QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Step-by-step example on how to predict the skin sensitisation potential approach of a chemical by read-across based on an analogue approach

Outlook

- **Background**
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow
- Save the prediction result

Background

- This is a step-by-step presentation designed to take the first-time user of the Toolbox through the workflow of a data filling exercise by read-across based on an analogue approach.

Outlook

- Background
- **Objectives**
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow
- Save the prediction result

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Identify analogues for a target chemical.
- Retrieve experimental results available for those analogues.
- Fill data gaps by read-across.

Outlook

- Background
- Objectives
- **Specific Aims**
- Read across and analogue approach
- The exercise
- Workflow
- Save the prediction result

Specific Aims

- To introduce to the first-time user the workflow of Toolbox.
- To familiarize the first-time user with the six modules of Toolbox.
- To familiarize the first-time user with the basic functionalities within each module.
- To explain to the first-time user the rationale behind each step of the exercise.

Outlook

- Background
- Objectives
- Specific Aims
- **Read-across and analogue approach**
- The exercise
- Workflow
- Save the prediction result

Read-across and Analogue Approach
Overview

- A 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.

Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- **The exercise**
- Workflow
- Save the prediction result

The Exercise

- In this exercise we will predict the skin sensitization potential for an untested compound, (4-nitrobenzoyl chloride) [CAS # 122-04-3], 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 mechanism of protein binding common to all the chemicals in the category.
- The prediction itself will be made by "read-across".

The Exercise
Theoretical considerations on Skin 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 a growing agreement that most organic chemicals must react covalently with skin proteins in order to behave as skin sensitizers.
- Therefore, the mechanisms by which organic chemicals bind with proteins are relevant to grouping chemicals that may be skin sensitizing agents.

Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
- Save the prediction result

Workflow

- **Toolbox has six modules, which are used in a sequential workflow:**
  - Chemical Input
  - Profiling
  - Endpoints
  - Category Definition
  - Filling Data Gaps
  - Report

Outlook

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

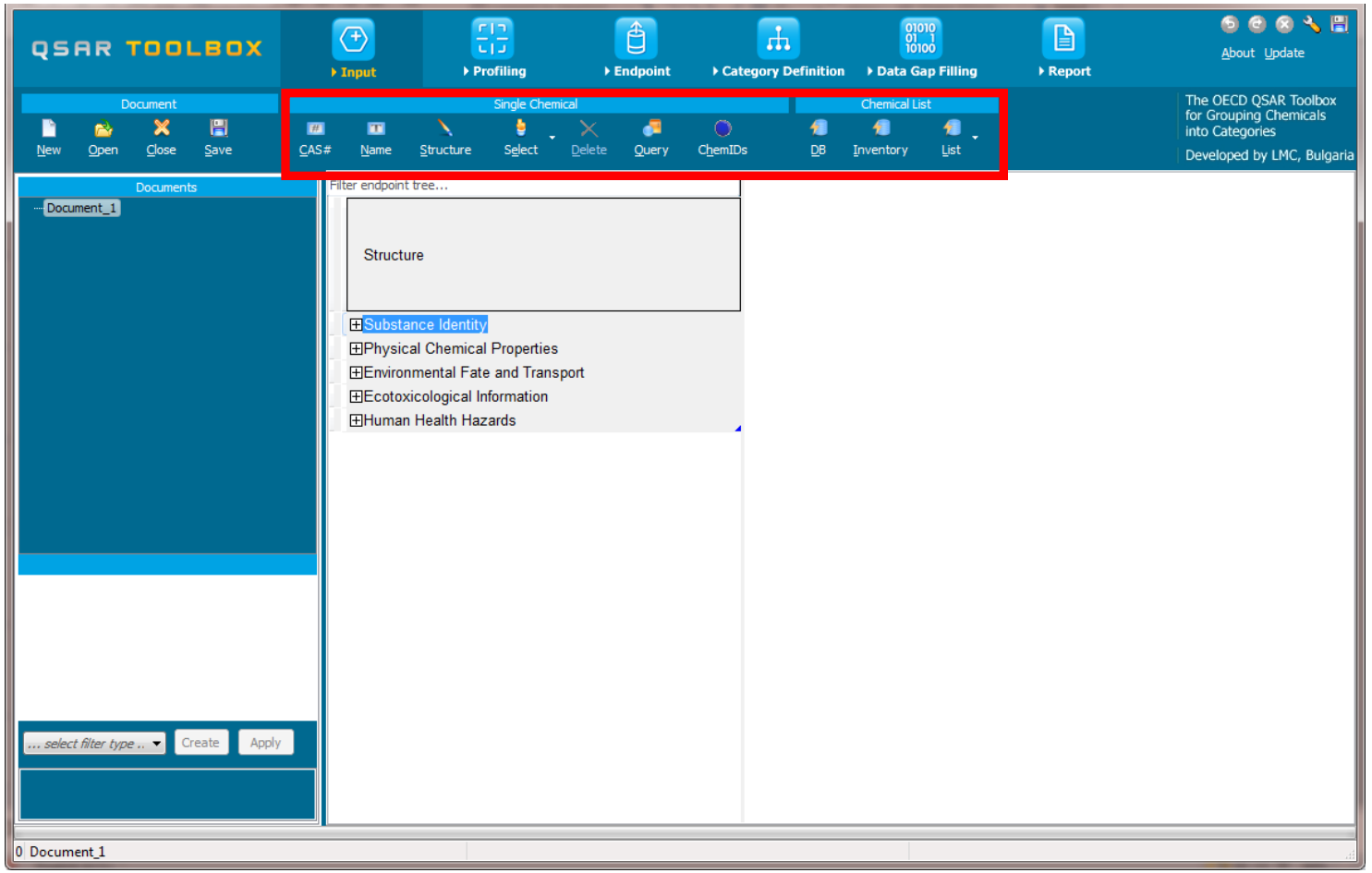
- User List/Inventory
- Specialized Databases

Getting Started

- Open the Toolbox.
- The six modules in the workflow are seen listed next to "QSAR TOOLBOX".
- Click on "Input" (see next screen shot)

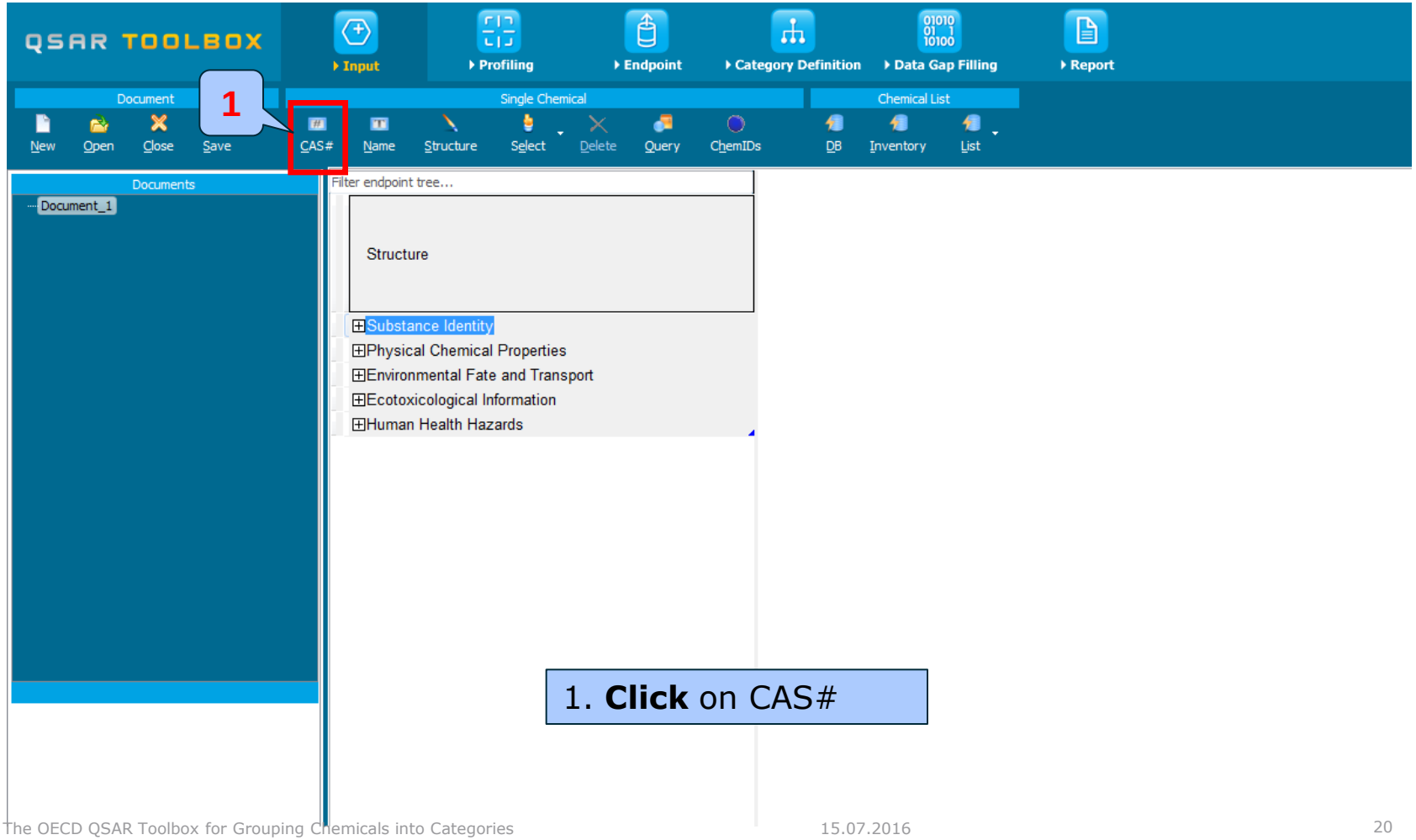
Chemical Input Screen

Input screen



Chemical Input Screen

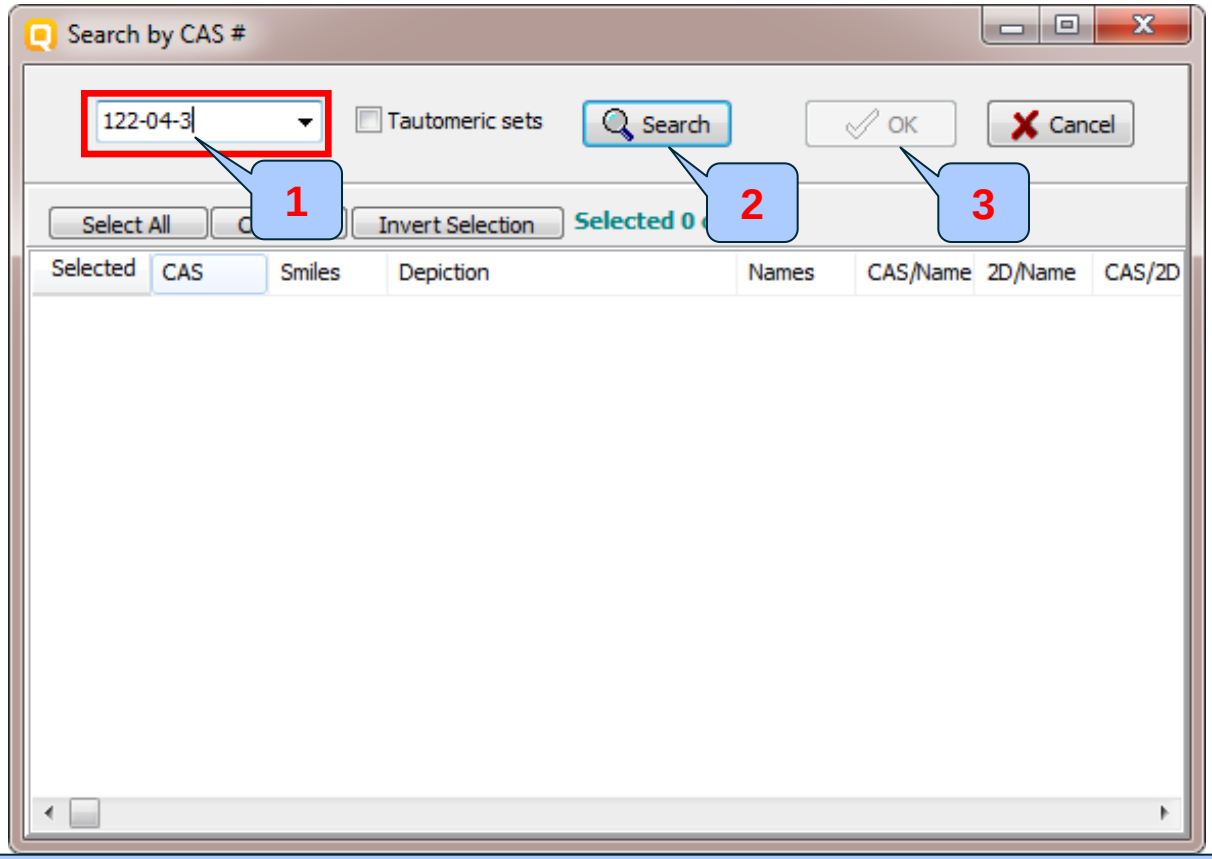
Input target chemical by CAS#



1. **Click** on CAS#

Chemical Input Screen

Enter CAS# of 4-nitrobenzoyl chloride

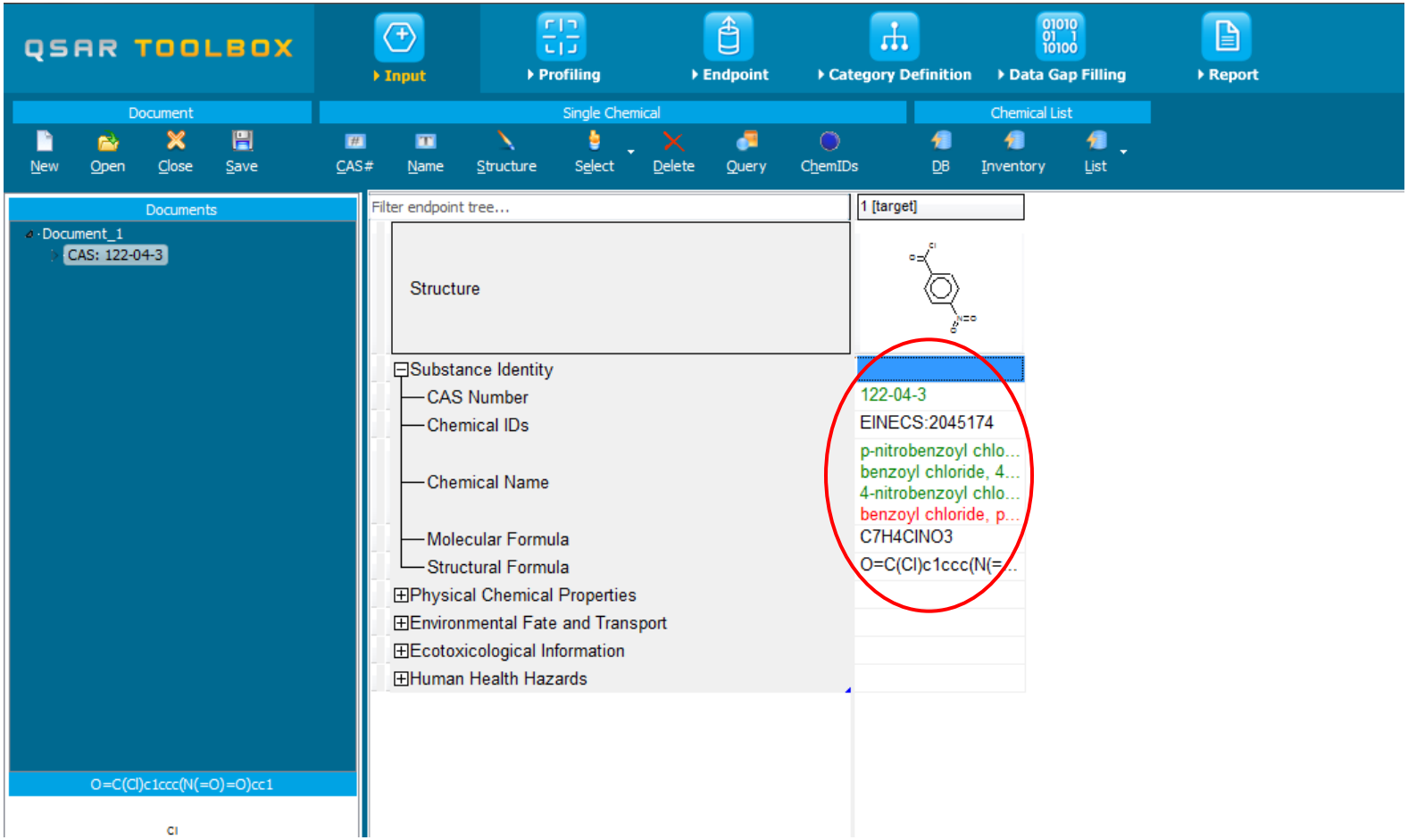


1. **Enter** the CAS# In the field; 2. **Click** Search button; 3. **Press** OK

Chemical Input
Target chemical identity

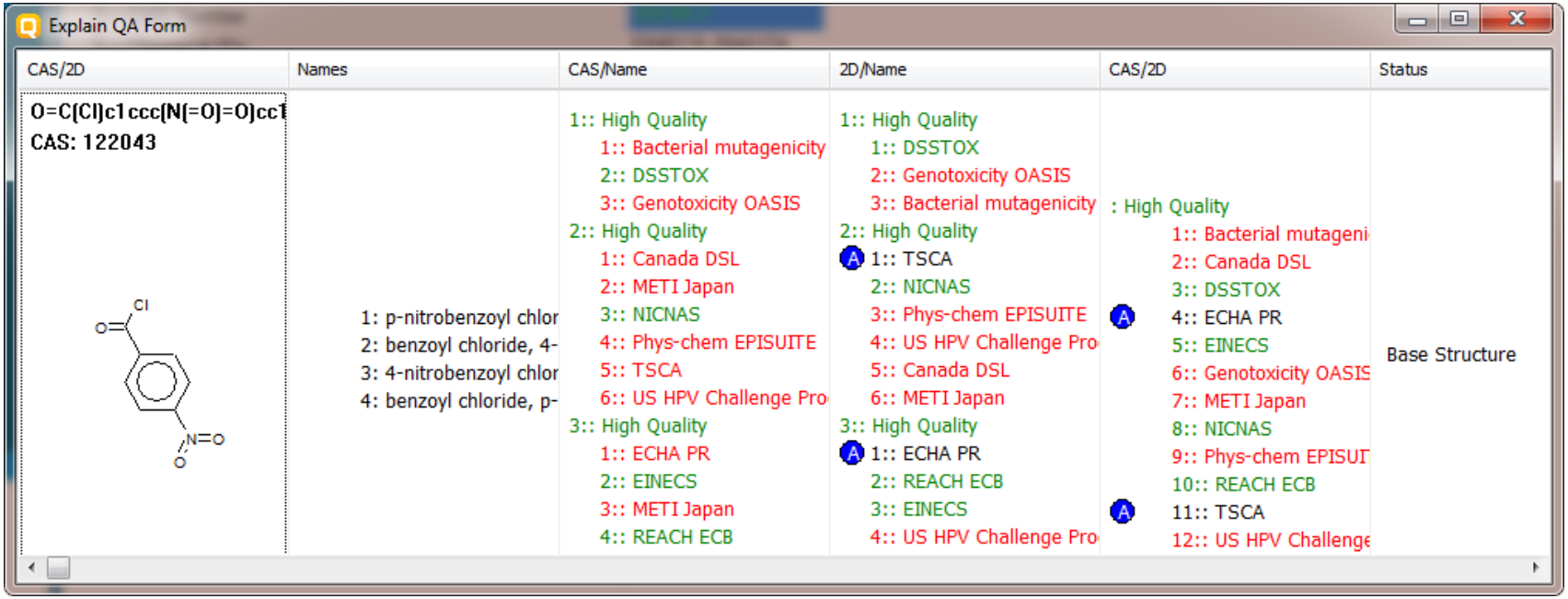
- Double click "Substance Identity"; this displays the chemical identification information.
- Note that existing in the Toolbox name of target chemical are in different colours (see next screen shot).
- The workflow on the first module is now complete, and the user can proceed to the next module.

Chemical Input
Target chemical identity



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
Chemical identity

The colour code indicates the reliability of the chemical identifier:

- **Green**: There is a high reliability 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

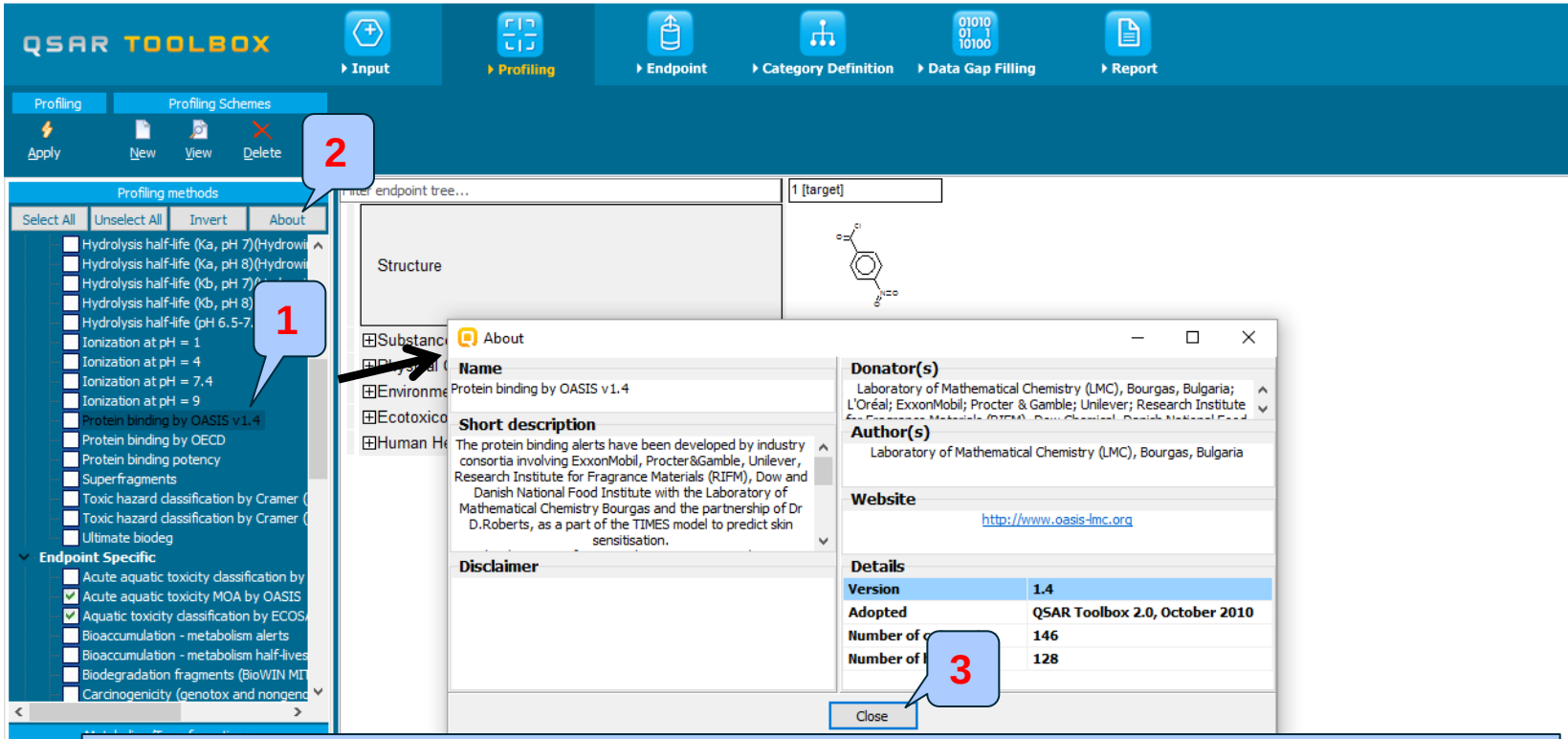
- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical Input
  - **Profiling**

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
Overview

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

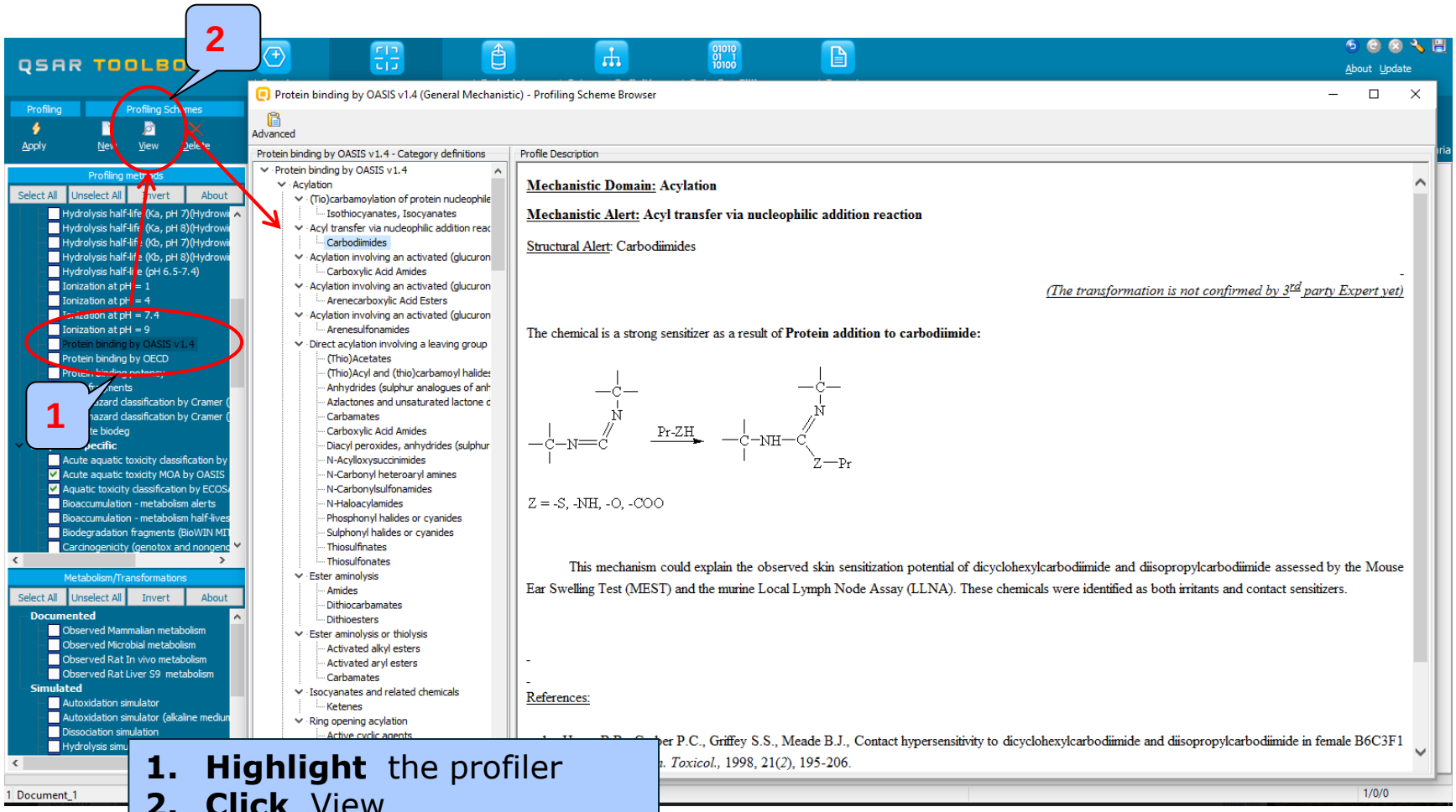


1. **Select** the name of the profiler, perform **right click** on it and then
2. **Select** About
3. **Close** before proceeding

Profiling
Overview

- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, Protein binding by OASIS) and clicking on "View" (see next screen shot).

Profiling
Overview



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

Profiling
Profiling the target chemical

- **Tick the** box of the selected profiling methods related to the target endpoint.

- This selects (a **green** check mark appears) or deselects (**green** check mark disappears) profilers.

- For this example, **tick** all the general mechanistic profilers and **click** on apply (see next screen shot).

Profiling

Profiling the target chemical: Example

- **Tick the** box of the selected profiling methods related to the target endpoint.

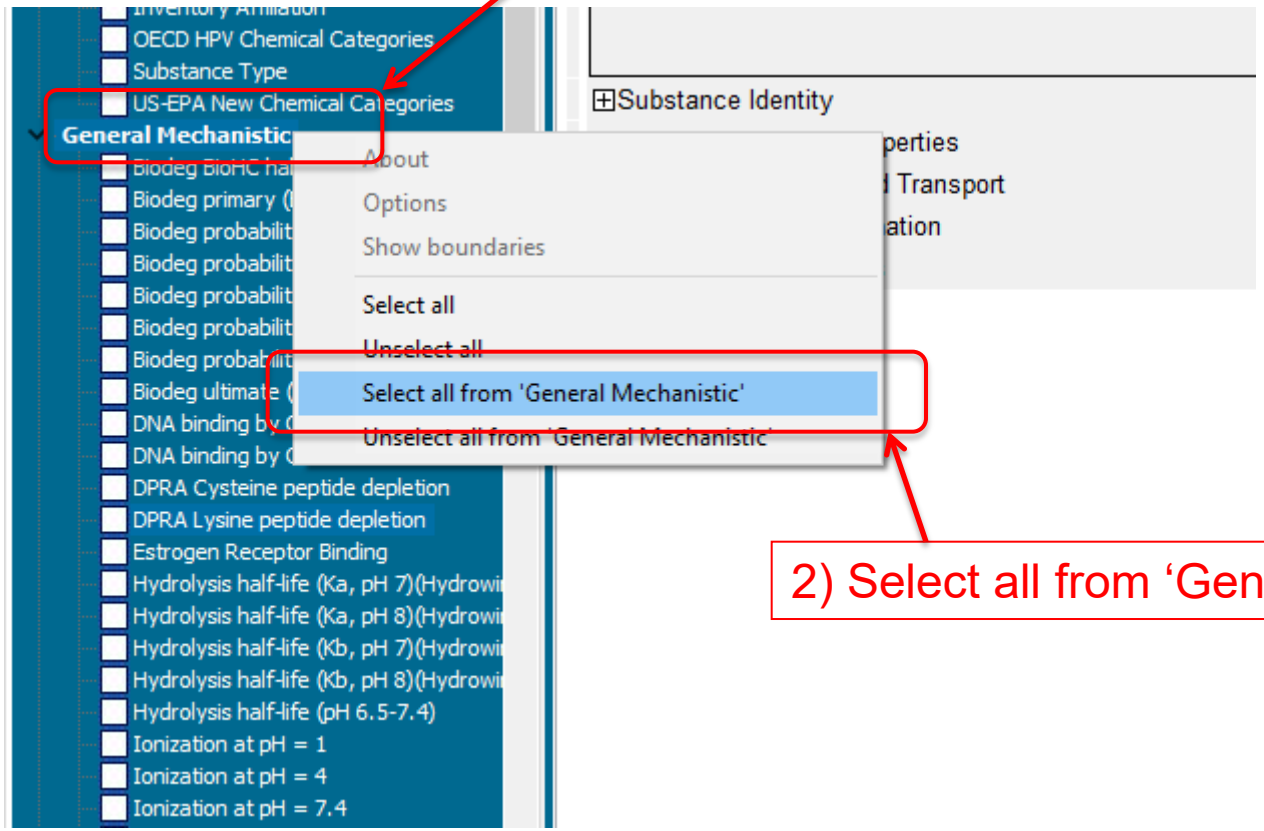
- This selects (a **green** check mark appears) or deselects (**green** check mark disappears) profilers.

- For this example, **tick** all the general mechanistic profilers and **click** on apply (see next screen shot).

Profiling

Profiling the target chemical: Example

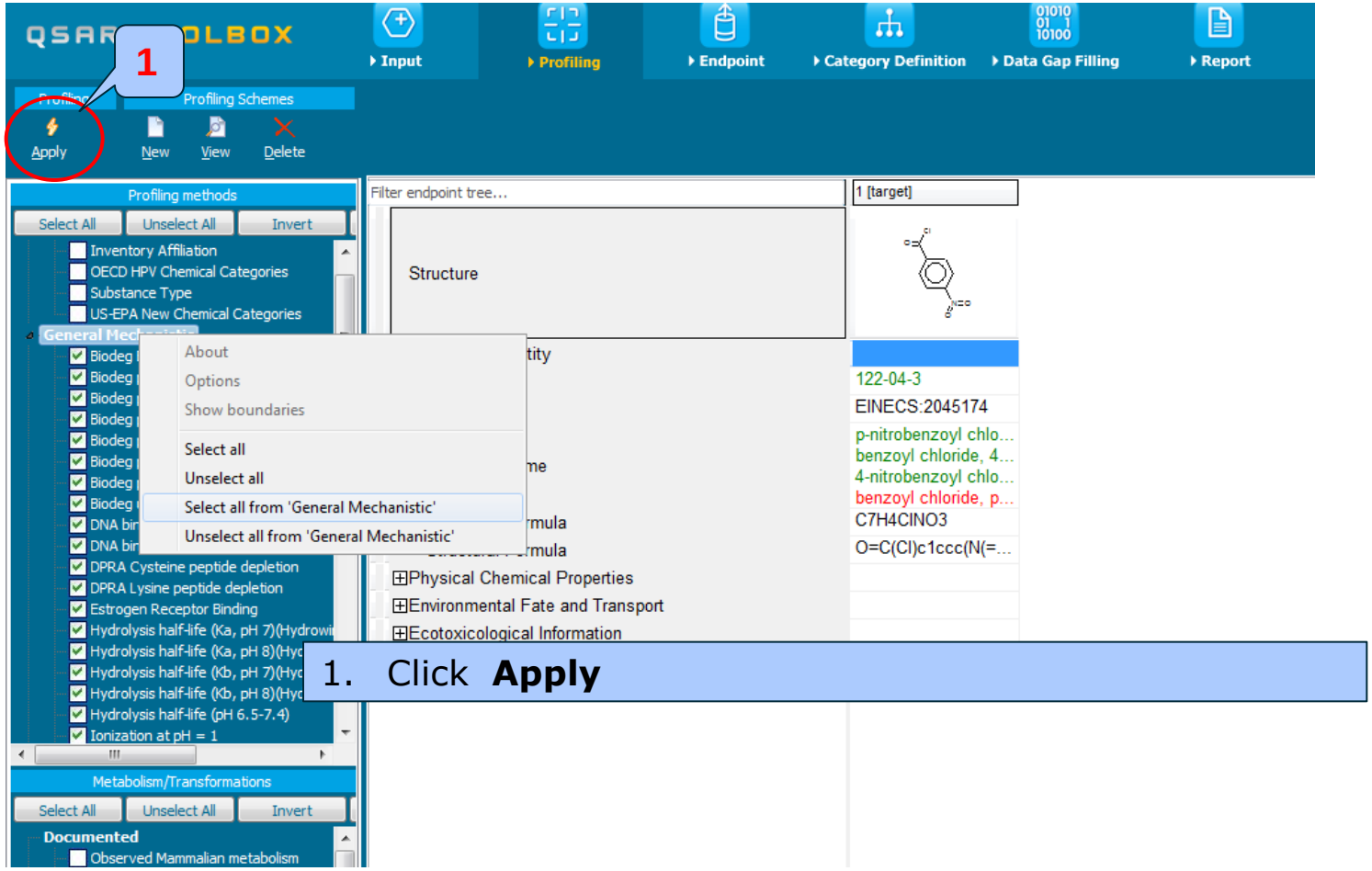
1) Right click on General mechanistic



2) Select all from ‘General Mechanistic’

Profiling

Profiling the target chemical

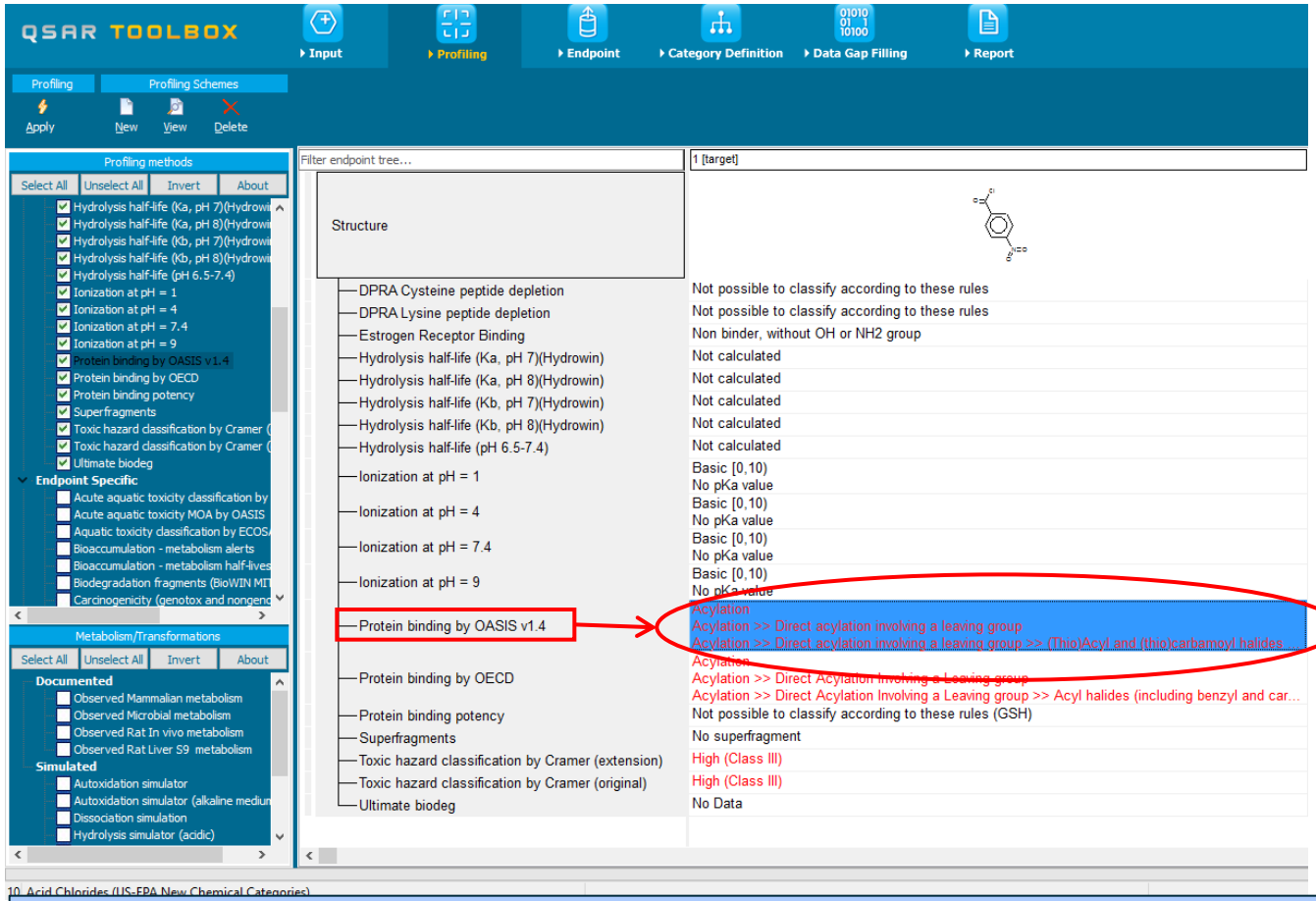


Profiling
Profiling the target chemical

- 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 specific protein-binding profiler – Protein binding by OASIS (see side-bar on sensitisation above).
- This result will be used to search for suitable analogues in the next steps of the exercise.

Profiling

Profiling the target chemical



1. Go to Protein binding by OASIS v1.4 to review the profiling results.

Outlook

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

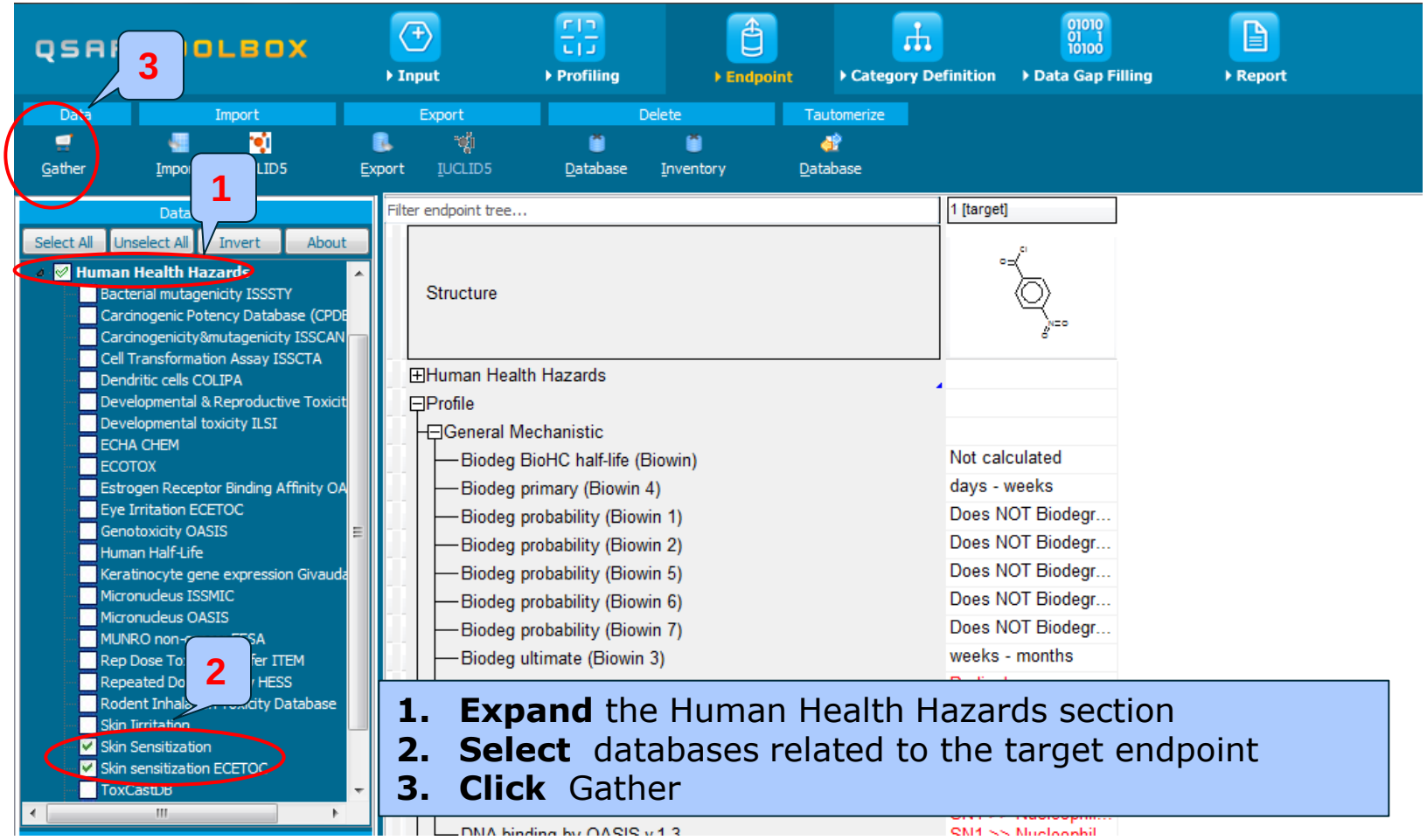
Endpoint
Overview

- "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
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).
- Click on "Endpoint" in the Toolbox workflow.
- Expand the "Human Health Hazards" section
- Click on the box to select that database.
- Click on "Gather data" (see next screen shot).

Endpoint
Gather data

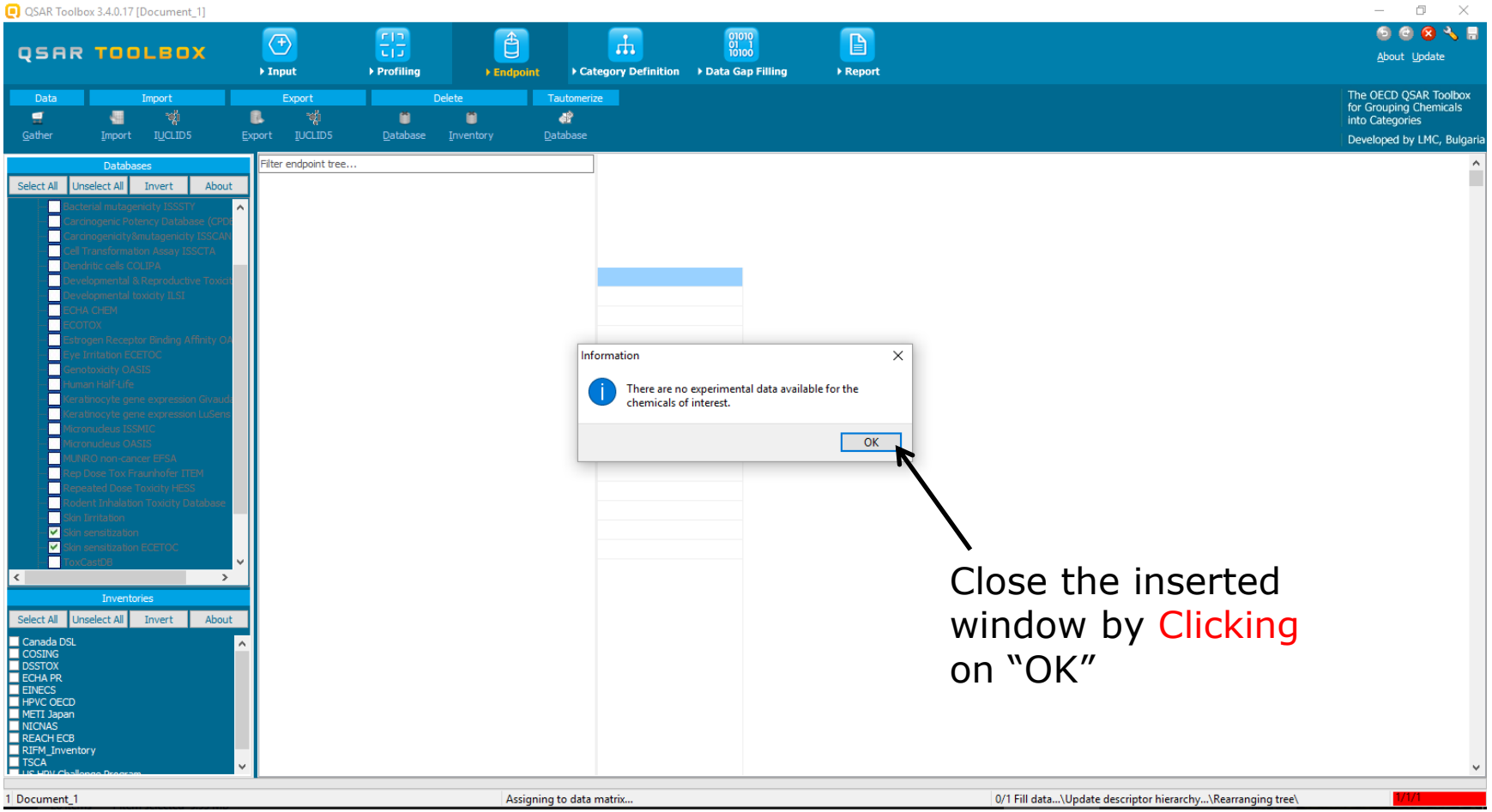


1. **Expand** the Human Health Hazards section
2. **Select** databases related to the target endpoint
3. **Click** Gather

Endpoint
Gather data

- Toxicity information on the target chemical is electronically collected from the selected dataset(s).
- It should be kept in mind that 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



Close the inserted window by Clicking on "OK"

Outlook

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

Recap

- In module one, you have entered the target chemical CAS RN in order to retrieve 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.

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.
- Detailed information about grouping chemicals could be found at the following link (Chapter 4). http://www.oecd.org/dataoecd/58/56/46210452.pdf
- For example, starting from a target chemical for which a specific protein binding mechanism is identified, analogues can be found which can bind to the proteins by the same mechanism and for which experimental results are available.

Category Definition
Protein binding by OASIS v.1.4 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
Background to Protein binding by OASIS v.1.4 categorization

- This scheme includes 146 categories organized in three level of information:

  - ✓ Level I: Mechanistic Domains
  - ✓ 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
  - ✓ Level III: A number of structural alerts specifying the substituents to a common reactive centre are made up for each mechanistic alert

Category Definition
Background to Protein binding by OASIS categorization

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

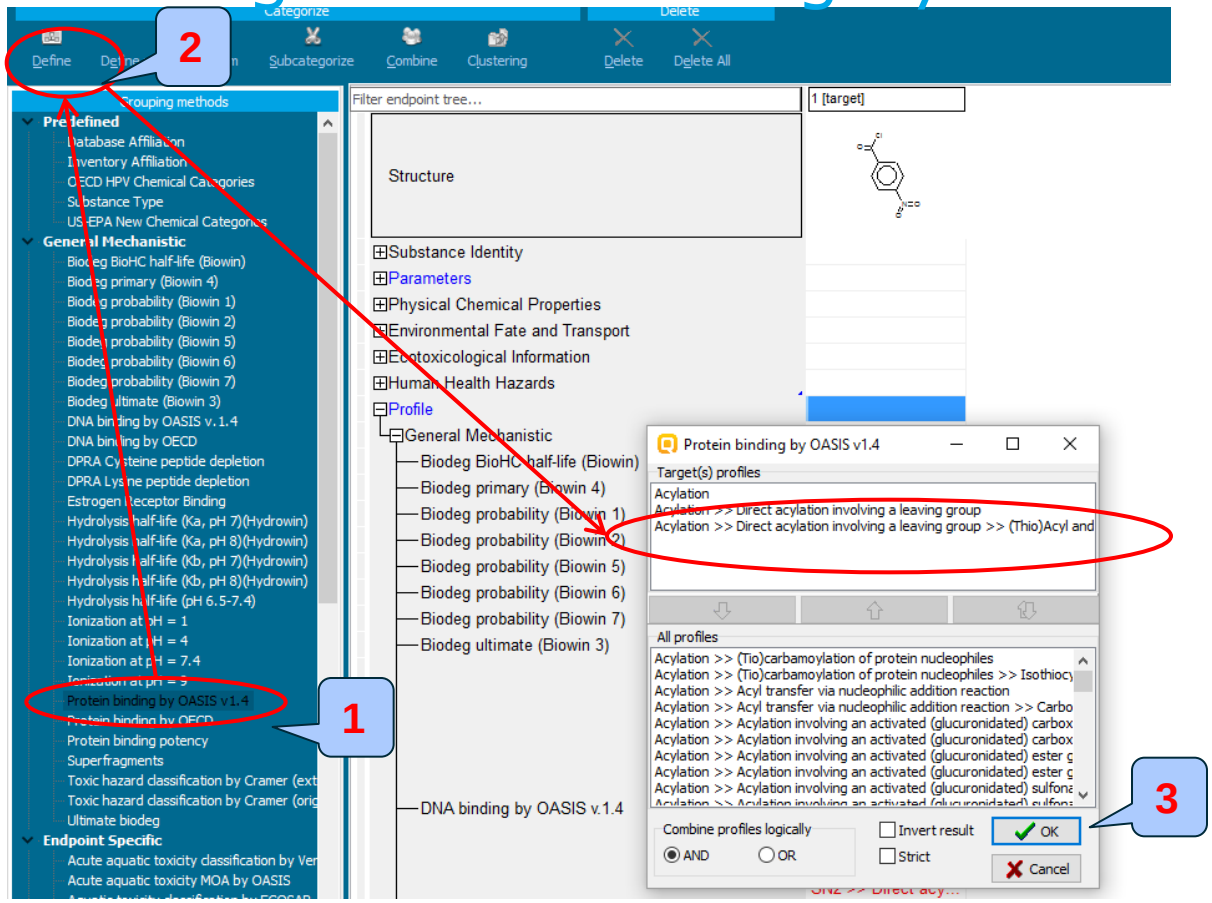
- The associated chemical reactions are in accordance with existing knowledge on electrophilic interaction mechanisms of various structural functionalities.

Category Definition
Background 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.

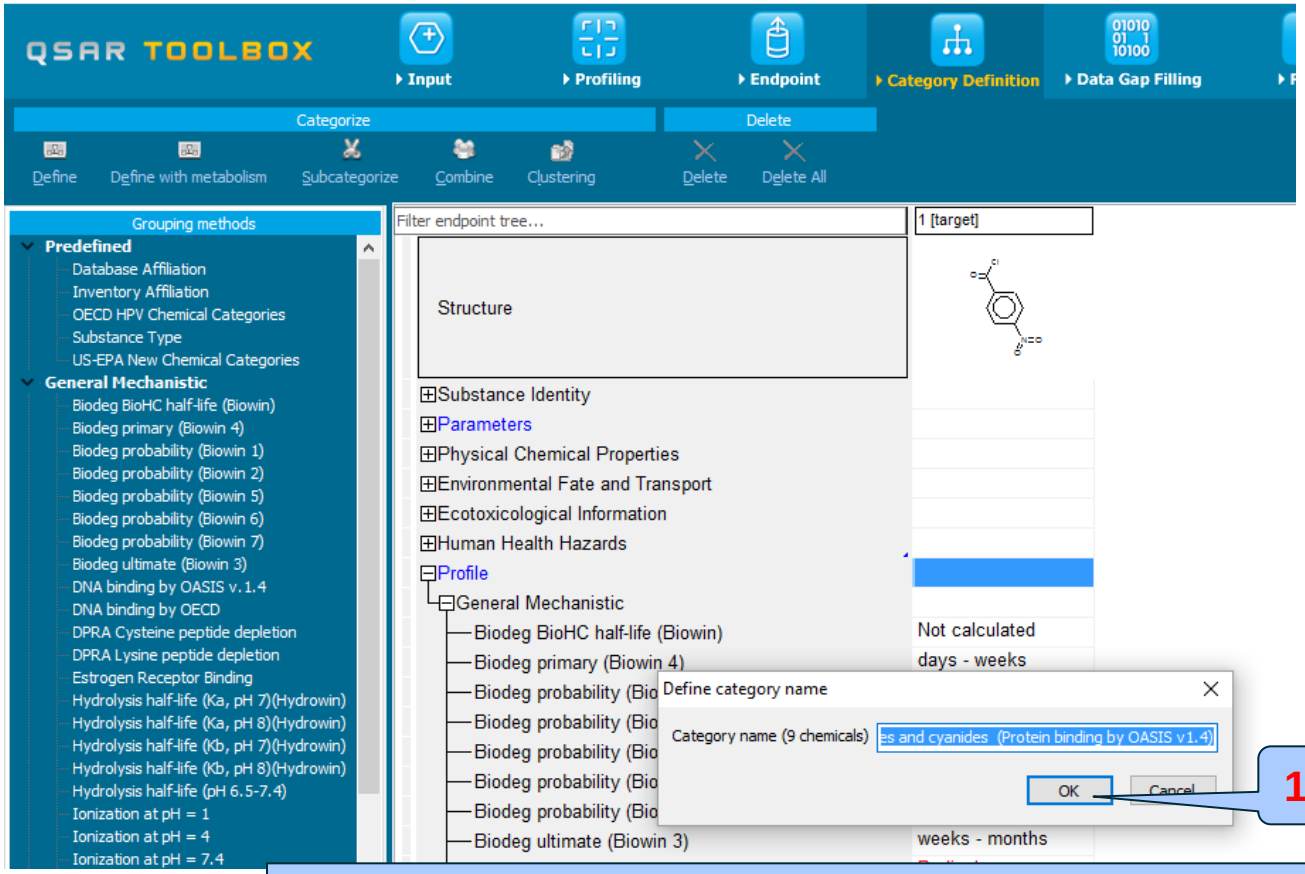
Category Definition

Defining Protein binding by OASIS v.1.4



1. **Highlight** the "Protein binding by OASIS v.1.4"; 2. **Click** Define; 3. **Click** OK to confirm the defined categories for the target chemical

Category Definition
Defining Protein binding by OASIS



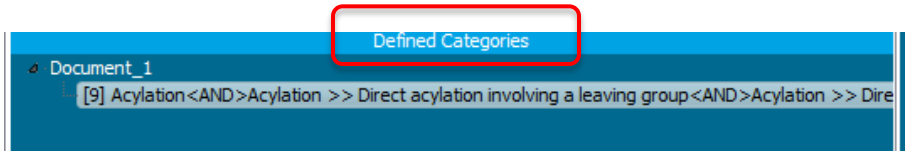
1. Click OK to confirm the name of the category

Category Definition
Analogues

- The data is automatically collected.

- Based on the defined category (Acylation<AND>Acylation >> Direct acylation involving a leaving group<AND>Acylation >> Direct acylation involving a leaving group >> (Thio)Acyl and (thio)carbamoyl halides and cyanides) 8 analogues have been identified

- The name of the category appears in the "Defined Categories" window, indicating the number of substances belonging to the category.



- In other words, these 9 compounds along with the target chemical form a category, which can be used for data filling. (see next slide)

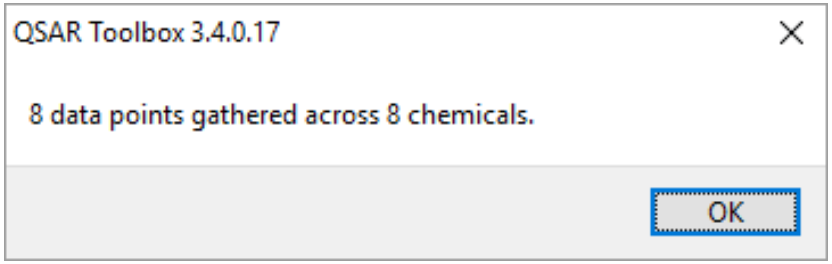
Category Definition
Read data for Analogues

- 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).

Category Definition
Read data for Analogues

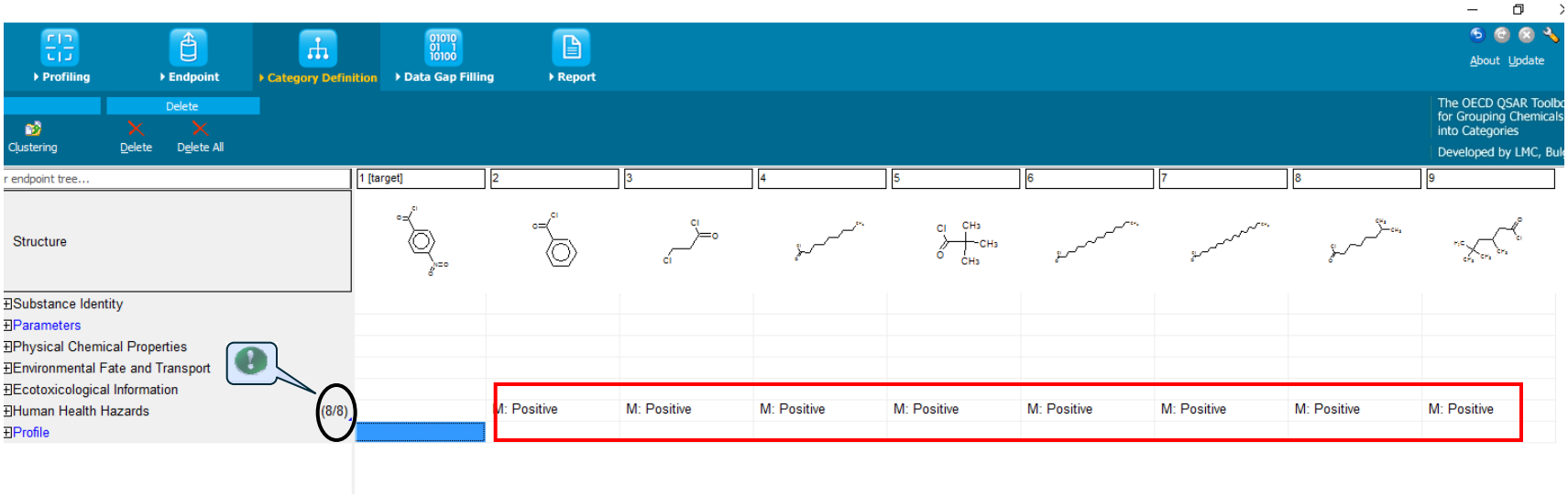
- Toolbox automatically informs the user for the number of gathered data points across the chemicals in the category



- Click OK to confirm the appeared message

Category Definition
Summary information for Analogues

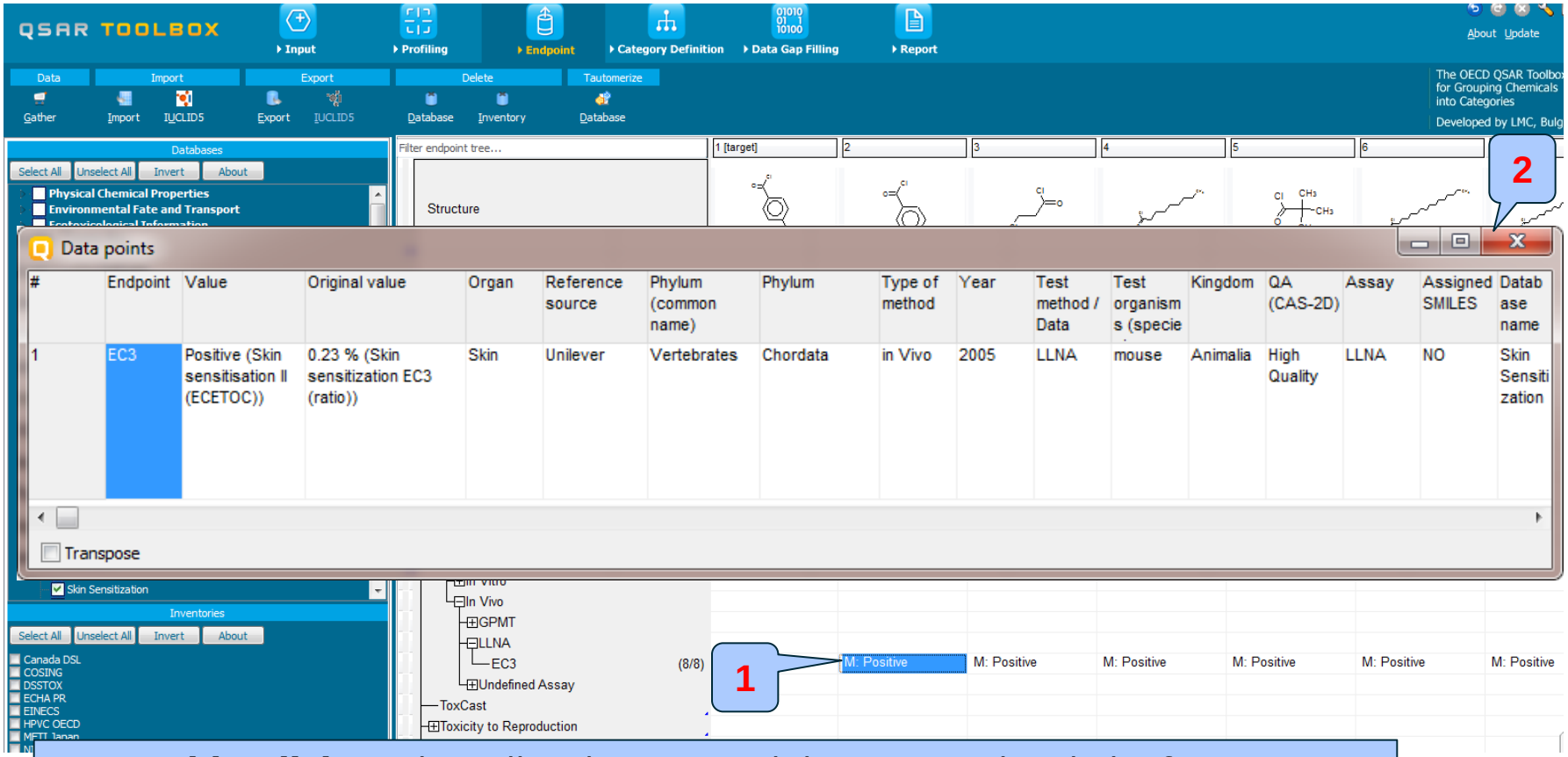
- The experimental results for the analogues are inserted into the matrix



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

Category Definition

Side bar of experimental data



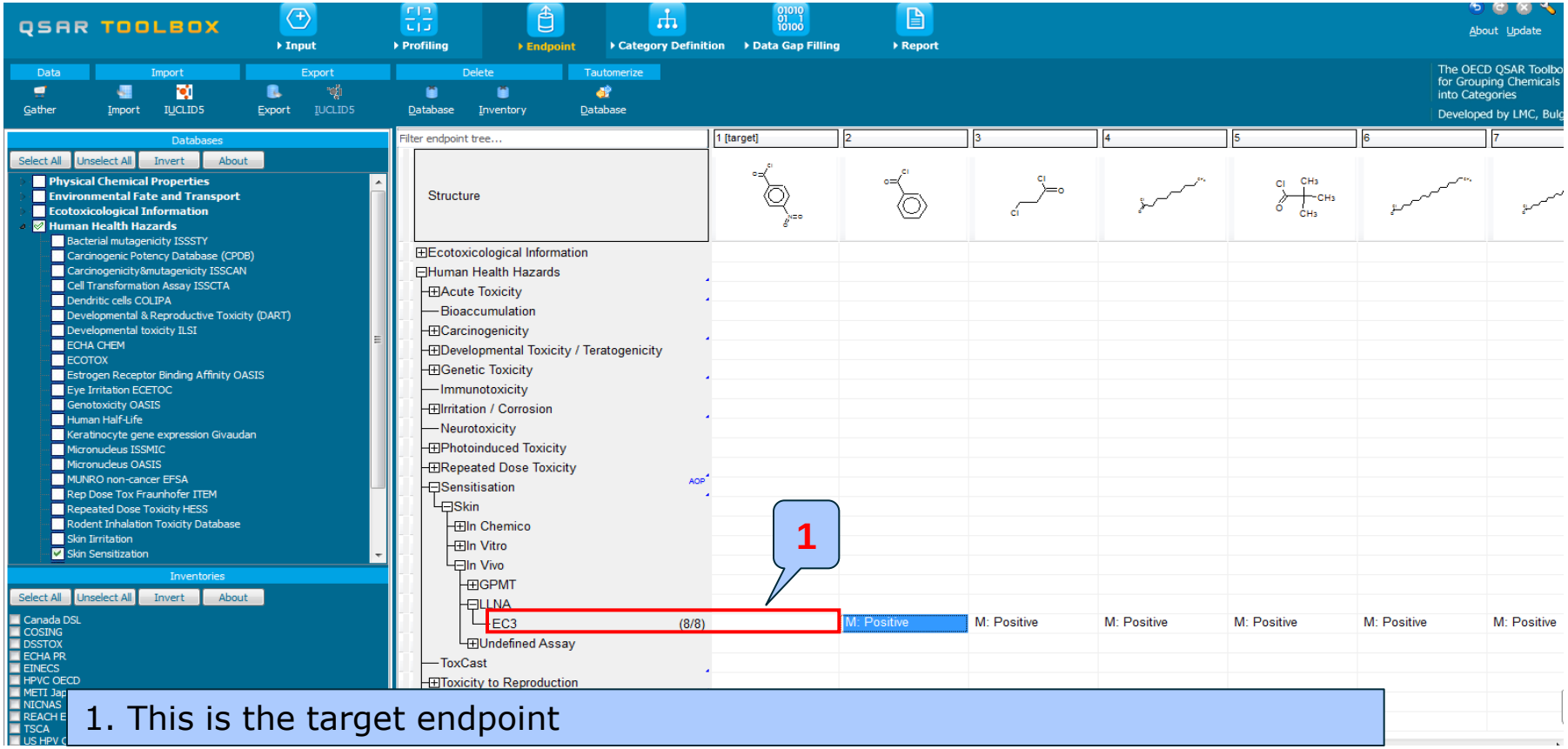
1. **Double-click** on the cell with measured data to see detailed information;
2. **Click** on the X to close the dropdown box.

Category Definition
Navigation through the endpoint tree

- The user can navigate through the data tree by closing or opening the nodes of the endpoint tree.
- Click on the plus sign next to Human Health Hazards then Sensitisation, followed by Skin, In Vivo and LLNA and finally EC3.
- Local lymph node assay is *in vivo* method for assessment of relative skin sensitization potential of chemicals. The potential is expressed as EC3 values.
- In this example, results from skin sensitisation testing for chemicals reacting via nucleophilic substitution of acyl halides are available (see next screen shot).

Category Definition

Navigation through the endpoint tree



1. This is the target endpoint

Outlook

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

Recap

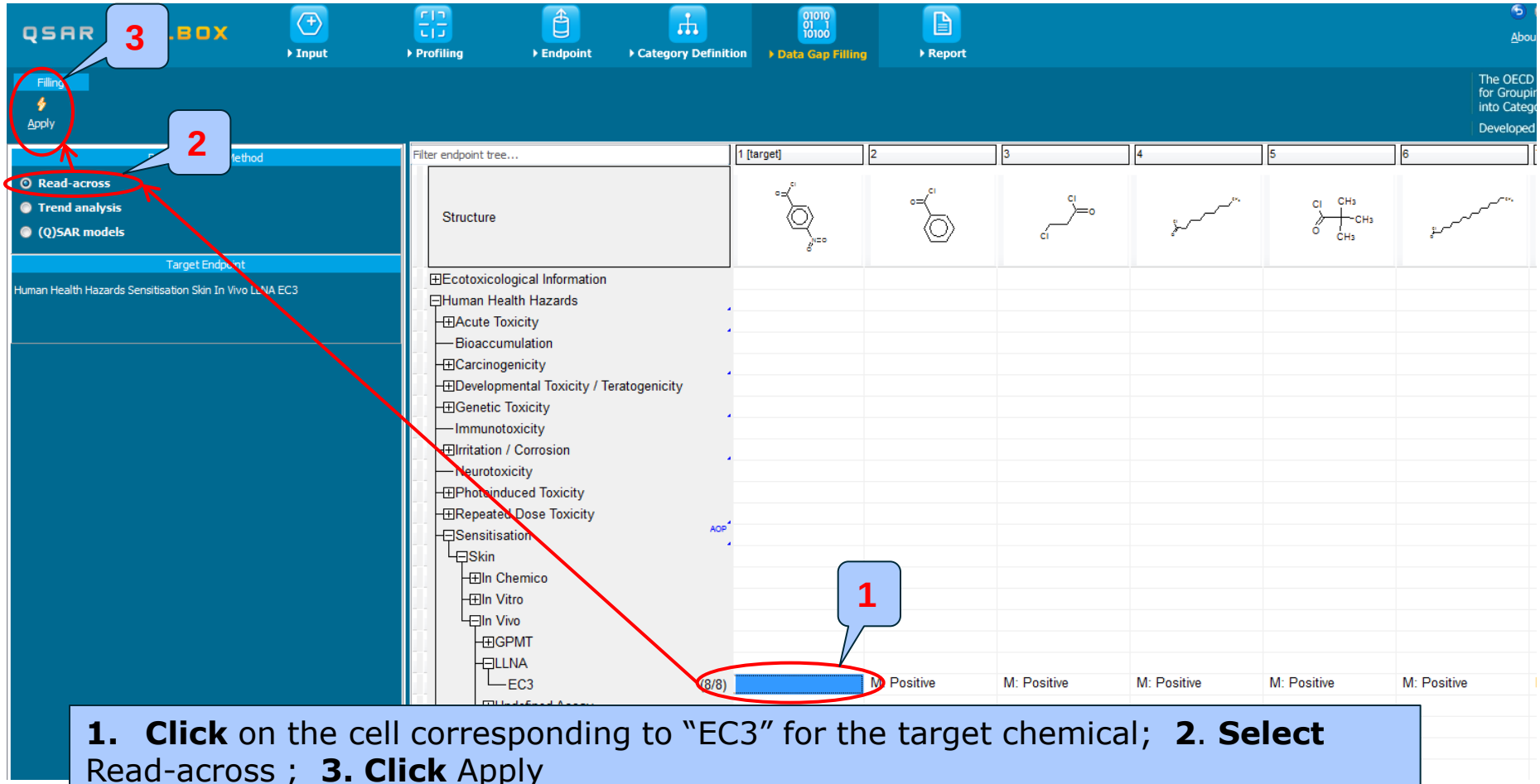
- You have identified a mechanistic category (Acylation<AND>Acylation >> Direct acylation involving a leaving group<AND>Acylation >> Direct acylation involving a leaving group >> (Thio)Acyl and (thio)carbamoyl halides and cyanides) for the target chemical (4-nitrobenzoyl chloride).
- You have now retrieved in the available experimental results on skin sensitisation (EC3) values for eight chemicals with the same mechanism of protein binding as the target compound, which were found in the “Skin Sensitisation” databases.
- The user can now proceed to the next module; click on ”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 use read-across.

Data Gap Filling
Apply Read across

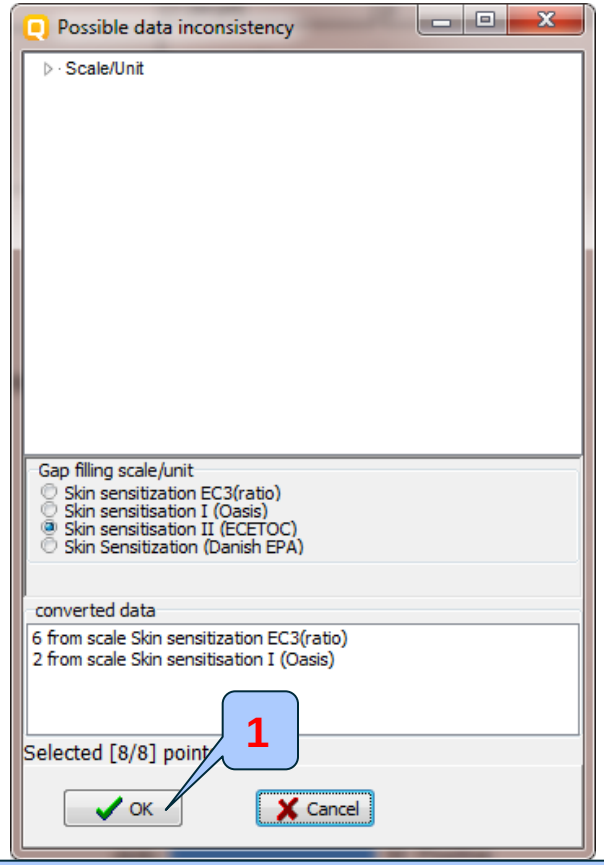


1. Click on the cell corresponding to "EC3" for the target chemical; **2. Select** Read-across ; **3. Click** Apply

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).
- 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*).
- 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 and Negative.

Data Gap Filling
Scale definition

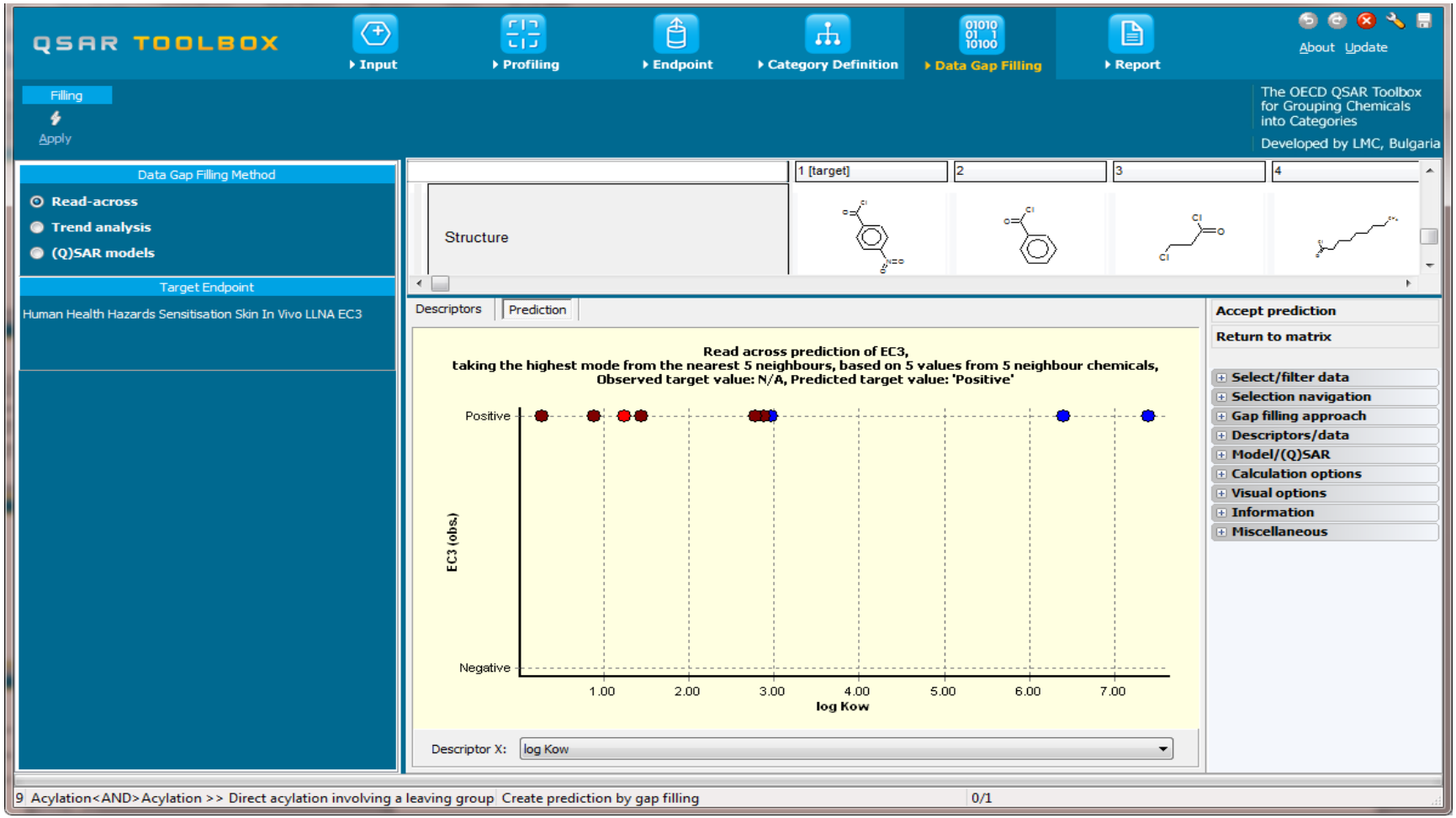


1. **Click** OK

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).
- The **RED** dot represents predicted results for the target chemical.
- The **BROWN** dots represent the experimental results available for the analogues that are used for the read-across.
- The **BLUE** dot represent the experimental results available for the analogues but not used for read-across.

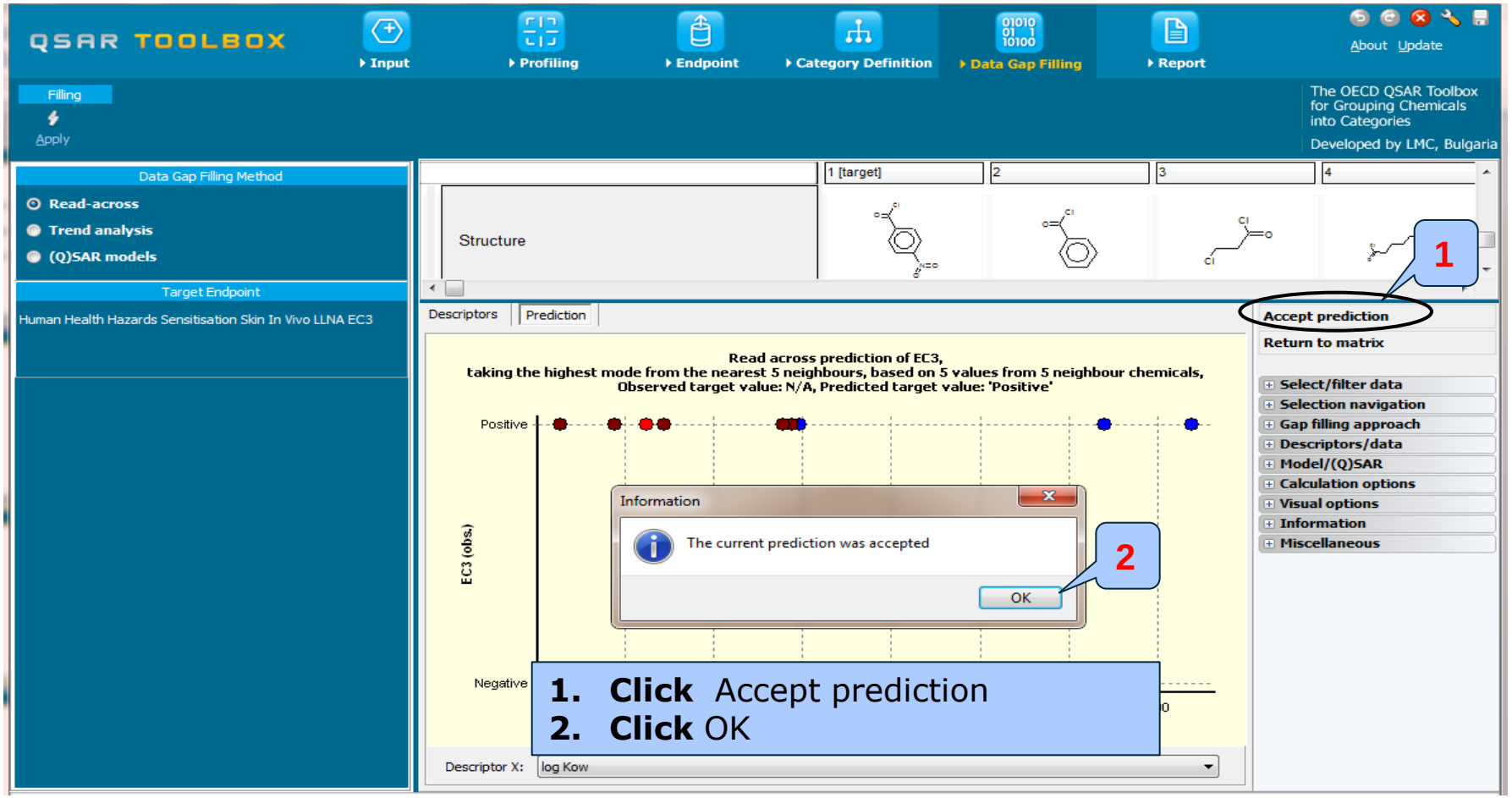
Data Gap Filling
Read-across



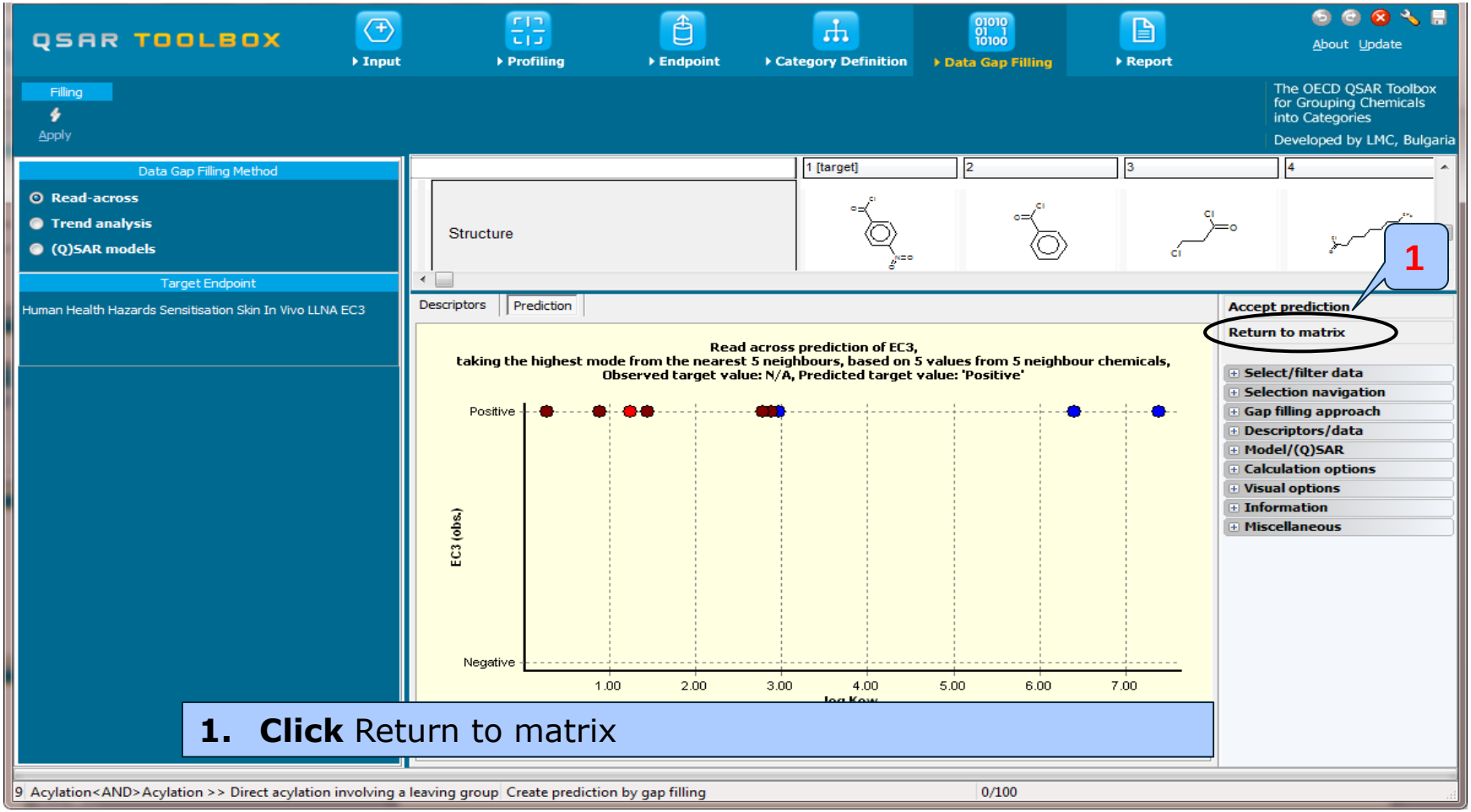
Data Gap Filling
Interpreting Read-across

- In this example, all results of the analogues are consistent; all the analogues are Skin sensitizers.
- The same positive sensitising potential is therefore predicted for the target chemical.
- Accept the prediction by clicking "Accept prediction" (see next screen shot).

Data Gap Filling
Accepting the predicted result



Data Gap Filling
Accepting the predicted result



Recap

- The read-across is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitisation. Since all the tested chemicals in the category were positive, it was easy to accept the positive predictions for the target chemical.
- You are now ready to complete the final module and to download the report.
- Click on "Report" to proceed to the last module.

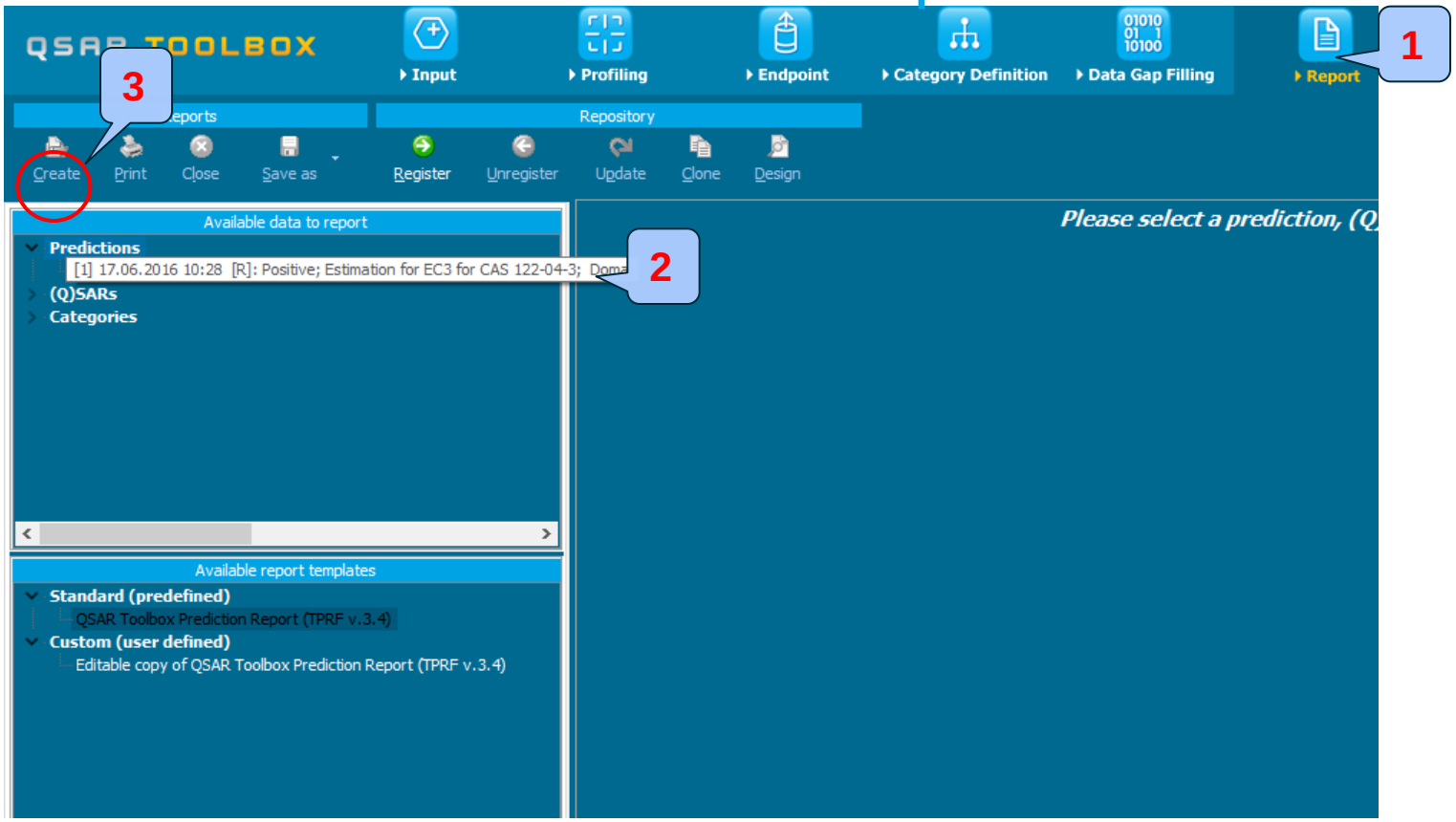
Outlook

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

Report
Overview

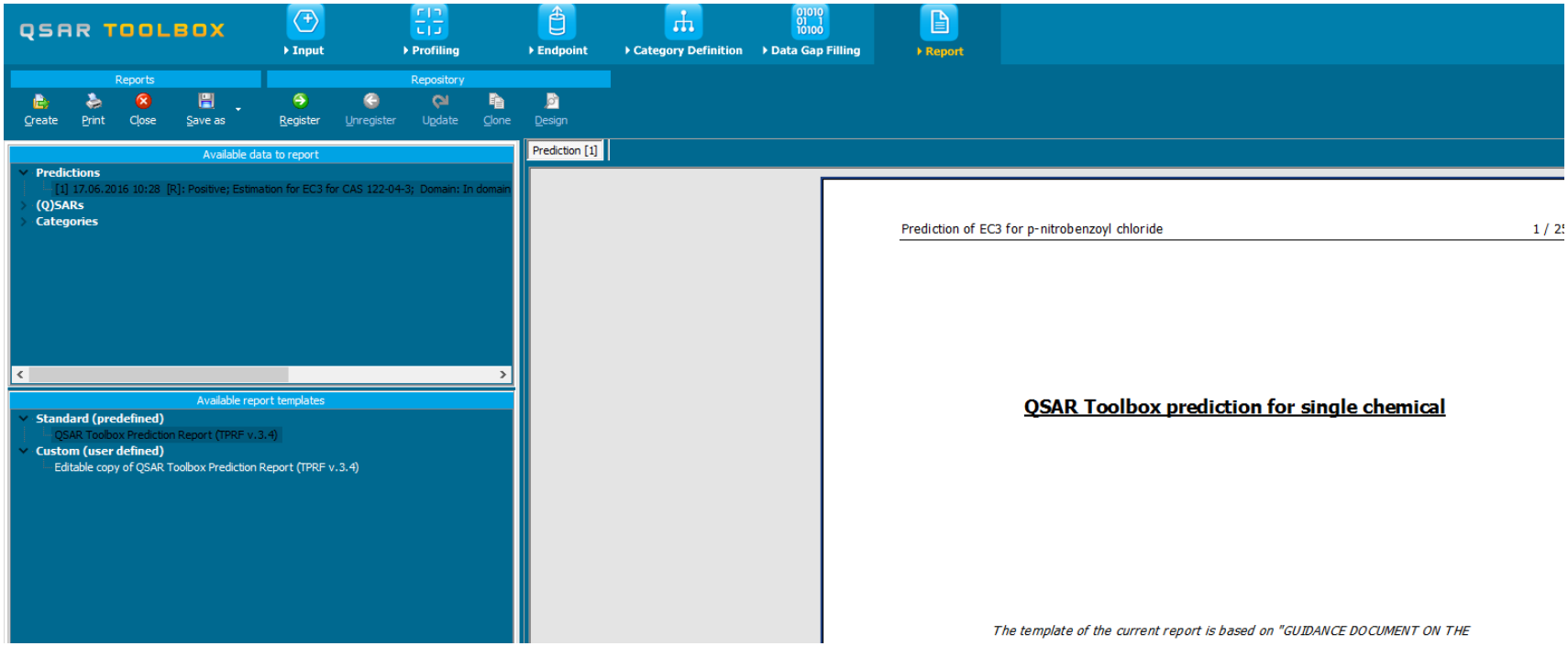
- The report module could generate report on any of predictions performed with the Toolbox.
- Report module contains predefined report templates as well as a template editor with which users can develop a user defined templates.
- The report can then be printed or saved in different formats.

Report
Generation report



1. **Go** to the Report section; 2. **Expand** Prediction in the "Available data to report" window; 3. **Click** Create

Report
Generation report



Outlook

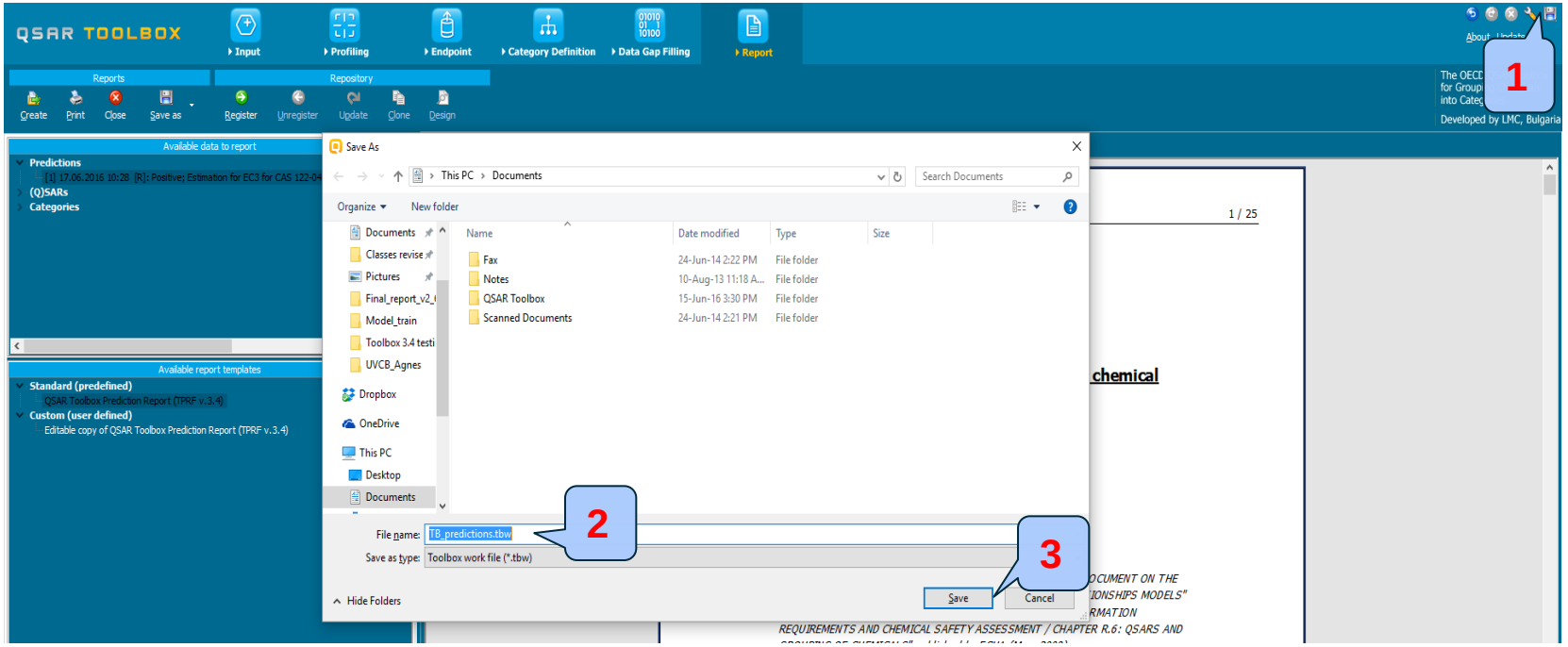
- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow
- **Save prediction**

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).

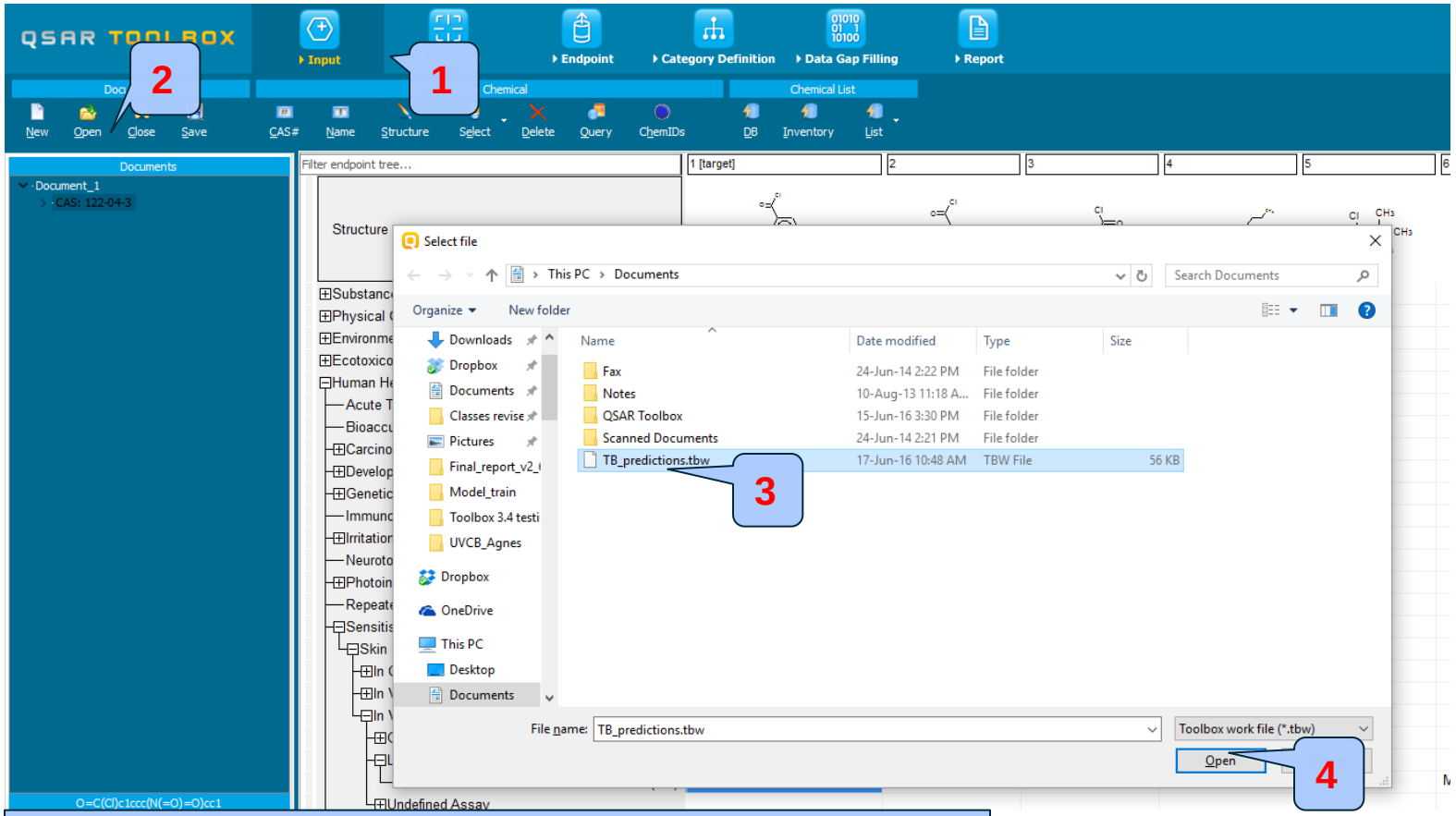
- Saving/Loading the file with TB prediction is shown on next screenshots

Saving the prediction result



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**

Congratulations

- You have now been introduced to the workflow of the Toolbox and completed the tutorial on data gap filling by read-across based on an analogue approach.
- You have been introduced to the six modules of the Toolbox, the basic functionalities within each module and the rationale behind each module.
- Note proficiency comes with practice.