QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Predicting skin sensitization potential of 3,4-dinitrophenol taking into account tautomerism

Outlook

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

Background

- This is a step-by-step presentation designed to take the user of the Toolbox through the workflow for filling data gap for skin sensitization taking into account tautomerism of target chemical.

Outlook

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

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Providing tautomeric set of target chemical
- Identify analogues for a set of tautomers
- Retrieve experimental results available for those analogues
- Fill data gaps by read across
- Save the prediction

Outlook

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

The Exercise

- In this exercise we will predict the skin sensitization potential for **(3,4-dinitrophenol) [CAS 577-71-9]**
- Set of simulated tautomers for the target chemical will be provided
- This prediction will be accomplished by collecting a set of similar analogues for set of target and its tautomers
- The initial category will be defined by Protein binding by OASIS v1.4
- Data gap will be filled by read-across

Outlook

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

Workflow

- **As you know the Toolbox has 6 modules which are typically used in sequence**:
  - Chemical Input
  - Profiling
  - Endpoint
  - Category Definition
  - Data Gap Filling
  - Reporting

Outlook

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

Chemical Input

There are two ways for simulating tautomers of chemicals

- During the process of entering the structure into the system (scenario 1)

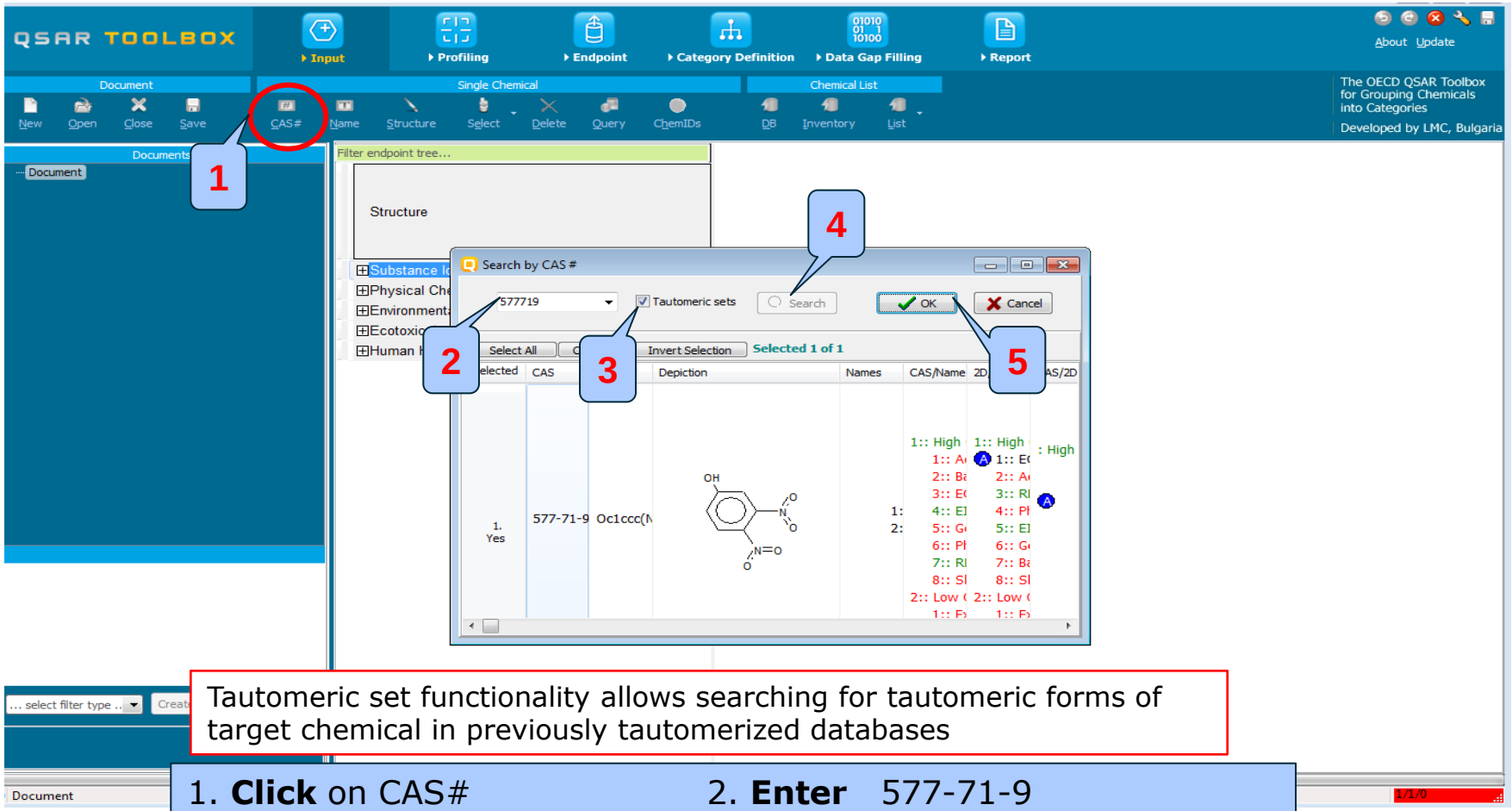
- Simulating tautomersim of already entered structure (scenario 2)

In order to accelerate the workflow all databases are preliminary tautomerized, calculated and profiled. The results are stored in the database.

See next screen shots

Chemical Input

Search by CAS# in tautomerized databases (scenario 1)

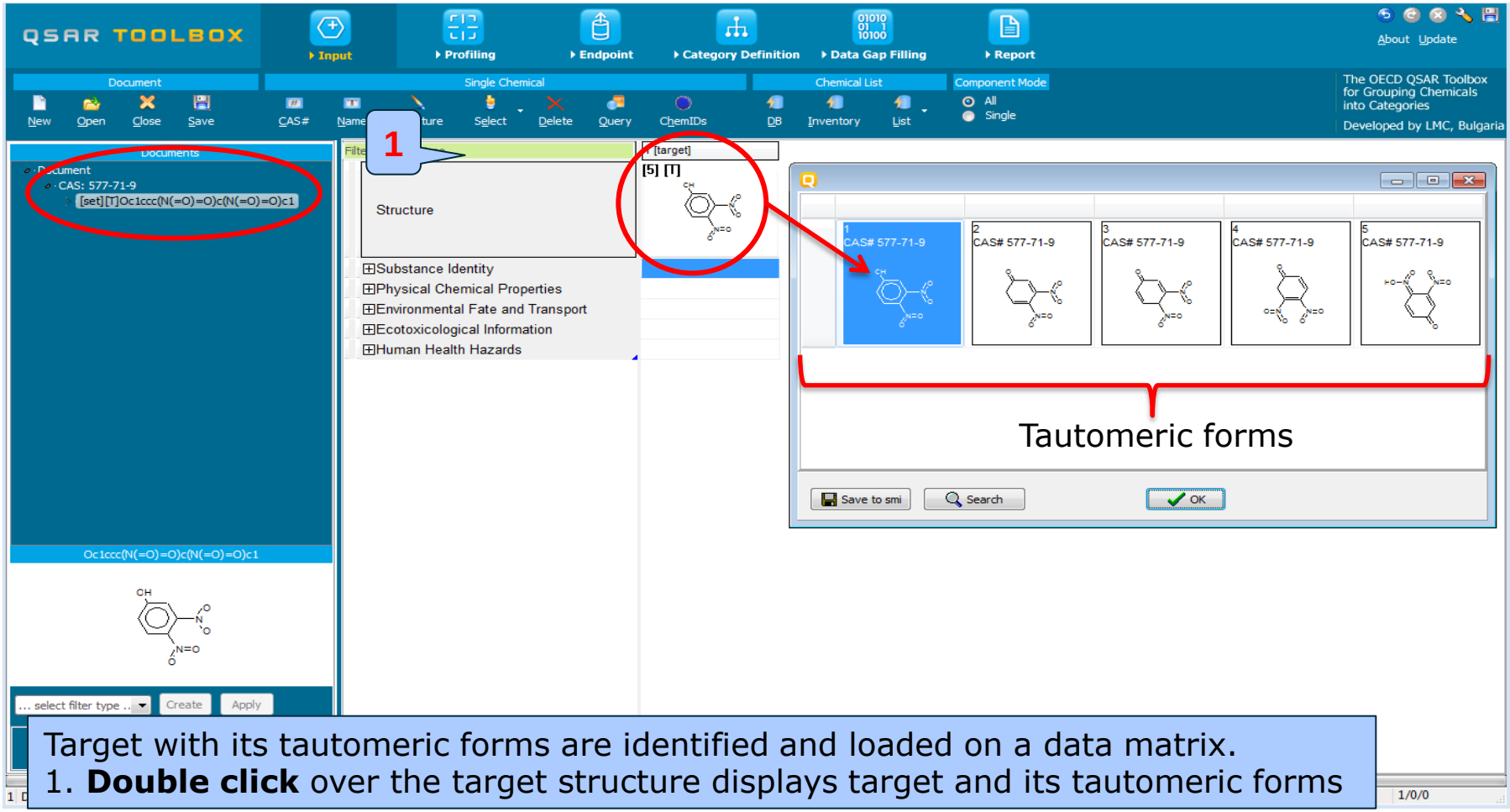


Tautomeric set functionality allows searching for tautomeric forms of target chemical in previously tautomerized databases

1. **Click** on CAS#
2. **Enter** 577-71-9
3. **Select** Tautomeric sets
4. **Click** Search
5. **OK**

Chemical Input

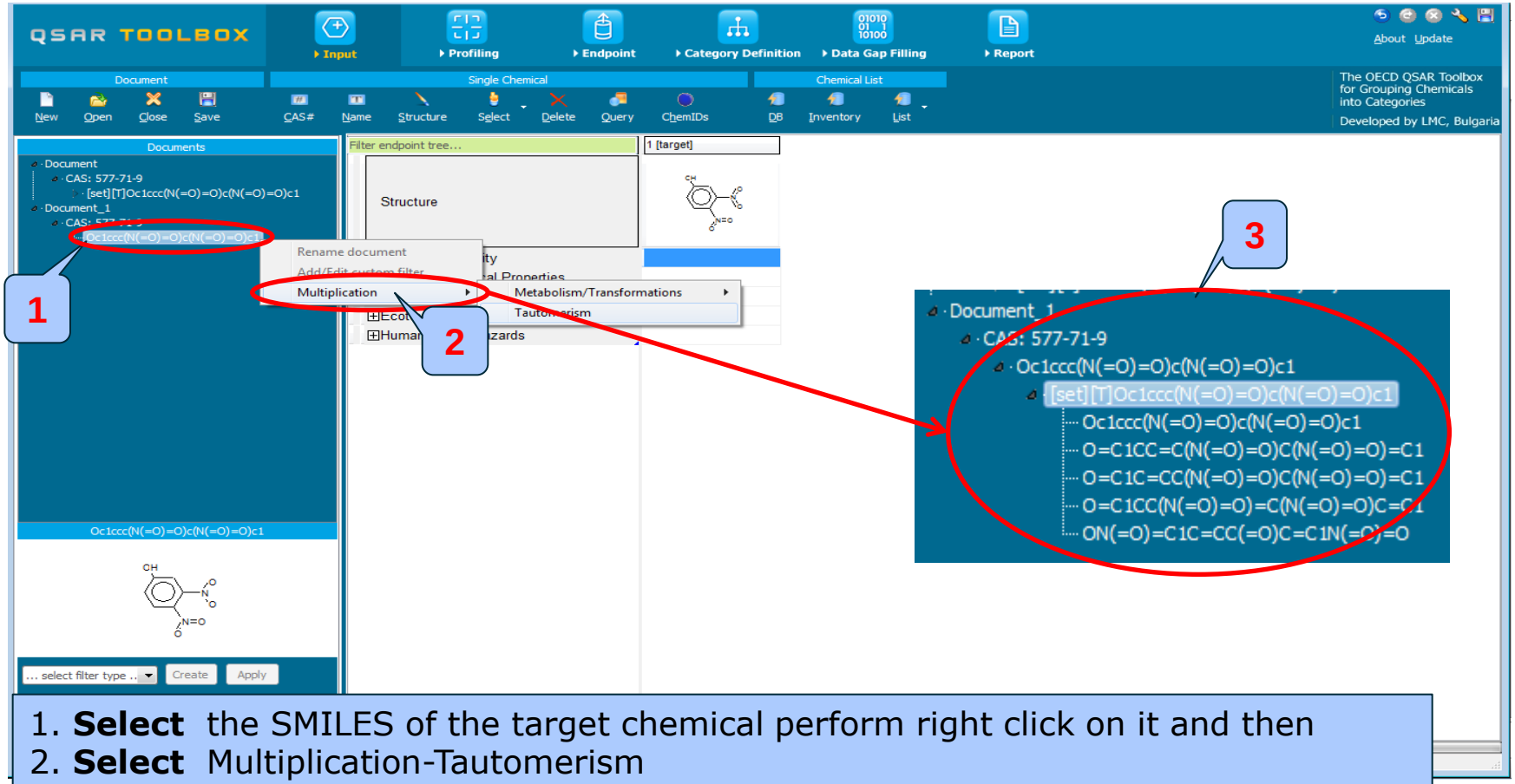
Target chemical identity (scenario 1)



Target with its tautomeric forms are identified and loaded on a data matrix.
1. **Double click** over the target structure displays target and its tautomeric forms

Chemical Input

Multiplication a tautomeric set of already defined target (scenario 2)

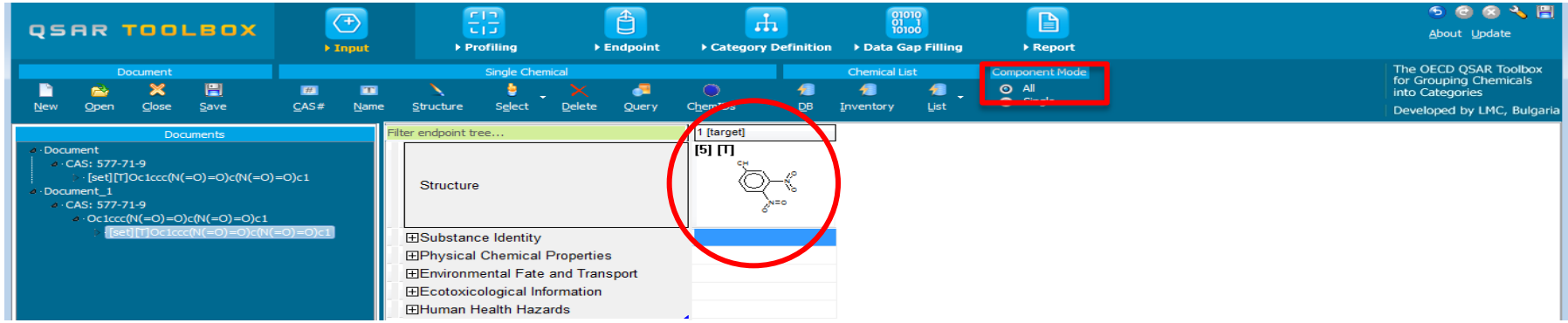


1. **Select** the SMILES of the target chemical perform right click on it and then
2. **Select** Multiplication-Tautomerism
3. **Generated** tautomers appear in tree like form

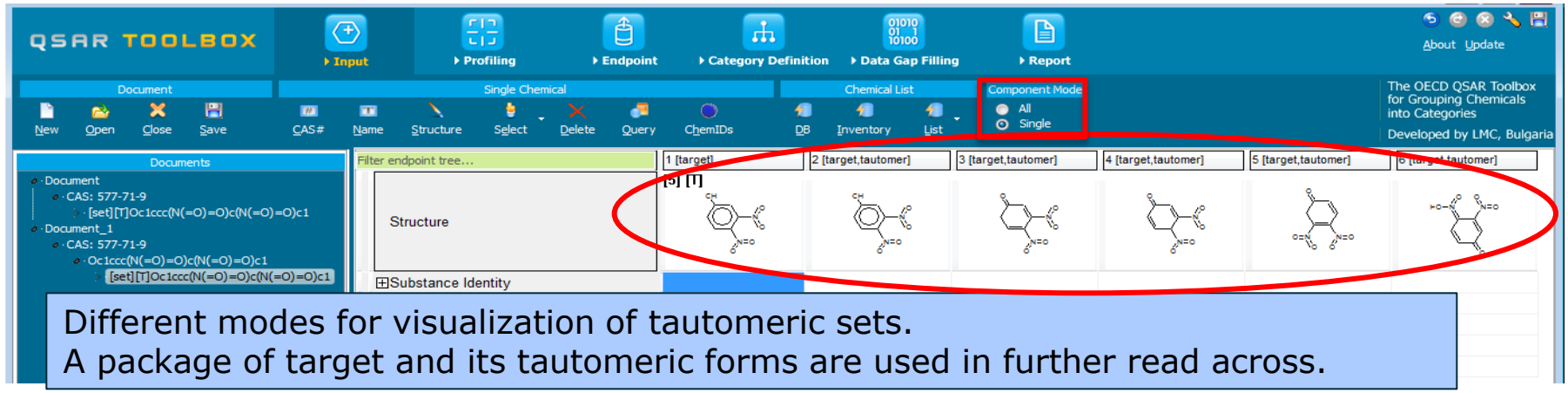
Chemical Input

Implementation of Modeling modes:

- **Component Mode All –** all tautomers are analyzed as a package



- **Component Mode Single** – each tautomer is analyzed individually



Different modes for visualization of tautomeric sets.
A package of target and its tautomeric forms are used in further read across.

Outlook

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

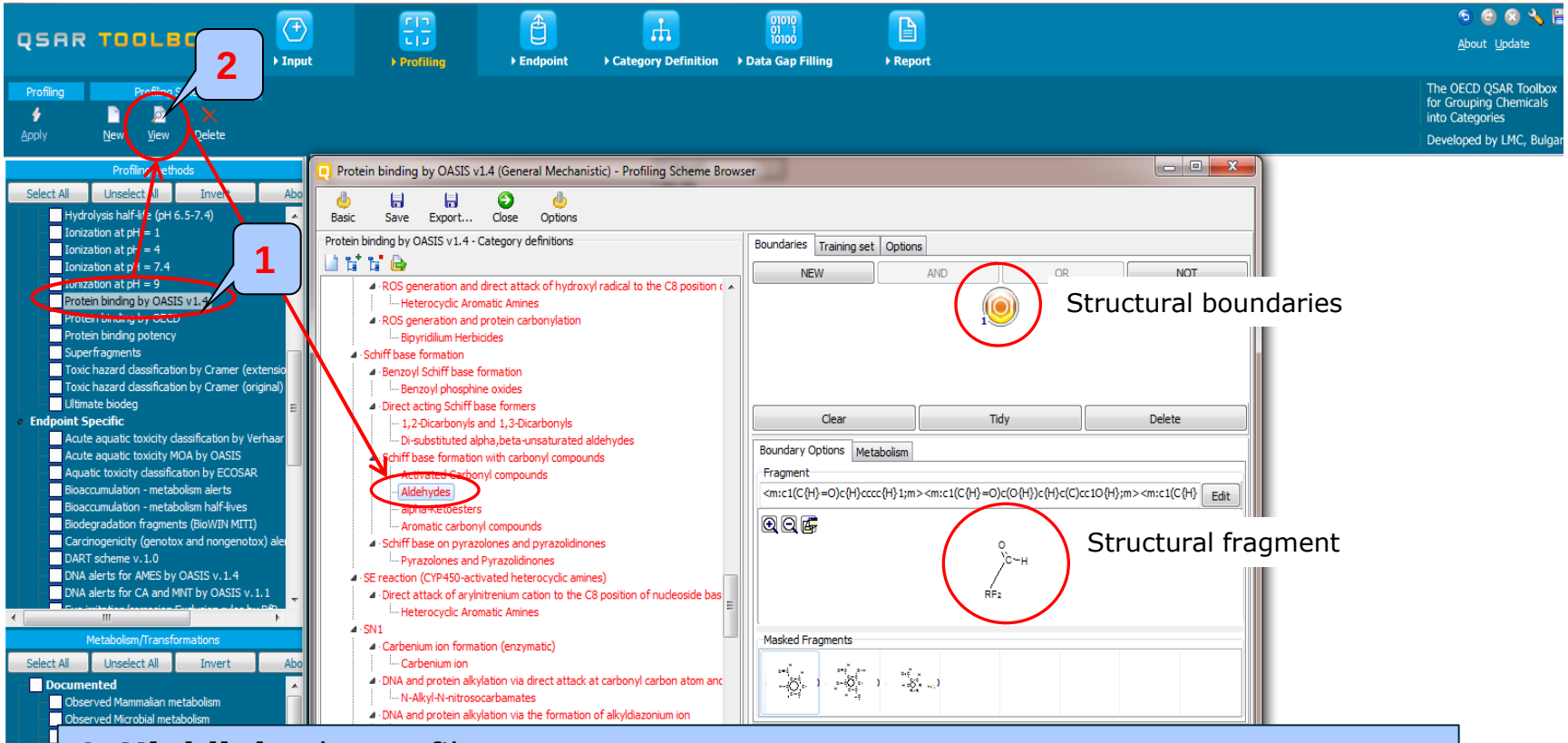
Profiling
Overview

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

Profiling
Side-Bar to Profiling

- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, Protein binding by OASIS v1.4 and clicking on “View” (see next screen shot).

Profiling
Side-Bar to Profiling



1. **Highlight** the profiler
2. **Click** View
3. **Select** "Aldehydes"

Profiling

Profiling the set of target and tautomers

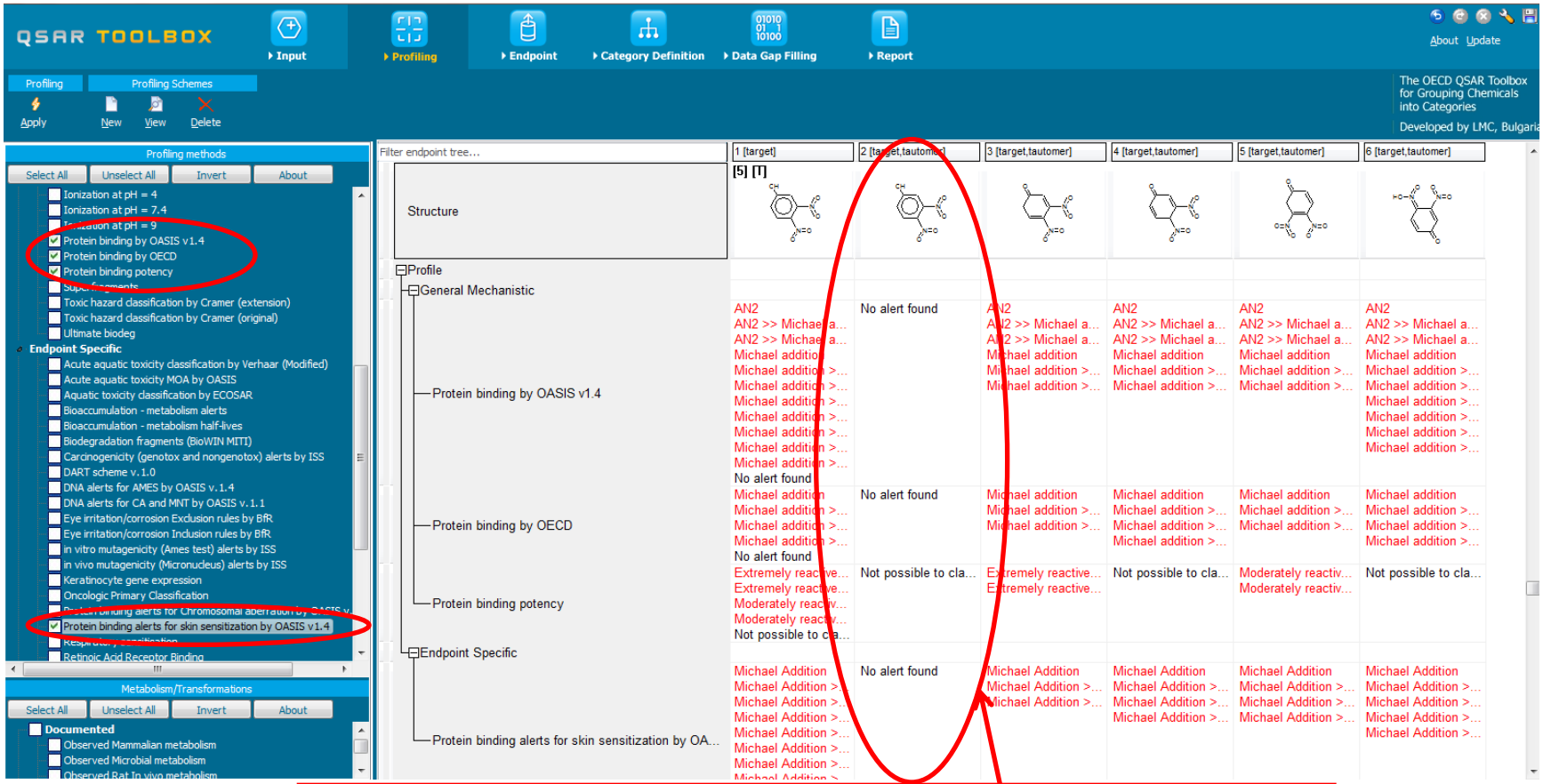
- For this example, the following profilers relevant to skin sensitization are used (see next screenshot):
  - Protein binding by OASIS v1.4
  - Protein binding by OECD
  - Protein binding potency
  - Protein binding for skin sensitization by OASIS v1.4

- Select the "Profiling methods" related to the target endpoint by clicking on the box next to the profilers name.

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

Profiling

Profiling the set of target and tautomers

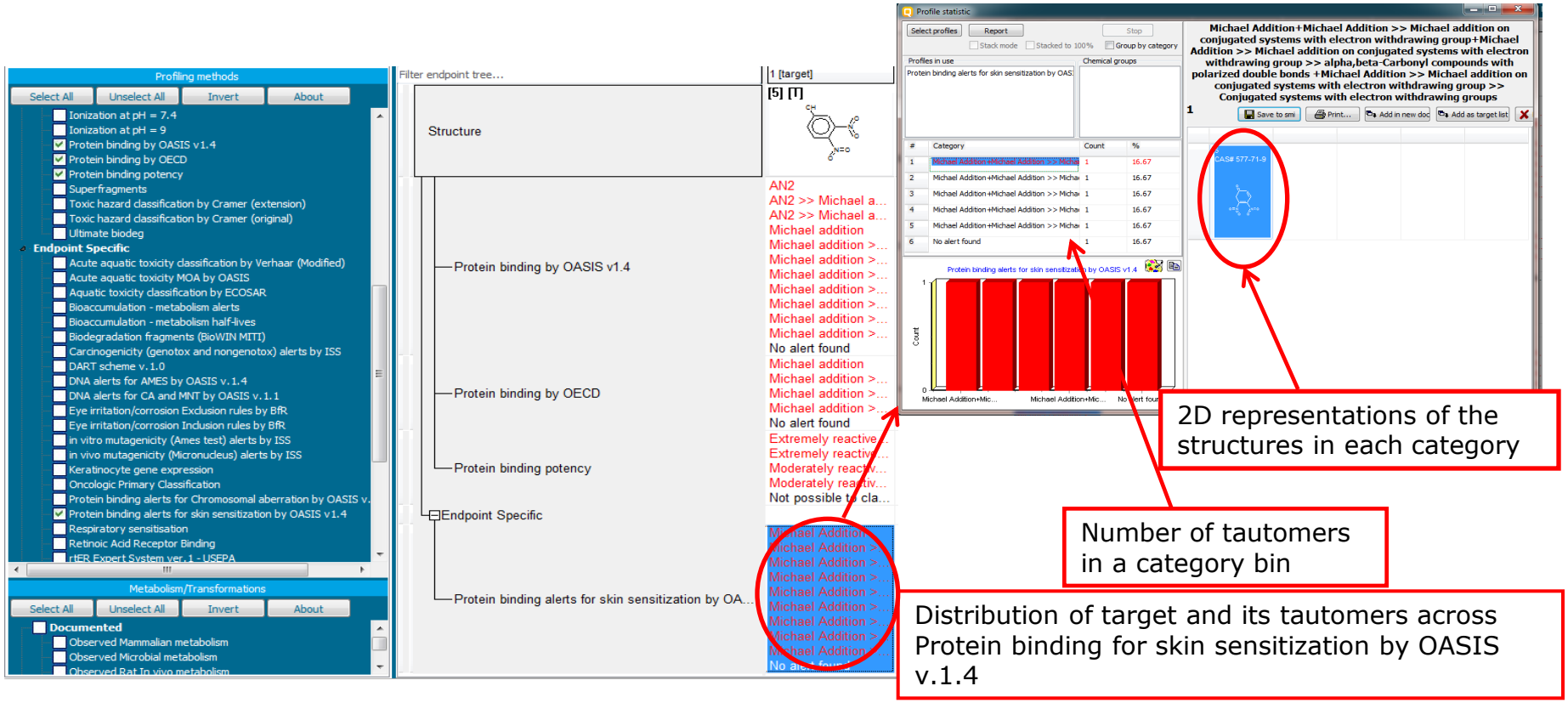


The target chemical has no protein binding alert – No Skin Sensitization effect is expected

Profiling

Profiling the set of target and tautomers

Profile statistic

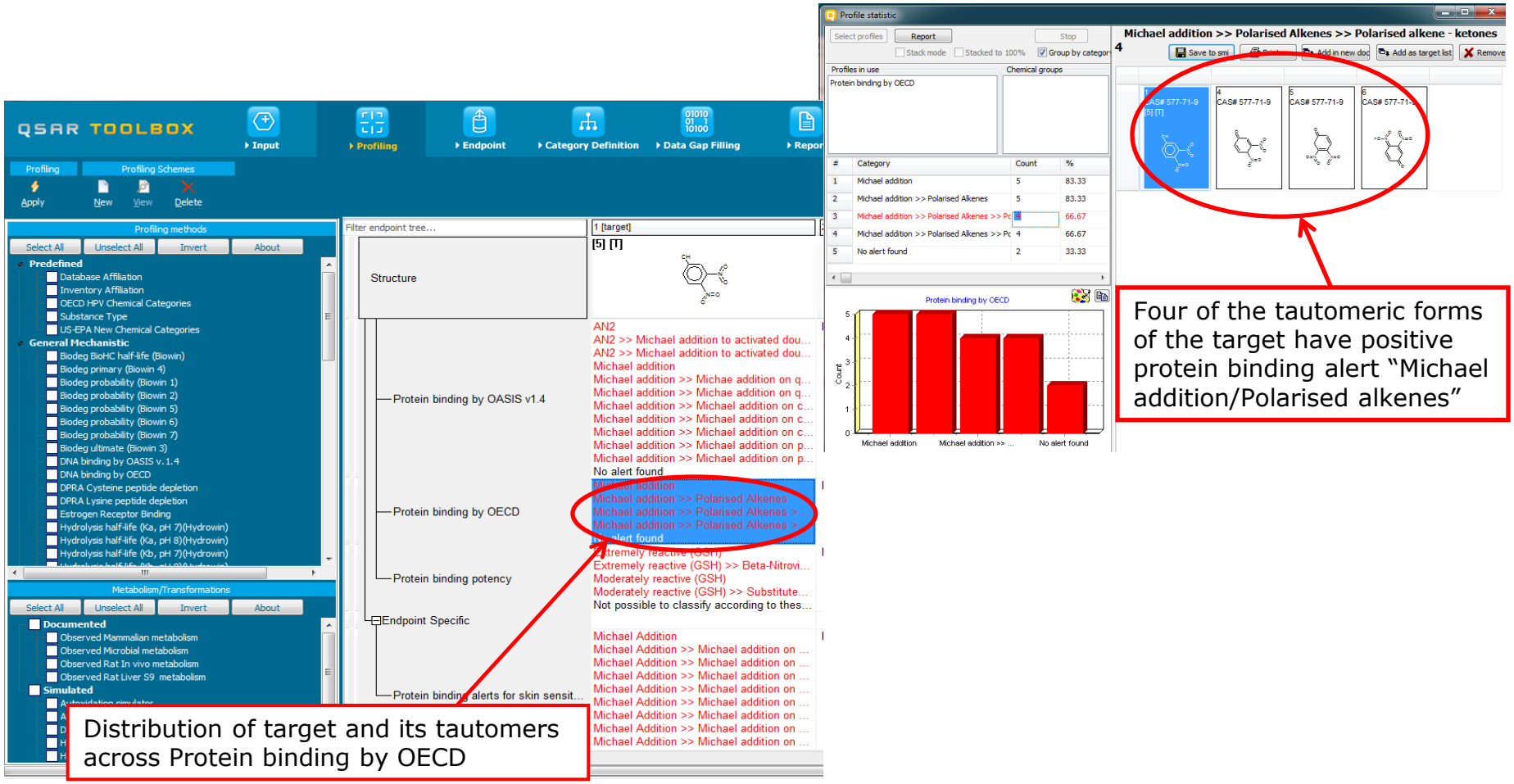


Four tautomeric forms of the target chemical have Protein binding alerts for skin sensitization: "Michael addition/Michael addition on conjugated systems with electron withdrawing group"

Profiling

Profiling the set of target and tautomers

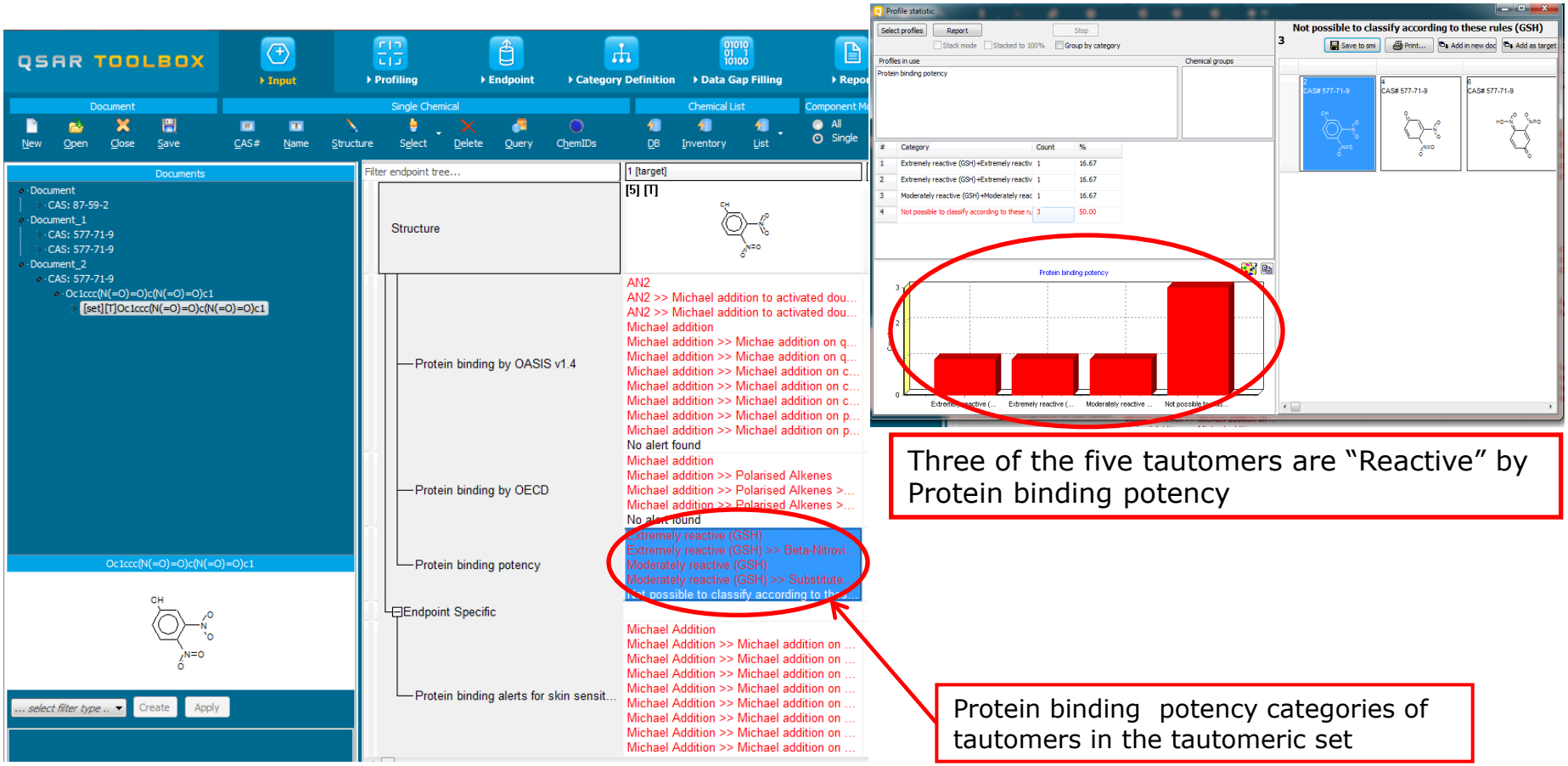
Profile statistic



Profiling

Profiling the set of target and tautomers

Profile statistic



Three of the five tautomers are “Reactive” by Protein binding potency

Protein binding potency categories of tautomers in the tautomeric set

Outlook

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

Endpoint

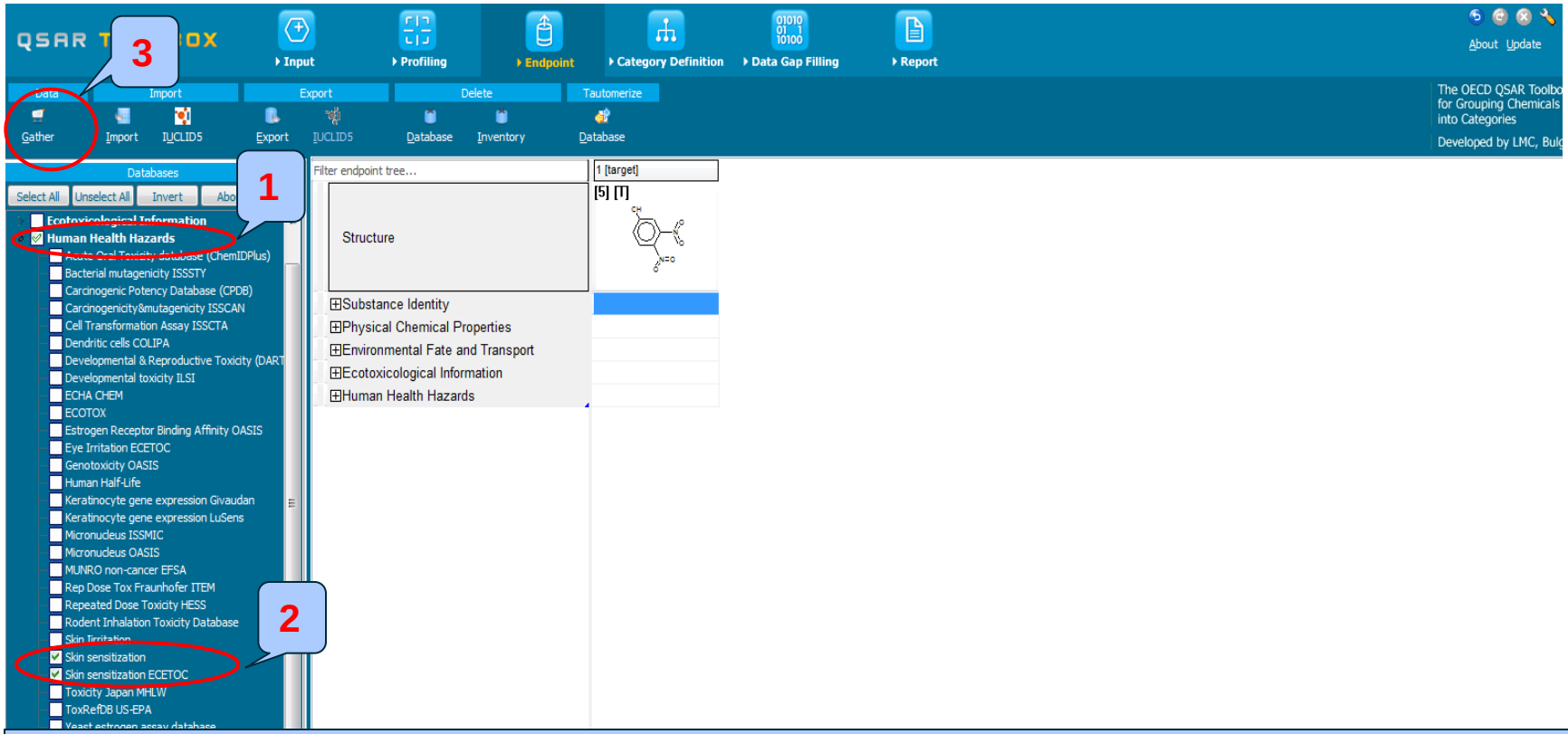
- "Endpoint" refer to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox database.
- Data gathering can be executed in a global fashion (i.e., collecting all data of all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).
- In this example, we limit our data gathering to the common skin sensititization from databases containing skin sensititization data
- Data for target chemical and its tautomeric forms is extracted from selected databases if available

Endpoint

- For this example, the following database are relevant to the skin sensitization(see next screen shot):

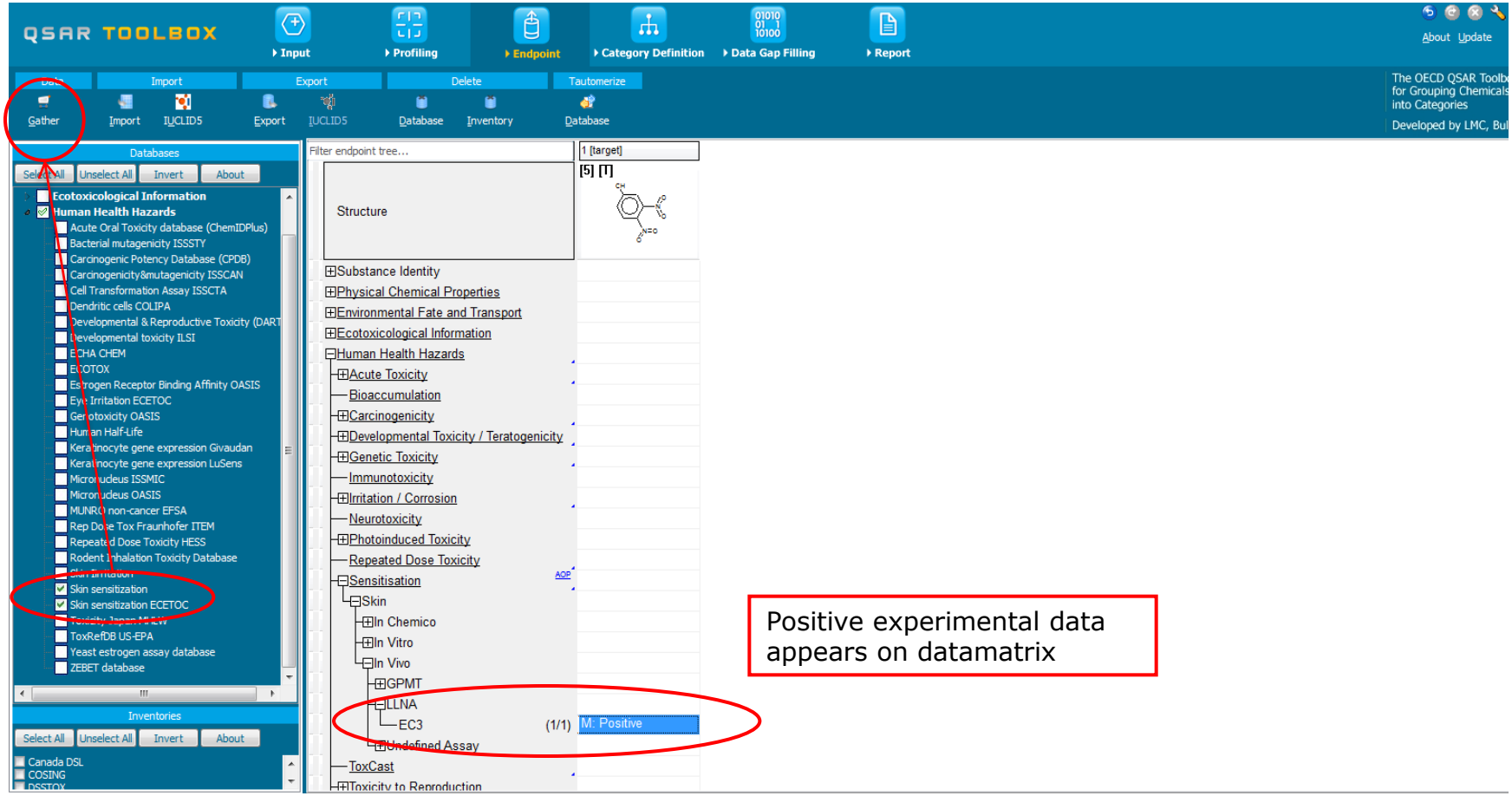
  - Skin sensitization
  - Skin sensitization ECETOC

Endpoint
Gather data



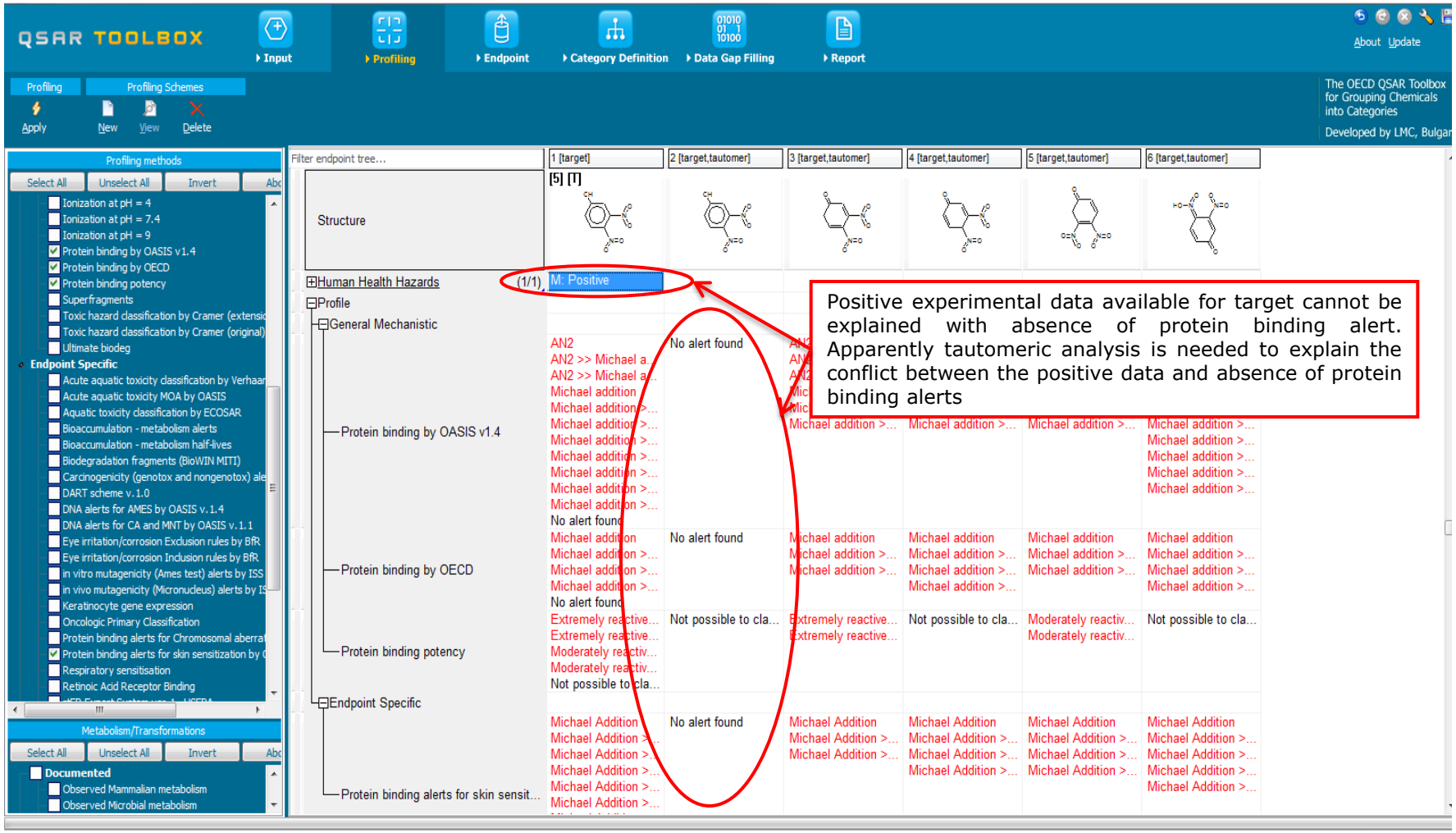
1. **Expand** the Human Health Hazards section
2. **Select** databases related to the target endpoint by adding a green check in the box before the database name.
3. **Click** Gather

Endpoint
Gather data



Positive experimental data appears on datamatrix

Endpoint
Gather data



Recap

- You have entered the target chemical
- You have simulate the tautomeric forms of target chemical
- You have profiled the tautomeric set of the target and identified no protein binding alert for the target. However, four tautomers have positive protein binding alerts
- You have gather data for chemical and its tautomeric forms and found positive experimental data for target.
- It is needed to verify the experimental data by searching for analogs having same functionalities
- Now you are ready to continue with next step of the workflow “Category definition”.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Category definition**

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".

- Detailed information about grouping chemical (Chapter 4) could be found in document "Manual for Getting started" published on OECD website:

  http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm

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
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

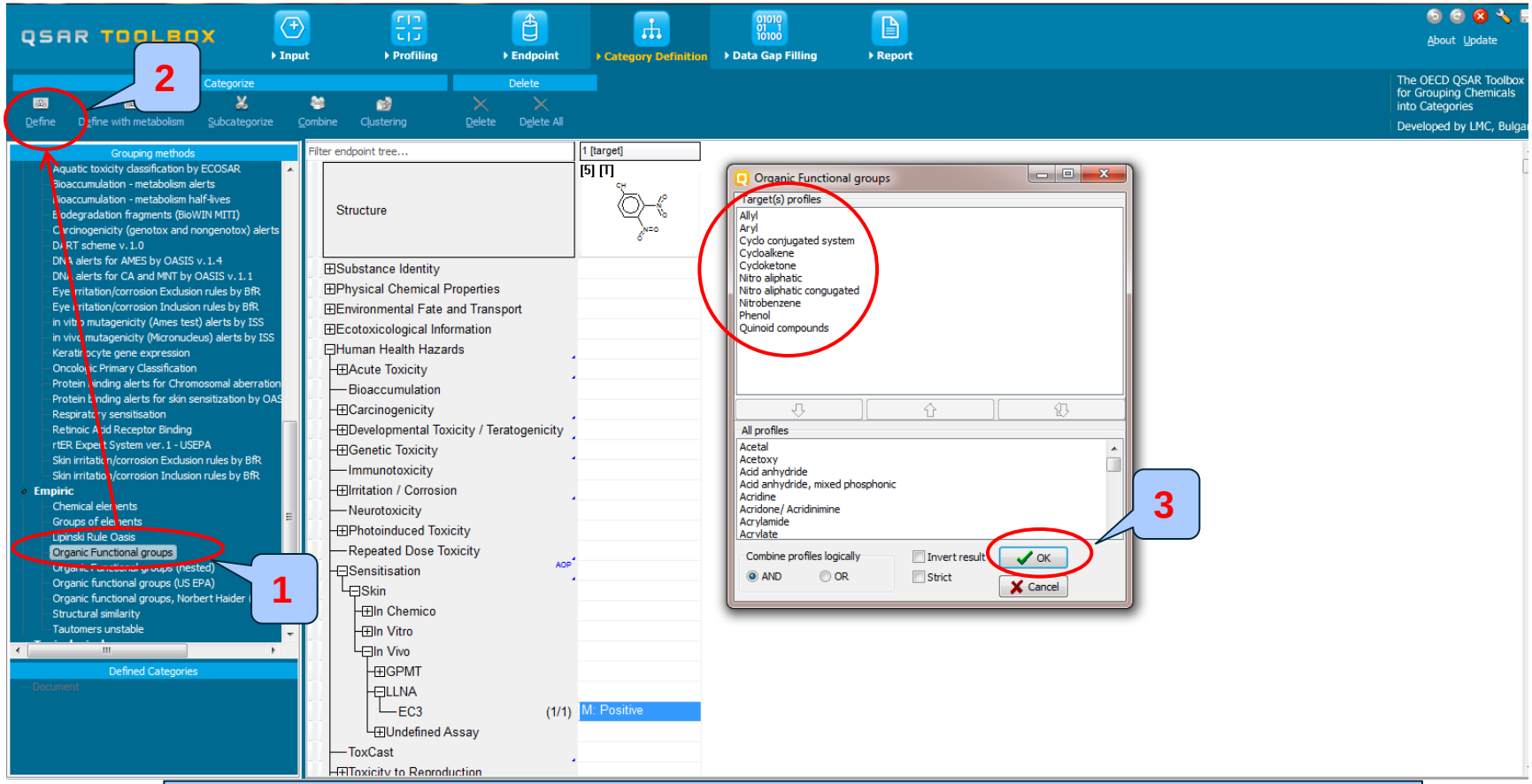
- For this example, the specific endpoint classification of target and its tautomers is identified by Organic Functional Group. Consistency of the category member is reached and phase I could be skipped (point 4 from performing categorization, slide #33).
- For this example initial group of analogues presented as tautomeric sets is identified by Organic Functional Group profiler
- Software search analogues presented as tautomeric sets having same protein binding distribution as those of the target tautomeric set

Category definition is a tool for grouping chemicals. For more details see tutorials posted on LMC and OECD website:

http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm

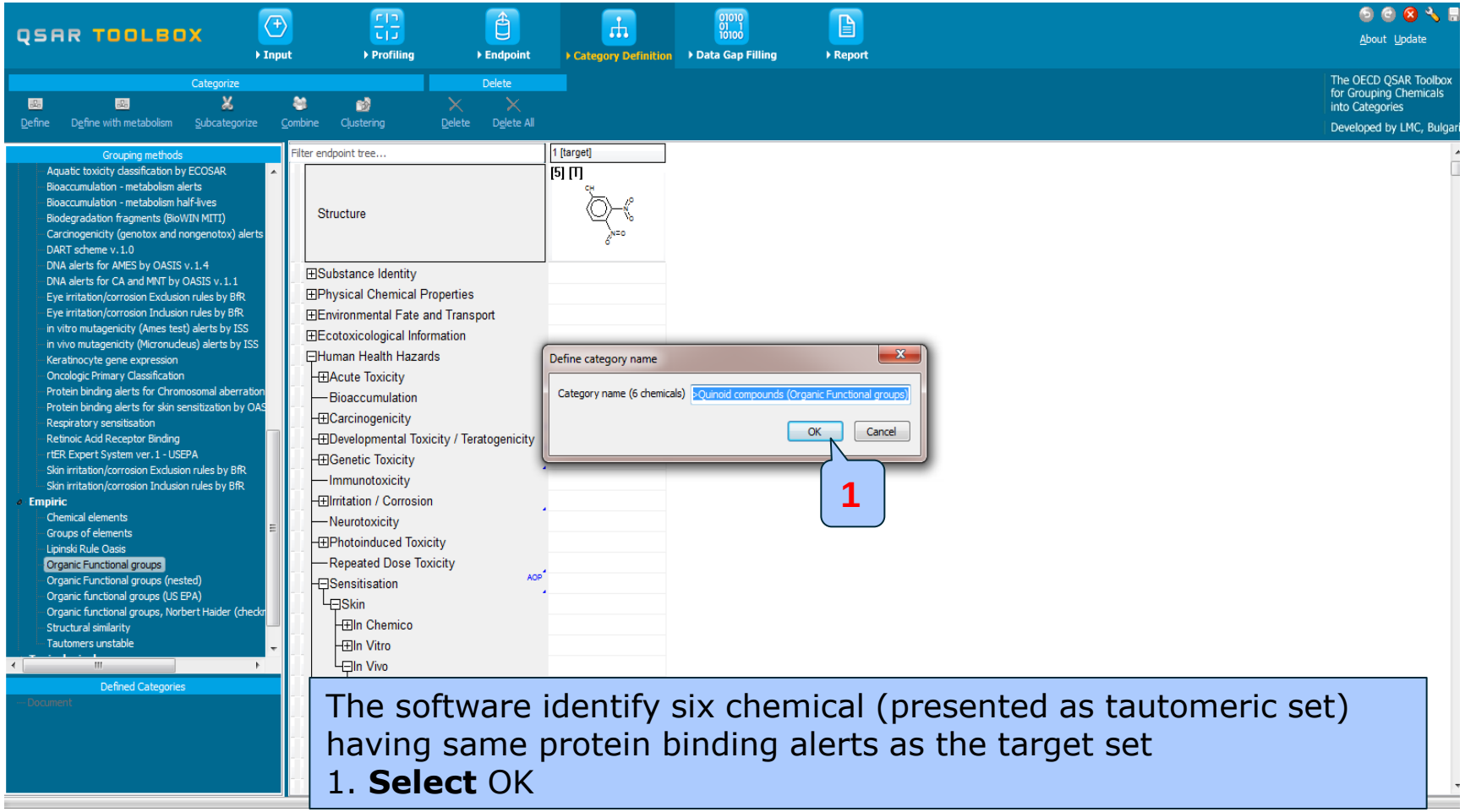
http://superhosting.oasis-lmc.org/products/software/toolbox/toolbox-support.aspx

Category Definition
Defining OFG



1. **Highlight** the "Organic Functional Group" 2. **Click** Define
3. **Click** OK to confirm the defined categories for the tautomeric set

Category Definition
Defining OFG



The software identify six chemical (presented as tautomeric set) having same protein binding alerts as the target set

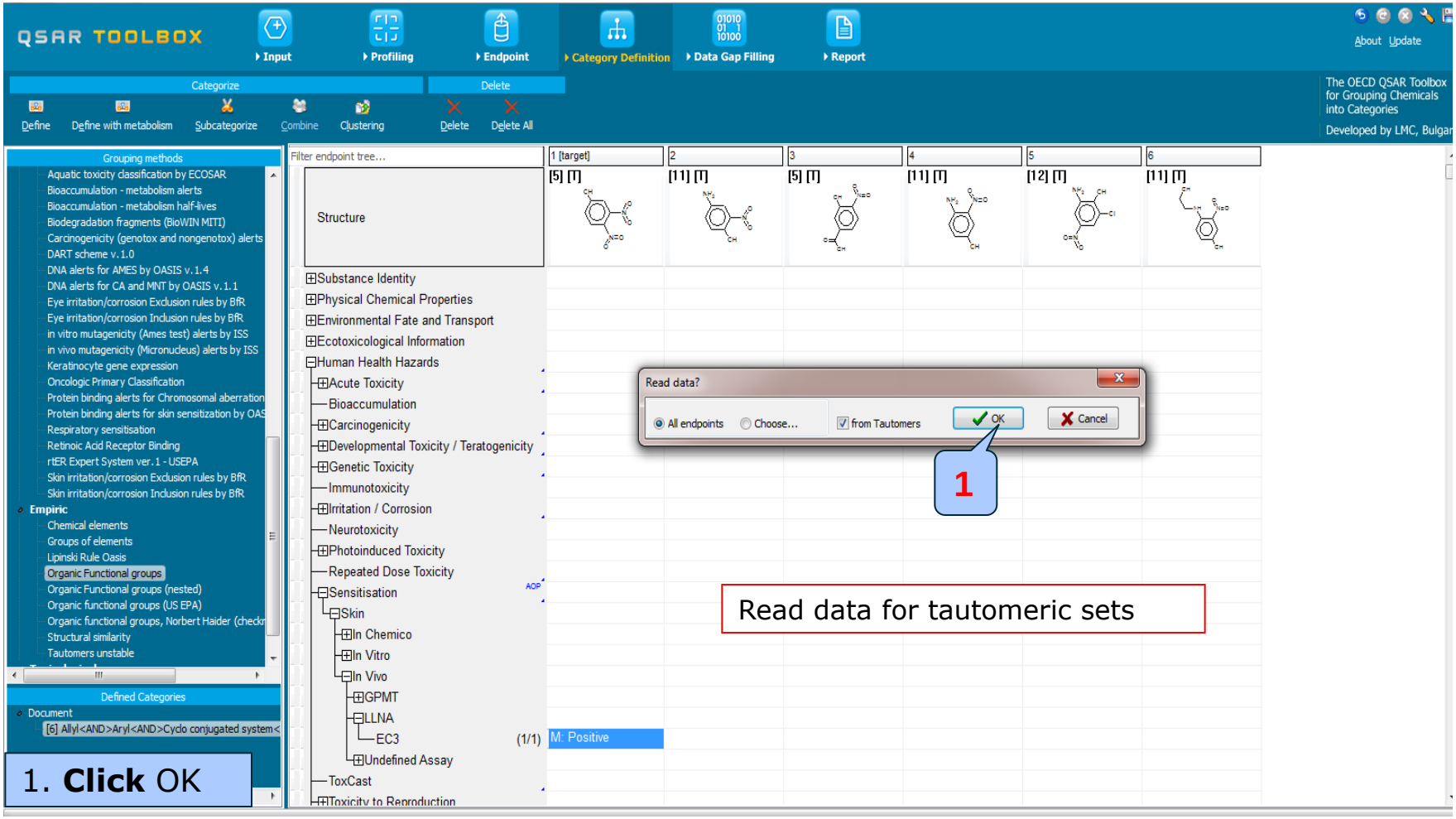
1. **Select** OK

Category Definition
Read data for Analogues

- The Toolbox automatically request 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).

- In this example, since only databases that contain information for human health hazards endpoints are selected, both options give same results.
- As the Toolbox must search the database, this may take some time.

Category Definition
Read data for Analogues



Recap

- You have identified a category of analogues presented as tautomeric sets having same distribution of functional groups as the target tautomeric set
- The available experimental results for these 6 analogues have been collected from the selected databases (Skin sensitization and Skin sensitization ECETOC)
- But before the user can proceed with the "Filling Data Gap" module, he/she should navigate through the endpoint tree and find the specific gap that will be filled

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.
- The data tree is extensive but logically constructed; it can be mastered with a practice.
- In this example, the "EC3" is the target endpoint.
- You can navigate through the endpoint tree: Double-click on the node next to Human Health Hazards then effect Sensitisation, followed by Skin, type of method In Vivo and assay LLNA and finally EC3 (see next screen shots)

Recap

- You have now retrieved the available skin sensitisation data for the four analogues represented by their tautomeric forms.

- You have identified the target endpoint of “Sensitization /Skin/In vivo/LLNA/EC3”.

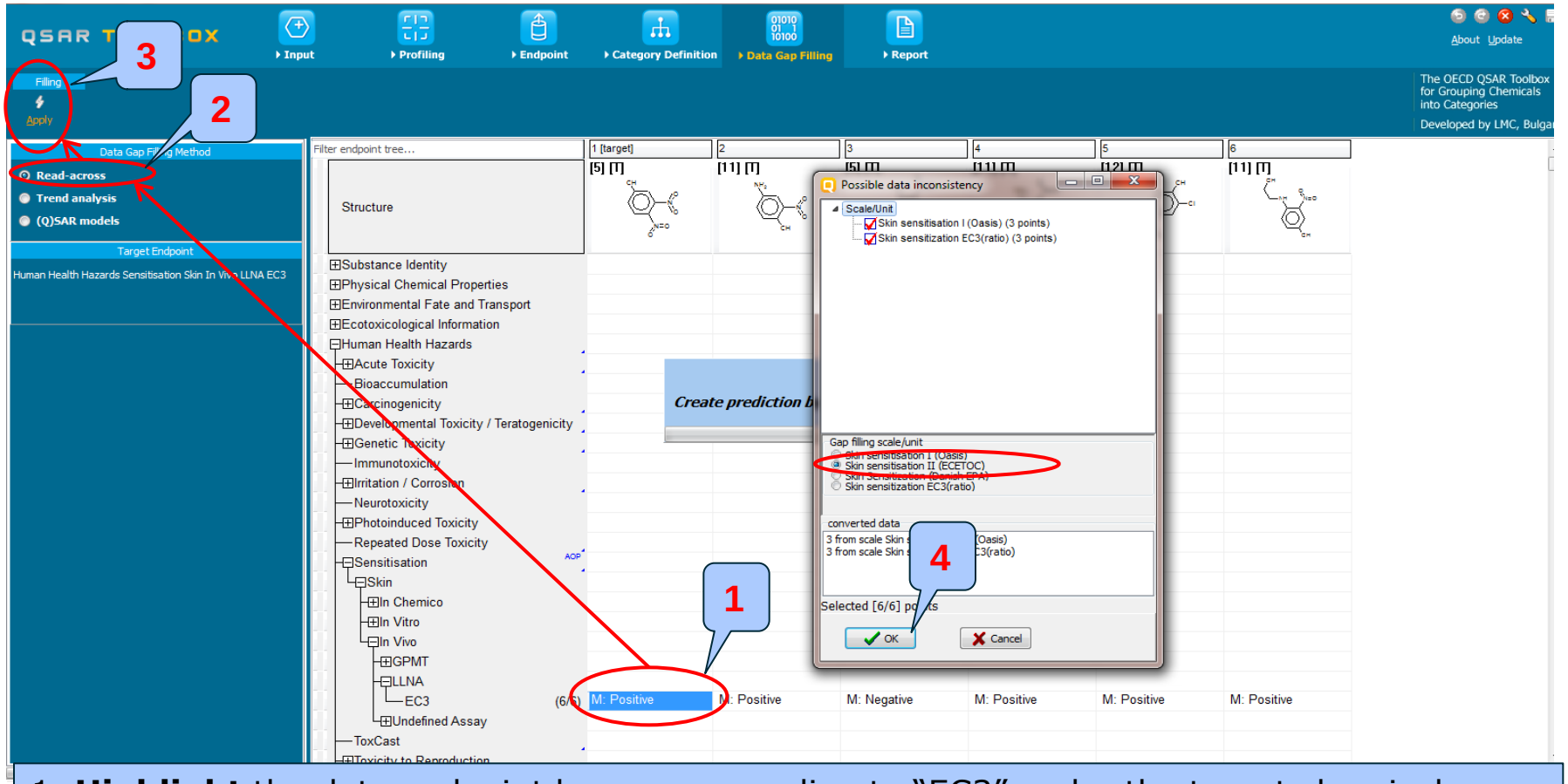
- You are ready to fill in the data gap, so click on “Data Gap Filling” (see next screen shot).

Outlook

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

Data Gap Filling

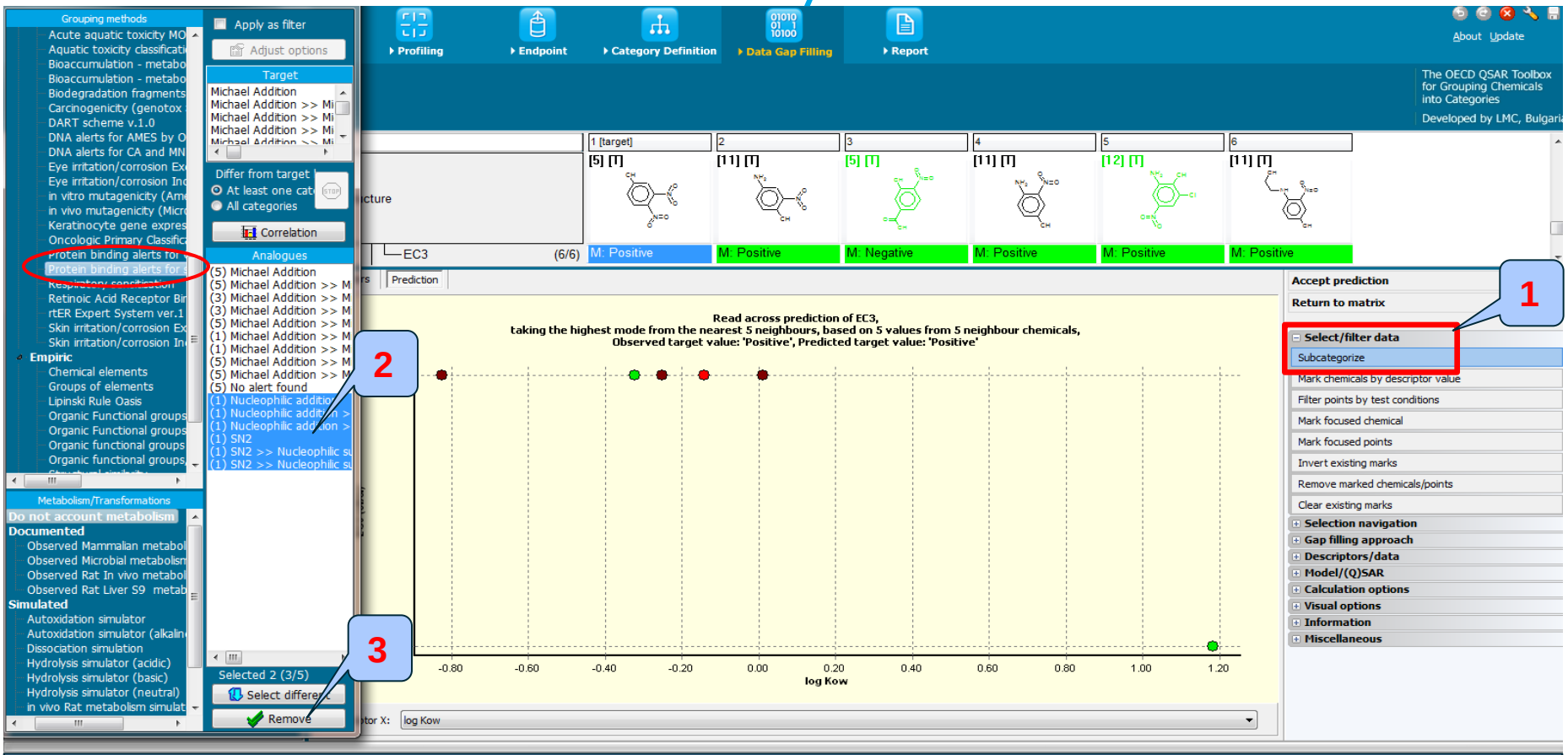
Apply Read across analysis



1. **Highlight** the data endpoint box corresponding to "EC3" under the target chemical.
2. **Select** Read-across 3. **Click** Apply 4. **OK**

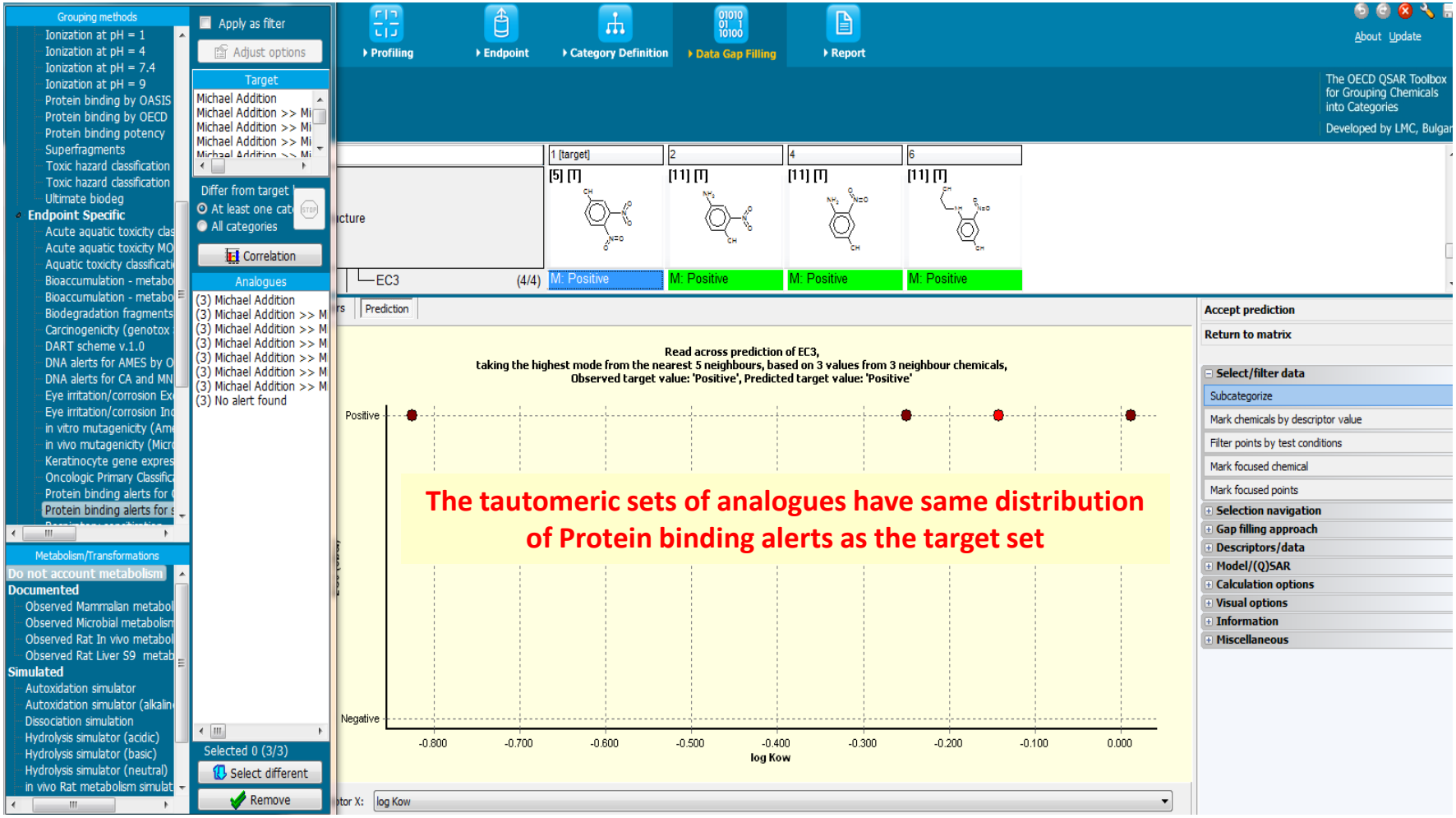
Data Gap Filling

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



1. **Select** Select/Filter data and then **Subcategorization**
2. **Select** Protein binding alerts for skin sensitization by OASIS v1.4
3. **Click** Remove

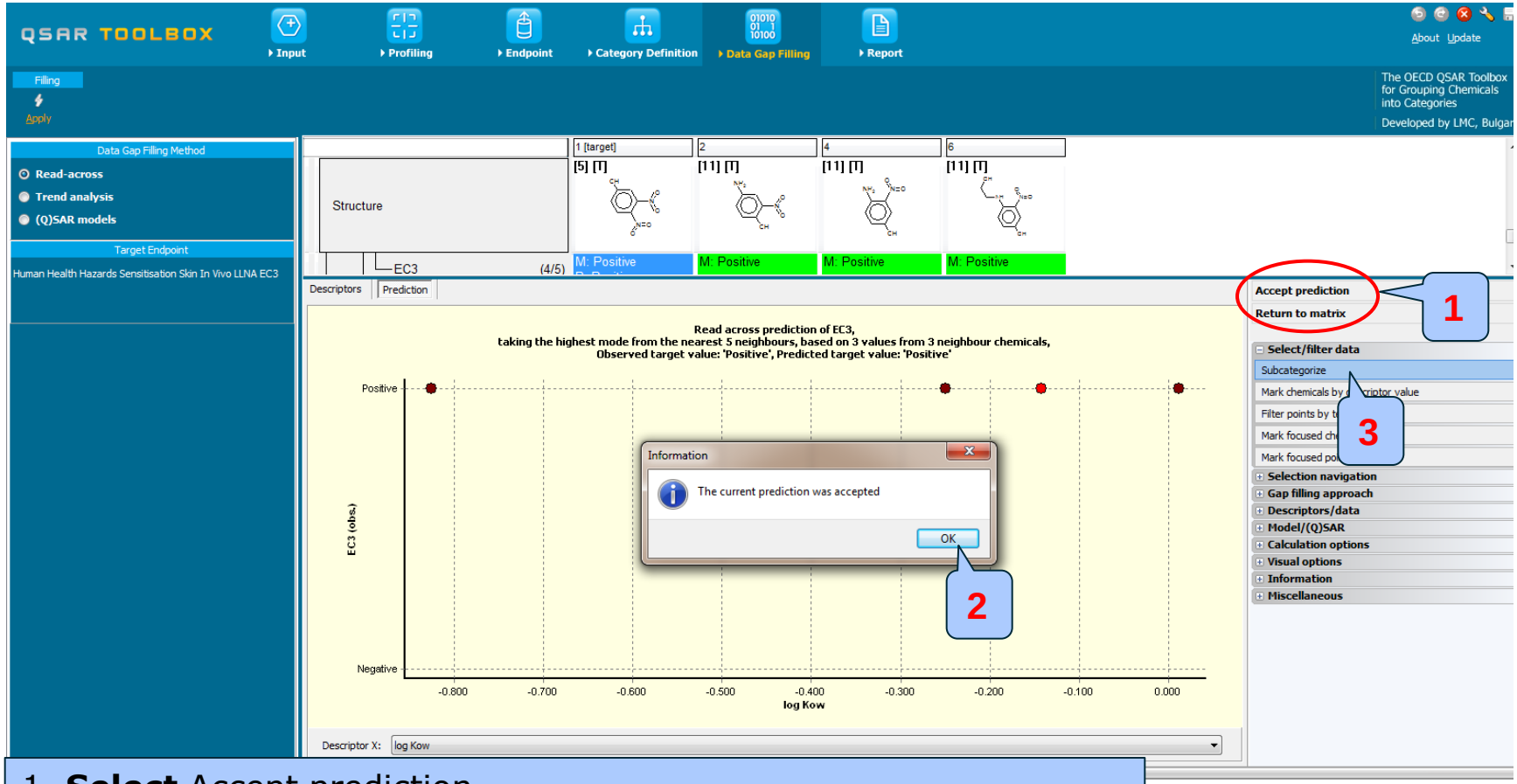
Data Gap Filling
Result of Subcategorisation



Data Gap Filling
Interpreting Read-across

- In this example, all results of the analogues are consistent; all the analogues are Skin sensitizers.
- The tautomeric sets of analogues have same distribution of Protein binding alerts as the target set.
- 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
Result of Read-across

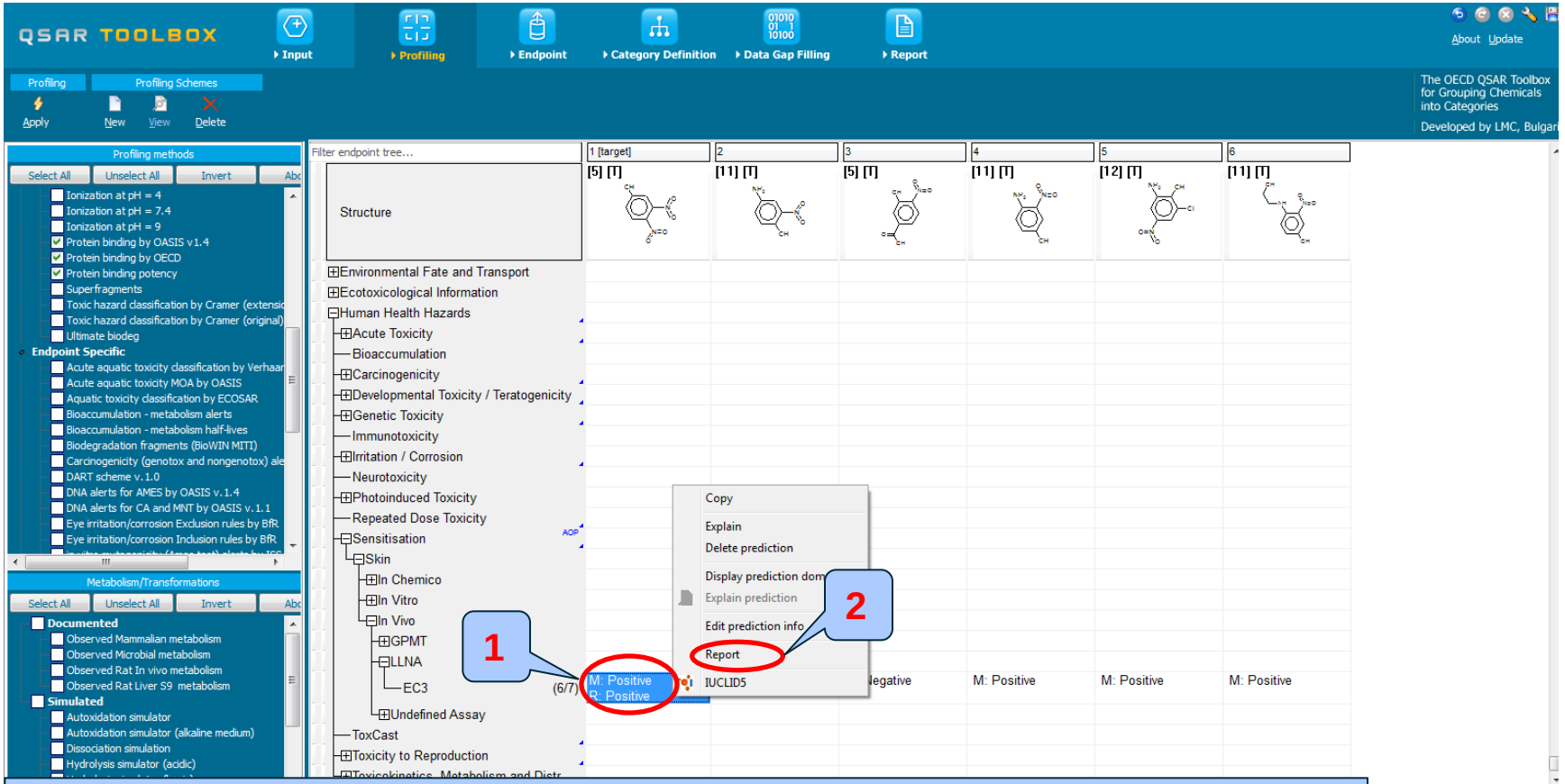


1. **Select** Accept prediction
2. **Click OK** and then 3. **Click** Return to matrix

Outlook

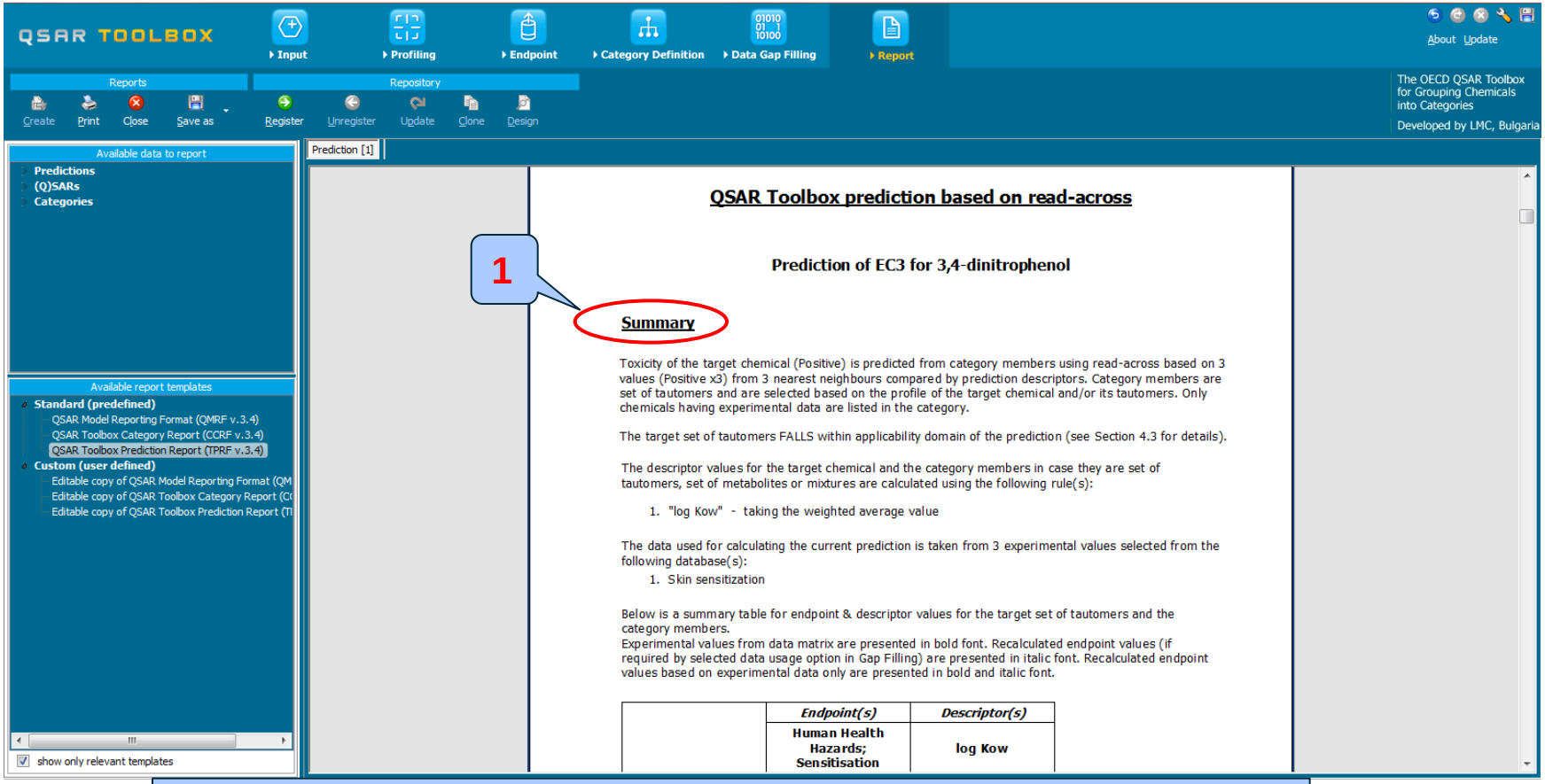
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Category definition
  - Data Gap Filling
  - **Report**

Report



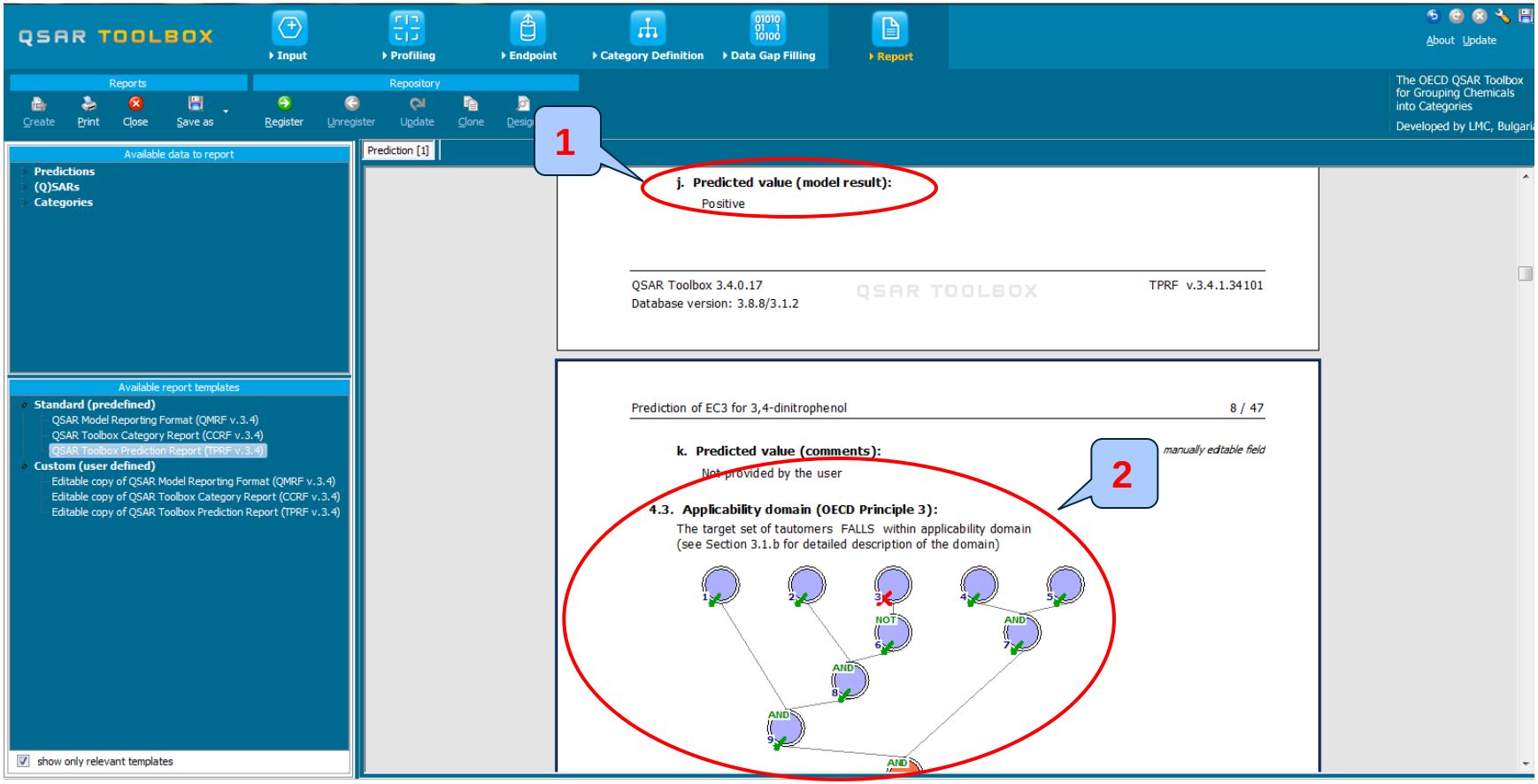
1. **Select** prediction
2. **Right Click** and **Select** Report

Report



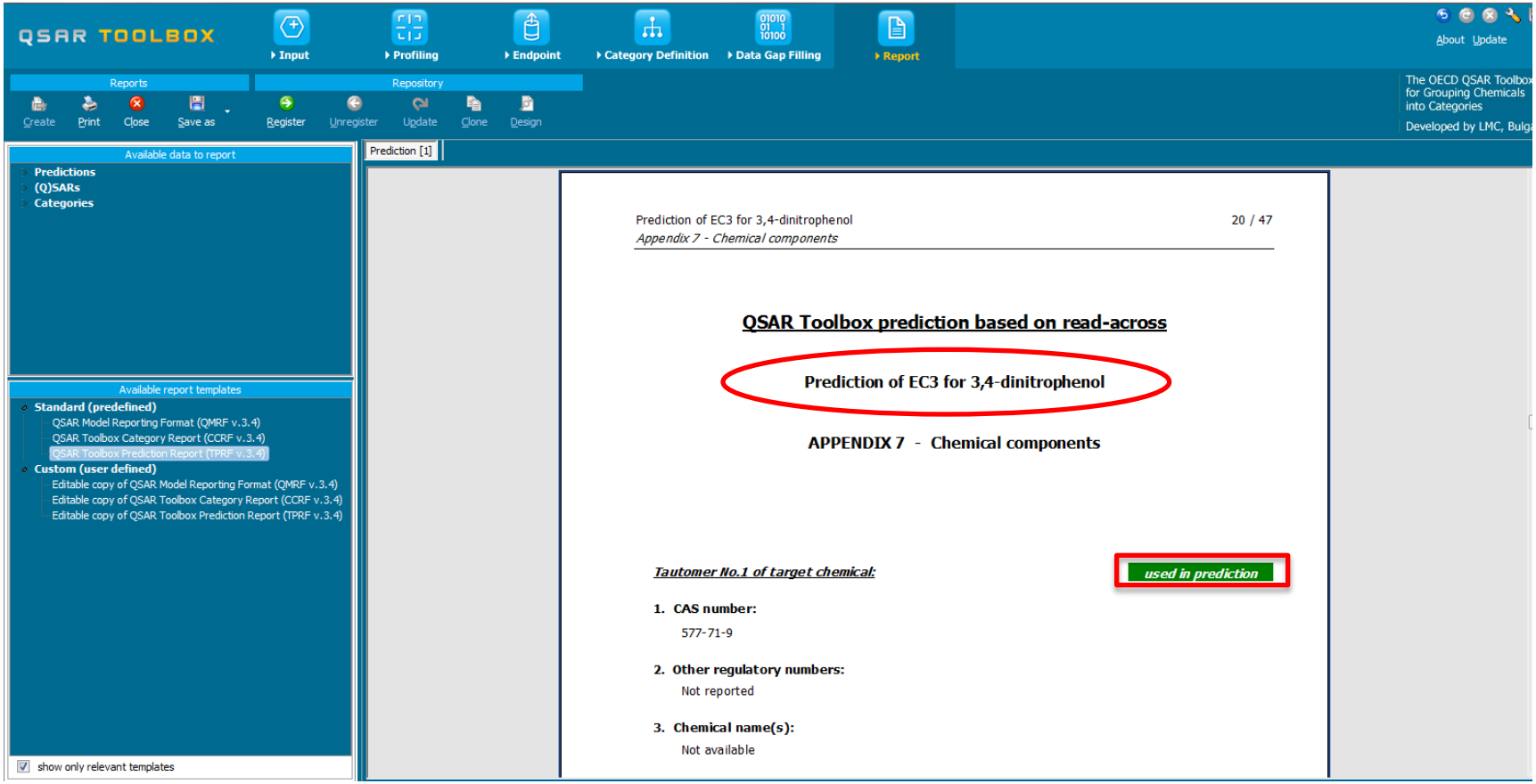
1. Summary information for the prediction of tautomeric set

Report



1. Predicted value 2. Applicability domain

Report



Additional Appendix 7 list tautomers of target and analogue chemicals used in read-across
Also an information about which tautomer is used in the RA prediction is provided.

Outlook

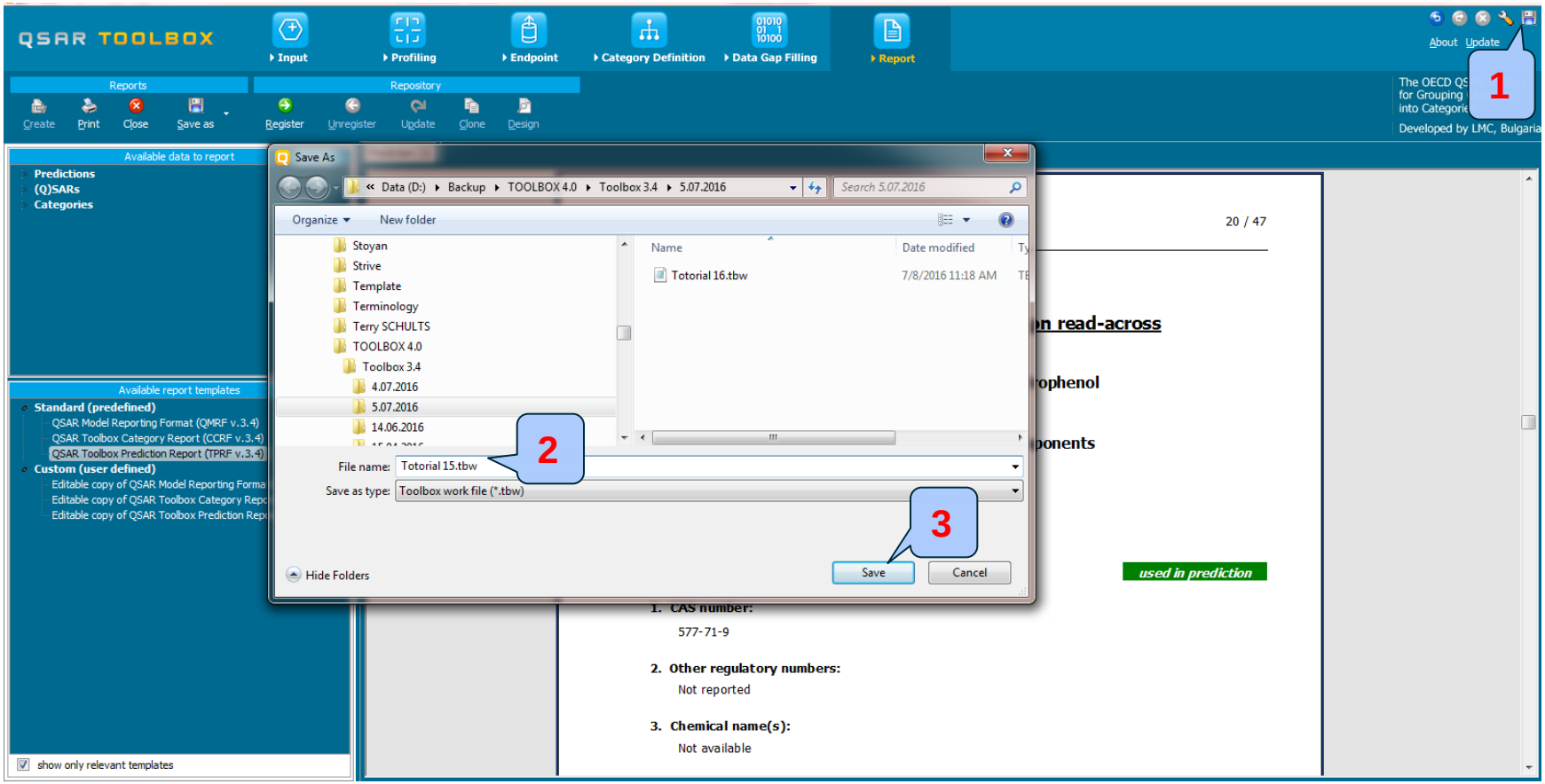
- Background
- Objectives
- The exercise
- Workflow
- **Save prediction**

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).

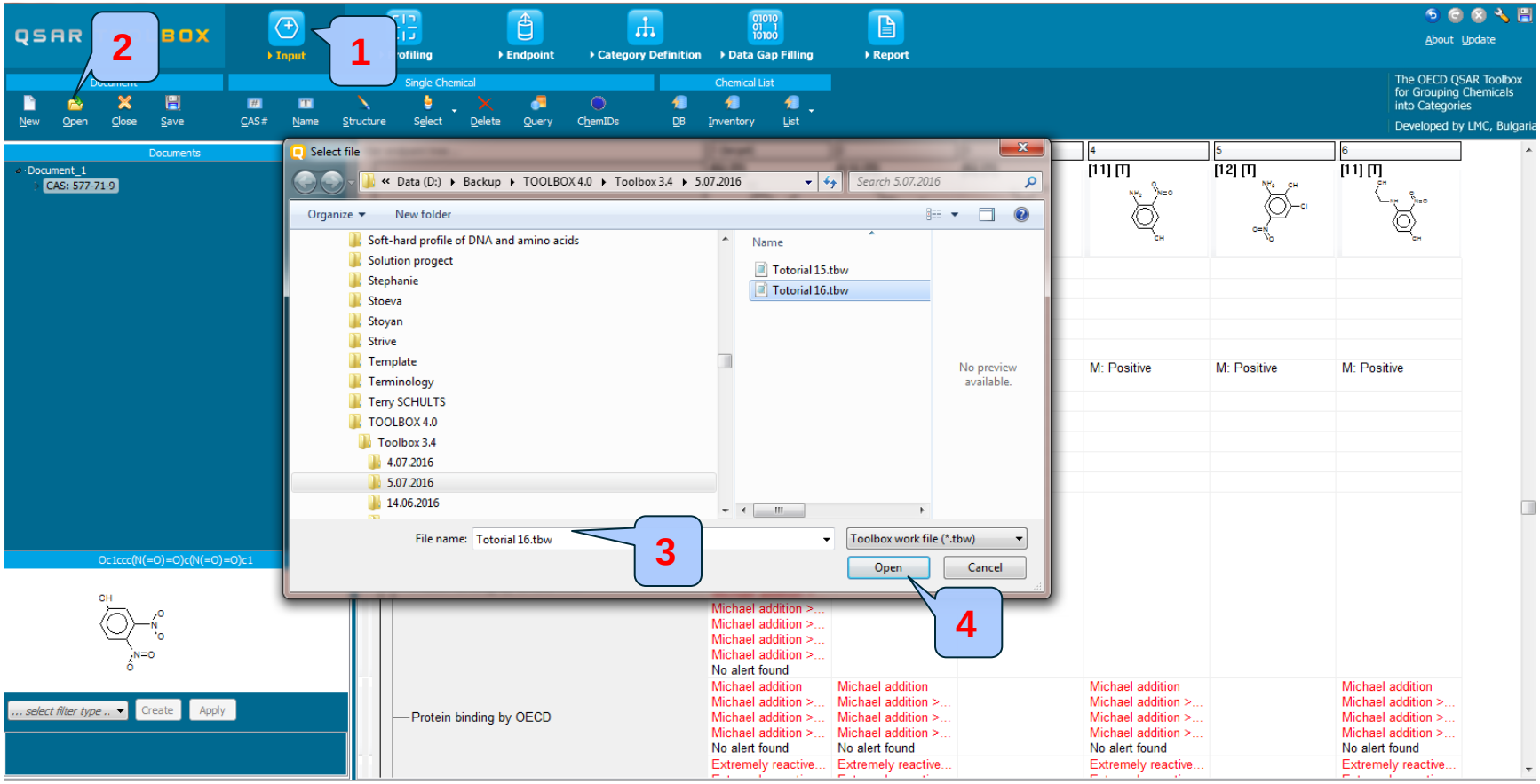
- Saving/Loading the file with TB prediction is shown on next screenshots

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