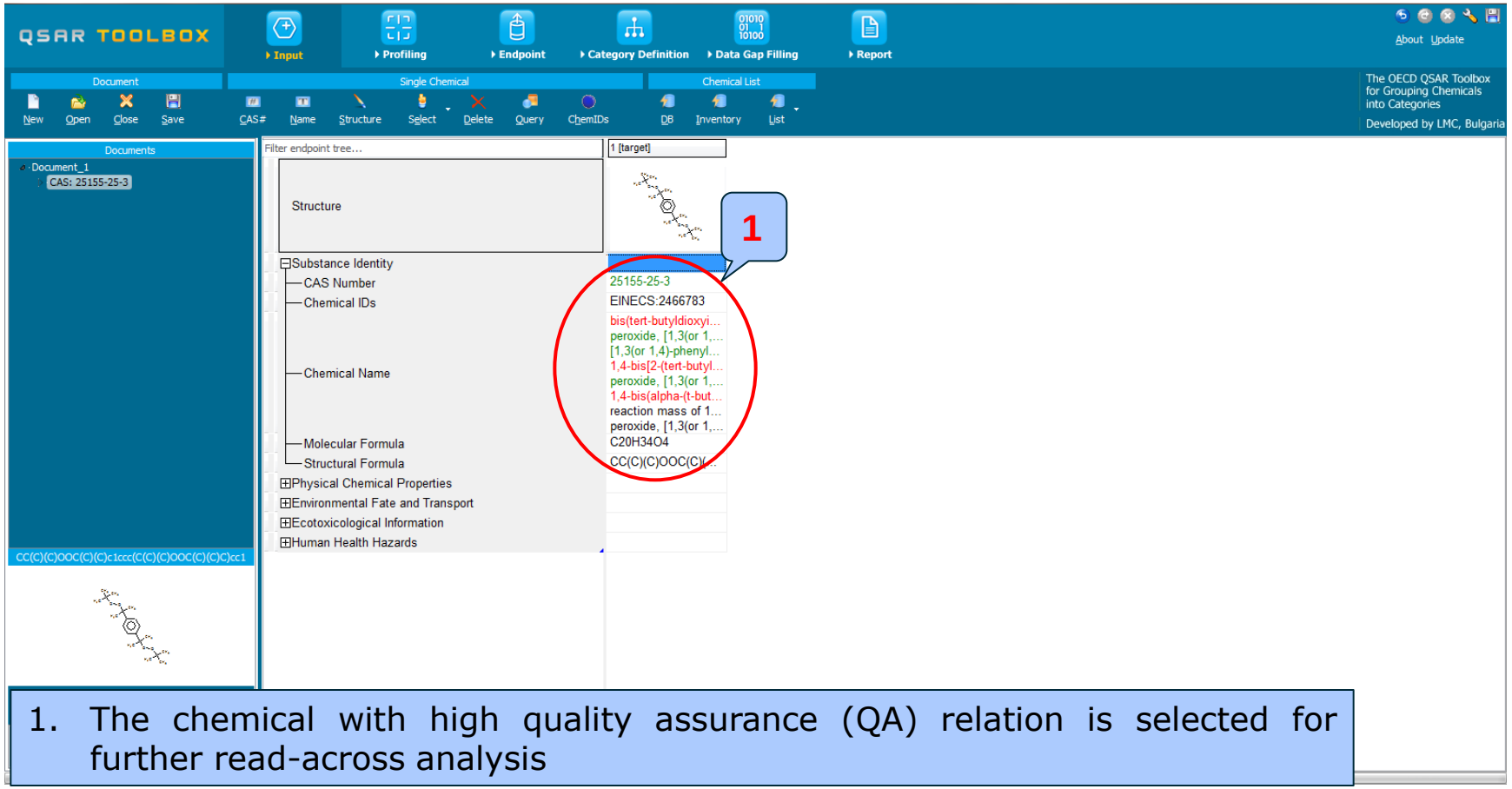
In case a structure has several CAS numbers or a structure could be related to more than one substance, more than one chemical identity could be retrieved. In this case the user can decide which substance is to be retained for the subsequent workflow.

Chemical Input
Target chemical identity



1. Double click "Substance Identity"; this displays the chemical identification information.
Note that existing names of the target chemical are in different colours (see next screen shot).

Chemical Input
Target chemical identity



1. The chemical with high quality assurance (QA) relation is selected for further read-across analysis

Chemical Input
Chemical identity

The colour code indicates the reliability of the chemical identifier:

- **Green**: There is a high consistency between the identifier and the structure. This colour is applied if the identifier is the same in several quality assured databases.
- **Yellow:** There is only a moderate reliability between the identifier and the structure. The colour is applied if the identifier is the same in several databases for which the quality assurance could not be established.
- **Red**: There is a poor reliability between the identifier and the structure. The colour is applied if the identifier is allocated to different structures in different databases.

Outlook

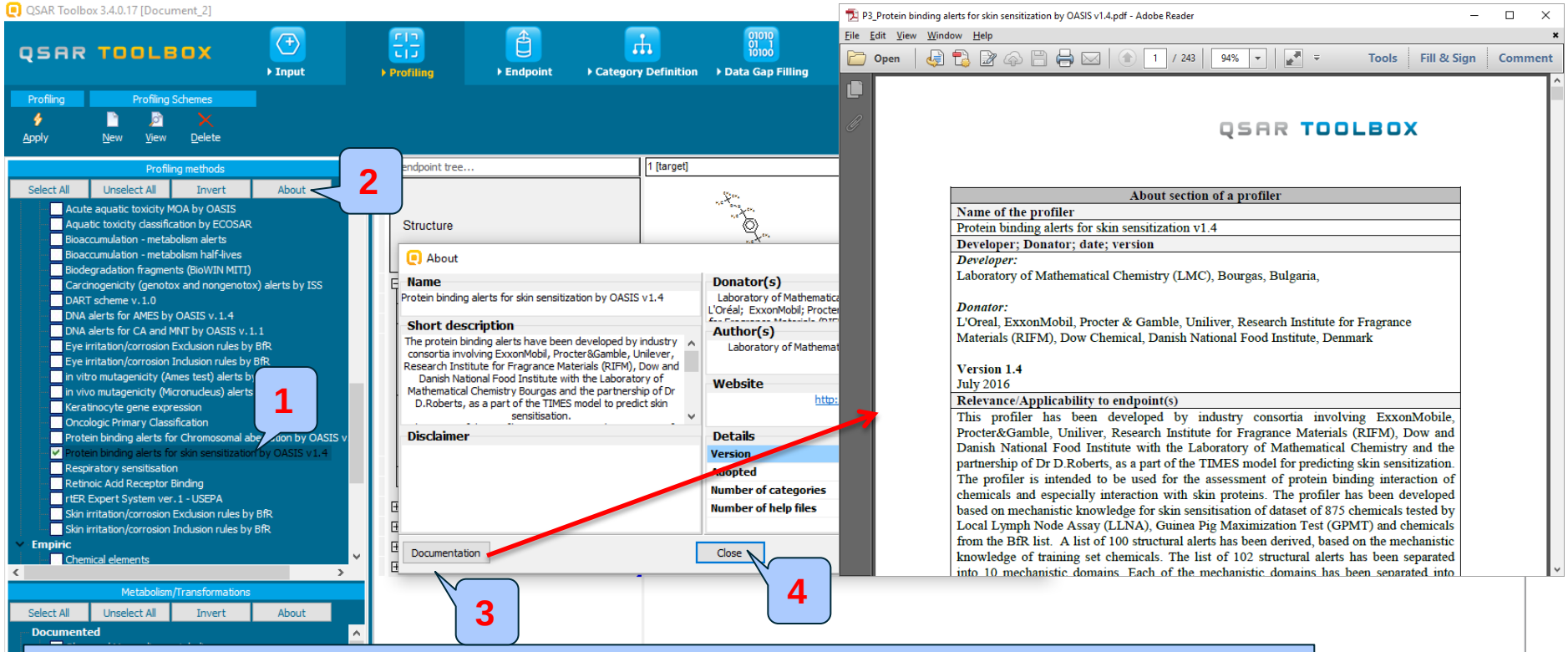
- Background
- The exercise
- **Workflow**
  - Chemical Input
  - **Profiling**

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
Side-Bar to Profiling

Summary information of the different profilers are provided in the "About"

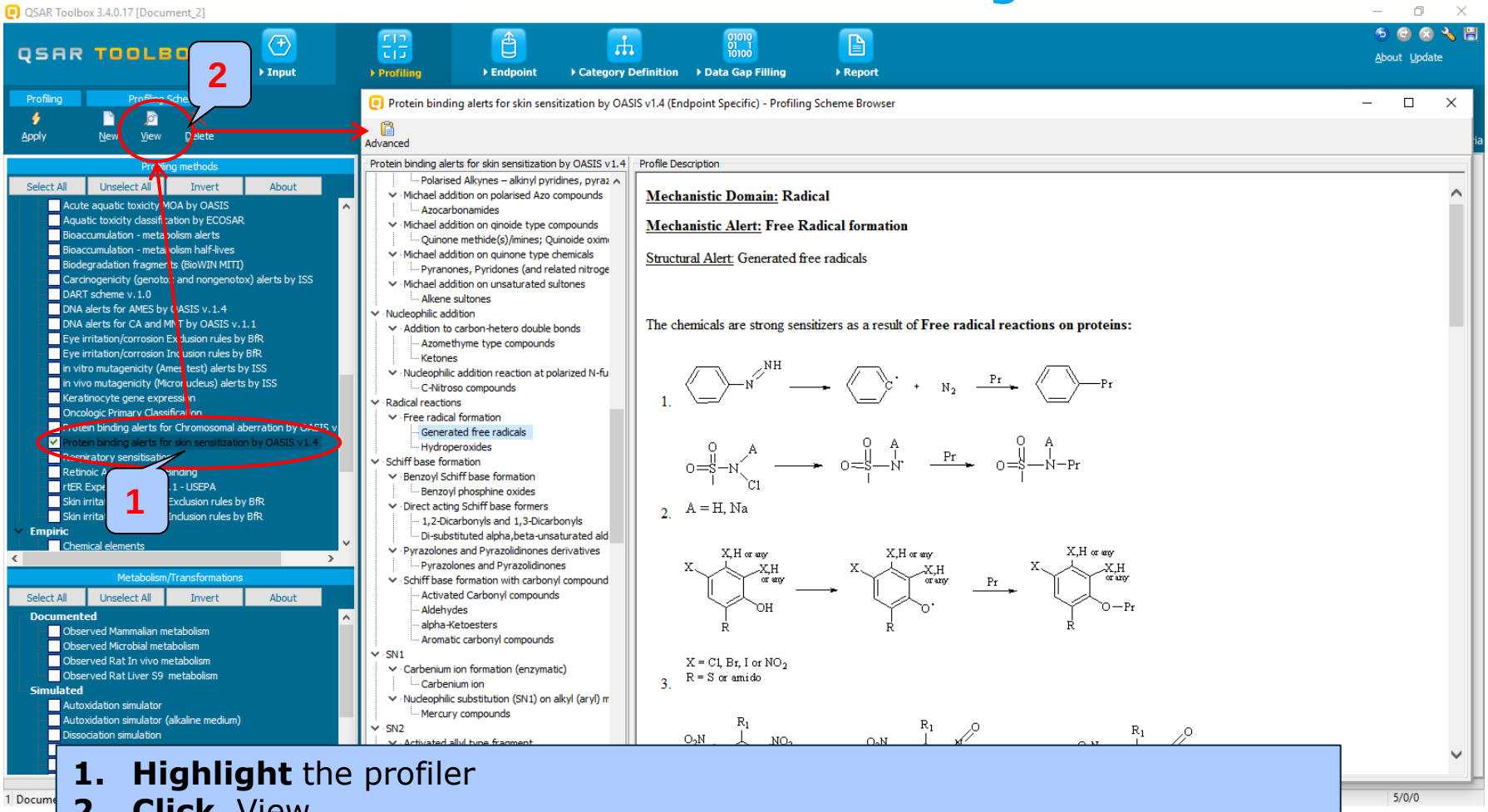


1. **Select** the name of the profiler, perform right click on it and then
2. **Select** About
3. Detailed documentation is available within **"Documentation"** button
4. **Close** before proceeding

Profiling
Overview

- Short description of the profilers is available within the About
- Details of the boundaries coded the rules of the categories along with textual description for each category is available within “View” functionality (shown on the next slide)

Profiling
Side-Bar to Profiling



1. **Highlight** the profiler
2. **Click** View

Profiling
Profiling the target chemical

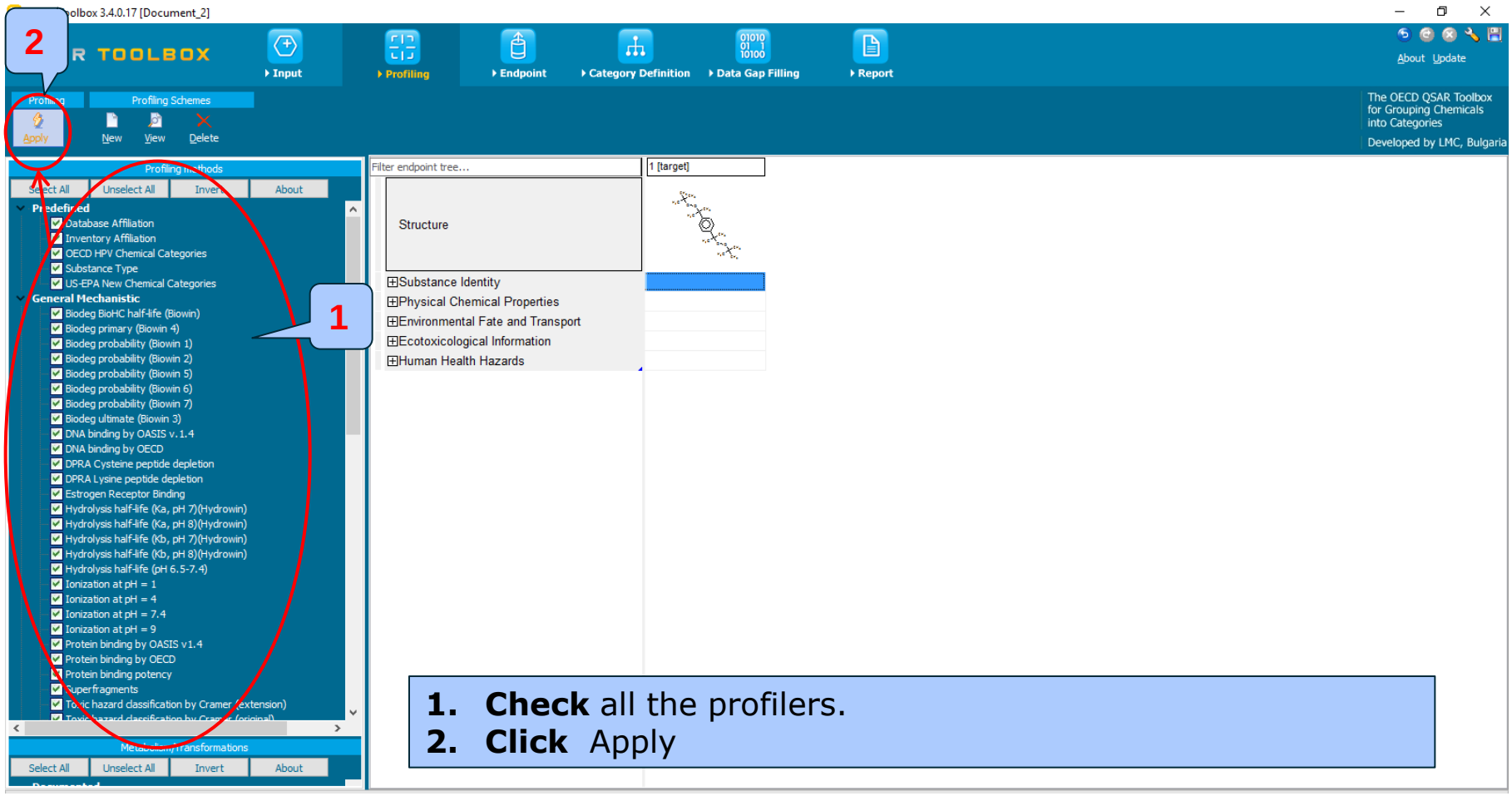
- The outcome of the profiling determines the most appropriate way to search for analogues (detailed information in Manual for getting started (Chapter 4). http://www.oecd.org/dataoecd/58/56/46210452.pdf

- Table 4 - 1 in chapter 4 (Manual for getting started) lists a selection of profilers and their relevance for different endpoints of regulatory relevance.

- For example the following mechanistic/endpoint specific profiling schemes are relevant to the Skin sensitization endpoint:
  - Protein binding by OASIS v.1.4 – mechanistic grouping
  - Protein binding by OECD – mechanistic grouping
  - Protein Binding Potency – mechanistic grouping
  - Protein binding alerts for skin sensitization by OASIS v1.4 – endpoint specific

Profiling

Profiling the target chemical



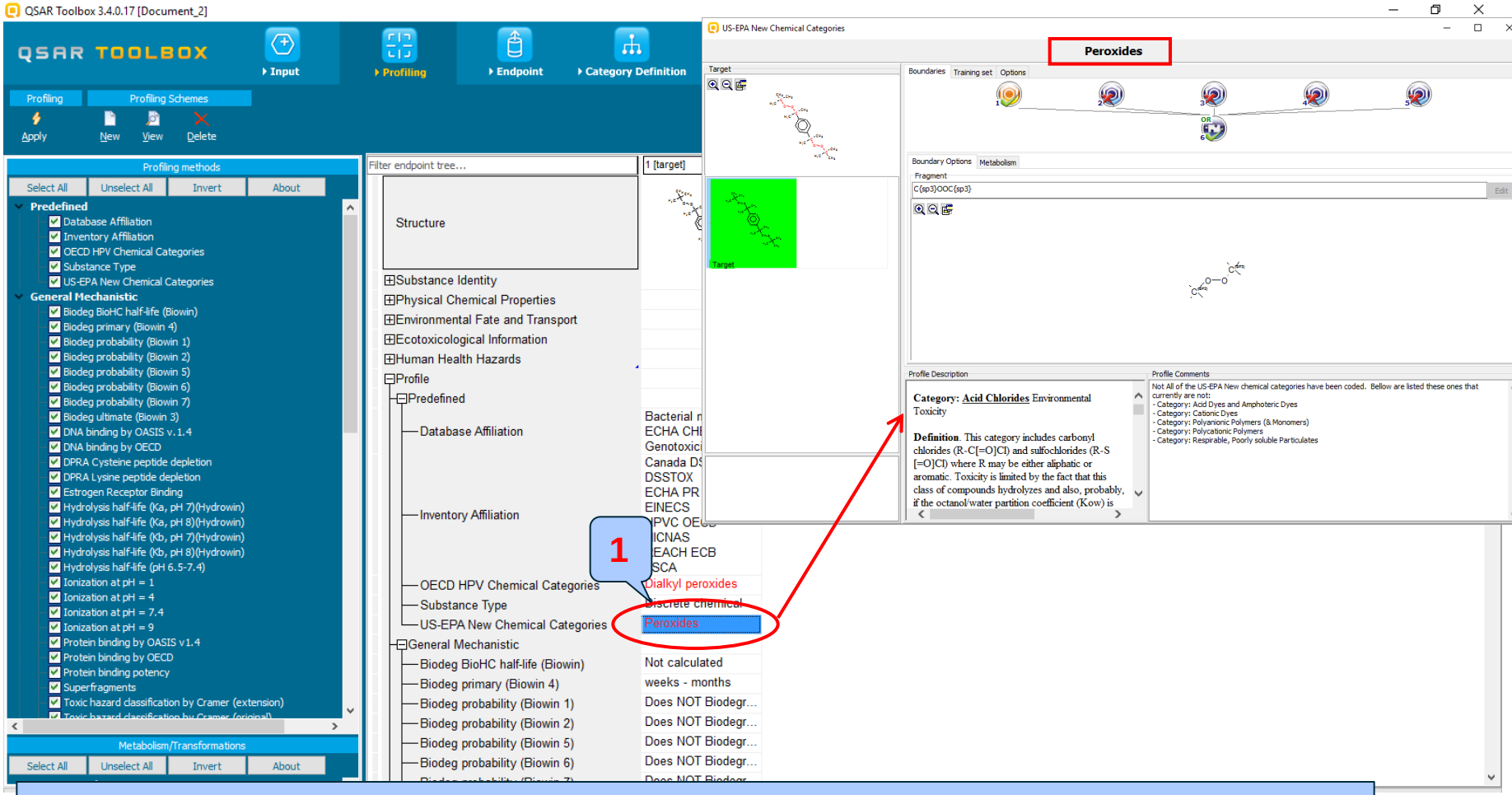
1. **Check** all the profilers.
2. **Click** Apply

Profiling
Profiling the target chemical

- The actual profiling will take up to several seconds depending on the number and type of profilers selected.
- The results of profiling automatically appear as a dropdown box under the target chemical (see next screen shot).
- Please note the profiling result based on US-EPA profiler
- This result will be used to search for suitable analogues in the next steps of the exercise.

Profiling

Profiling the target chemical



1. **Right** click over the profiling result to review it.

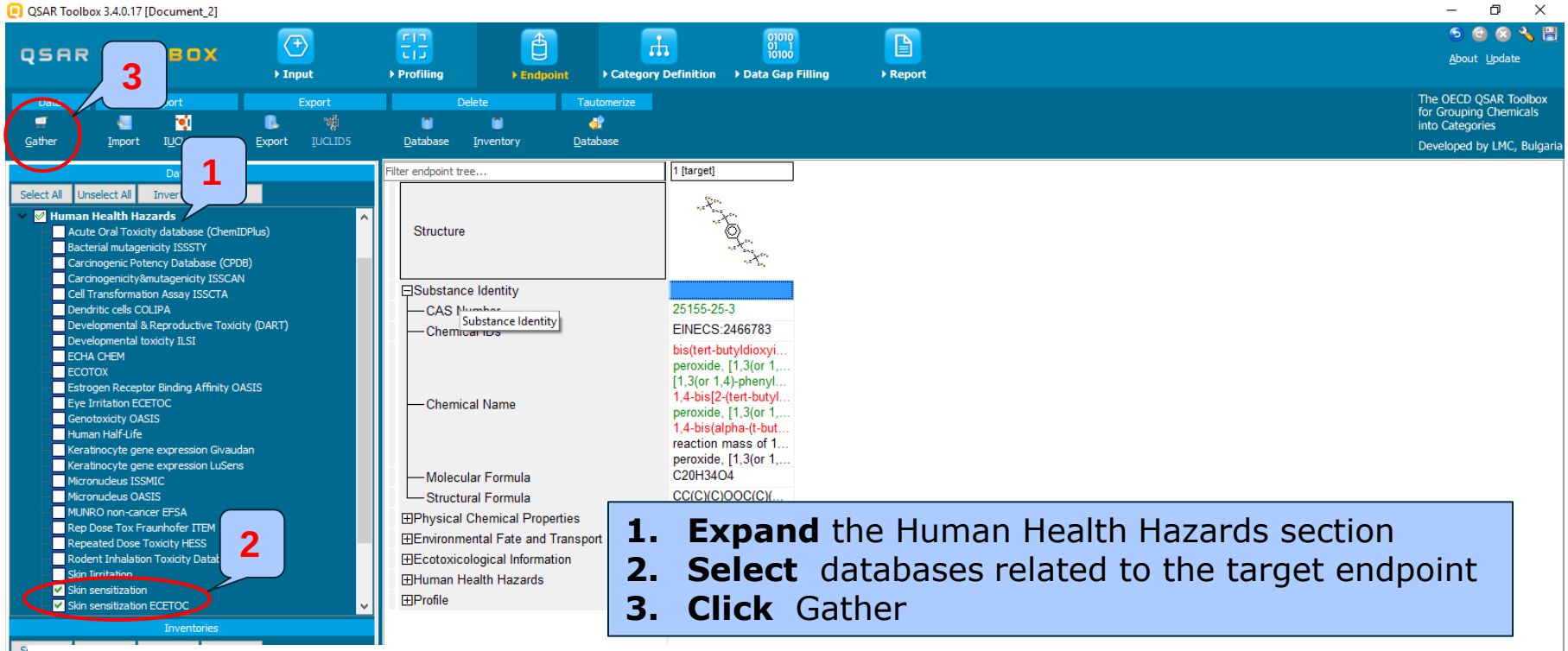
Outlook

- Background
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - **Endpoint**

Endpoint
Overview

- "Endpoint" 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).

Endpoint
Gather data



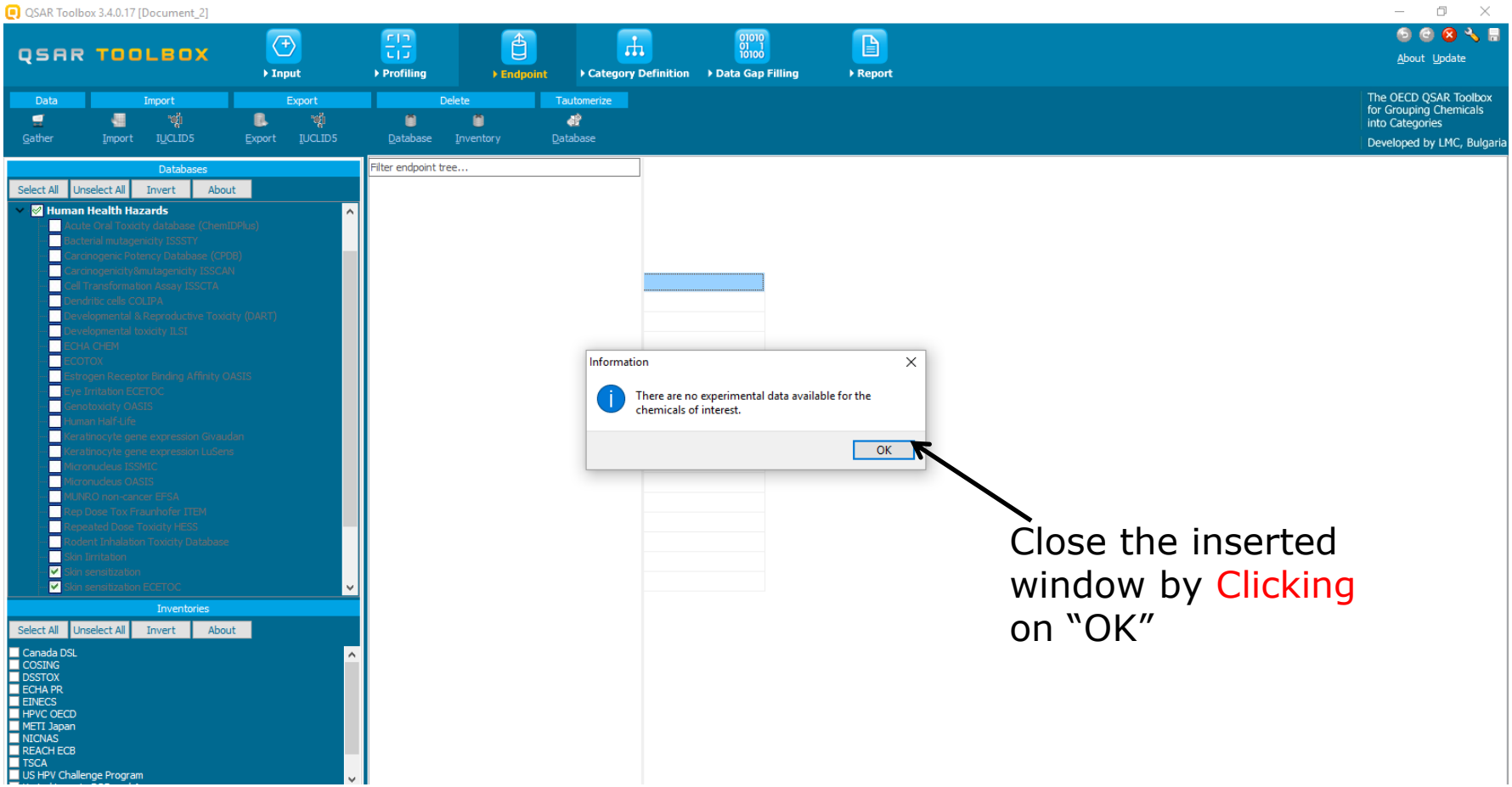
1. **Expand** the Human Health Hazards section
2. **Select** databases related to the target endpoint
3. **Click** Gather

- 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).

Endpoint
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 .
- In this example, an insert window appears stating there was “no data found” for the target chemical (see next screen shot).

Endpoint
Gather data



Recap

- In module one, you have entered the target chemical being sure of the correct structure.
- In the second module, you have profiled the target chemical.
- In the third module, you have found that no experimental data is currently available in the Toolbox for this 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.

Outlook

- Background
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - **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.

- Detailed information about grouping chemicals on the following link (Chapter 4). http://www.oecd.org/dataoecd/58/56/46210452.pdf

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
2. Endpoint specific profilers (for sub-cat)
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 performed (endpoint dependant)
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

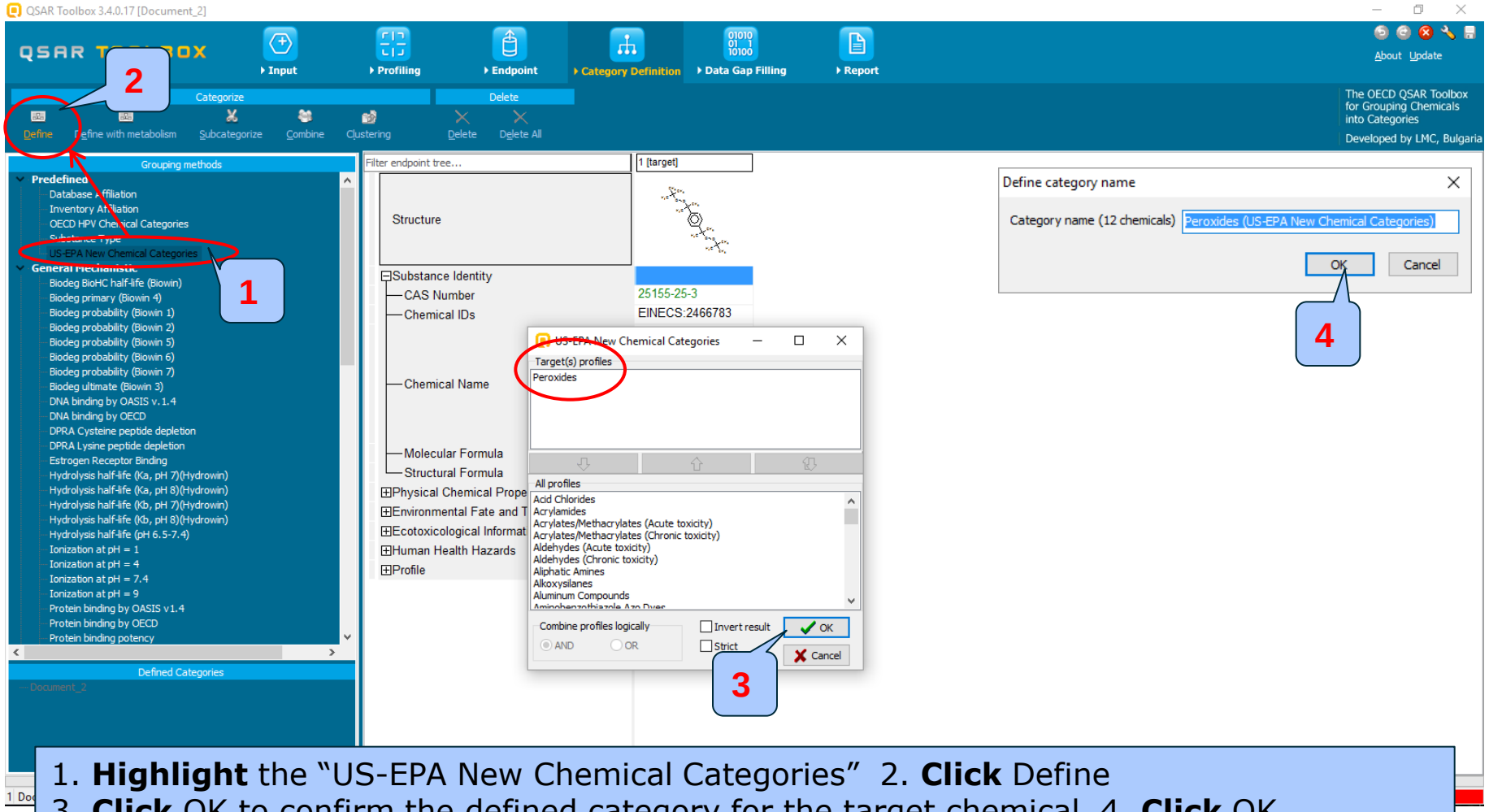
Phase III. Eliminating dissimilar chemicals

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

Subcategorization
Endpoint Specific

Category Definition

Defining US-EPA New Chemical categories

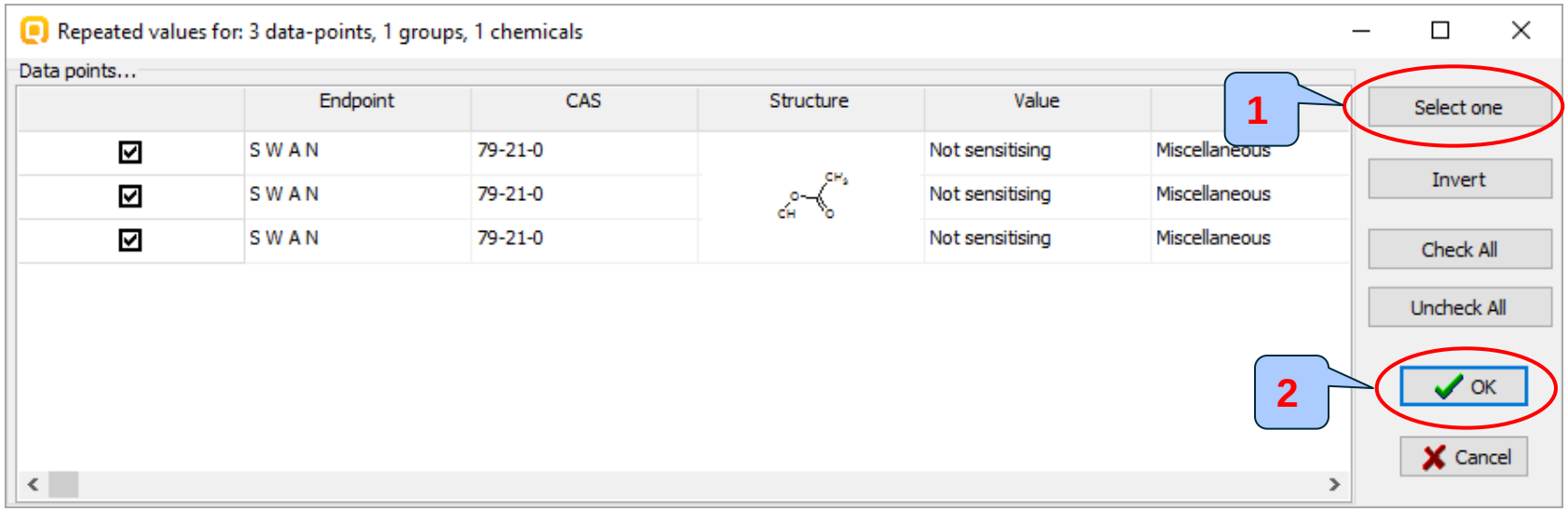


1. **Highlight** the "US-EPA New Chemical Categories" 2. **Click** Define
3. **Click** OK to confirm the defined category for the target chemical 4. **Click** OK

Category Definition

Defining US-EPA New Chemical categories

Due to overlap between the Toolbox databases for intersecting chemicals the same data may be found simultaneously. Data redundancies are identified and the user has the opportunity to select either a single data value or all data values.

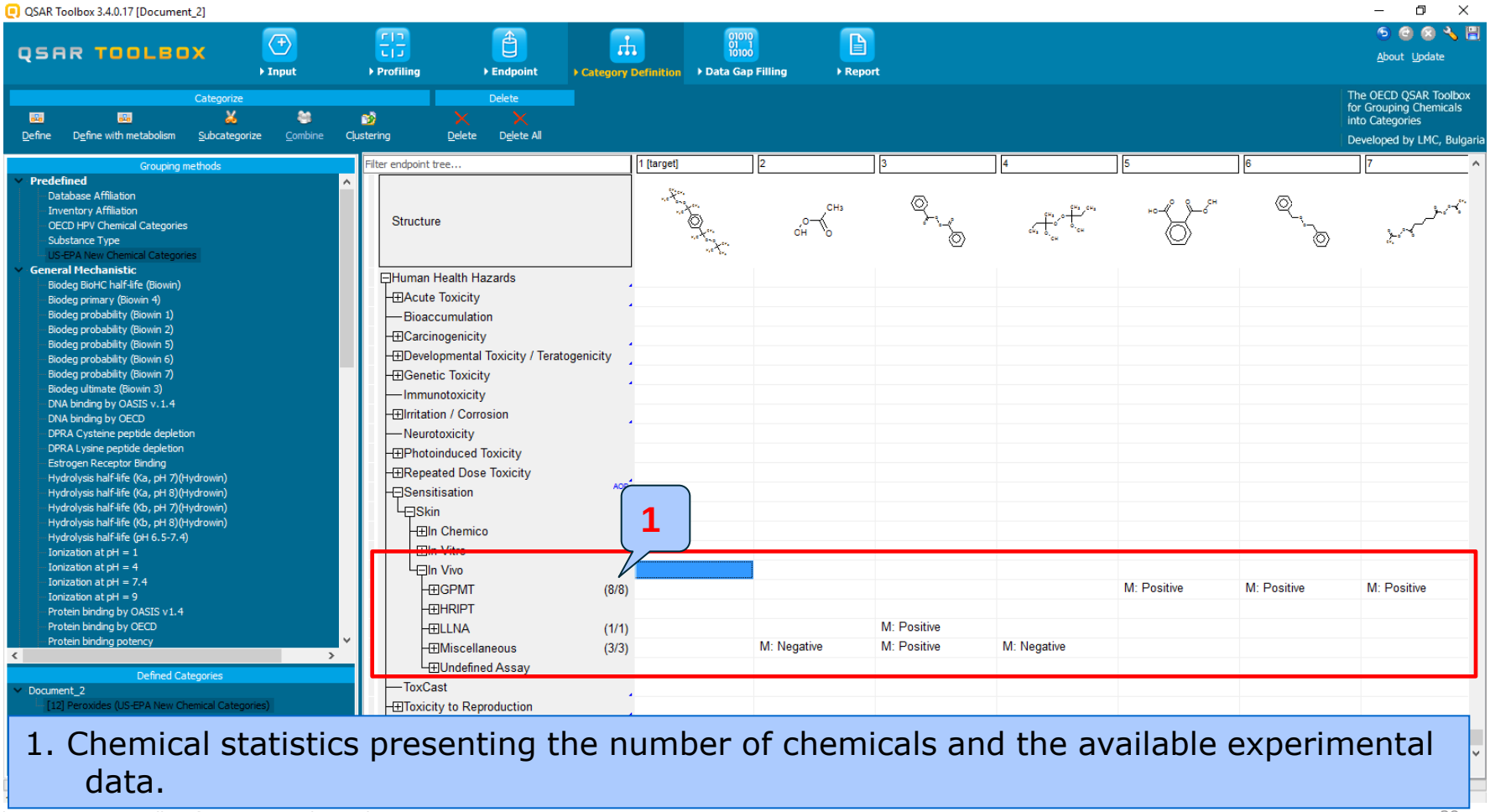


1. **Click Select one**
2. **OK** to retrieve all available experimental data

Category Definition

Defining US-EPA New Chemical categories

The experimental results for the analogues appeared on datamatrix



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

Outlook

- Background
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - Category definition
  - **Data Gap Filling**

Data Gap Filling
Overview

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

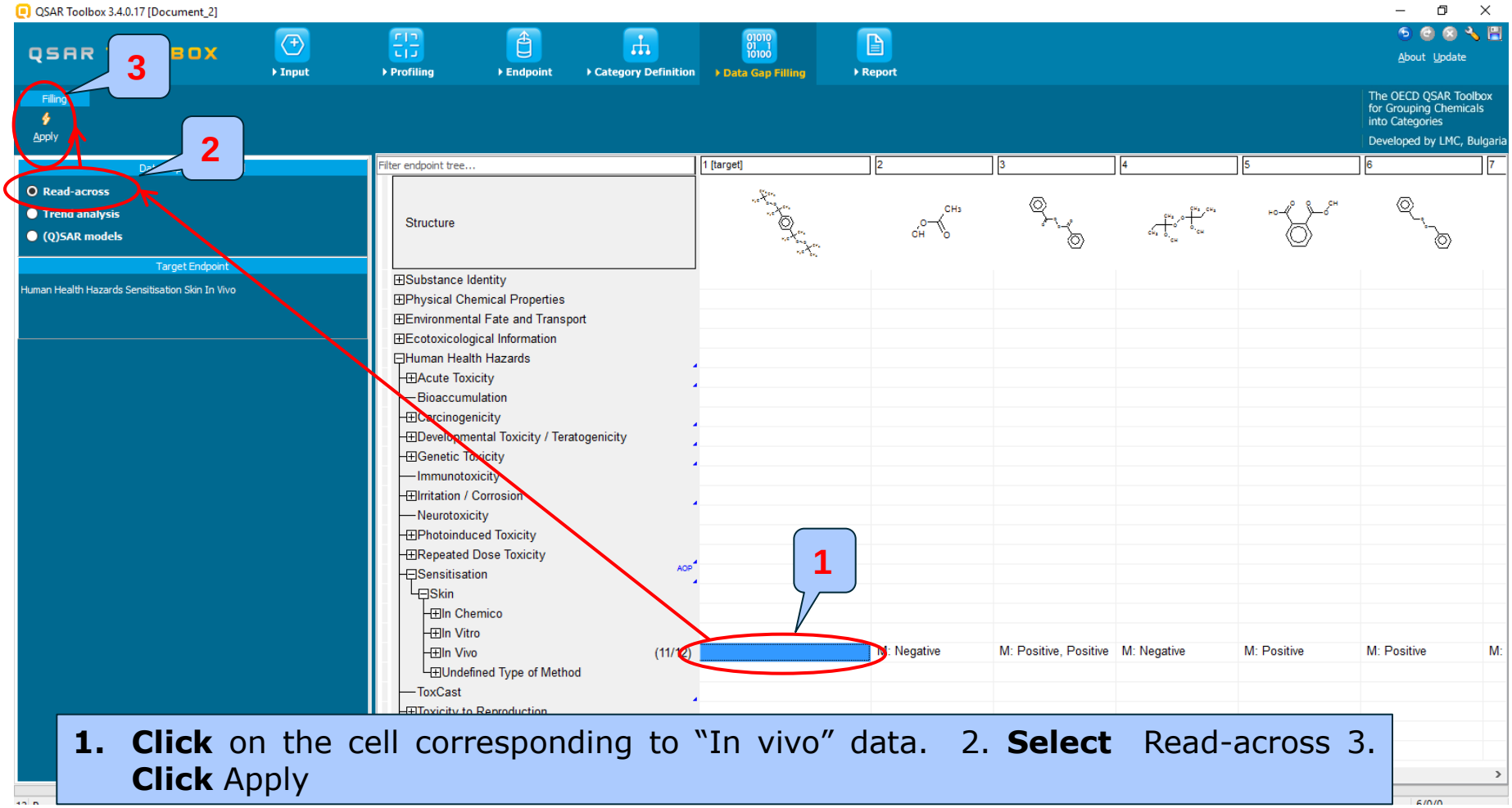
- Depending on the situation, the most relevant data gap mechanism should be chosen, taking into account the following considerations:
  - Read-across is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitisation or mutagenicity for which a limited number of results are 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.

- In this example, we apply read-across.

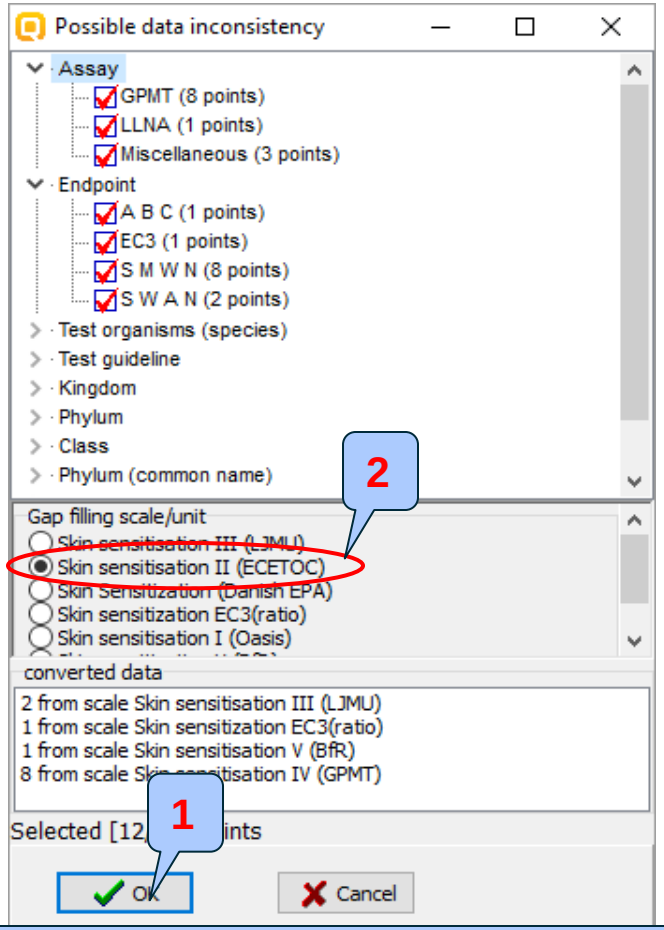
Data Gap Filling
Scale definition

- Skin sensitisation is "qualitative" endpoint for which the results are presented with categorical data (for example: positive; negative; weak sensitizer; strong sensitizer ,etc).
- Skin sensitisation potential of the chemicals came from different authors coded with different names (for example: data from John Moores University of Liverpool are: *Strongly sensitizing, Moderately sensitizing etc*.; data from European centre for Ecotoxicology and Toxicology of chemicals are: *Positive, Negative, and Equivocal*).
- In order to unify the different skin sensitization potential values grouped in two or three different values (negative; weak sensitizer; strong sensitizer) the scale definitions and respectively scale conversions have been developed
- The default scale for Skin Sensitisation is "Skin Sensitisation ECETOC". It converts all skin data into: Positive, Negative, and Equivocal.

Data Gap Filling
Apply Read across



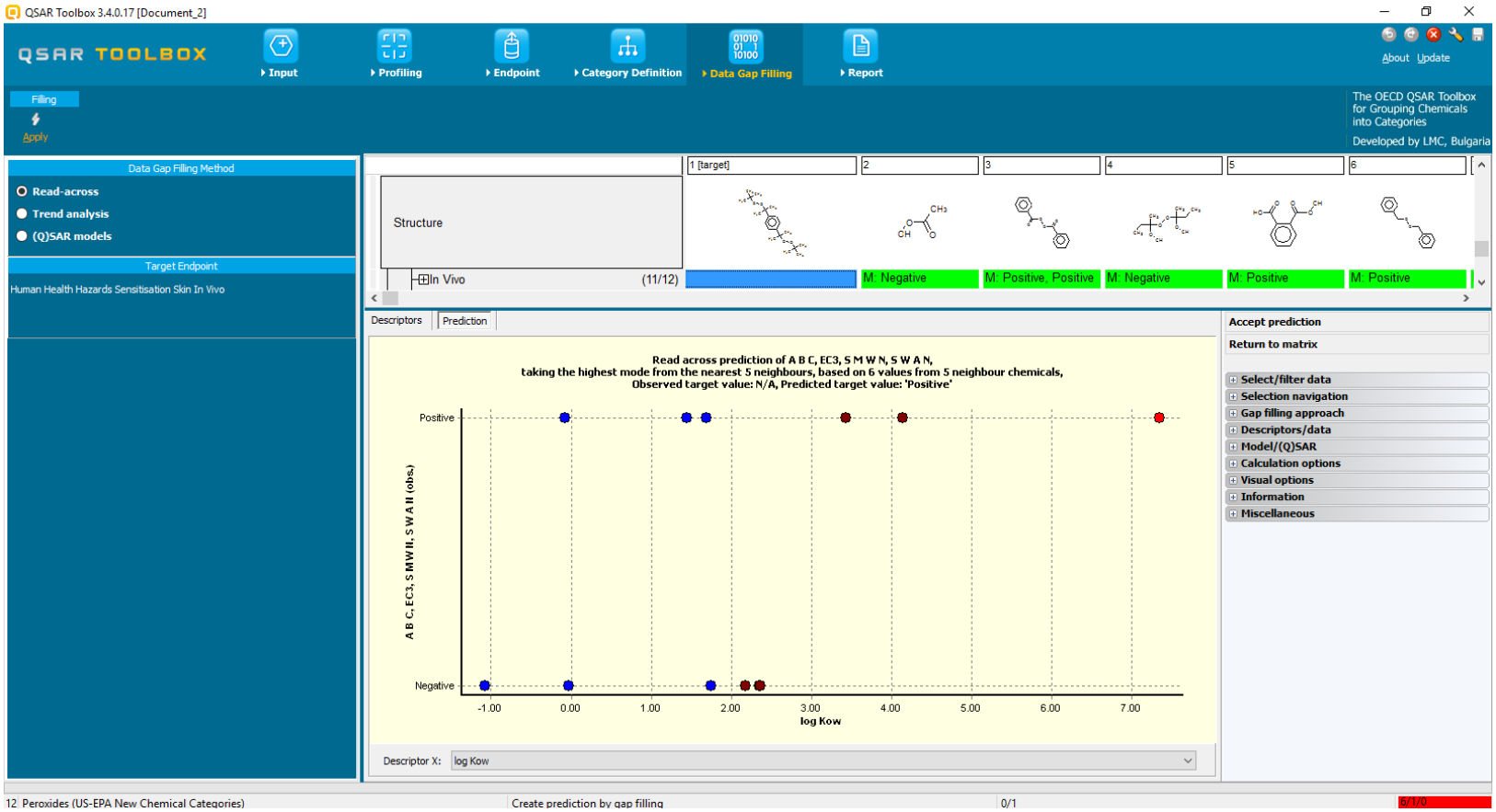
Data Gap Filling
Scale definition



1. **Click** OK
2. Scale Skin sensitisation II(ECETOC) is selected by default as being default scale

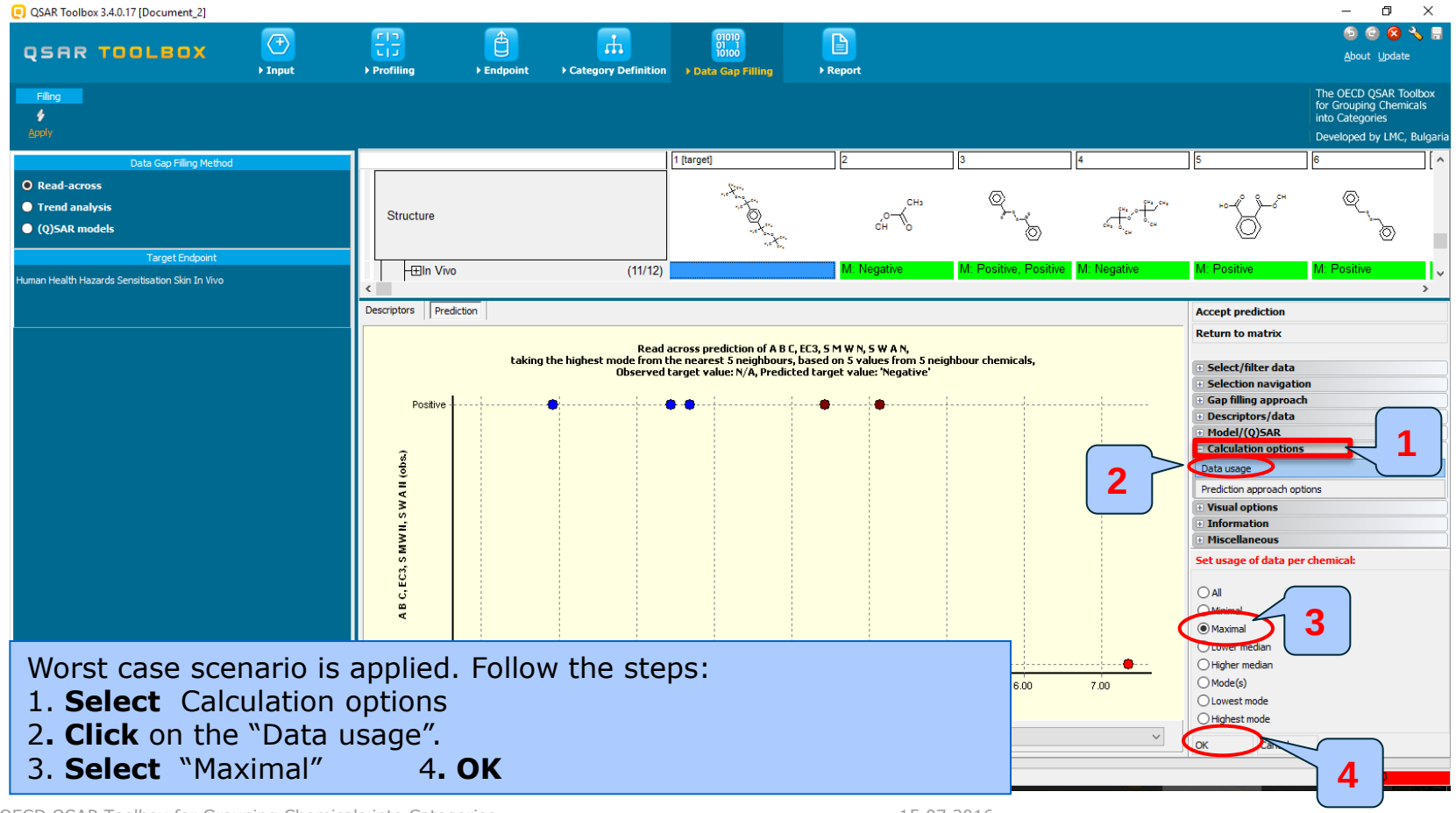
Data Gap Filling
Read-across

The resulting plot is experimental results of all analogues (Y axis) according to a descriptor (X axis) with the default descriptor of log Kow (see next screen shot).



Data Gap Filling
Read-across

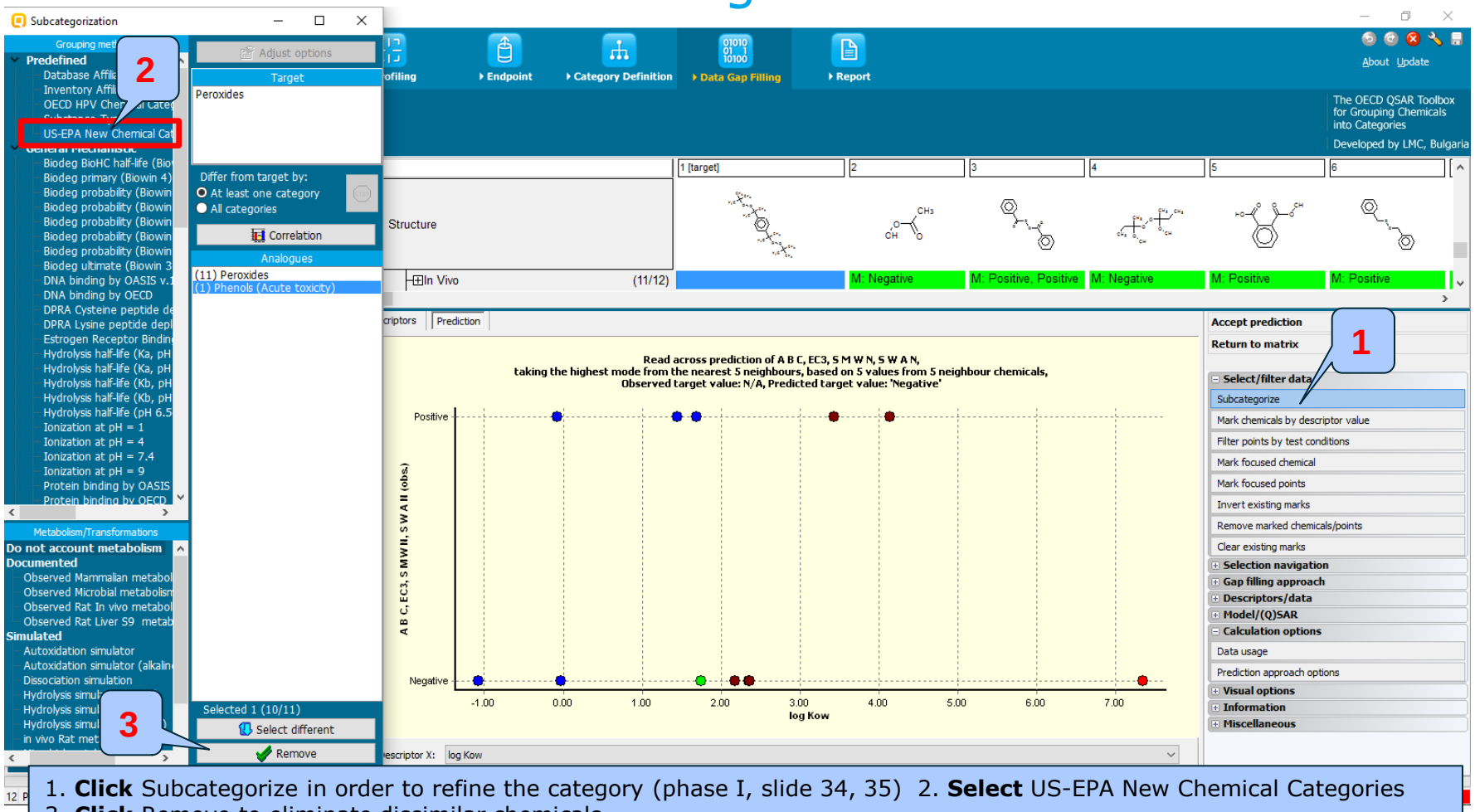
The resulting plot is experimental results of all analogues (Y axis) according to a descriptor (X axis) with the default descriptor of log Kow (see next screen shot).



Worst case scenario is applied. Follow the steps:
1. **Select** Calculation options
2. **Click** on the "Data usage".
3. **Select** "Maximal" 4. **OK**

Data Gap Filling

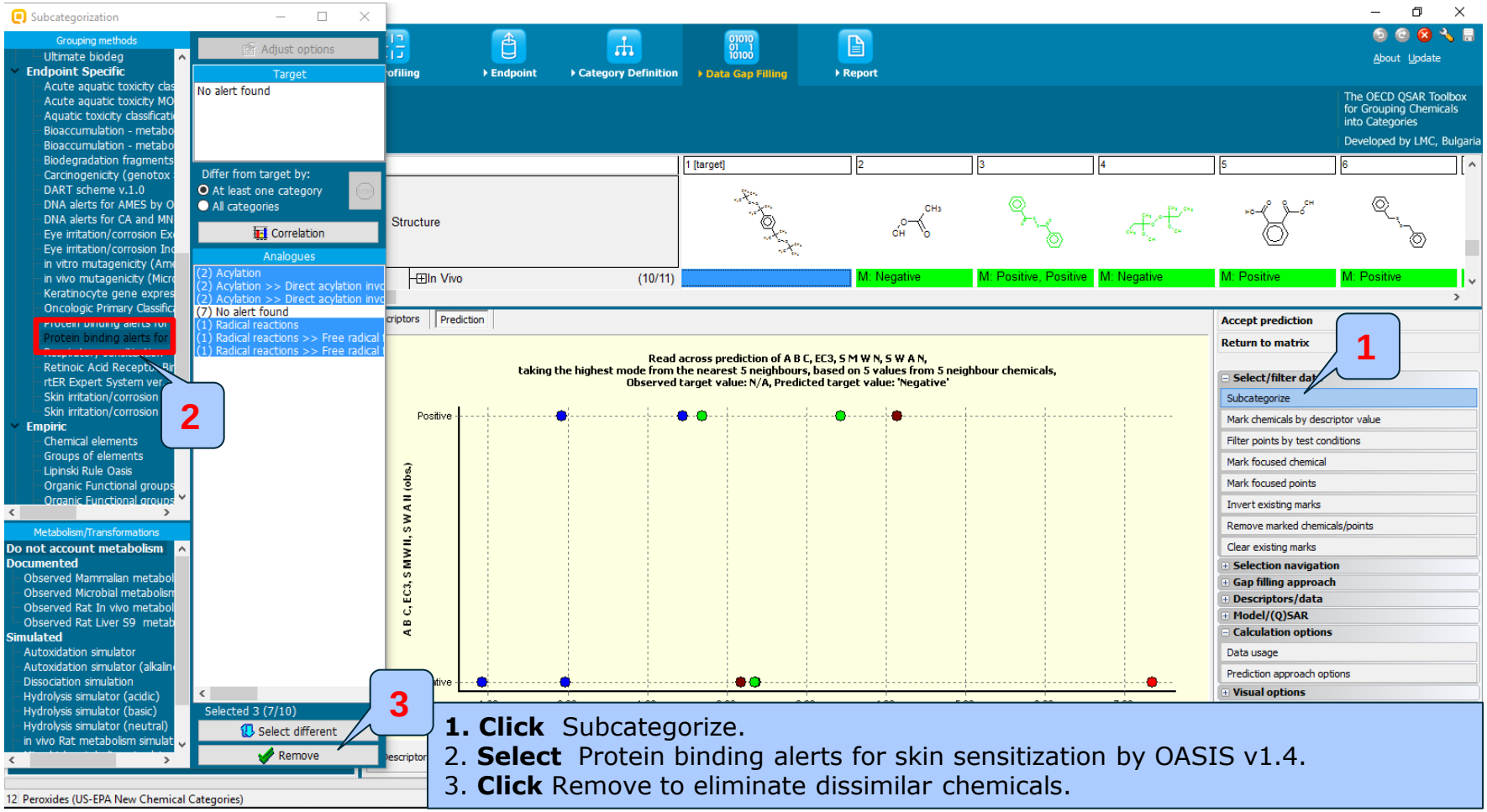
Subcategorization by US-EPA New Chemical Categories



1. **Click** Subcategorize in order to refine the category (phase I, slide 34, 35) 2. **Select** US-EPA New Chemical Categories
3. **Click** Remove to eliminate dissimilar chemicals.

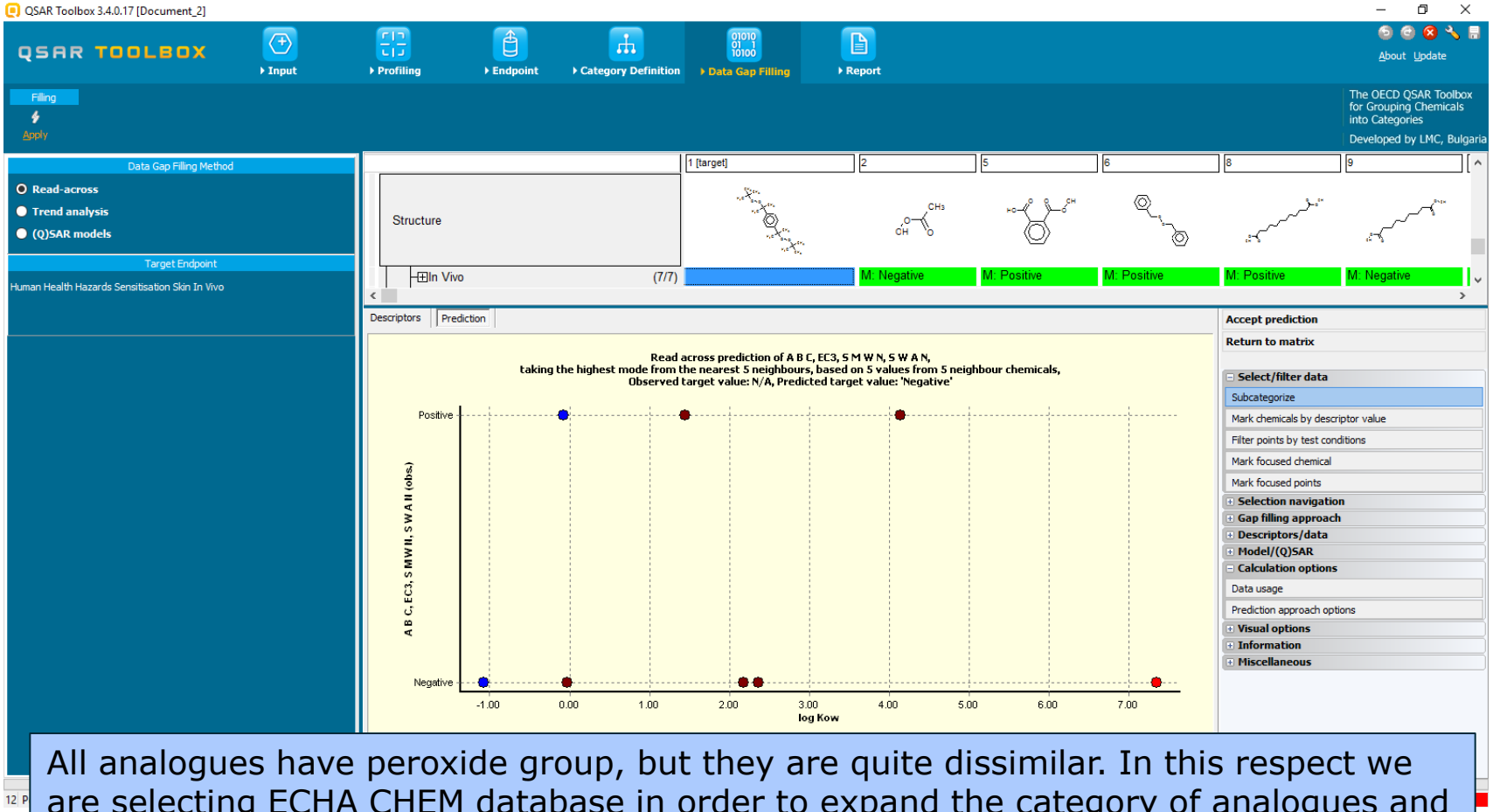
Data Gap Filling

Subcategorization by Protein binding alerts for skin sensitization by OASIS v1.4



1. Click Subcategorize.
2. **Select** Protein binding alerts for skin sensitization by OASIS v1.4.
3. **Click** Remove to eliminate dissimilar chemicals.

Data Gap Filling
Results after subcategorization

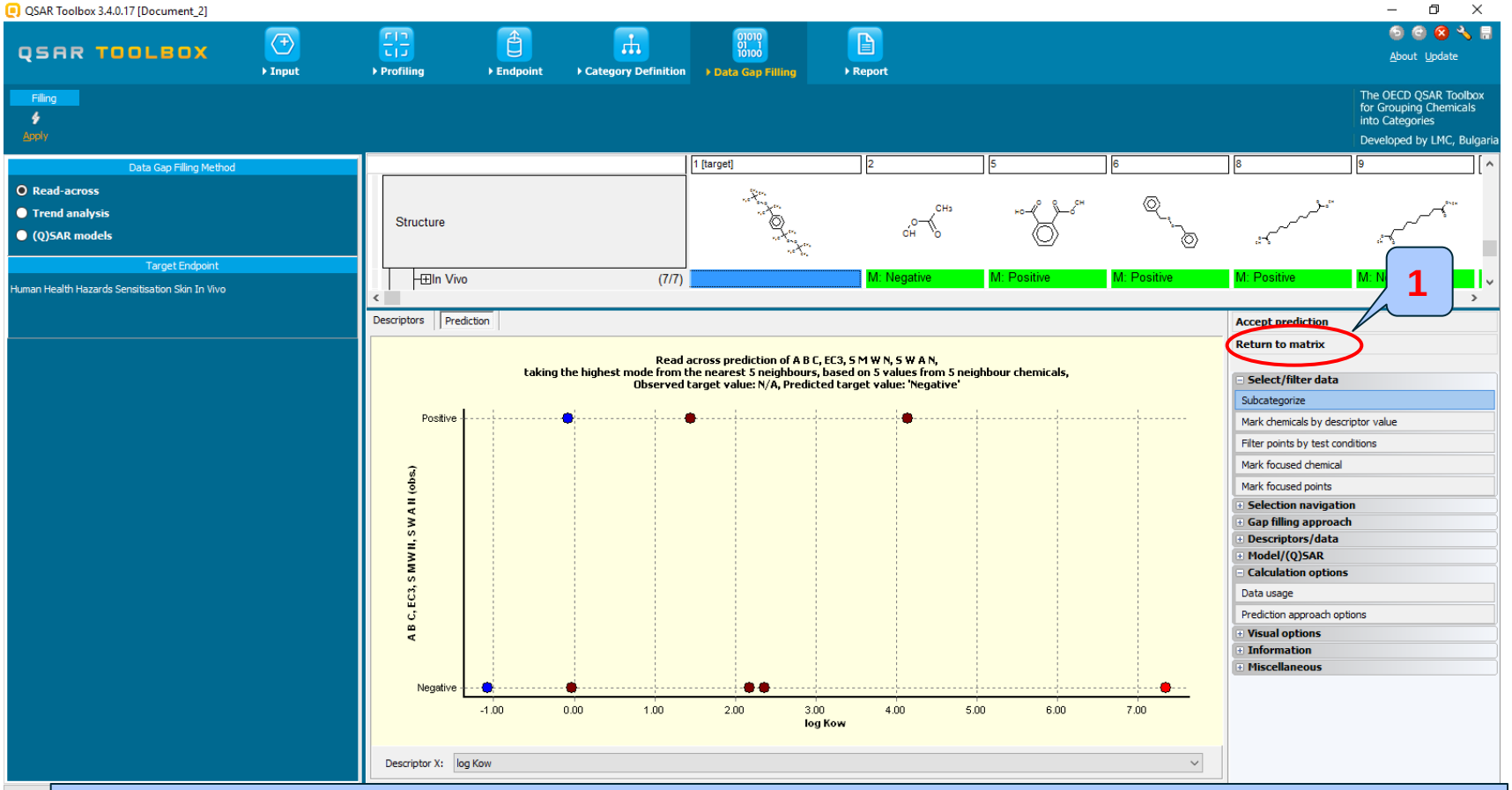


All analogues have peroxide group, but they are quite dissimilar. In this respect we are selecting ECHA CHEM database in order to expand the category of analogues and to identify more suitable analogues for the current read-across analysis.

Data Gap Filling
Interpreting Read-across

- In this example, all analogues have peroxide group, but they are very structurally dissimilar
- The prediction is not reliable due to the structurally dissimilar analogues and could not be accepted
- In order to expand the category of analogues and identify more similar analogues to the target chemical, data extracted from ECHA Chem database is used in the further read-across analysis (see next screen shots).

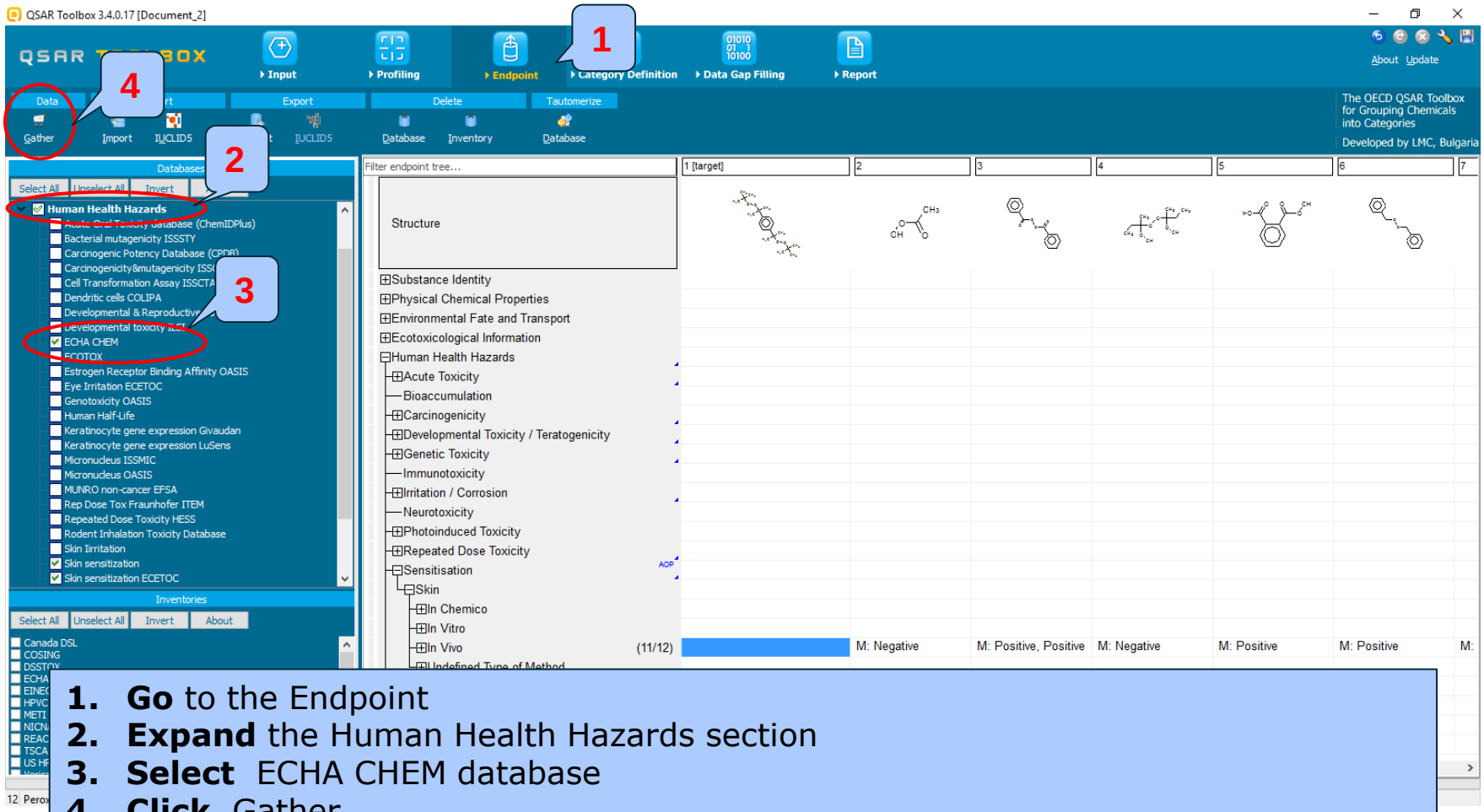
Data Gap Filling
Return to the matrix



1. Select Return to matrix

Endpoint

Gather data from ECHA Chem database

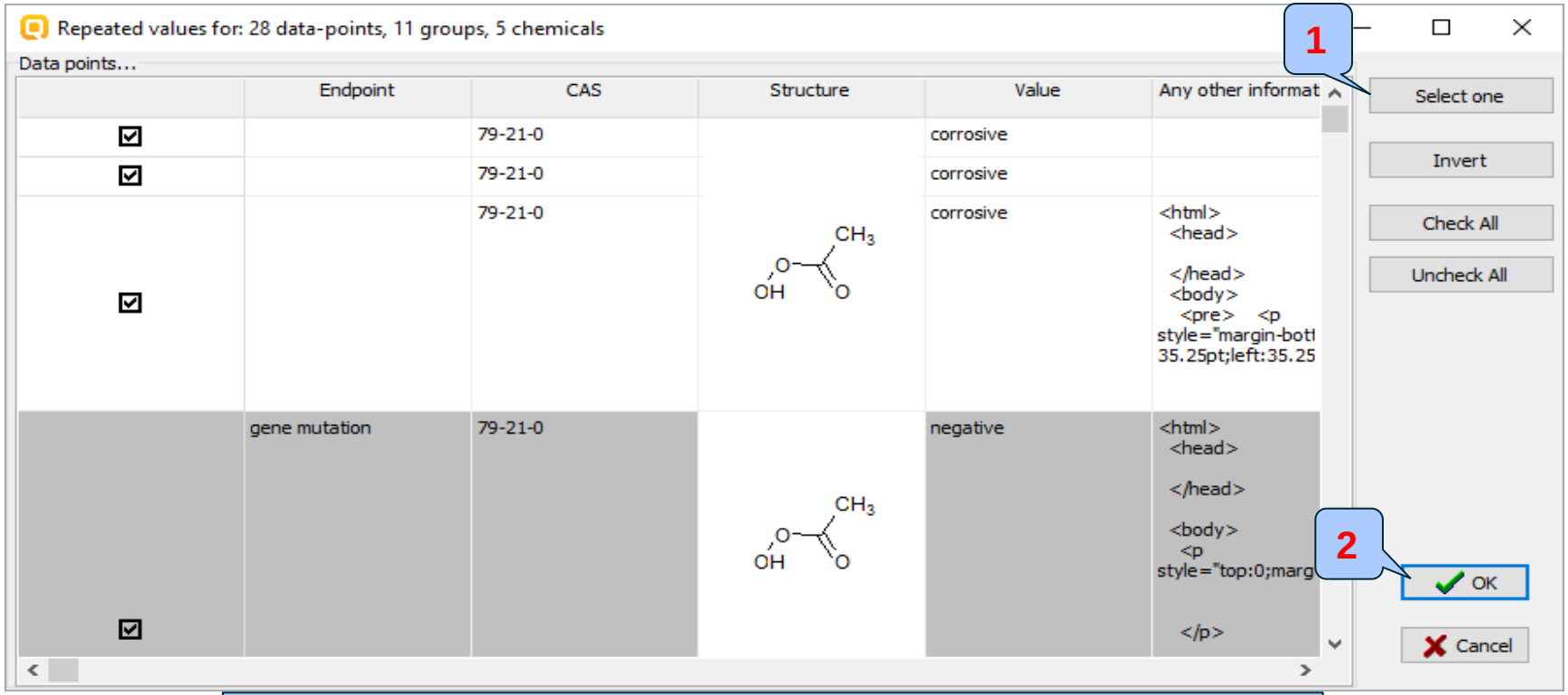


1. **Go** to the Endpoint
2. **Expand** the Human Health Hazards section
3. **Select** ECHA CHEM database
4. **Click** Gather

Endpoint
Gather data

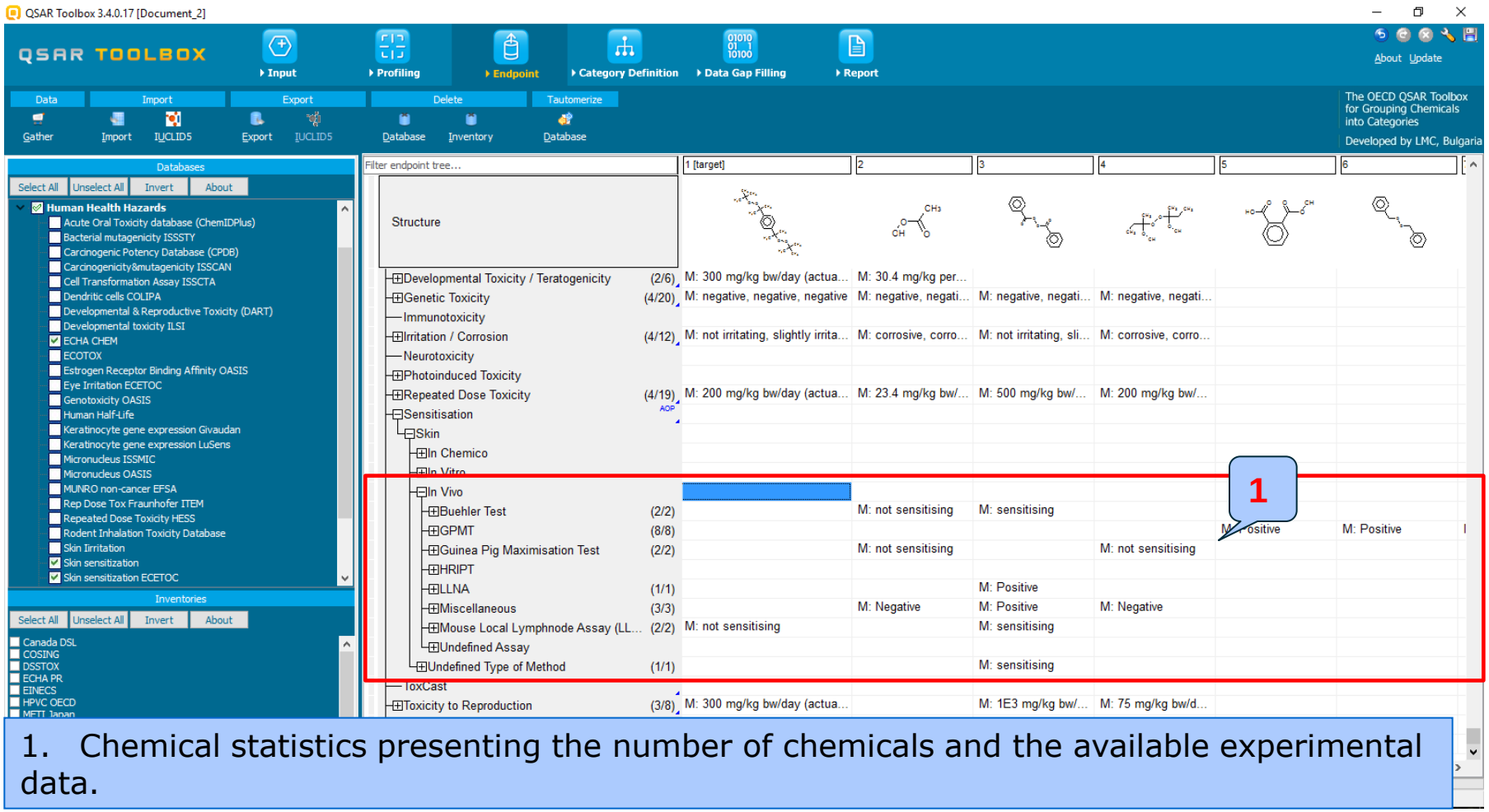
- Toxicity information on the target chemical is electronically collected from the selected dataset(s).
- It should be kept in mind that ECHA CHEM database has been selected
- In this case there is experimental data extracted for the target chemical (see next screen shot).

Endpoint
Gather data



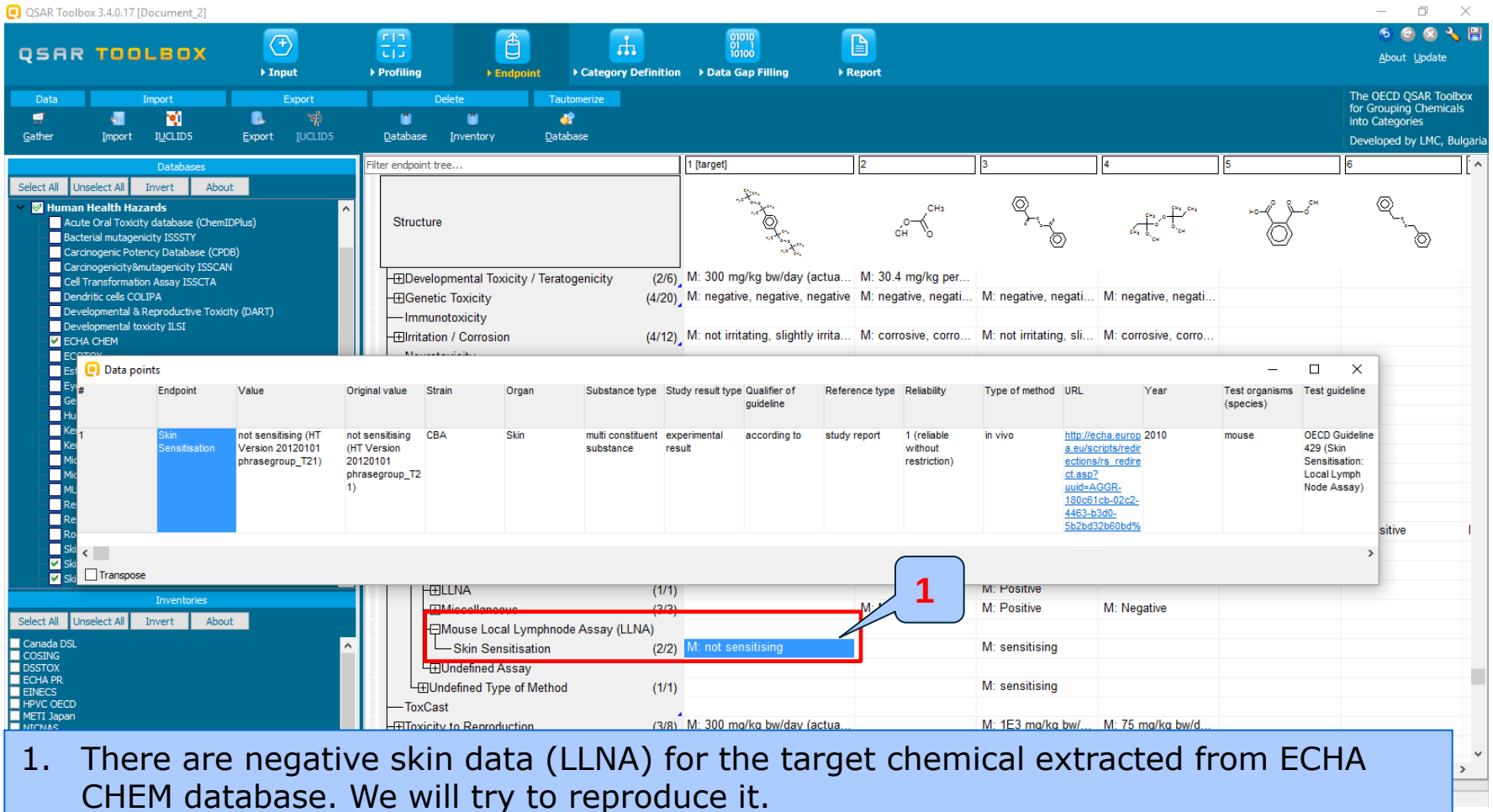
1. **Select** one
2. **Click** OK

Endpoint
Gather data



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

Endpoint
Gather data



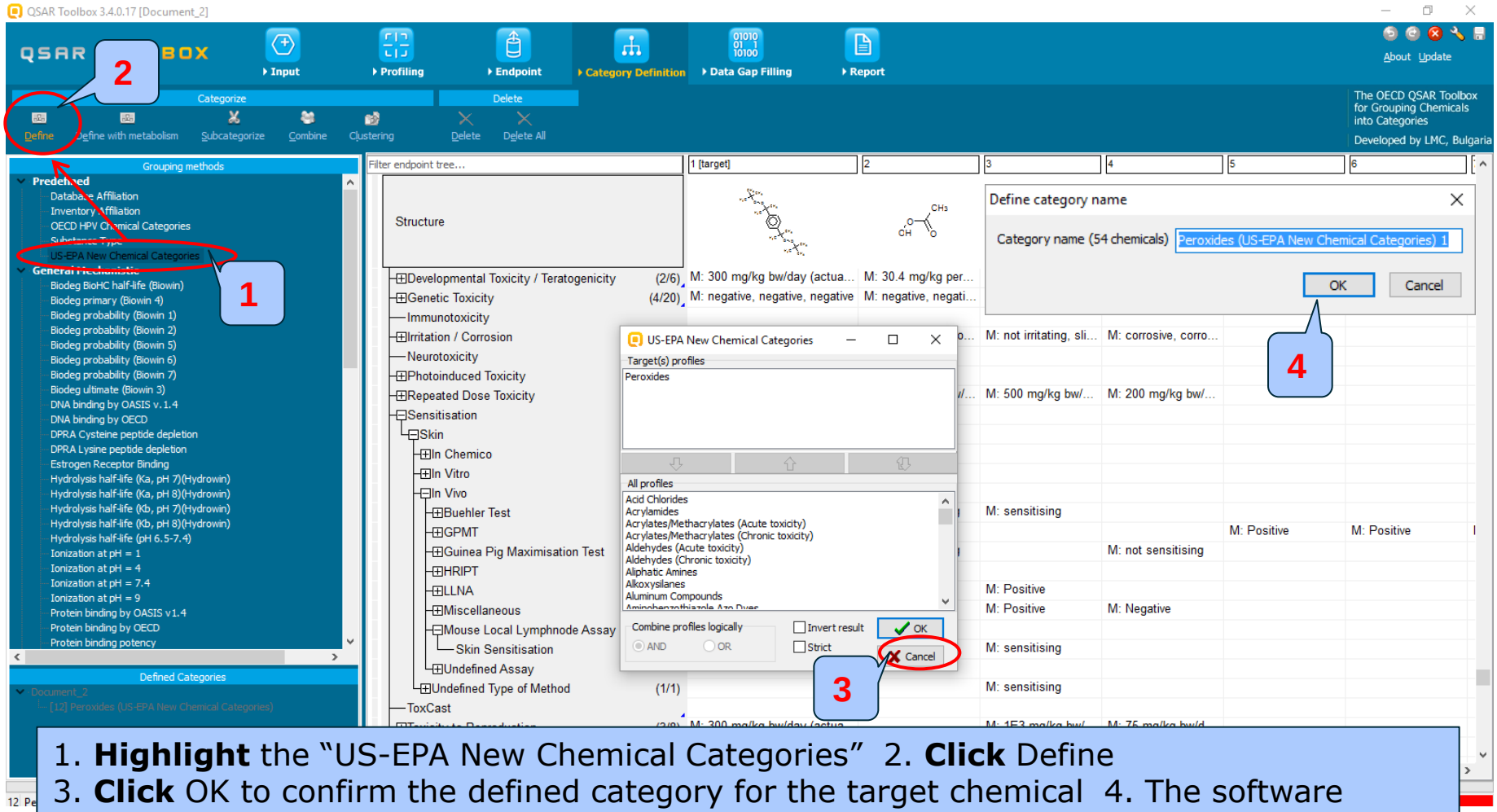
1. There are negative skin data (LLNA) for the target chemical extracted from ECHA CHEM database. We will try to reproduce it.

Expand the defined category

- The next step of the exercise is to identify new analogues. The procedure of defining the category should be repeated in a same manner because the software should search for analogues within the newly selected ECHA CHEM database

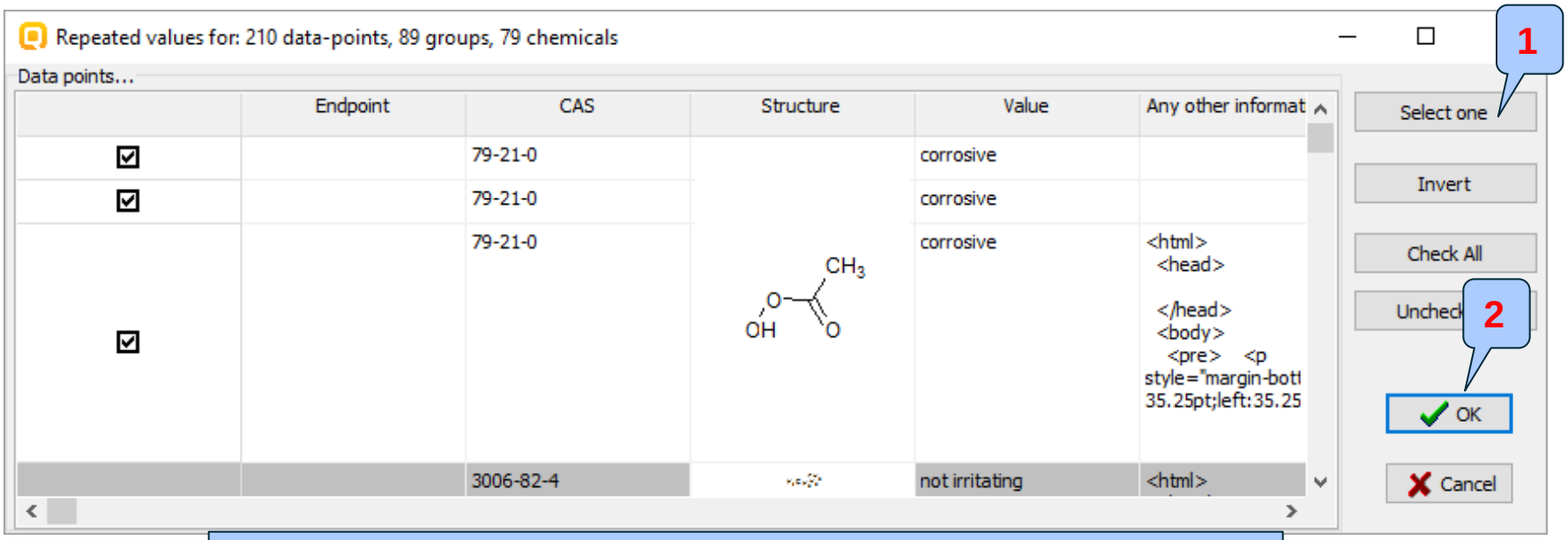
Category Definition

Defining US-EPA New Chemical categories



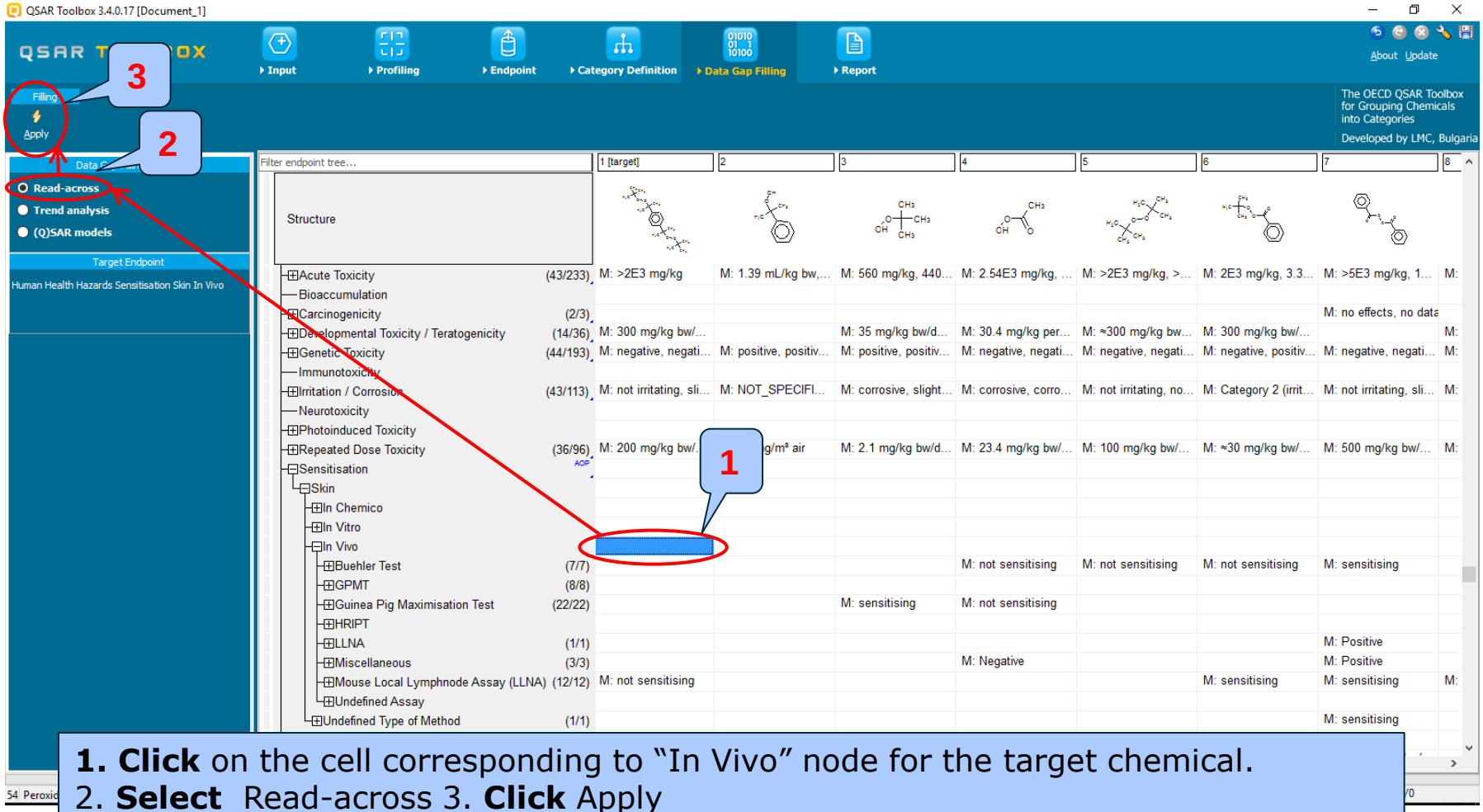
1. **Highlight** the "US-EPA New Chemical Categories" 2. **Click** Define
3. **Click** OK to confirm the defined category for the target chemical 4. The software identify 53 analogues **Click** OK

Endpoint
Gather data



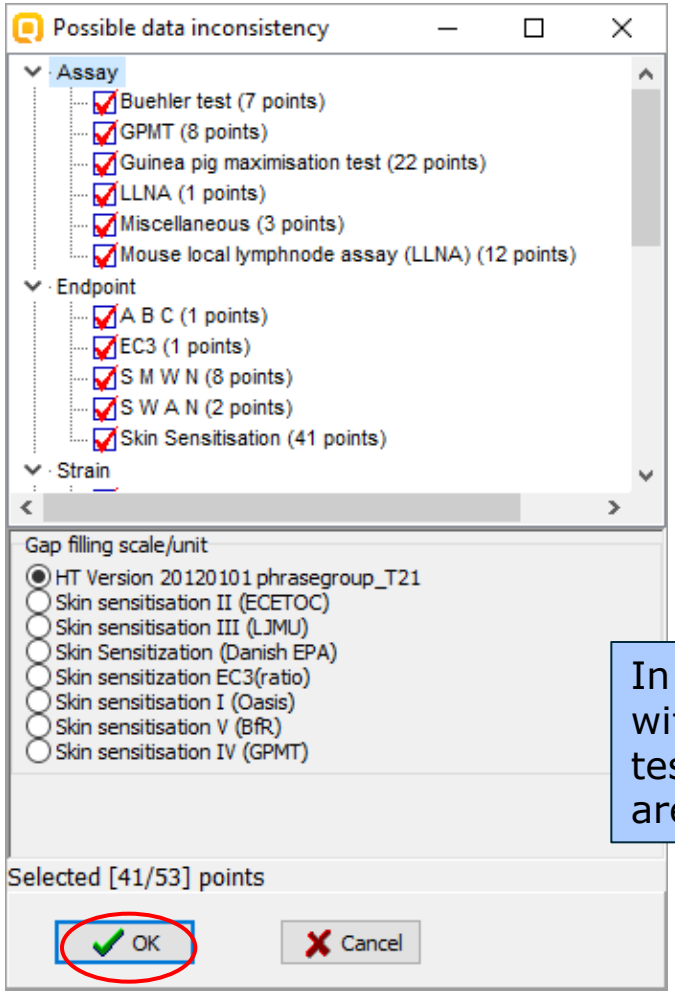
1. **Select** one
2. **Click** OK

Data Gap Filling
Apply Read across



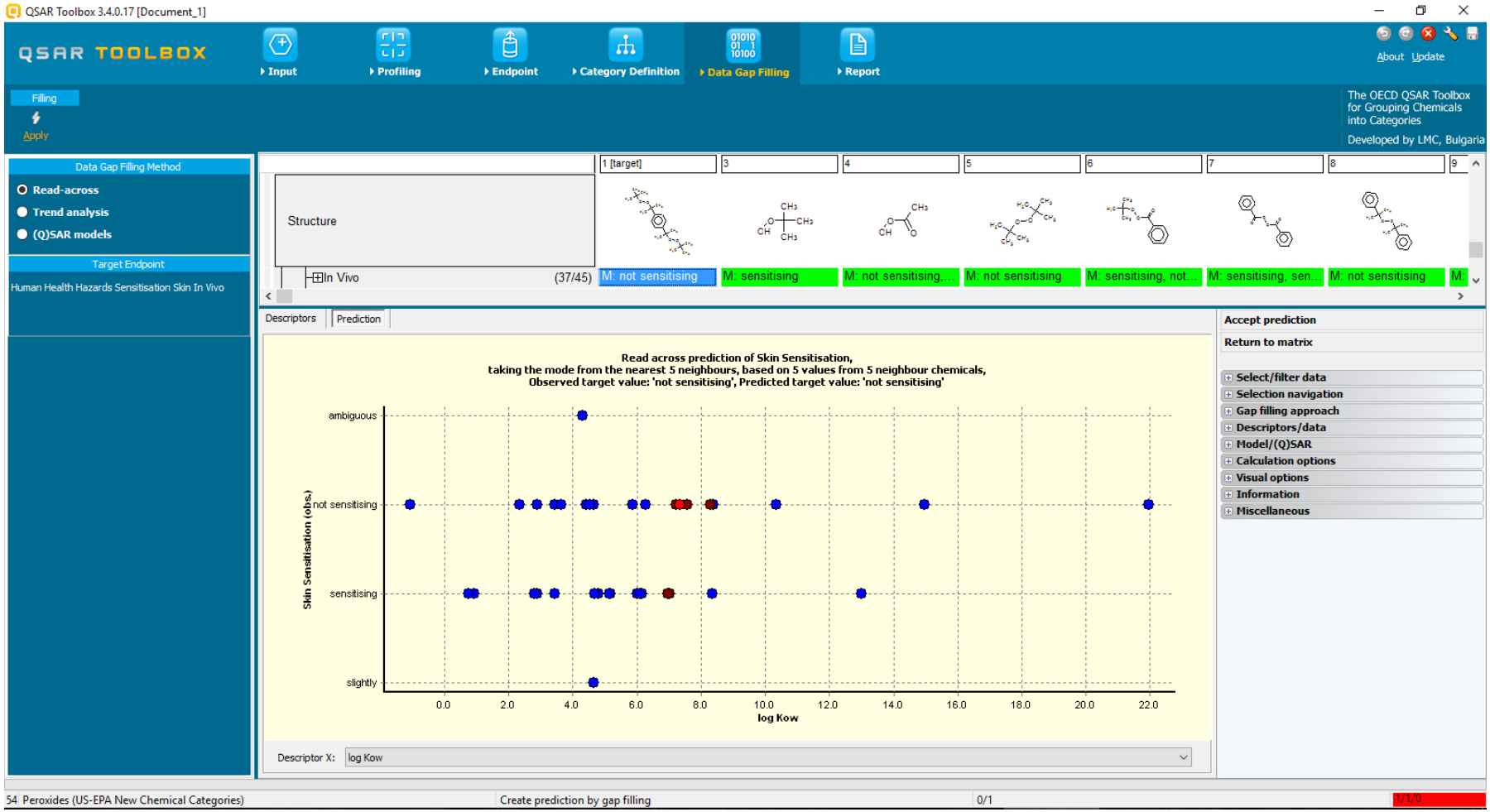
1. Click on the cell corresponding to "In Vivo" node for the target chemical.
2. **Select** Read-across 3. **Click** Apply

Data Gap Filling
Read-across



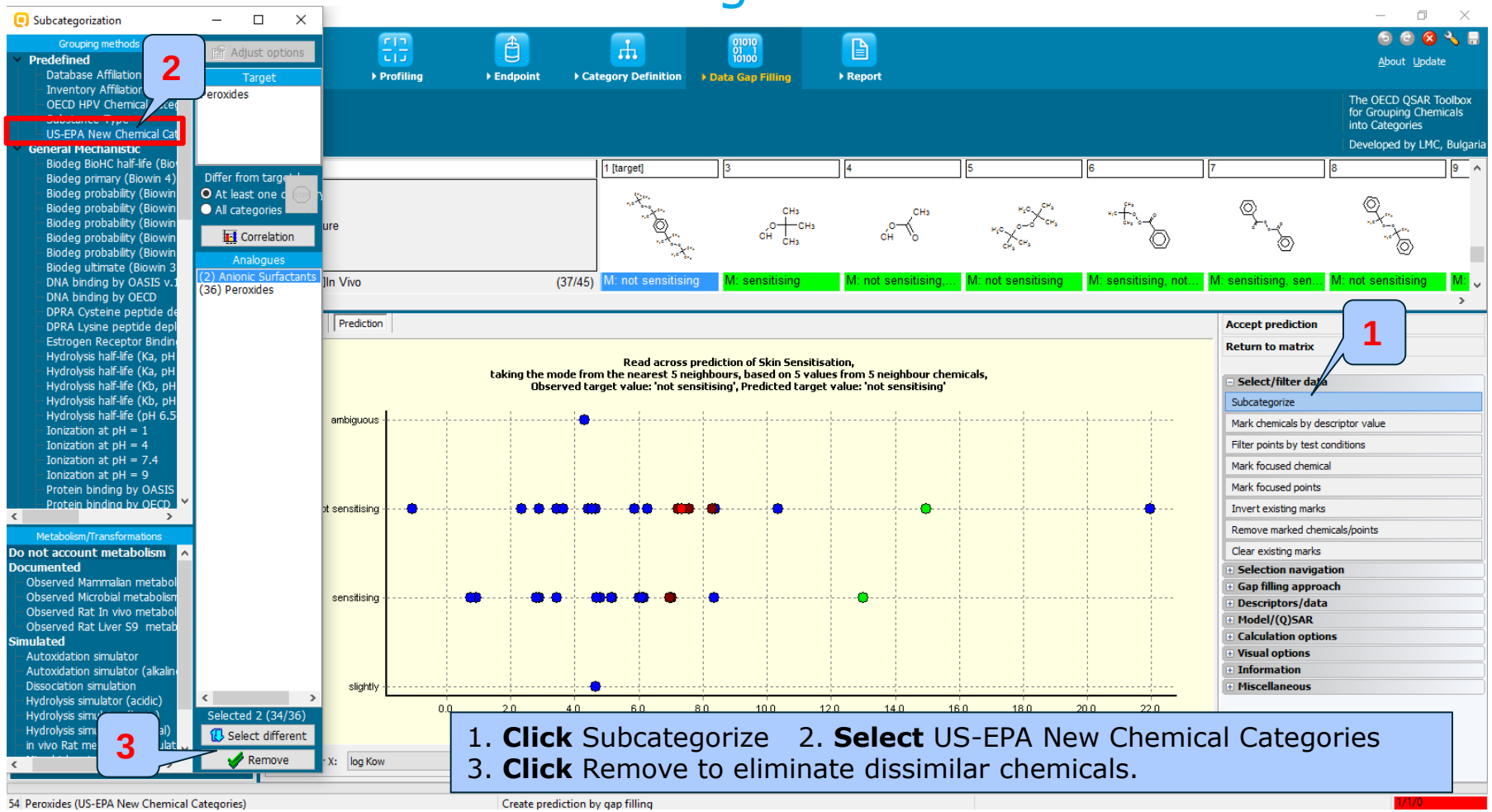
In this case we are mixing data with different strain, species test type. In this exercise we are using "HT" scale.

Data Gap Filling
Read-across



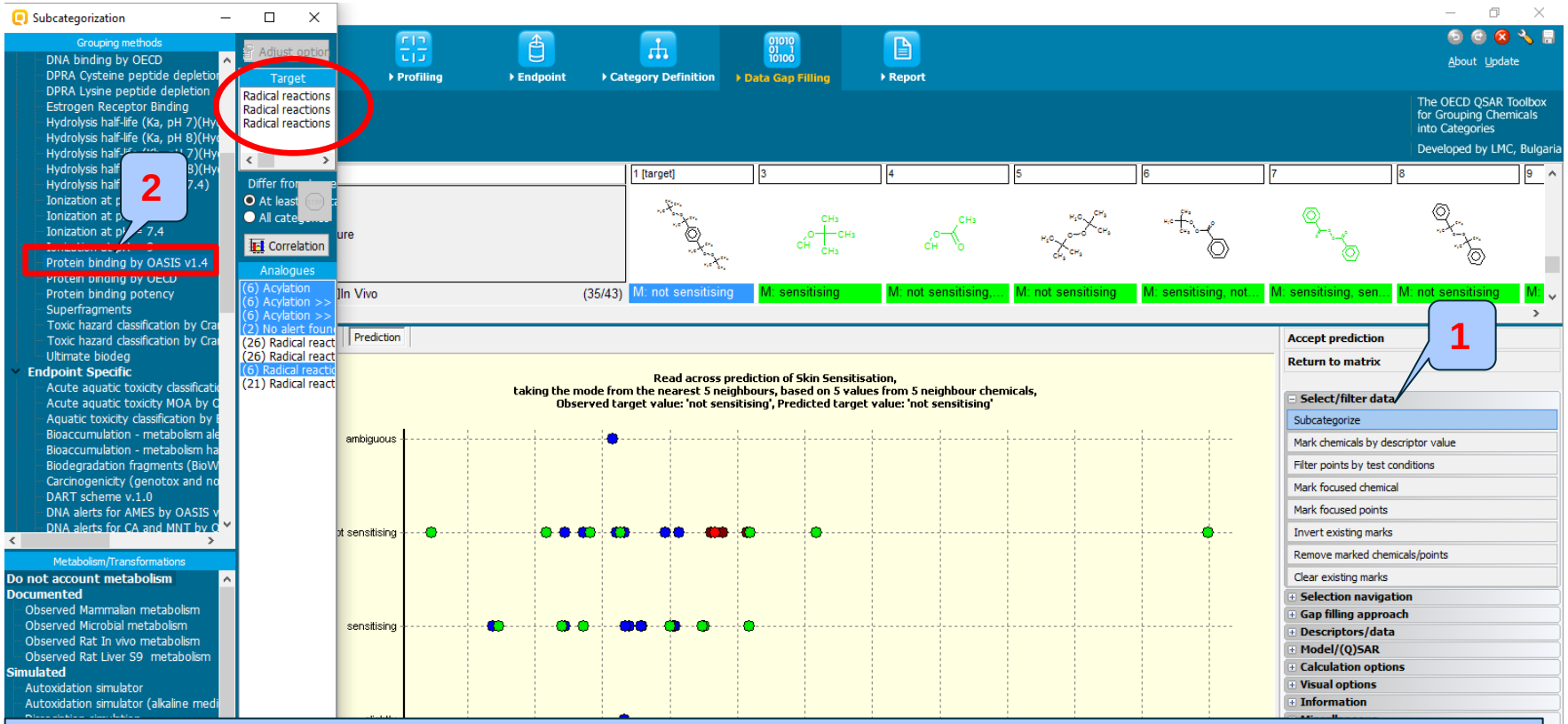
Data Gap Filling

Subcategorization by US-EPA New Chemical Categories



Data Gap Filling

Subcategorization by Protein binding by OASIS v1.4

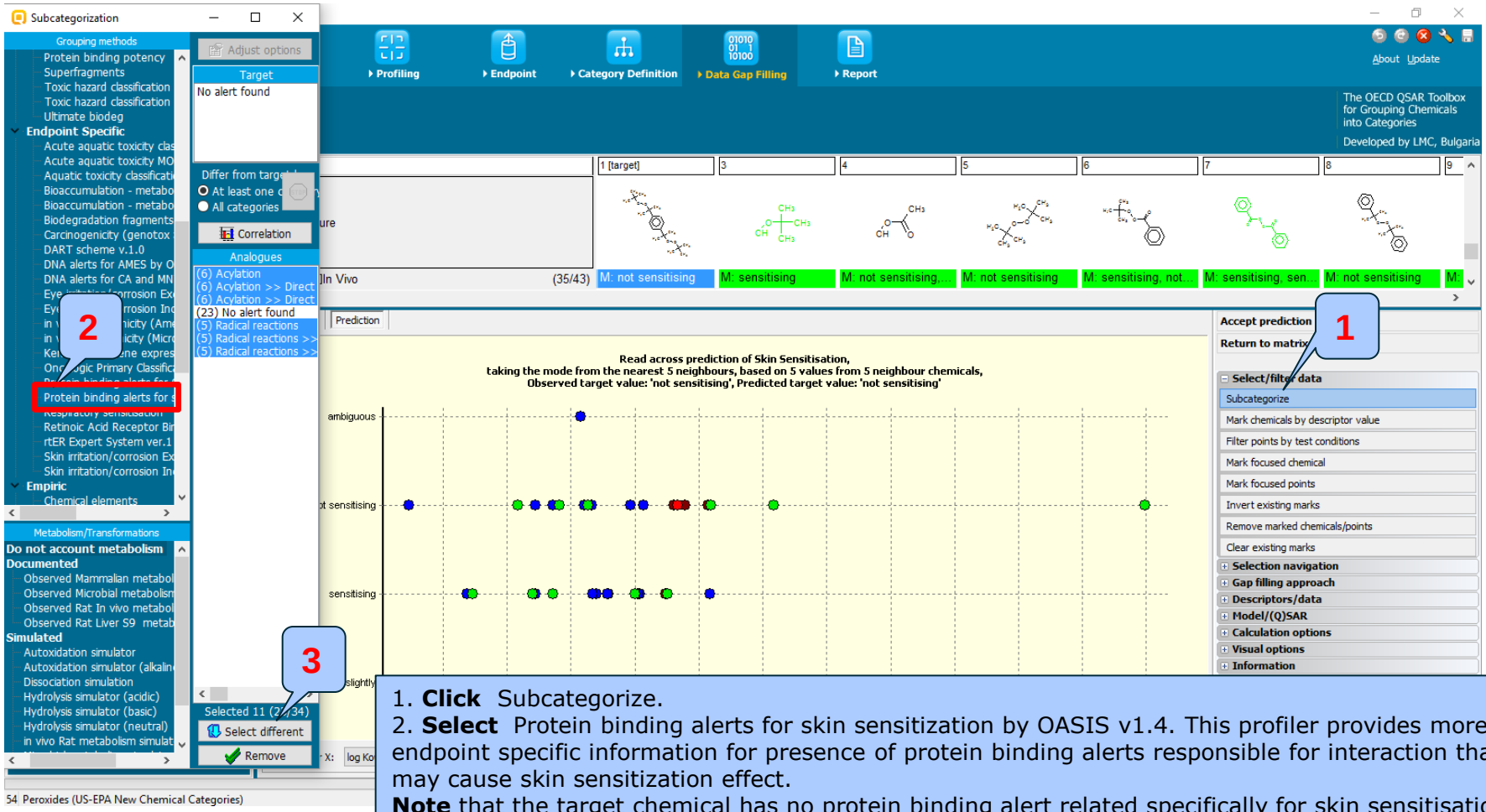


1. **Click** Subcategorize.
2. **Select** general mechanistic profiler Protein binding alerts by OASIS v1.4. This profiler provides information for presence of alert responsible for protein binding interaction independent from a particular endpoint.

Note that the target and the chemicals within the category have a general mechanistic alert for Protein Binding interaction indicating that sub-categorization with an endpoint specific profiler could be more appropriate (see next slide).

Data Gap Filling

Subcategorization by Protein binding alerts for skin sensitization by OASIS v1.4



1. **Click** Subcategorize.
2. **Select** Protein binding alerts for skin sensitization by OASIS v1.4. This profiler provides more endpoint specific information for presence of protein binding alerts responsible for interaction that may cause skin sensitization effect.

Note that the target chemical has no protein binding alert related specifically for skin sensitisation and sub-categorization was performed eliminating dissimilar chemicals (Click Remove, step 3)

Data Gap Filling

Interpretation of Protein binding result obtained by both protein binding alerts related to Peroxides category

- Positive protein binding alert (Organic peroxides) has been found for the target chemical by general mechanistic "Protein binding by OASIS v1.4" profile
- No positive Protein binding alert has been found for the target chemical based on endpoint specific "Protein binding alerts for skin sensitization by OASIS v1.4" profile. The reason for this is that organic peroxide could interact with proteins without eliciting skin sensitization effect. Additional information for both protein binding profilers is provided on the next slide
- The obtained analogues after protein binding elimination were not very similar to the target chemical. In this respect subcategorization by OFG (US-EPA) is applied as next subcategorization step (see slide 67)

Description of general mechanistic and endpoint specific Protein binding by OASIS profilers

General mechanistic: Protein binding by OASIS

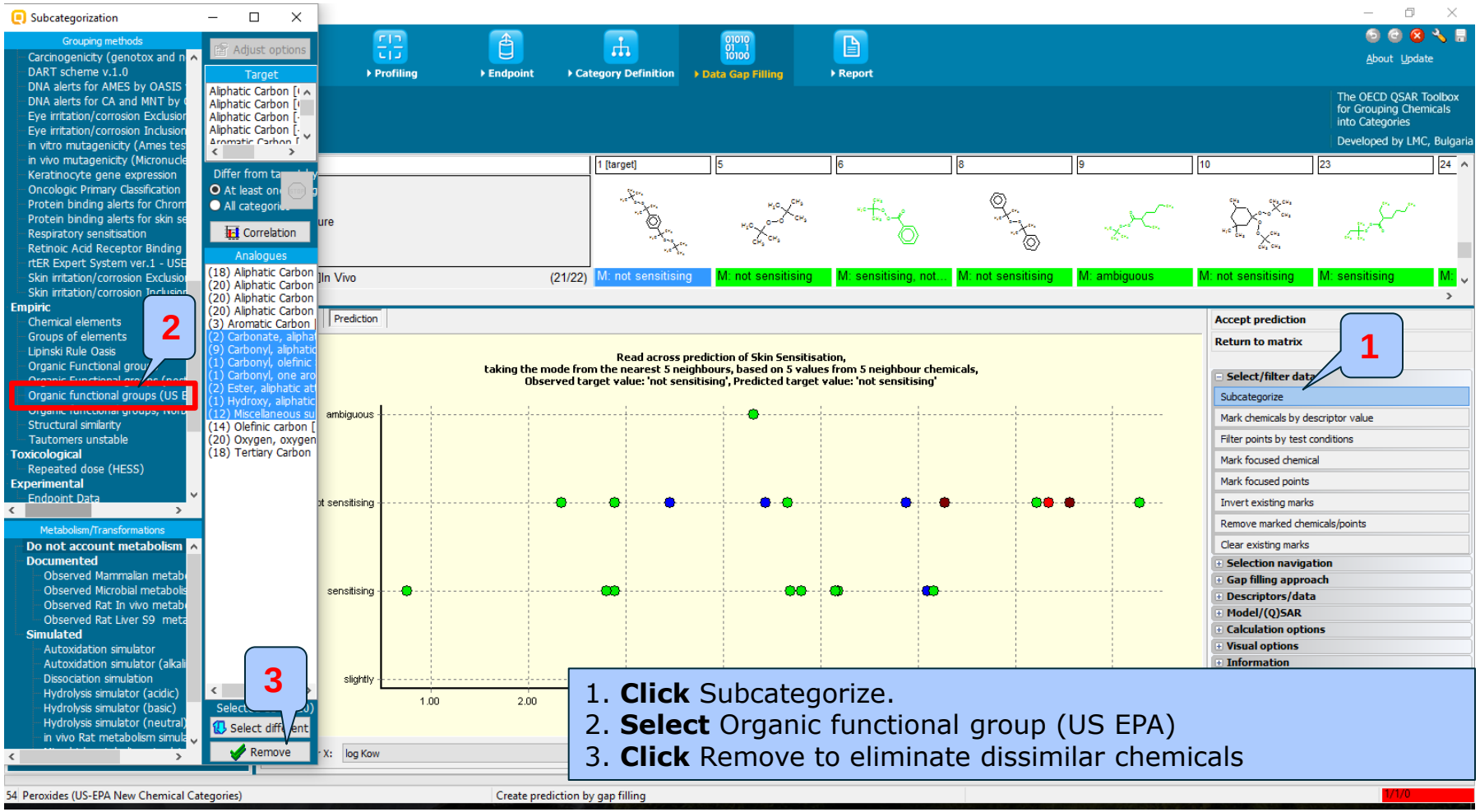
The protein binding alerts have been developed by industry consortia involving ExxonMobil, Procter&Gamble, Unilever, Research Institute for Fragrance Materials (RIFM), Dow and Danish National Food Institute with the Laboratory of Mathematical Chemistry, Bourgas and the partnership of Dr D.Roberts, as a part of the TIMES model to predict skin sensitization. The scope of the profiler is to investigate presence of alerts within target molecules responsible for interaction with proteins.

Endpoint specific: Protein binding alerts for skin sensitization by OASIS

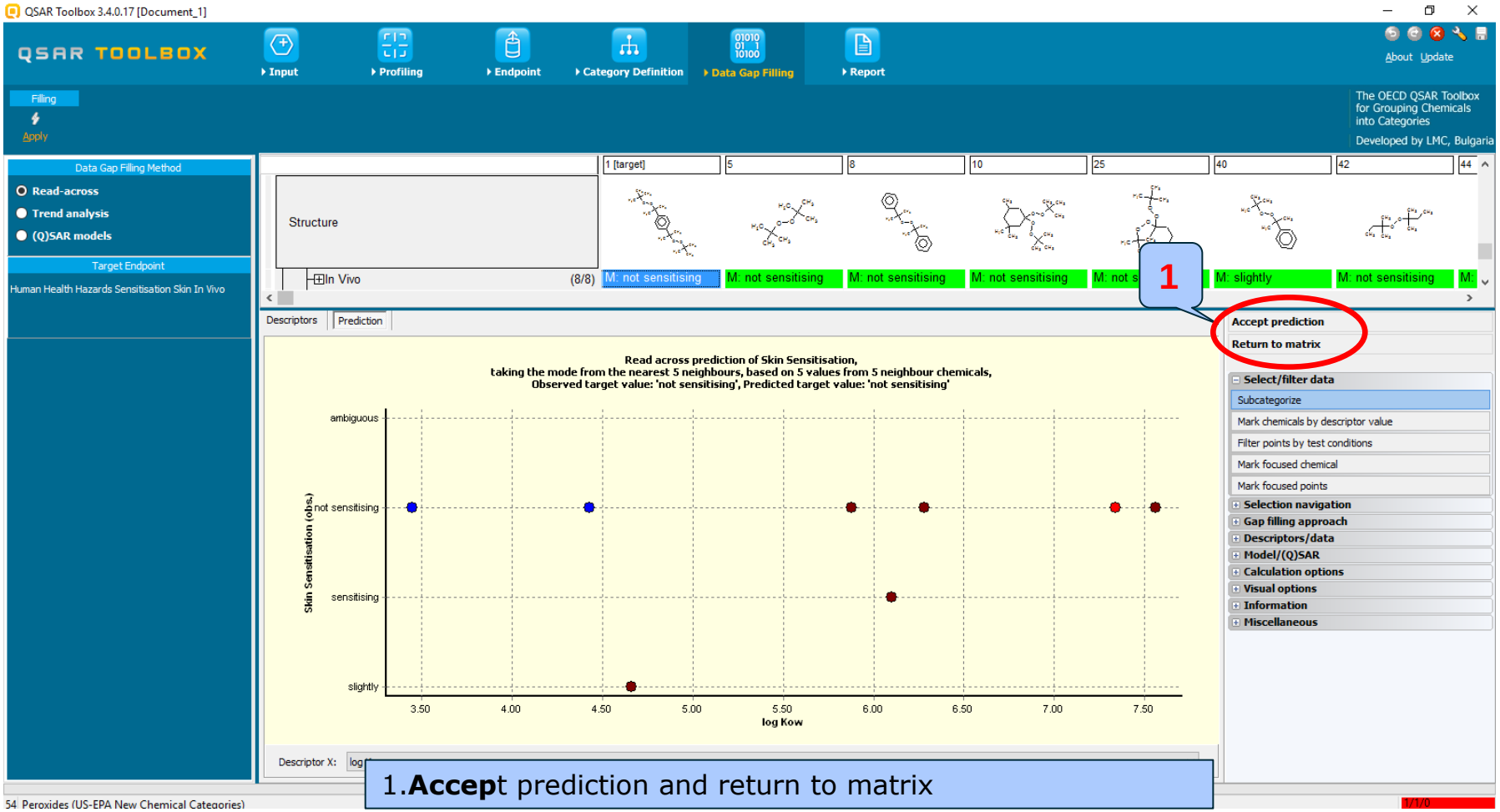
The scope of this profiler is to investigate the presence of alerts within the target molecules responsible for interaction with proteins and especially with skin proteins. This profiler accounts for incapability of some chemicals having an alert to interact with skin due to electronic and steric factors. This is explicitly defined by inhibition masks associated with some alerts.

Data Gap Filling

Subcategorization by OFG (US-EPA)



Data Gap Filling
Read-across



Recap

- Read-across is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitisation. All the tested chemicals in the category except two were not sensitizing, the negative prediction for the target chemical could be accepted.
- You are now ready to complete the final module and to download the report.
- Click on "Report" to proceed to the last module.

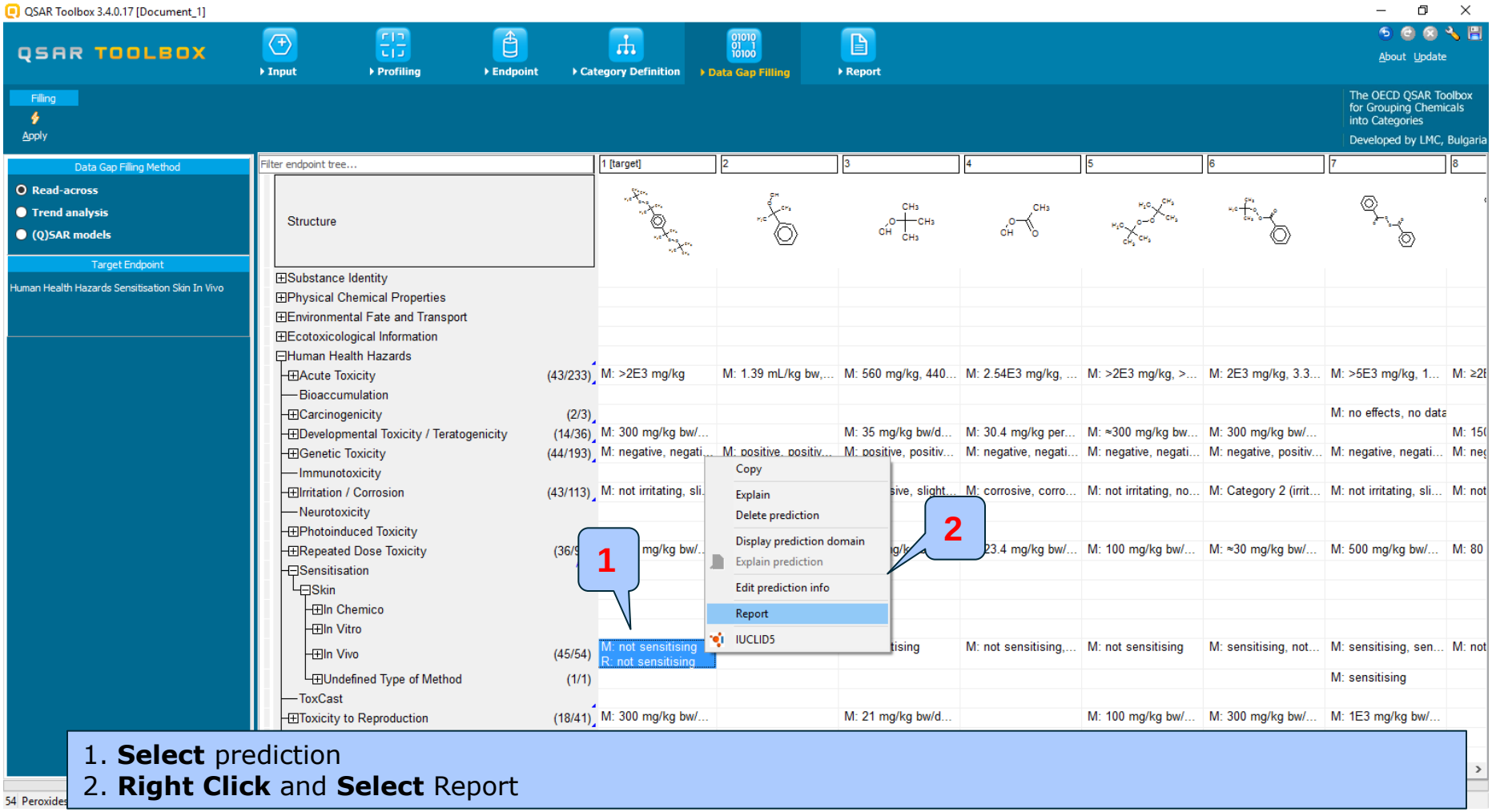
Outlook

- Background
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - Category definition
  - Data Gap Filling
  - **Report**

Report
Overview

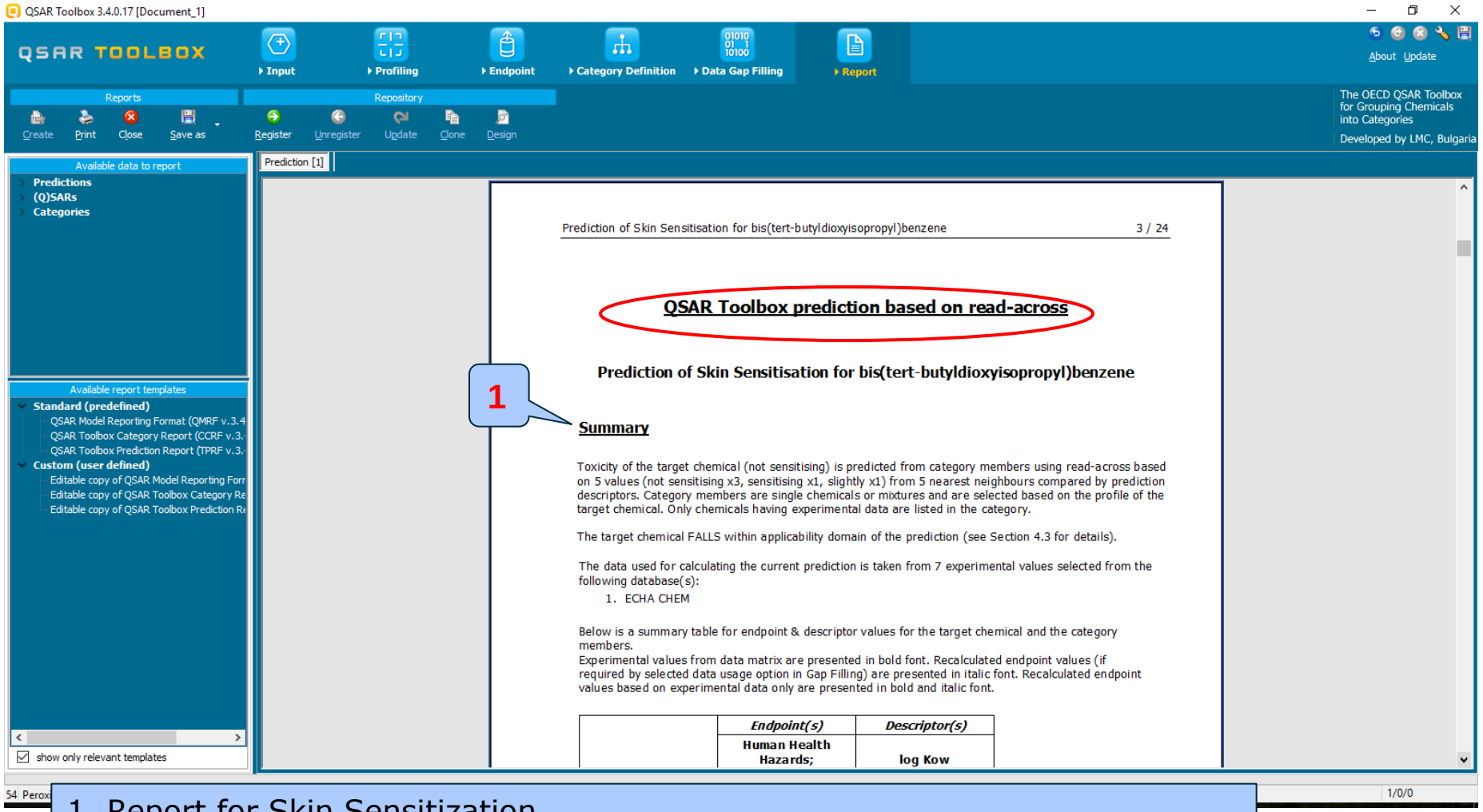
- Report module could generate report on any of predictions performed with the Toolbox.
- Report module contains predefined report templates as well as a template editor with which users can develop a user defined templates.
- The report can then be printed or saved in different formats.

Report



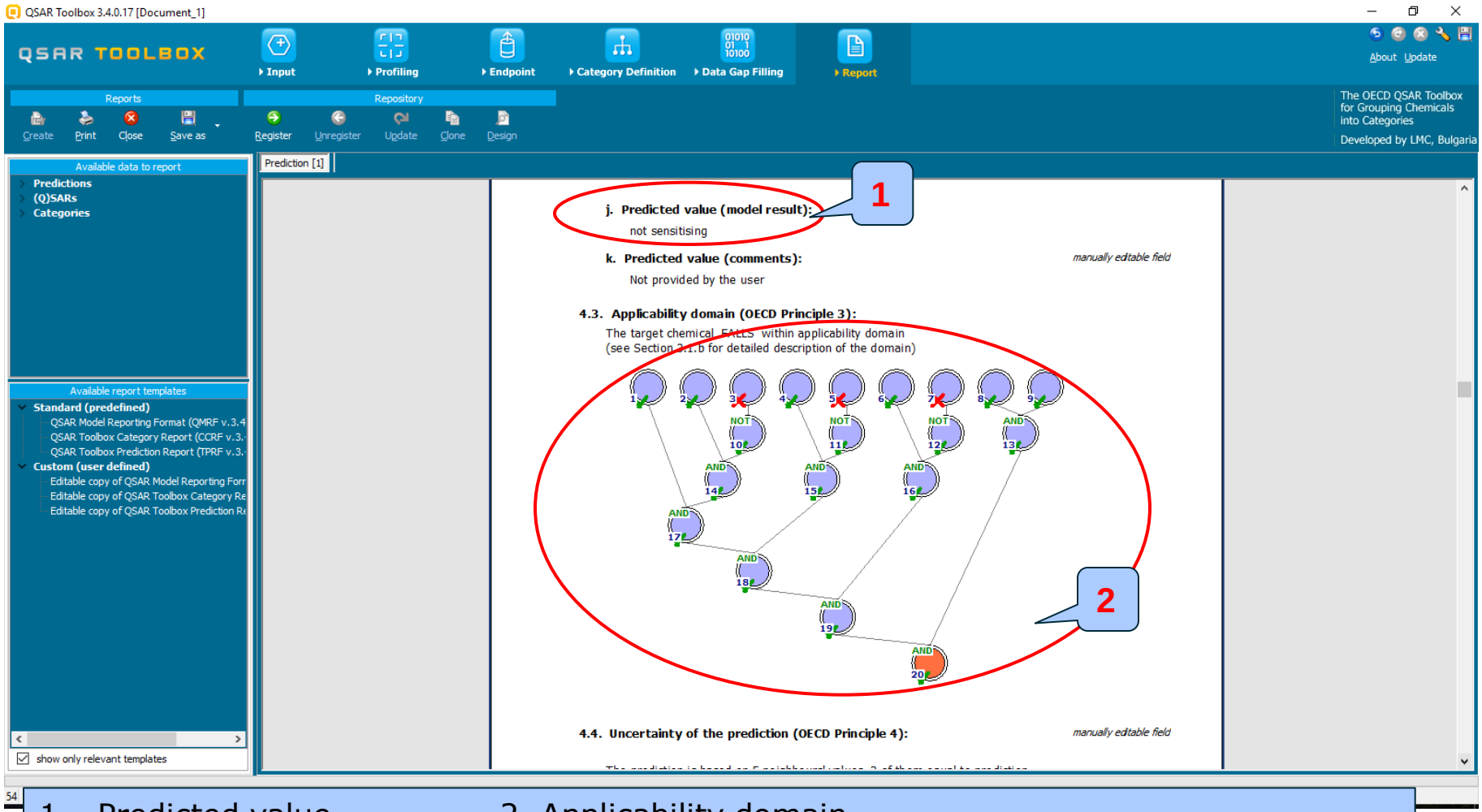
1. **Select** prediction
2. **Right Click** and **Select** Report

Report



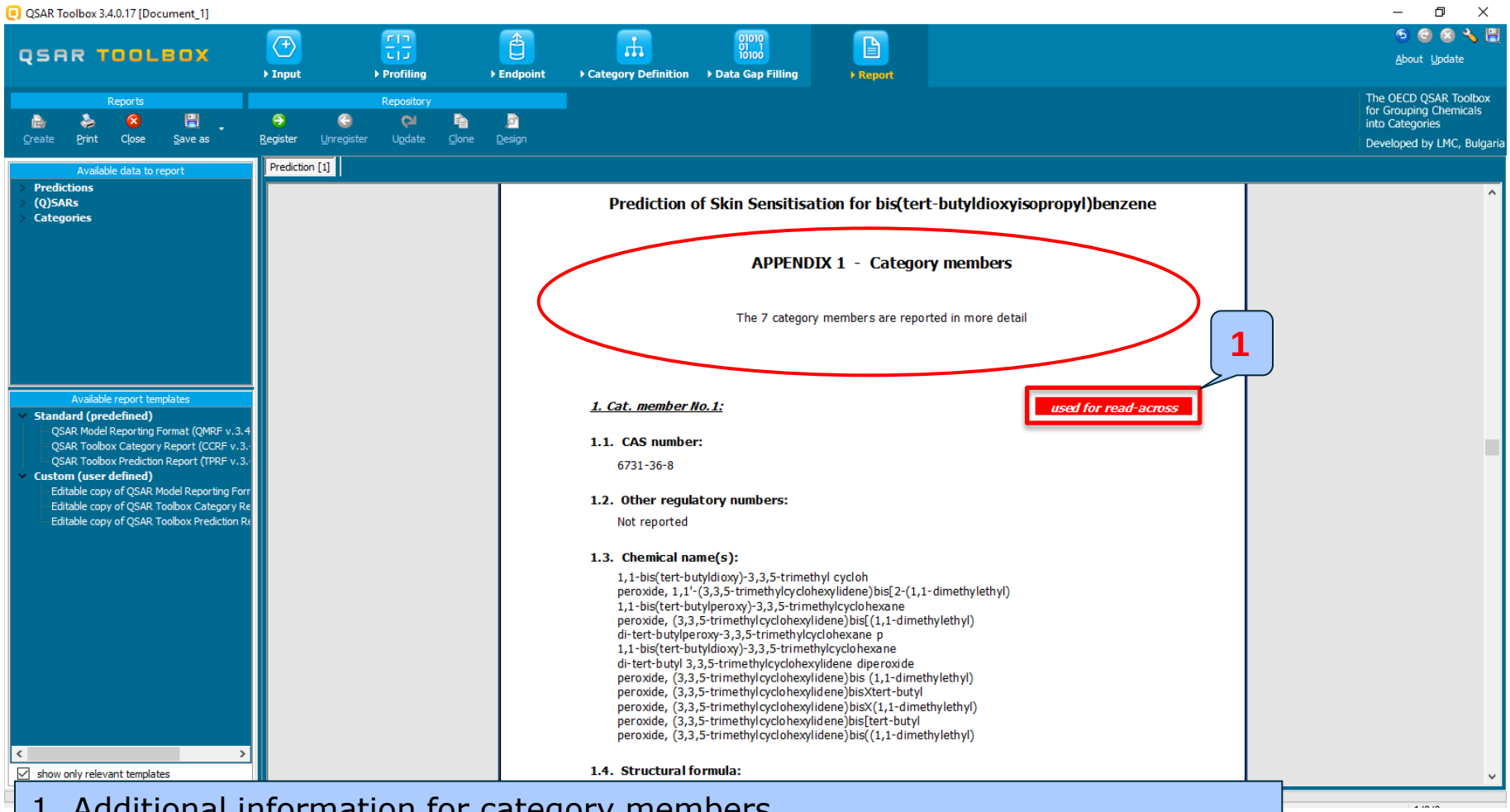
1. Report for Skin Sensitization

Report



1. Predicted value 2. Applicability domain

Report



1. Additional information for category members

Outlook

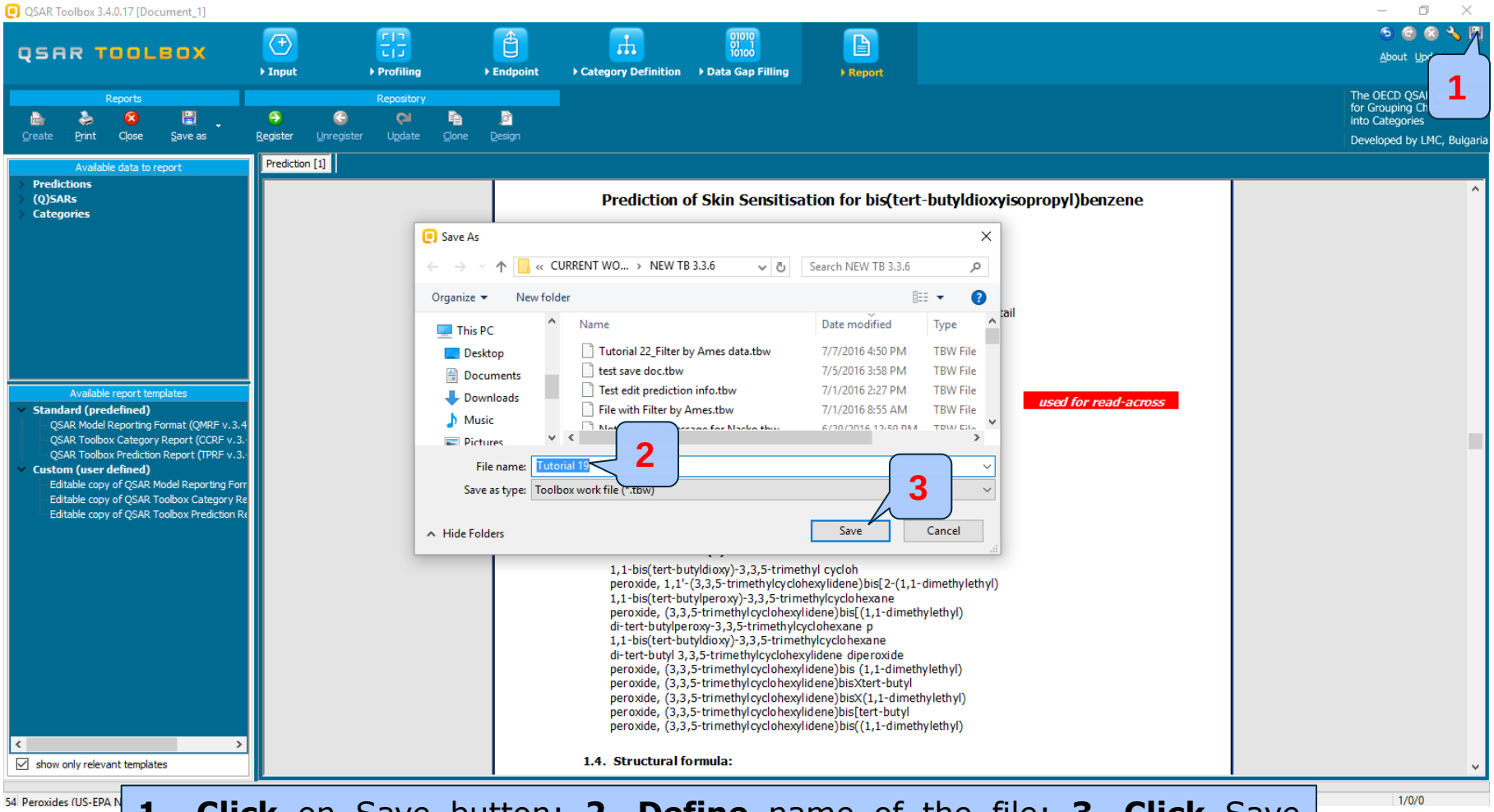
- Background
- The exercise
- Workflow
  - Chemical Input
  - Profiling
  - Endpoint
  - Category definition
  - Data Gap Filling
  - Report
- **Save predictions**

Saving the prediction result

- This functionality allow 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).

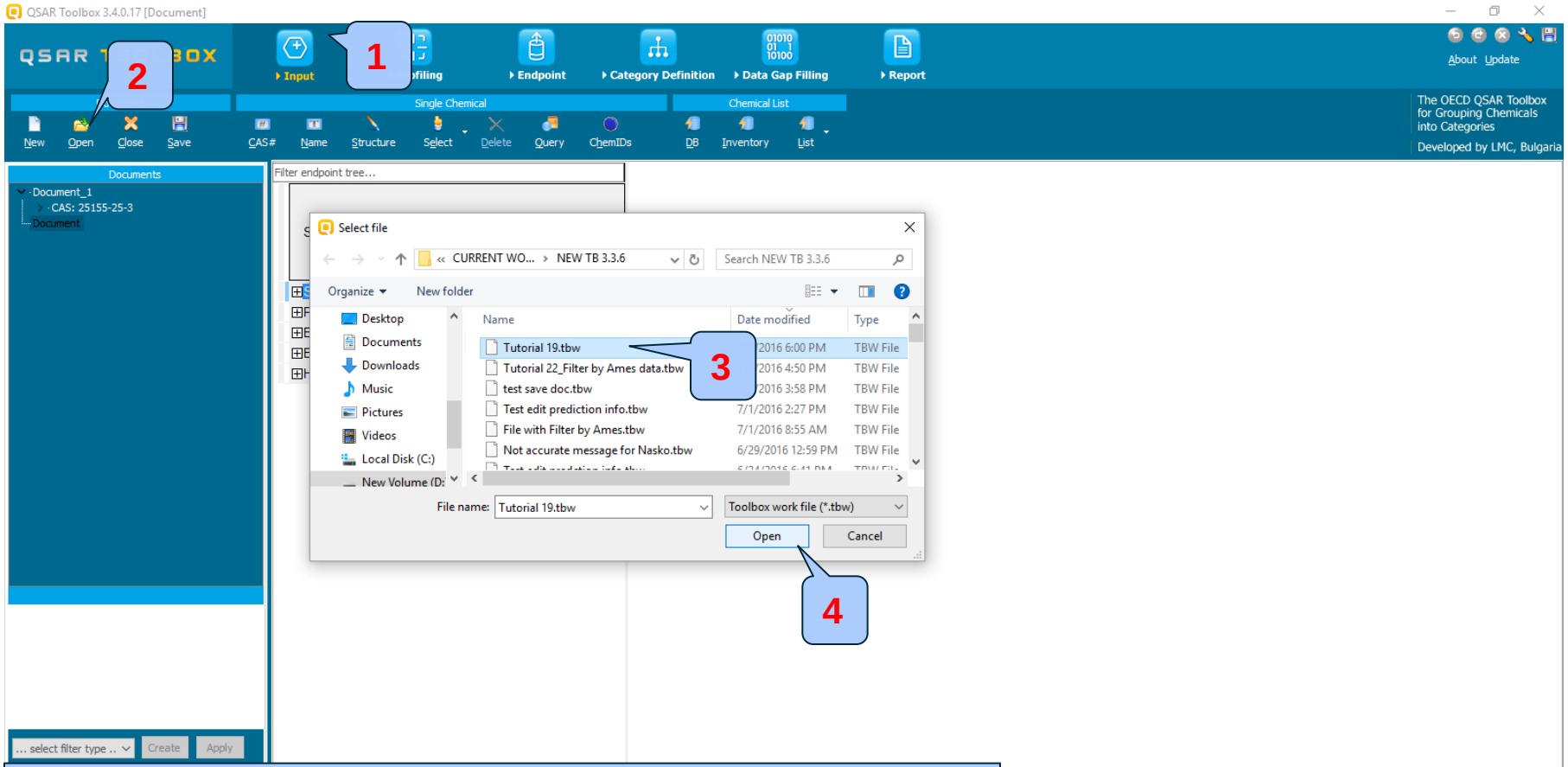
- Saving/Loading the file with TB prediction is shown on next screenshots

Saving the prediction



1. **Click** on Save button; **2.** **Define** name of the file; **3.** **Click** Save button

Open saved file



Once the file has been saved **1. Go** to Input; **2. Click** Open; **3. Find** and **select file**; **4. Click** Open