QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Example for predicting Repeated dose toxicity of 2,3-dimethylaniline

Outlook

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

Background

- This is a step-by-step presentation designed to take the user through the workflow for filling data gap for Repeated dose toxicity by read-across based on an analogue approach.

Outlook

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

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

The exercise

- In this exercise we will predict the repeated dose toxicity of **2,3-dimethylaniline CAS 87-59-2**
- Define initial category of similar analogues based on US-EPA New chemical categories.
- Gather available experimental data for the target chemical and identified analogues
- Apply read across prediction based on analogue approach

Outlook

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

Workflow

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

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 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
- Query Tool

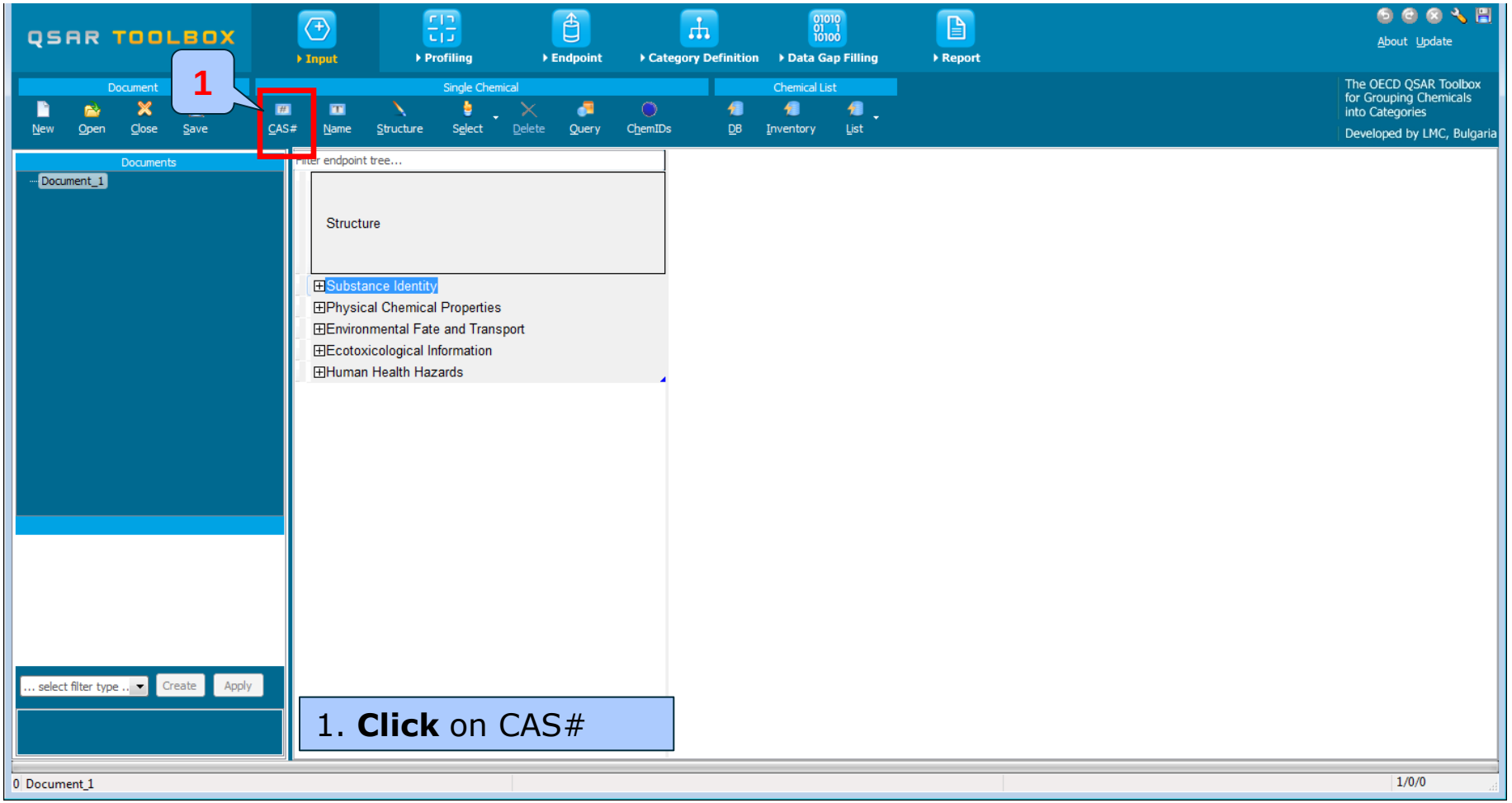
B. Group of chemicals
- User List/Inventory
- Specialized Databases

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)

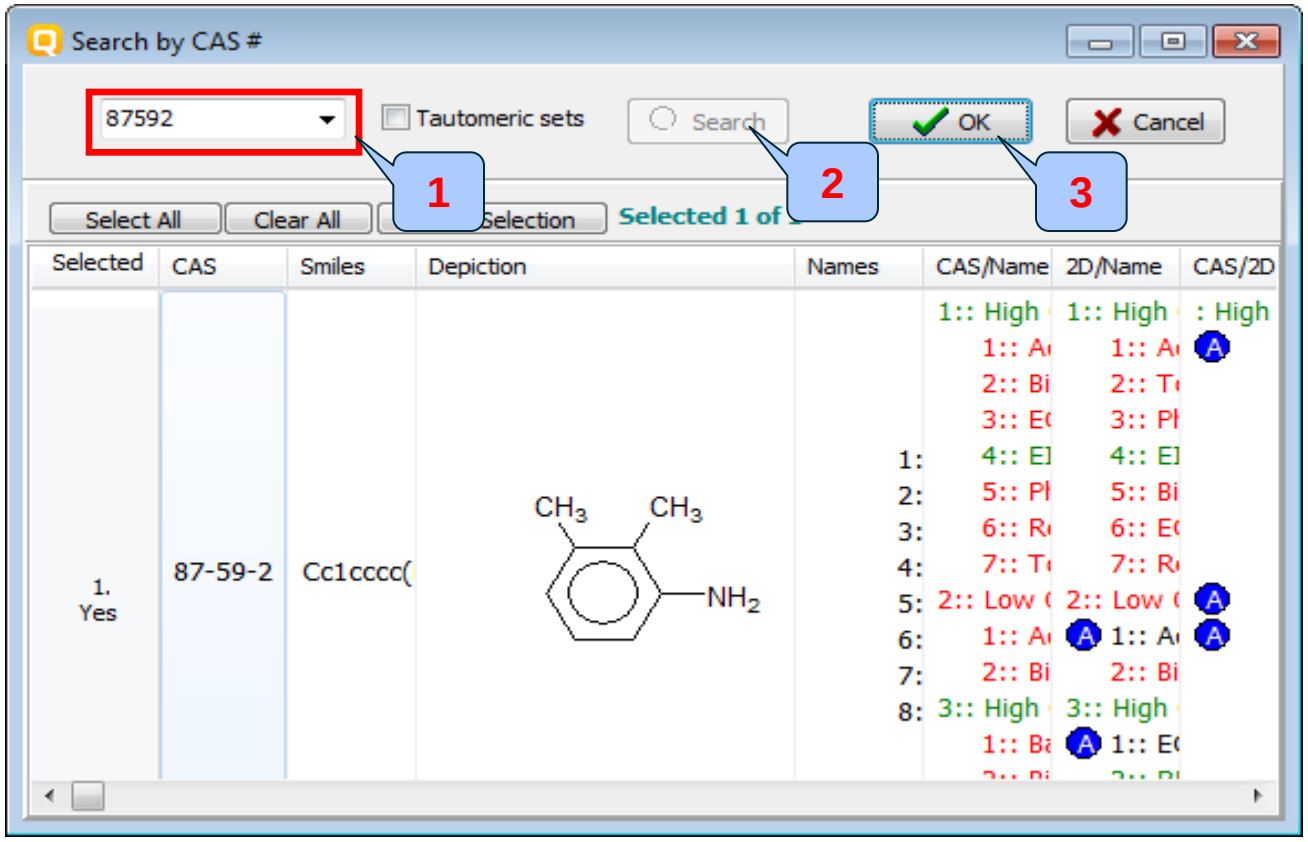
Chemical Input Screen

Input target chemical by CAS#



Chemical Input Screen

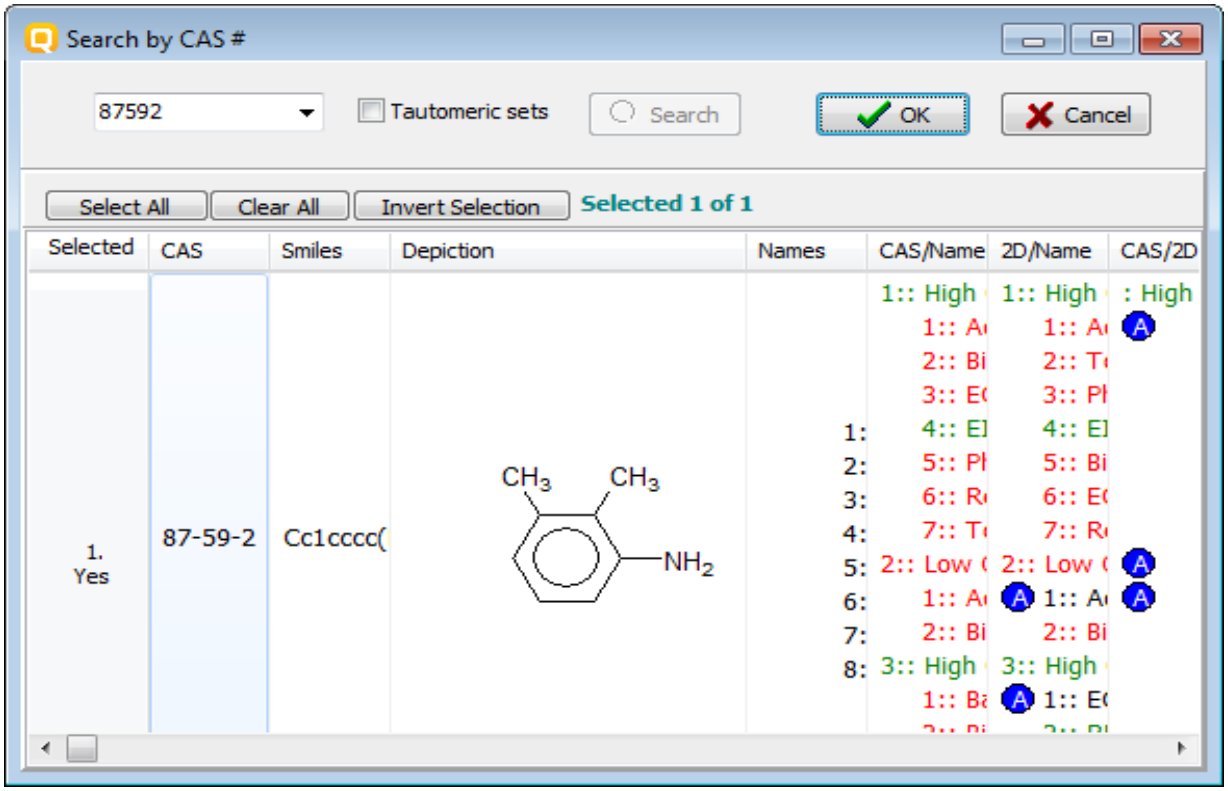
Enter CAS# 87-59-2



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

Chemical Input
Target chemical identity

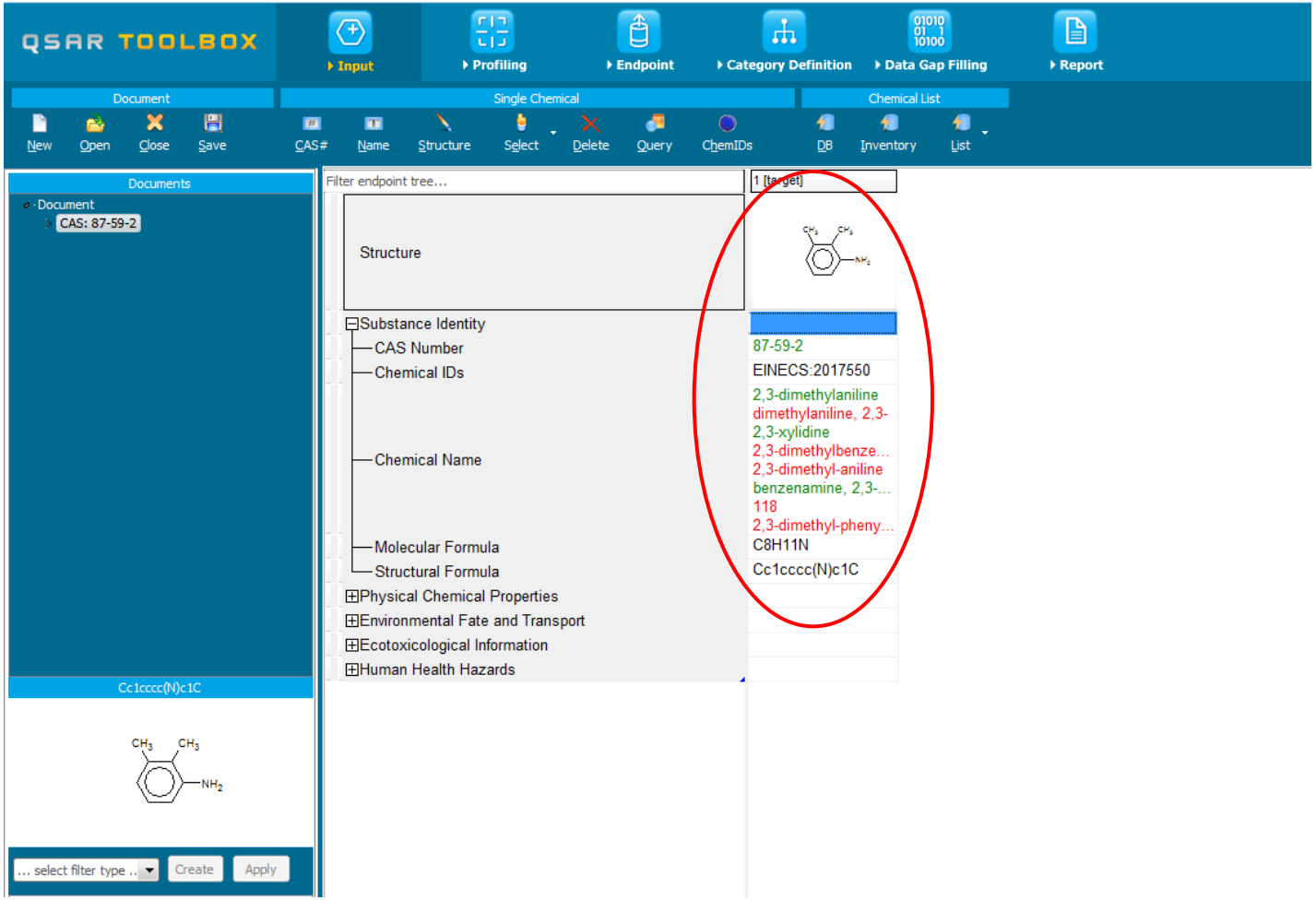
The Toolbox now searches the databases to find out if the CAS# you entered is linked to a molecular structure stored in the Toolbox. It is displayed as a 2-demensional depiction.



Chemical Input
Target chemical identity

- Double click "Substance Identity" displays the chemical identification information.
- The user should note that existing names of the target chemical are presented in different colours. This indicates the reliability of relation CAS-Name for the target chemical(see next screen shots).
- The workflow on the first module is now complete, and the user can proceed to the next module.

Chemical Input
Target chemical identity



Outlook

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

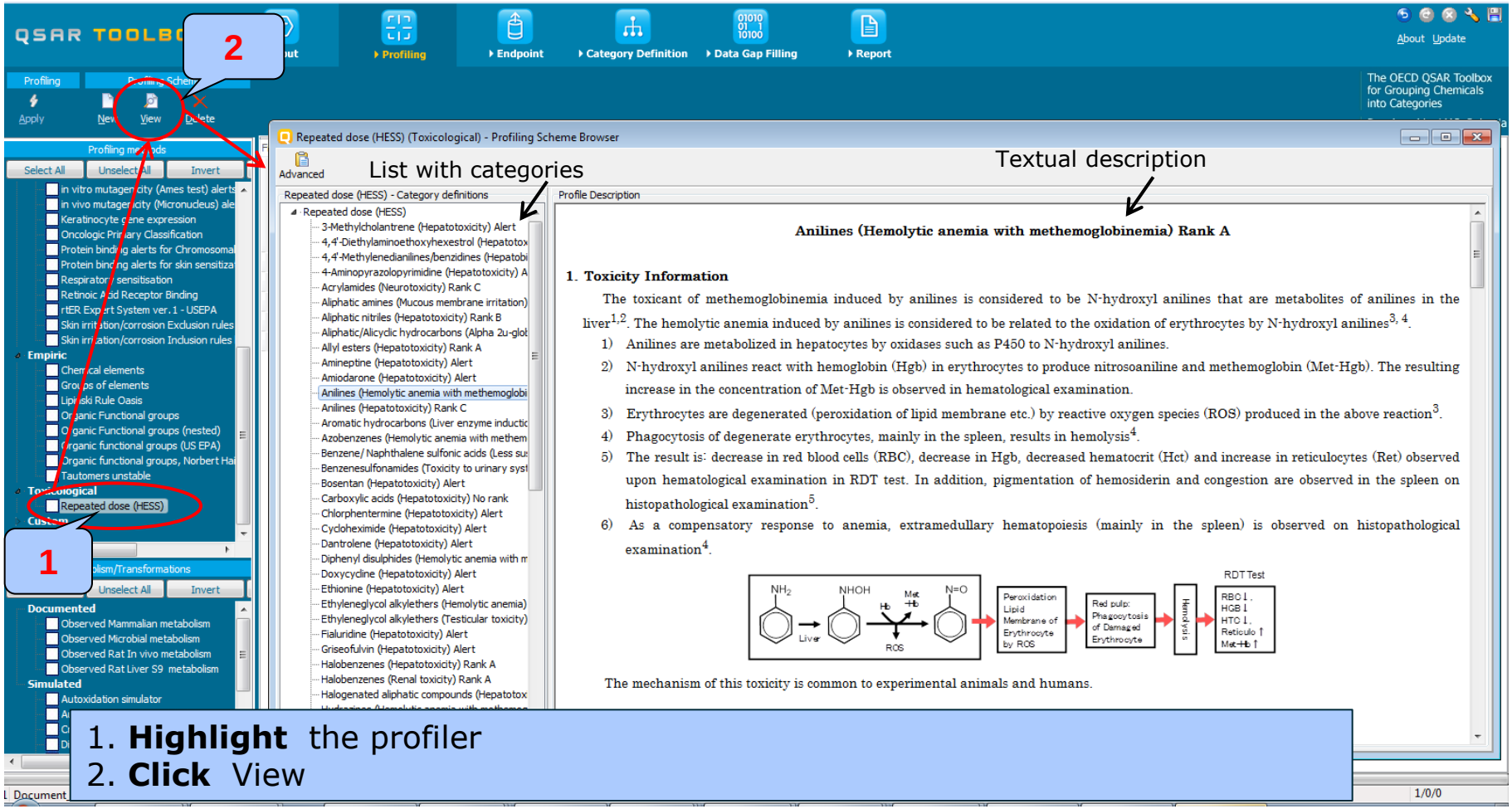
Profiling
Overview

- "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.

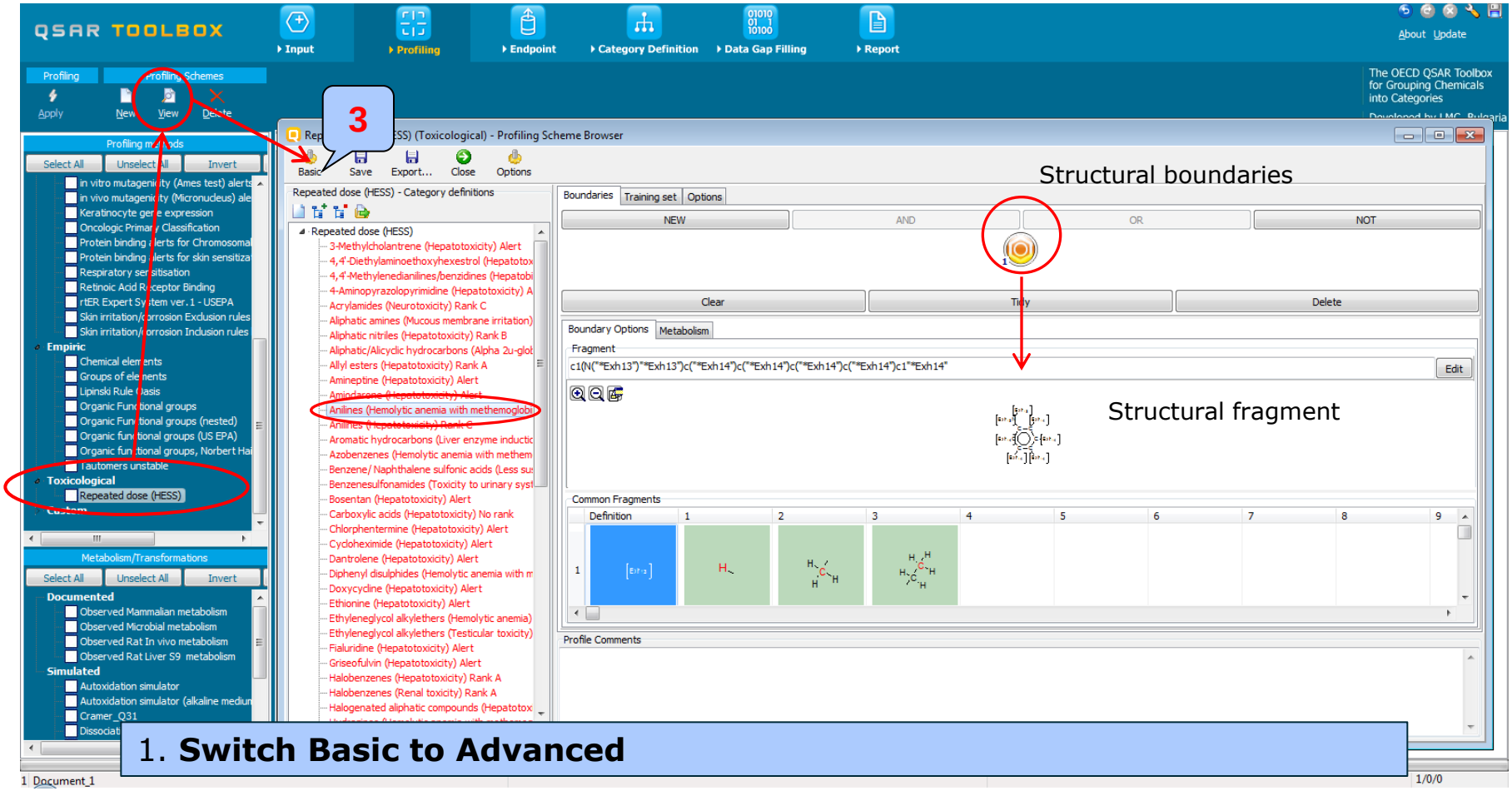
Profiling
Side-Bar to Profiling

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

Profiling
Side-Bar to Profiling

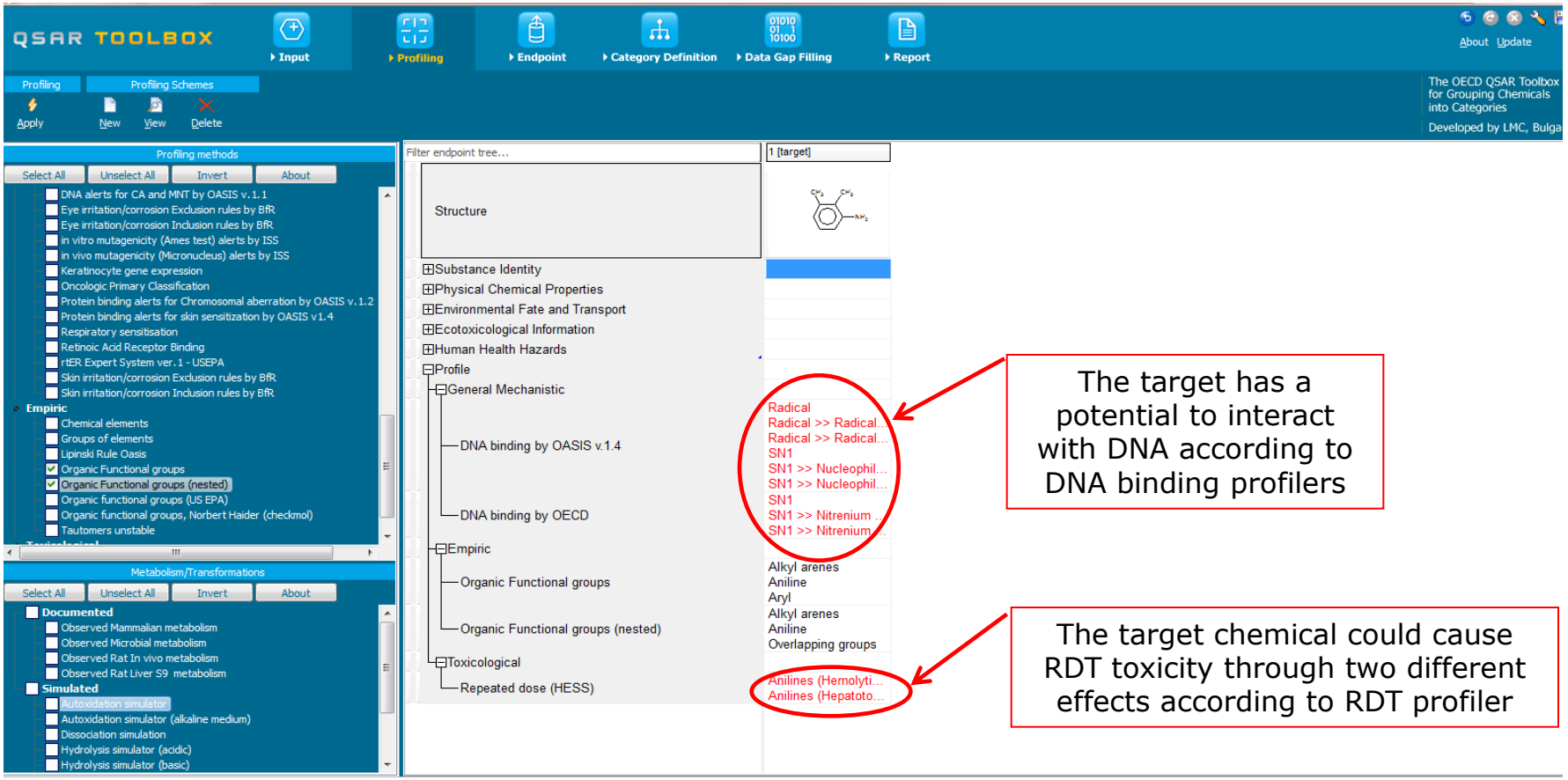


Profiling
Side-Bar to Profiling

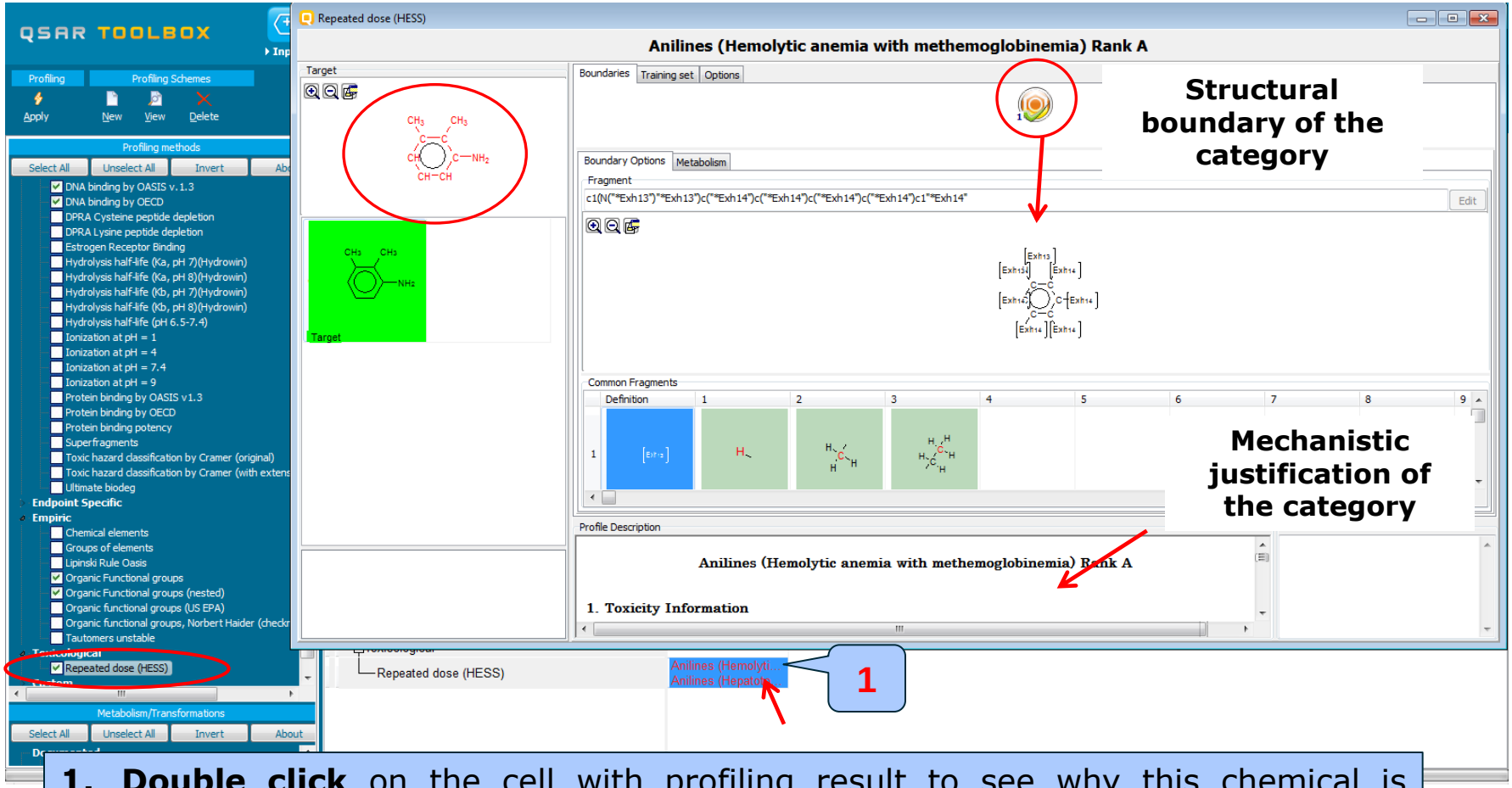


Profiling

Side-Bar to Profiling results



Profiling
Side-Bar to Profiling



1. **Double click** on the cell with profiling result to see why this chemical is classified as Anilines

Outlook

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

Endpoint
Overview

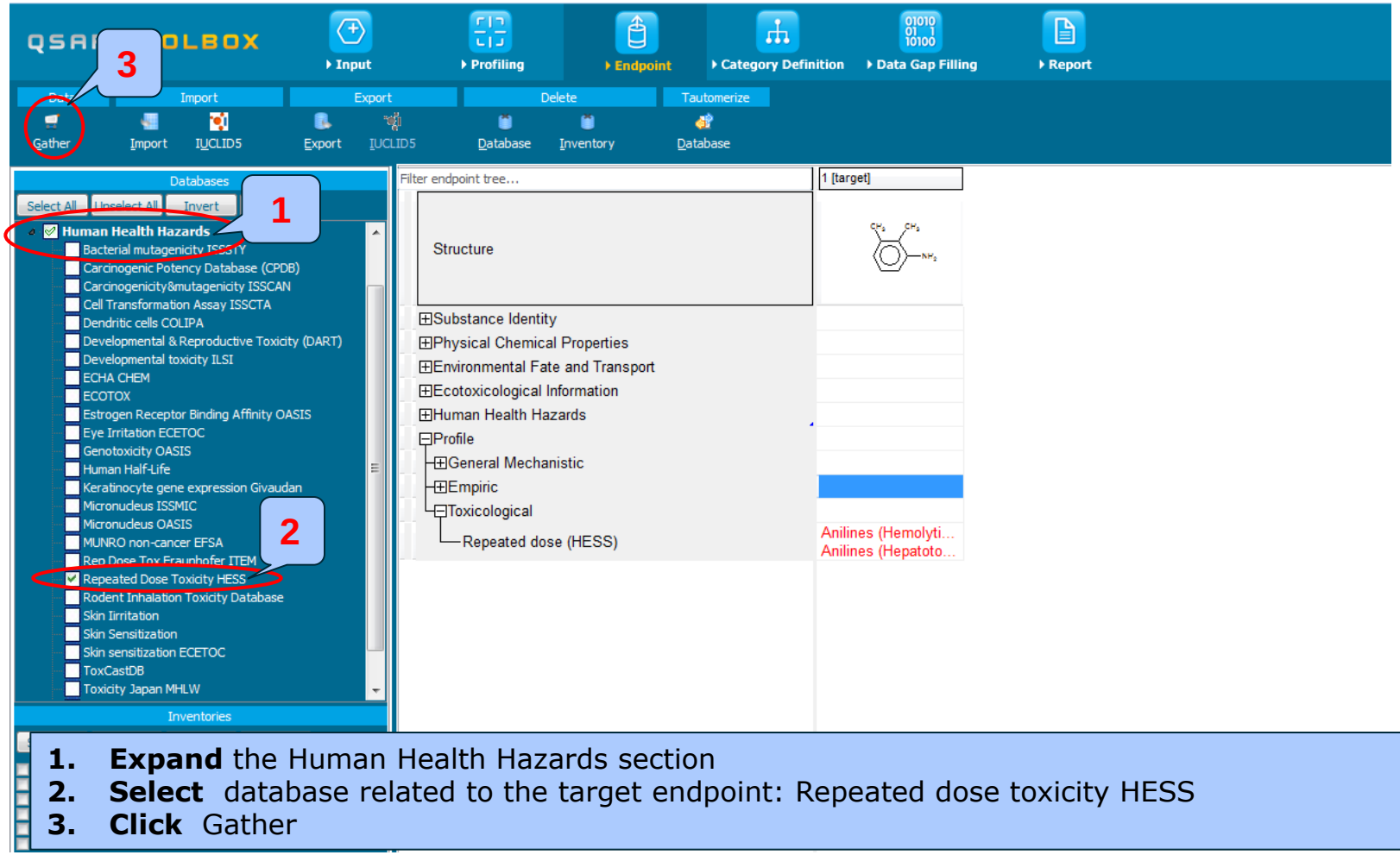
- "Endpoint" refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.

- 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

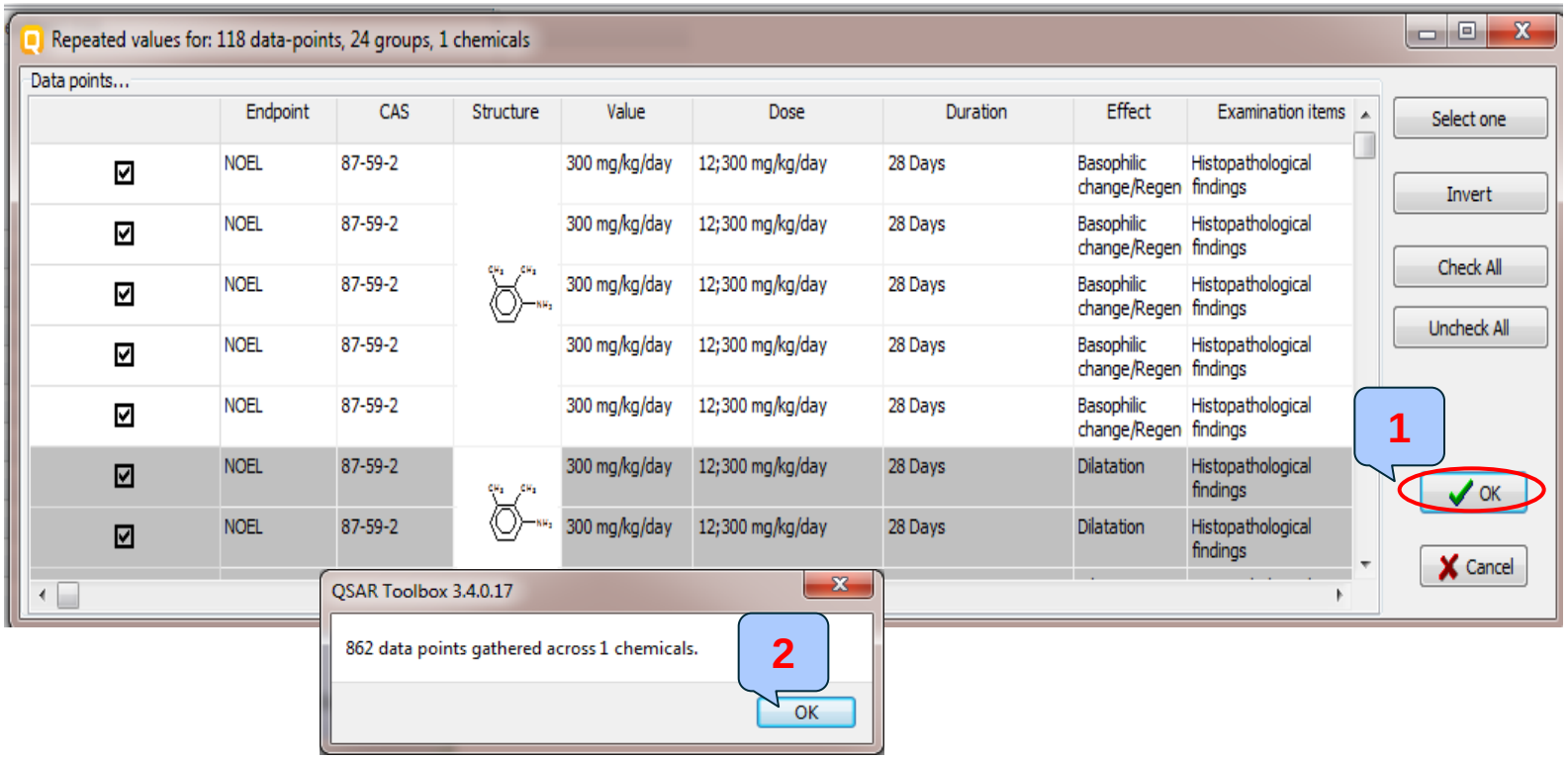
- In this example, we limit our data gathering to a single toxicity endpoint: repeated dose toxicity
- In this example, we collect data from the databases containing experimental results for Repeat dose toxicity (Repeated Dose Toxicity (HESS)).
- Click on "Endpoint" in the Toolbox workflow.
- Expand the "Human Health Hazards" section
- Click on the box to select that database.
- Click on "Gather data" (see next screen shot).

Endpoint
Gather data



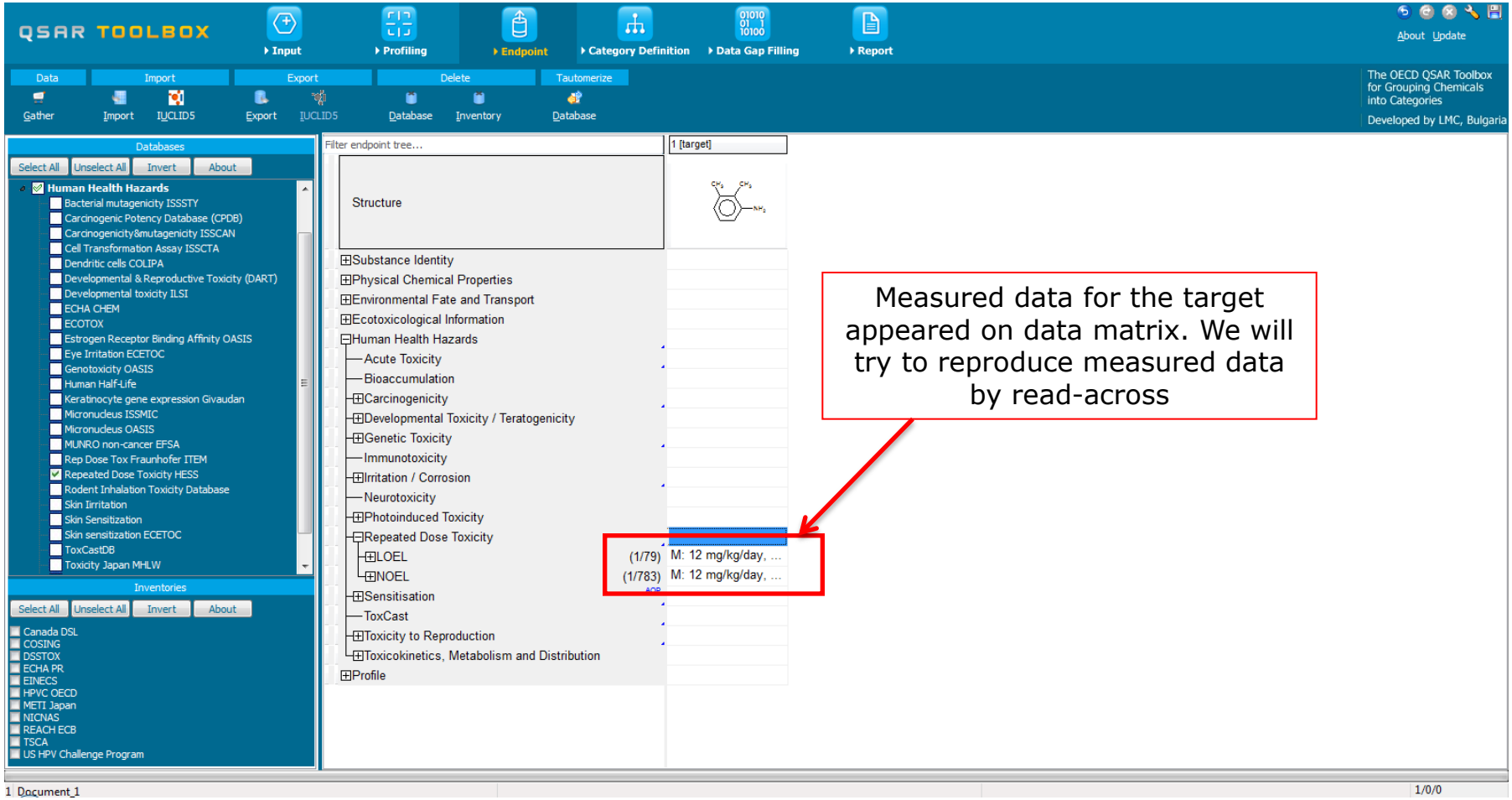
1. **Expand** the Human Health Hazards section
2. **Select** database related to the target endpoint: Repeated dose toxicity HESS
3. **Click** Gather

Endpoint
Gather data



1. **Select** OK.
2. The message informs you for number of retrieved data points. **Click** OK

Endpoint
Gather data



Recap

- In the first module, you have entered the target chemical being sure of the correctness of the structure.
- In the second module, you have profiled the target chemical and found that the target could cause RDT toxicity through two different effects
- In the third module, you have found that there is an experimental RDT data for the target structure. We will try to reproduce it using read across analysis
- But before the user can proceed with the "Filling Data Gap" module, he/she should define a category with similar analogues
- Click on "Category Definition" to move to the next module.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - 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 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
- Repeated dose profiler (NITE)

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

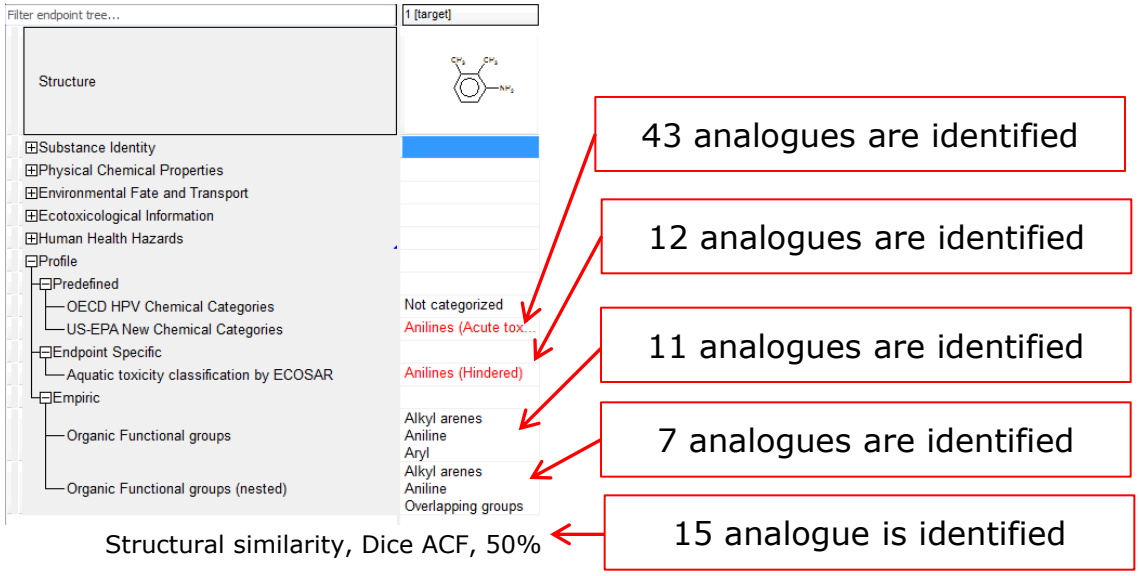
Suitable Categorization/Assessment Phases

Phase I. Structure based

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

Broad grouping
Endpoint Non-specific

Phase I categorization in Toolbox



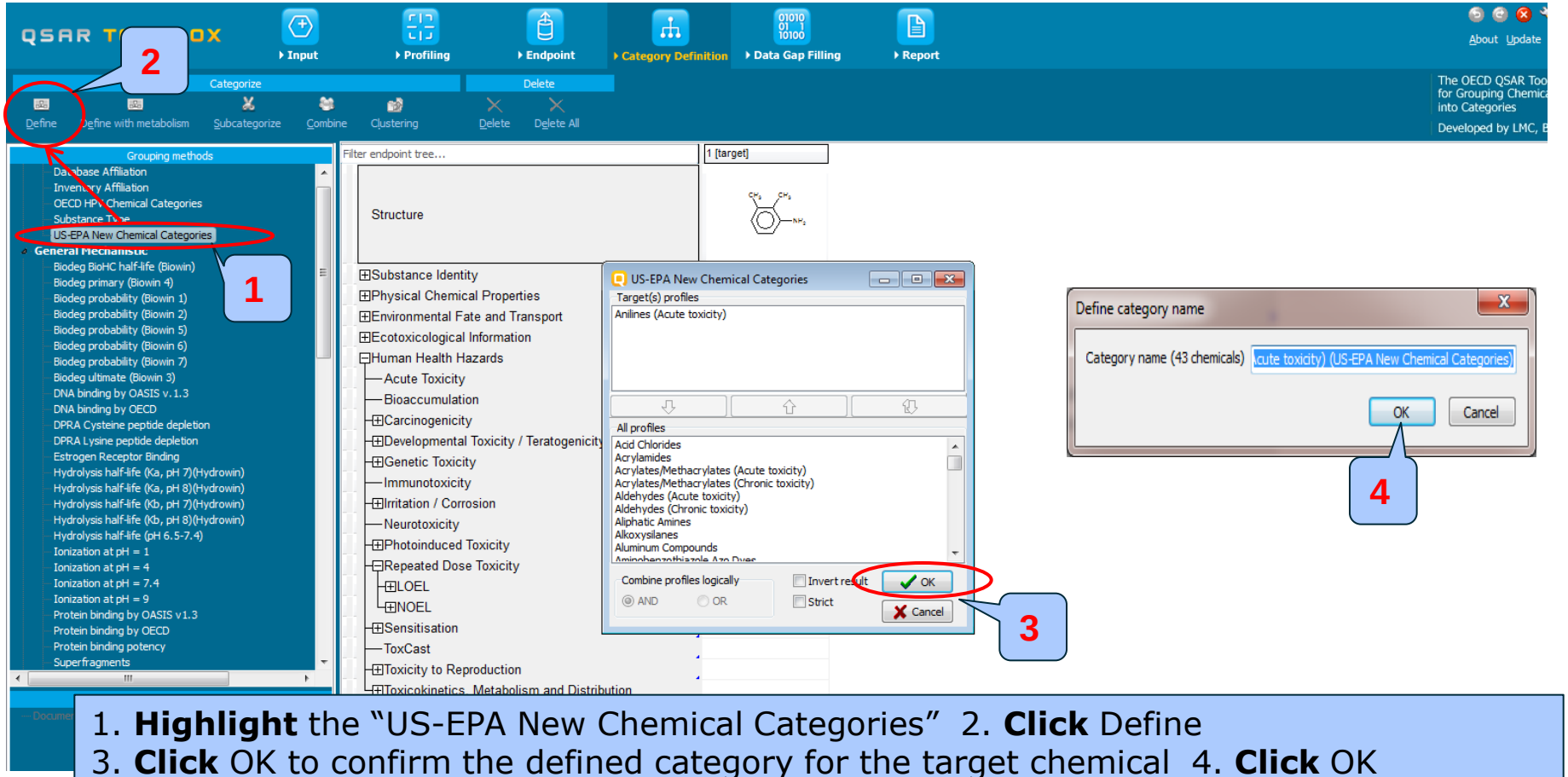
Category Definition
Grouping methods

- Based on these classifications and basic guidance for grouping chemicals explained on the previous slides the US-EPA (as broader group: 43 analogues) is used for defining initial group of analogues (phase I)

- For refinement of category and eliminating dissimilar chemicals a sequence of endpoint specific and structural profilers are applied (phase II):
  - US-EPA New chemical categories
  - Repeated dose (HESS)
  - Chemical elements
  - Structural similarity

Category Definition

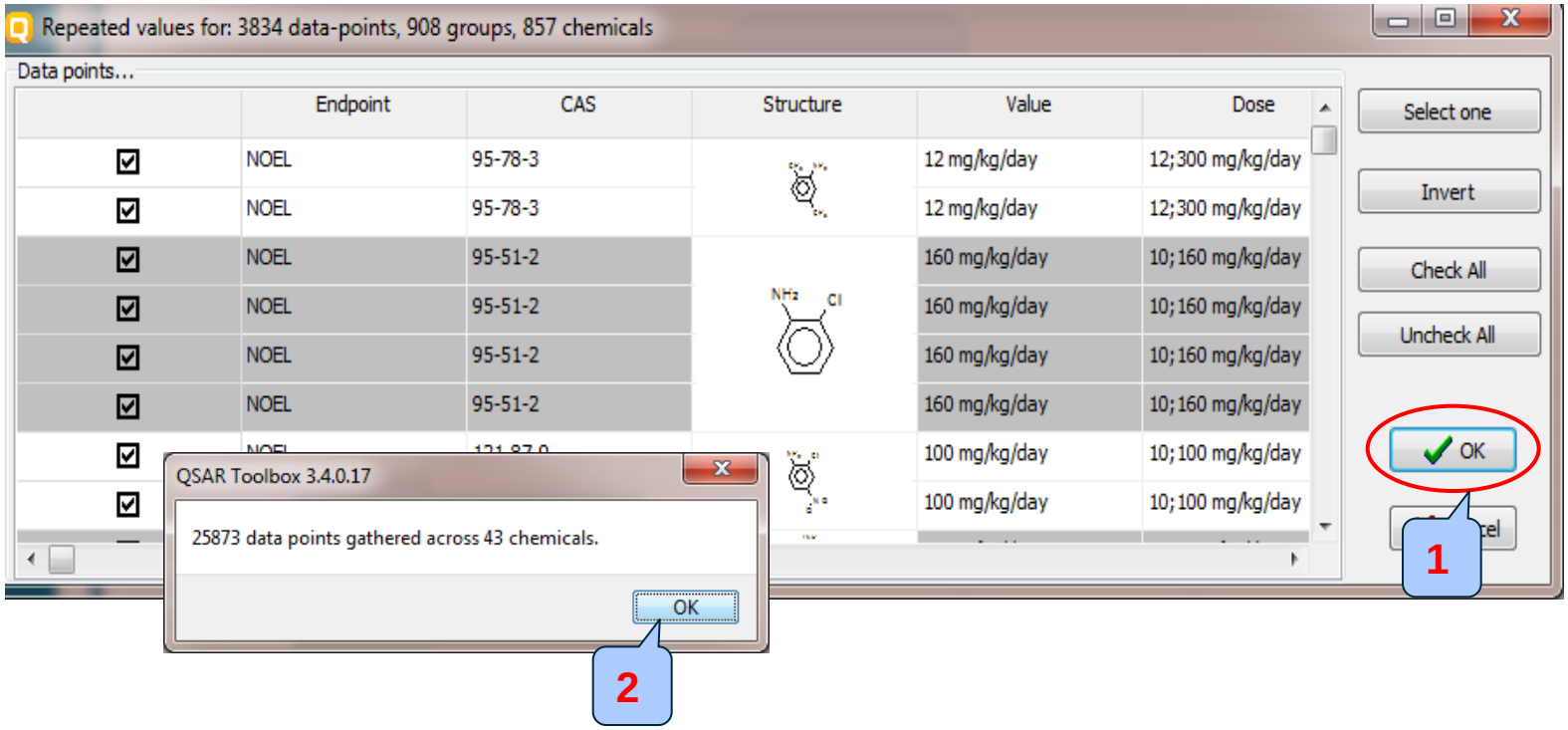
Defining US-EPA New Chemical categories



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

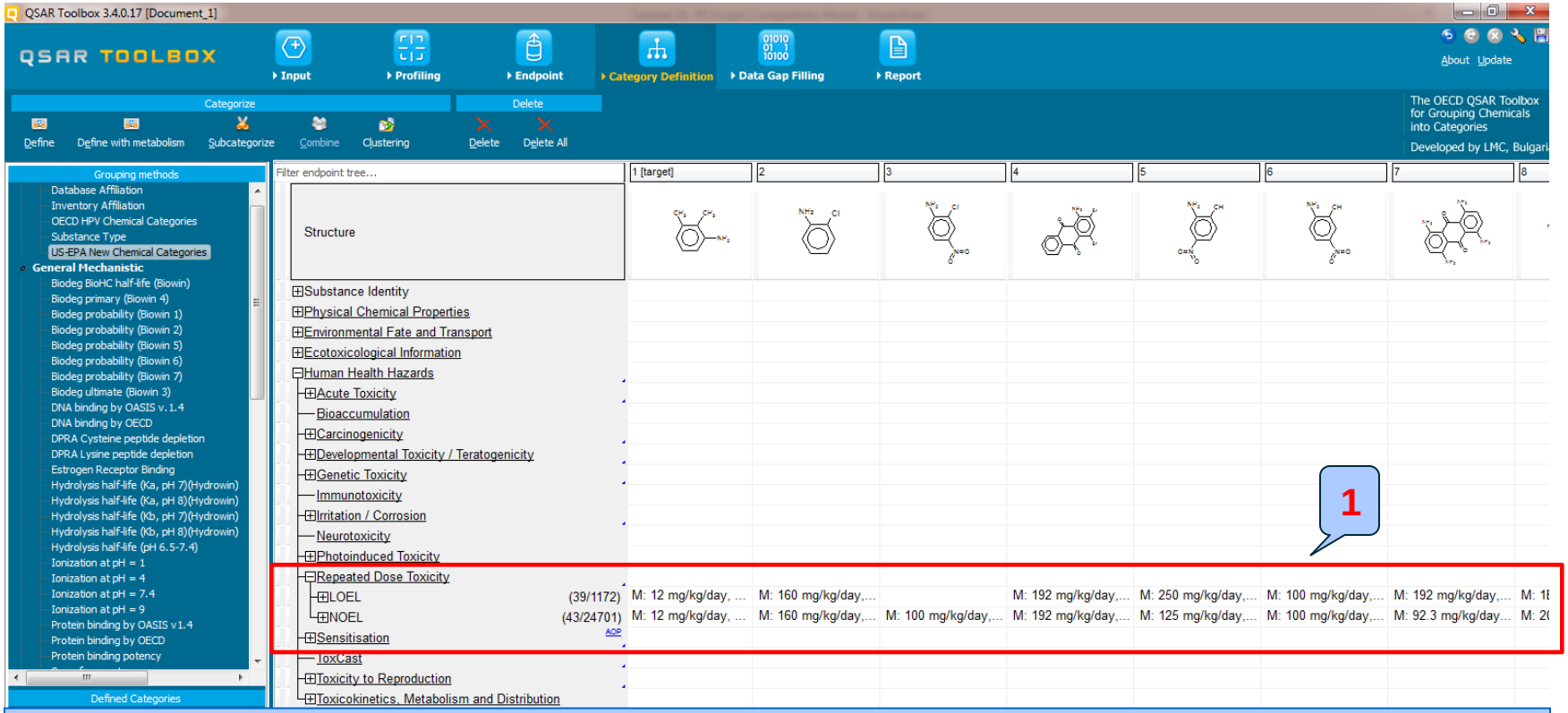
Defining US-EPA New Chemical categories



1. **Click** OK to retrieve all available experimental data
2. The message informs you for number of retrieved data points. **Click** OK

Defining US-EPA New Chemical categories

The experimental results for the analogues appeared on datamatrix

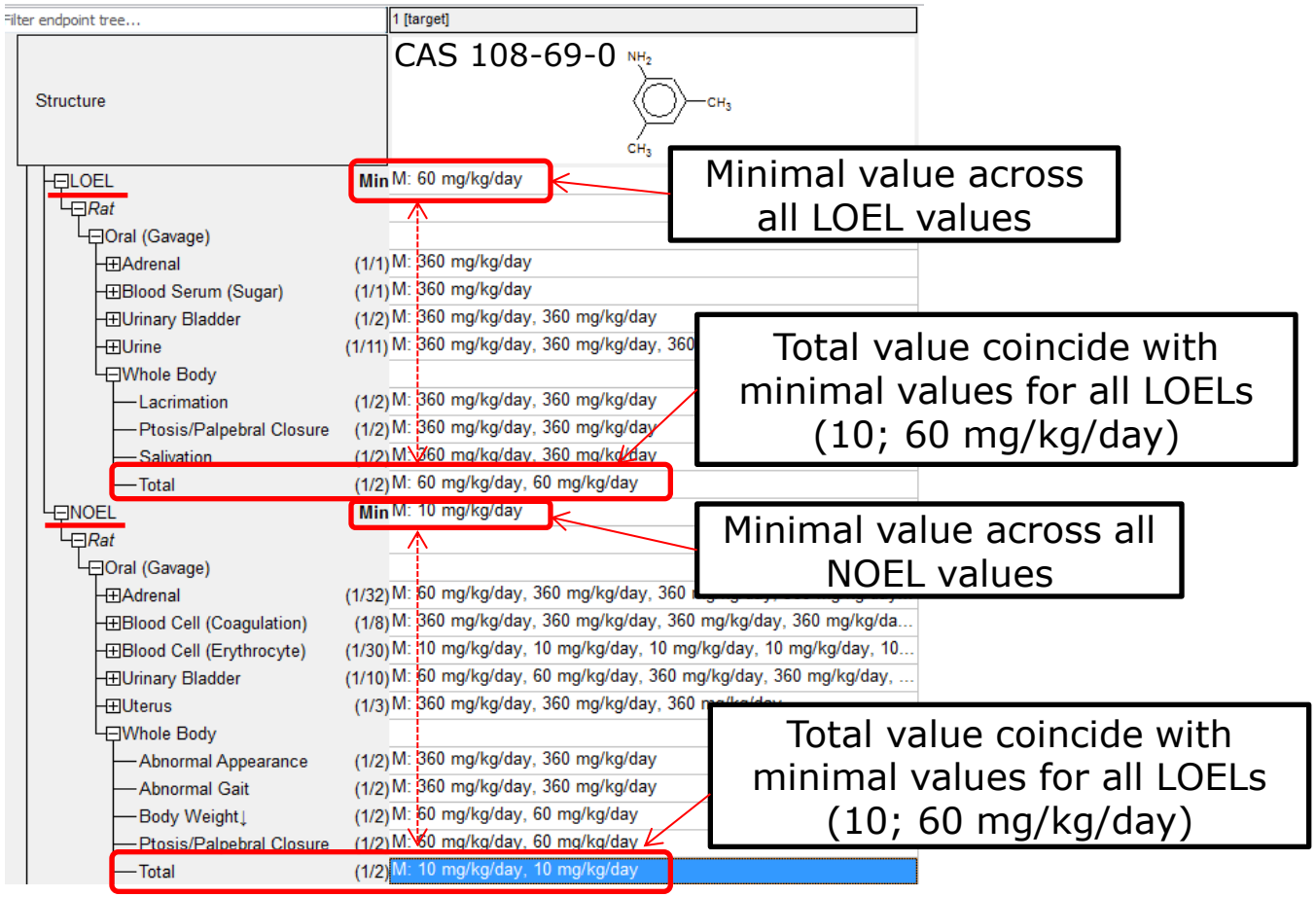


1. Chemical statistics presenting the number of chemicals and the available experimental data for the two endpoints.

Recap

- In this module, you have defined the category of similar analogues.
- In the next module, you should apply read across in order to fill in data gap
- But before the user can proceed with the "Filling Data Gap" module, he/she should navigate to the target endpoint: In our case we will predict RDT of target for two endpoints: Total NOEL and Total LOEL; Route: Oral (gavage)
- Total NOEL and Total LOEL values coincide with minimal values for all LOELs (NOELs) of the current chemical (more info could be found on next snapshot)
- Click on "Data Gap Filling" to move to the next module.

Total LOEL/NOEL



Now you are ready to continue with next module data gap filling

Outlook

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

Data Gap Filling
Overview

- "Data Gap Filling" module give 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) in case 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
Interpreting Read-across

- In this example, all the analogues have repeated dose toxicity data (LOEL and NOEL values)

- Predicted values for the target compound is based on initial group of Anilines defined by US-EPA New Chemical categories

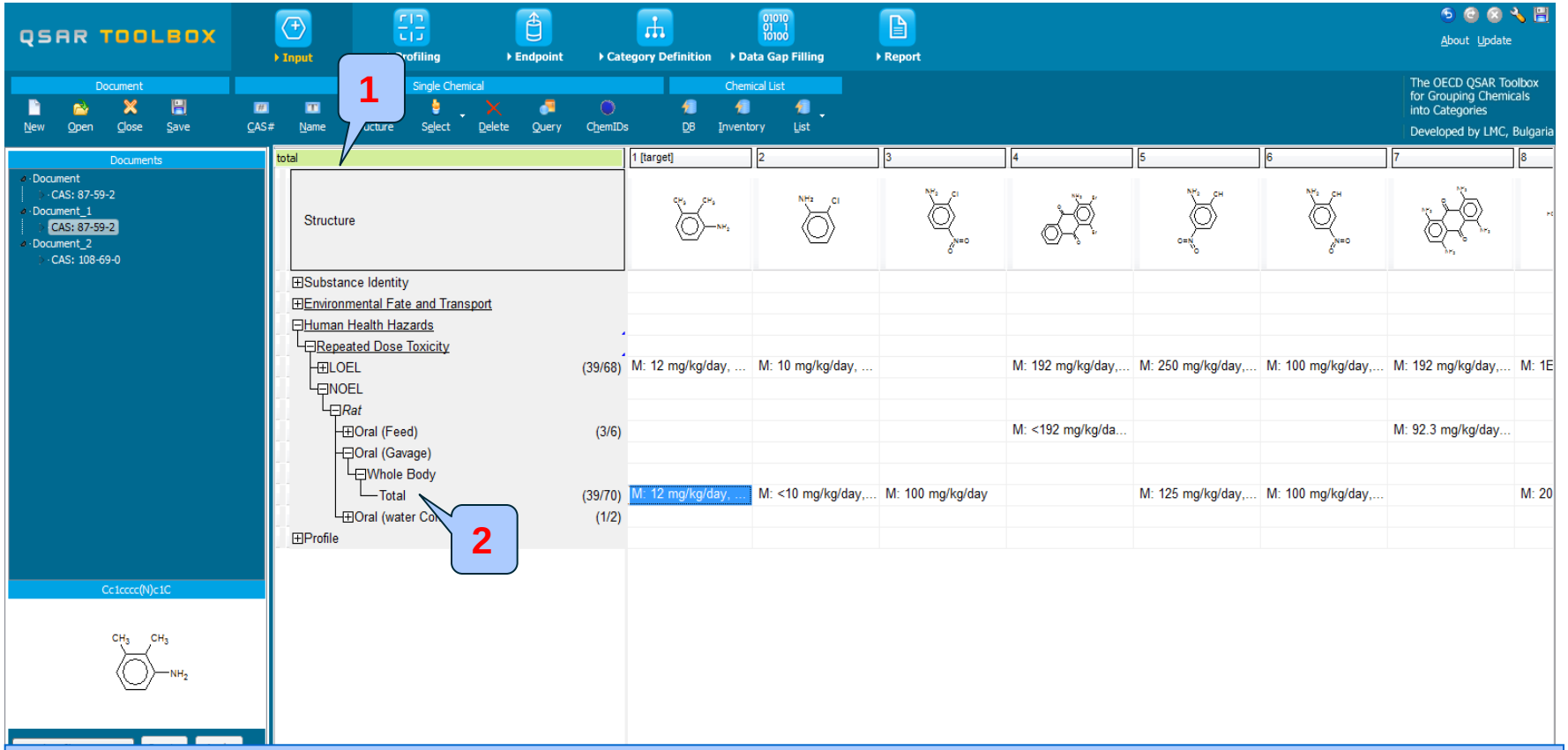
- The following subcategorizations are used for filtering the initial group of analogues:
  - US-EPA New chemical categories
  - Repeated dose (HESS)
  - Chemical elements
  - Structural similarity

- Before applying the read across, we should navigate to the target endpoint Total NOEL

See next screen shots

Data Gap Filling

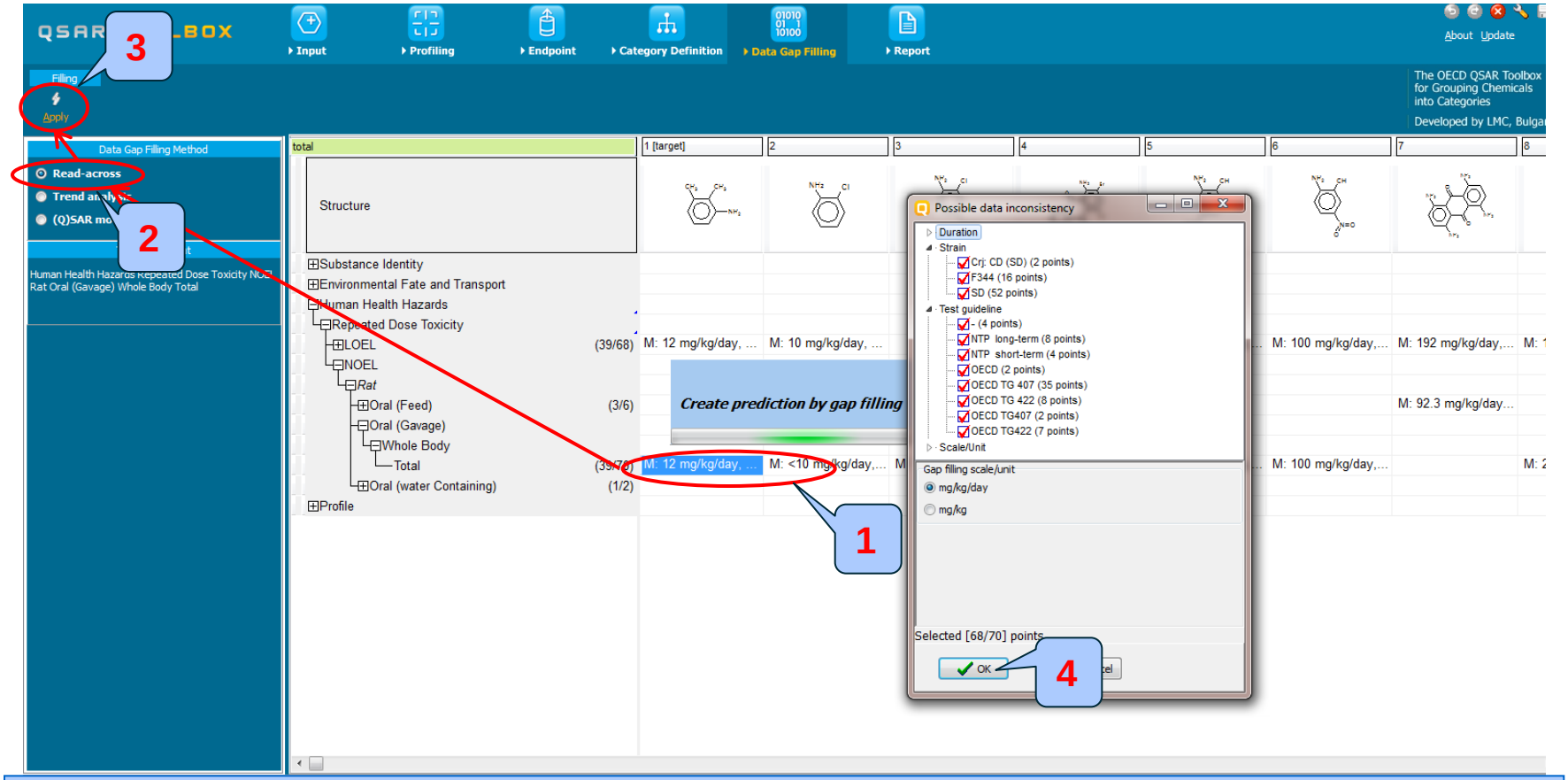
Navigation of endpoint tree: Repeated dose toxicity/**NOEL**/oral gavage/Total



1. **Type** "Total" in the filter field
2. **Expand** the tree to "Total" node by single left click on the "+" sign

Data Gap Filling

Apply read across for Total NOEL



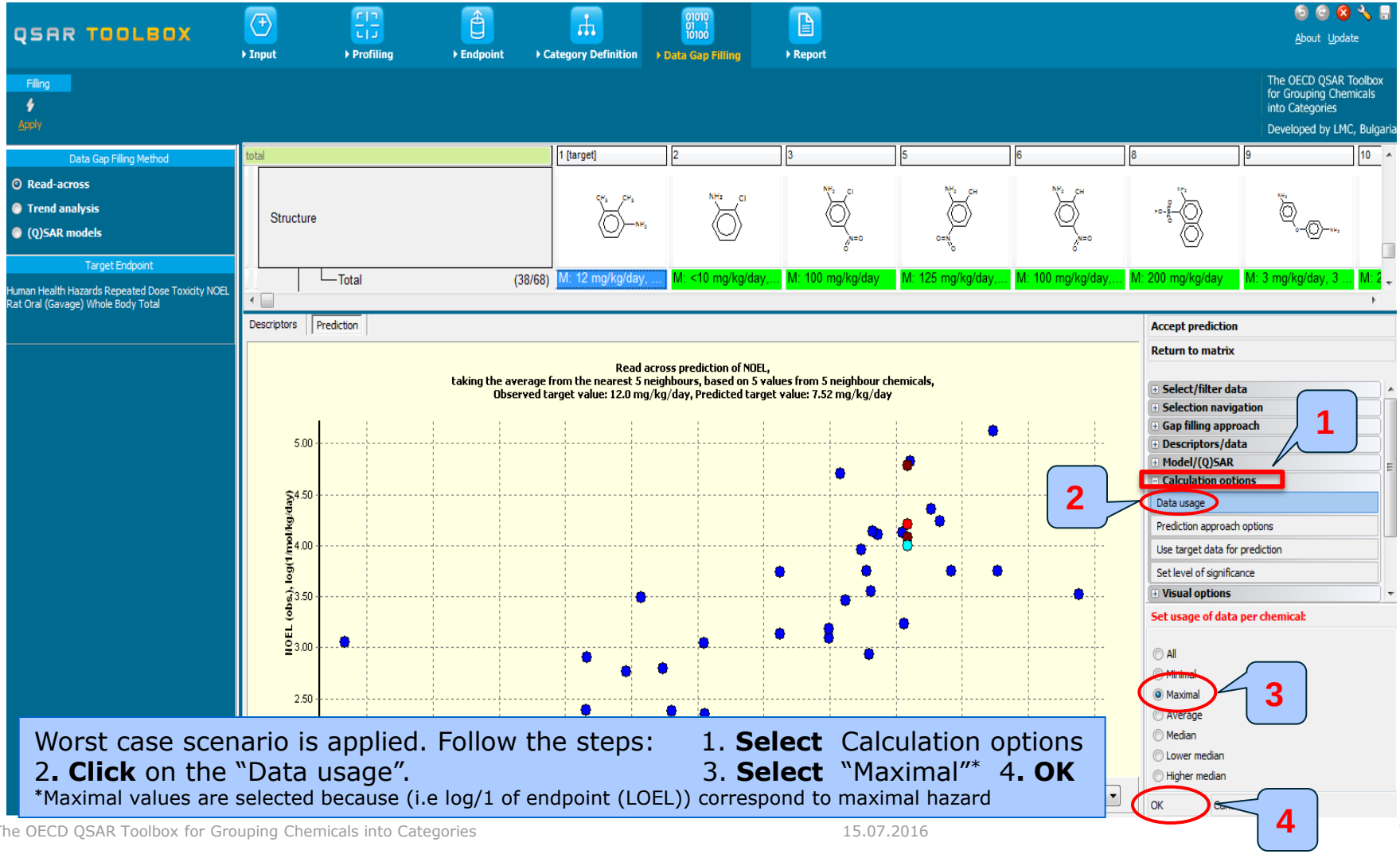
1. **Click** on the cell corresponding to "NOEL" total value for the target chemical.
2. **Select** Read-across
3. **Click** Apply
4. **OK**

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.

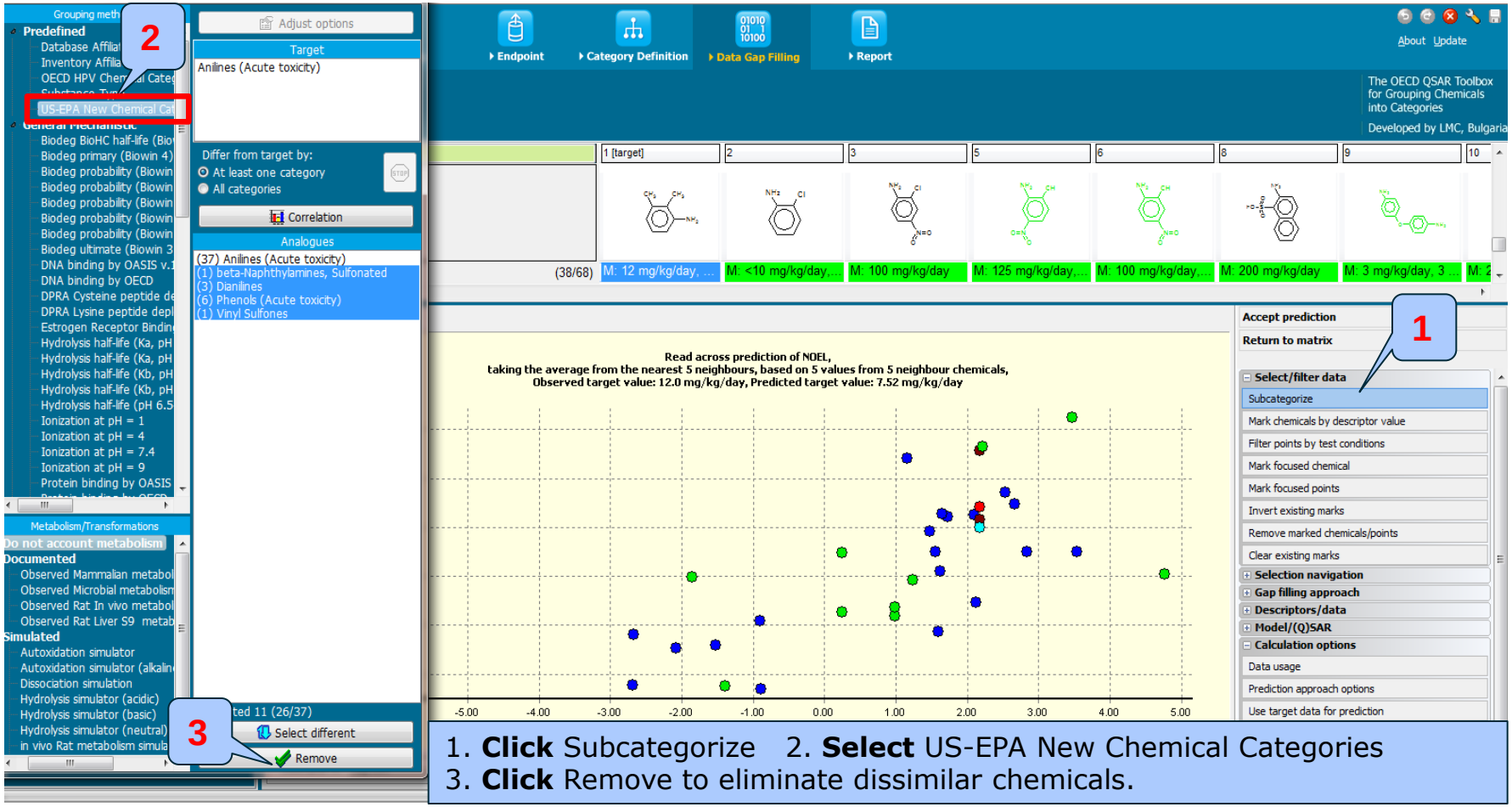
Data Gap Filling
Read-across for NOEL



Worst case scenario is applied. Follow the steps: 1. **Select** Calculation options
2**. Click** on the "Data usage". 3. **Select** "Maximal"* 4**. OK**
*Maximal values are selected because (i.e log/1 of endpoint (LOEL)) correspond to maximal hazard

Data Gap Filling

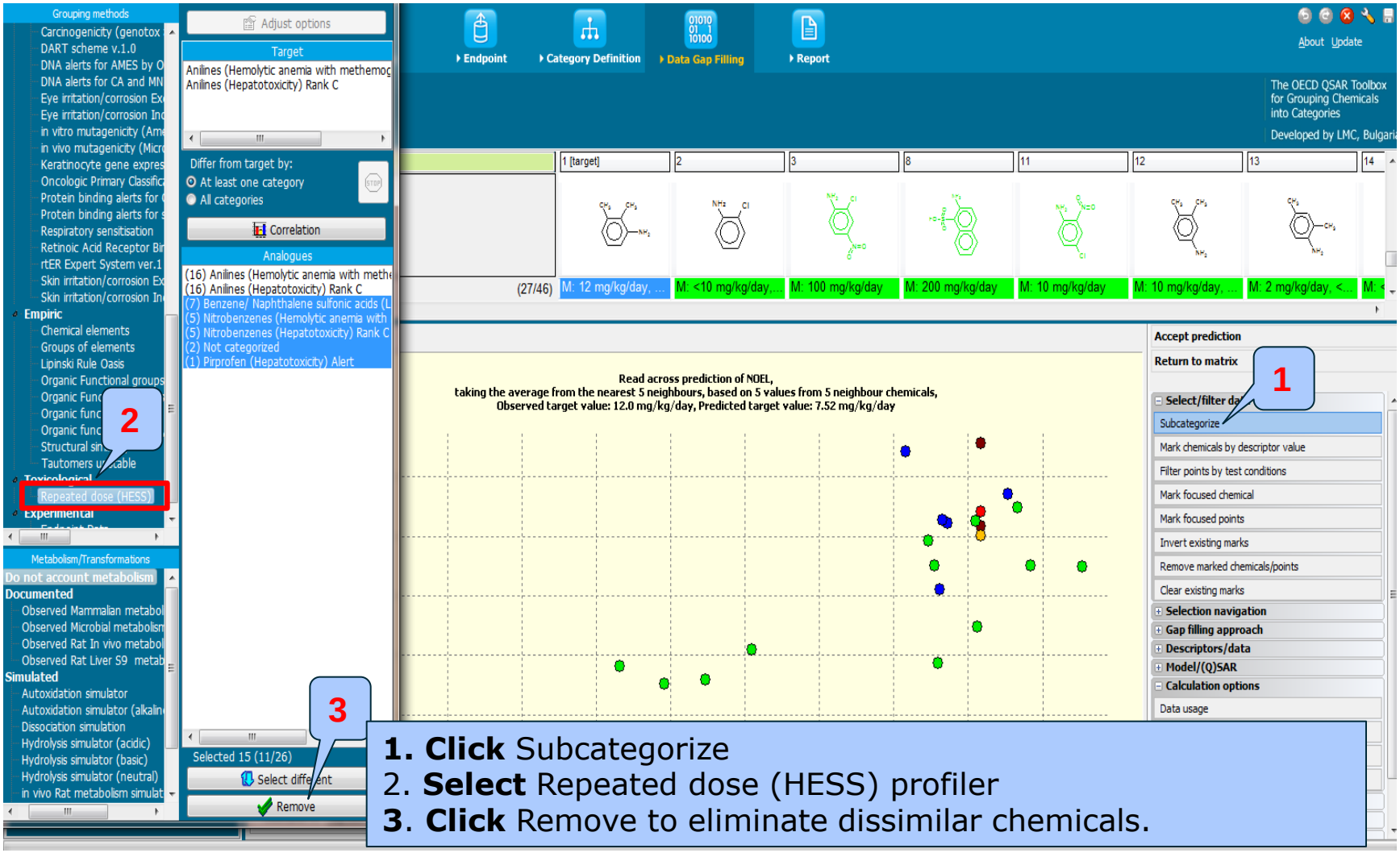
Subcategorization by US-EPA New Chemical Categories



1. **Click** Subcategorize 2. **Select** US-EPA New Chemical Categories
3. **Click** Remove to eliminate dissimilar chemicals.

Data Gap Filling

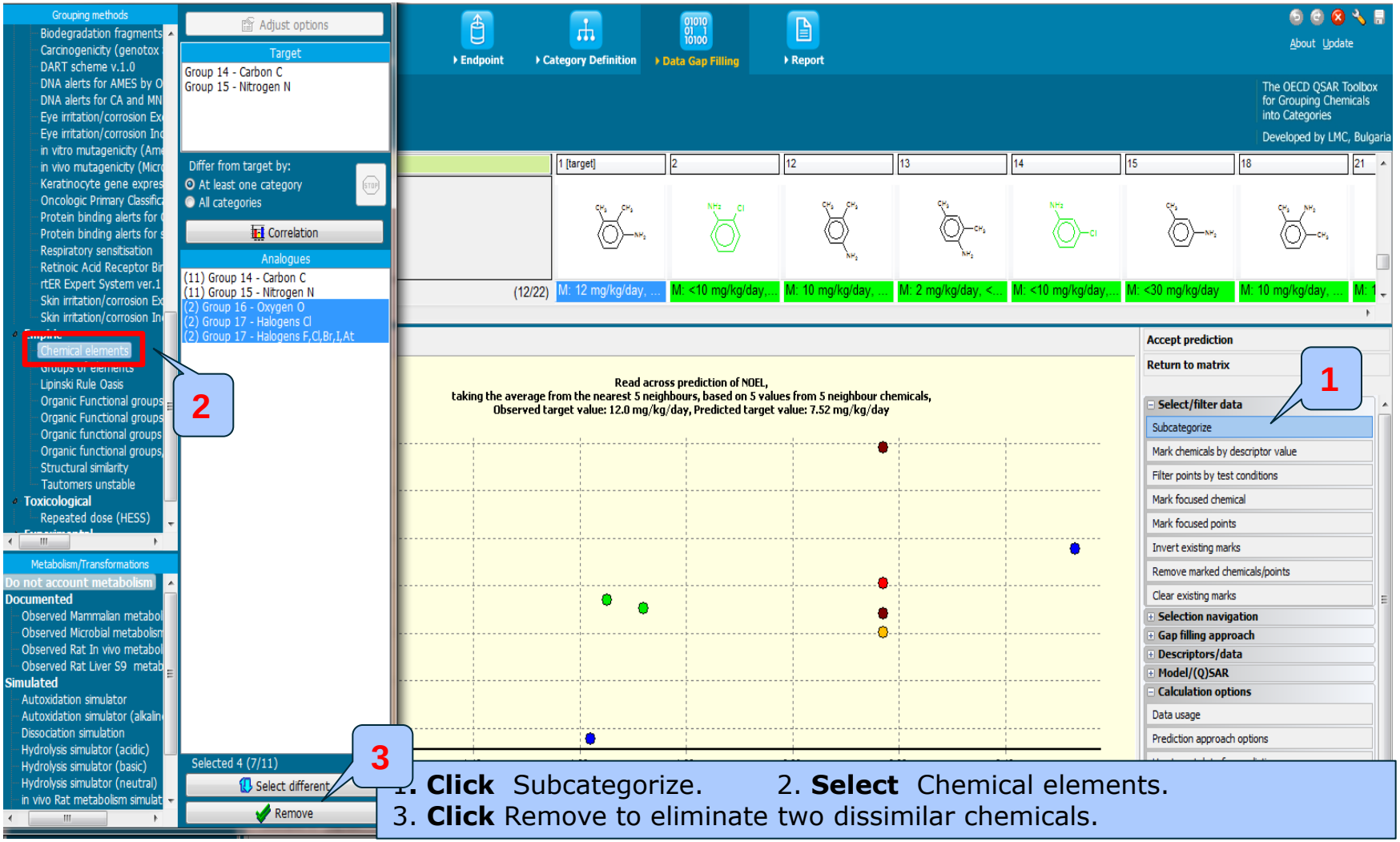
Subcategorization by Repeated dose (HESS)



1. **Click** Subcategorize
2. **Select** Repeated dose (HESS) profiler
3. **Click** Remove to eliminate dissimilar chemicals.

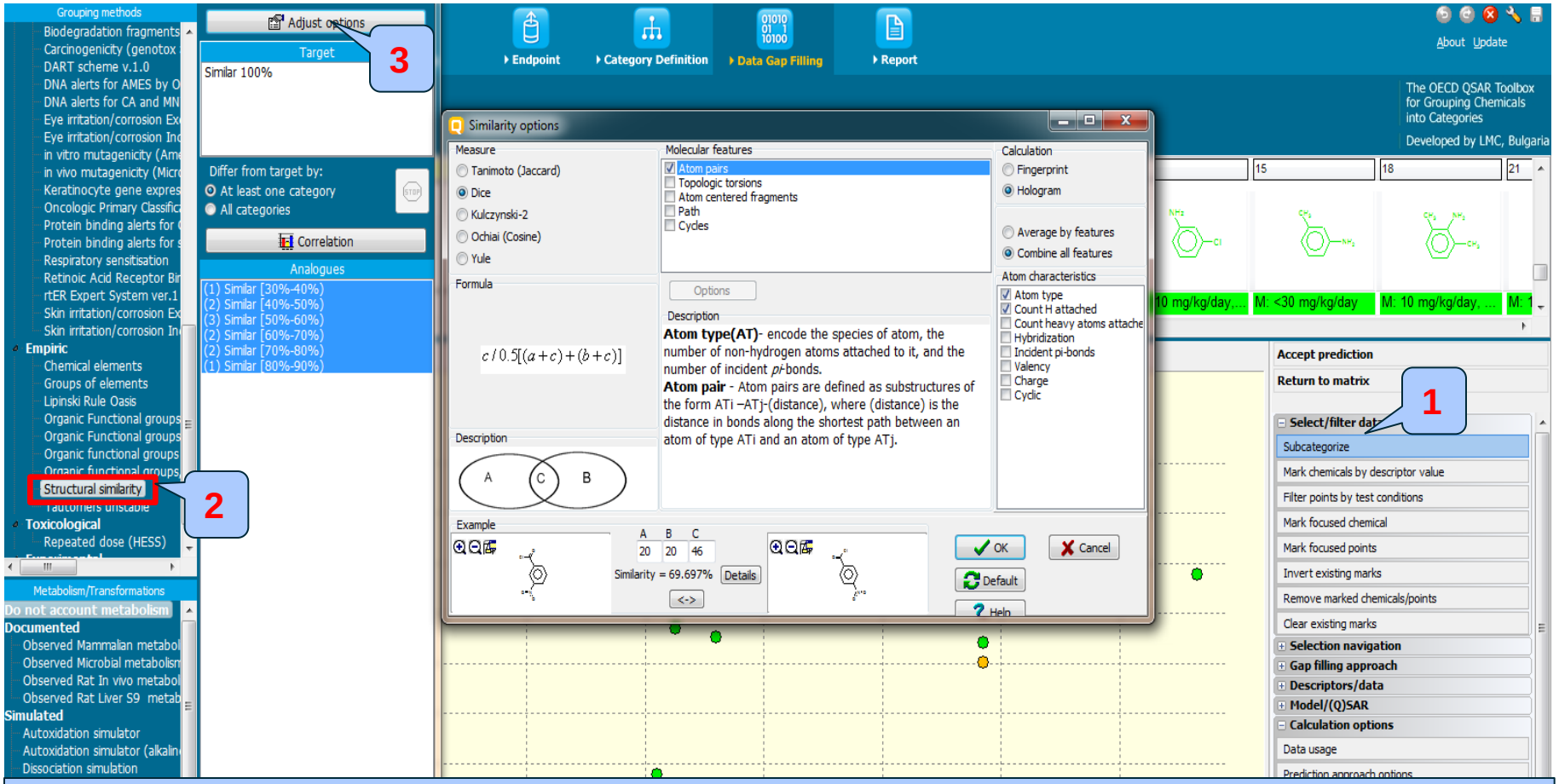
Data Gap Filling

Subcategorization by Chemical elements



Data Gap Filling

Subcategorization by Structural similarity

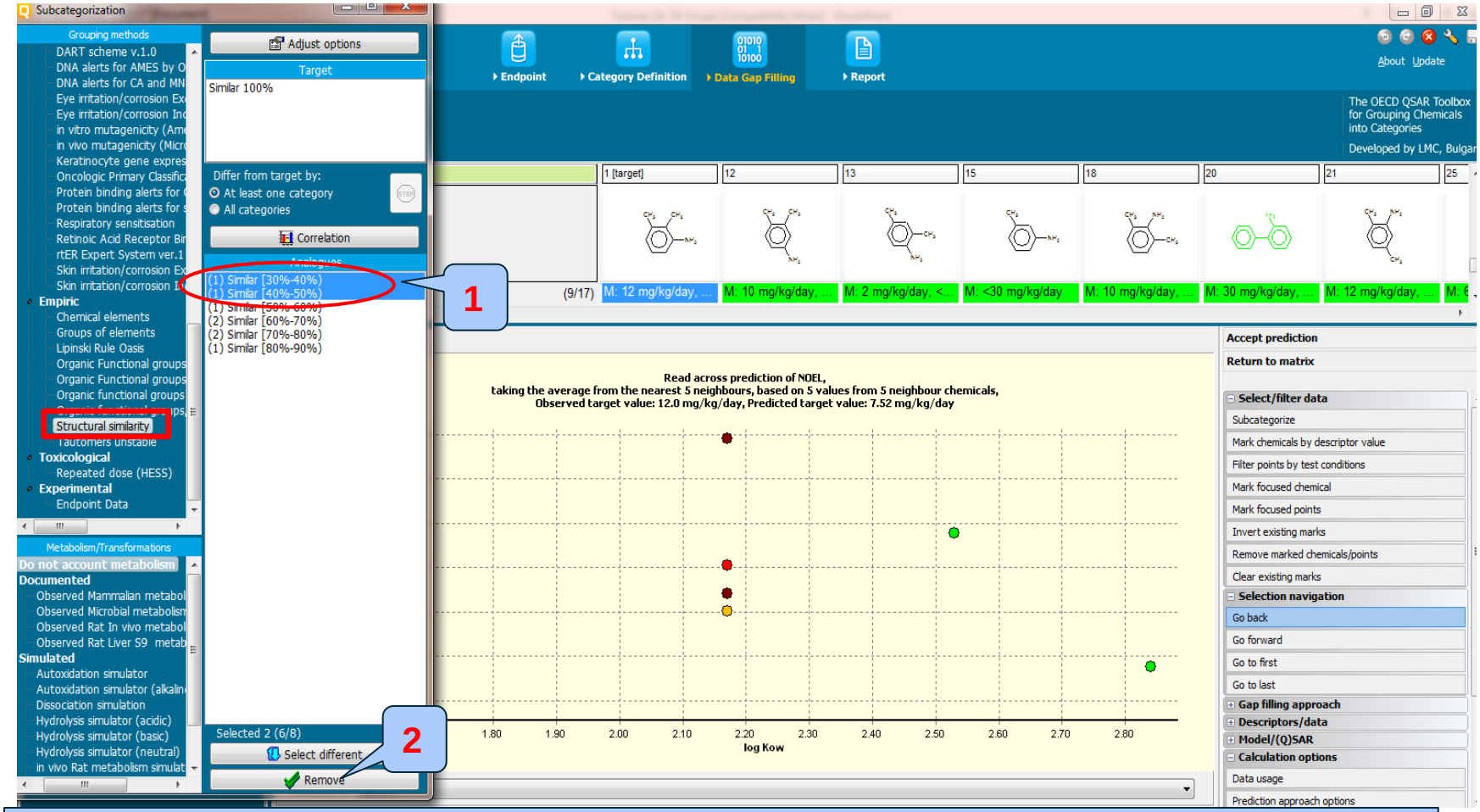


Structural similarity is applied in order to refine the category to the most similar analogues

1. **Click** Subcategorize.
2. **Select** Structural similarity and modify default options (3) to the following: Dice, Atom pairs, Atom type; Count H attached.

Data Gap Filling

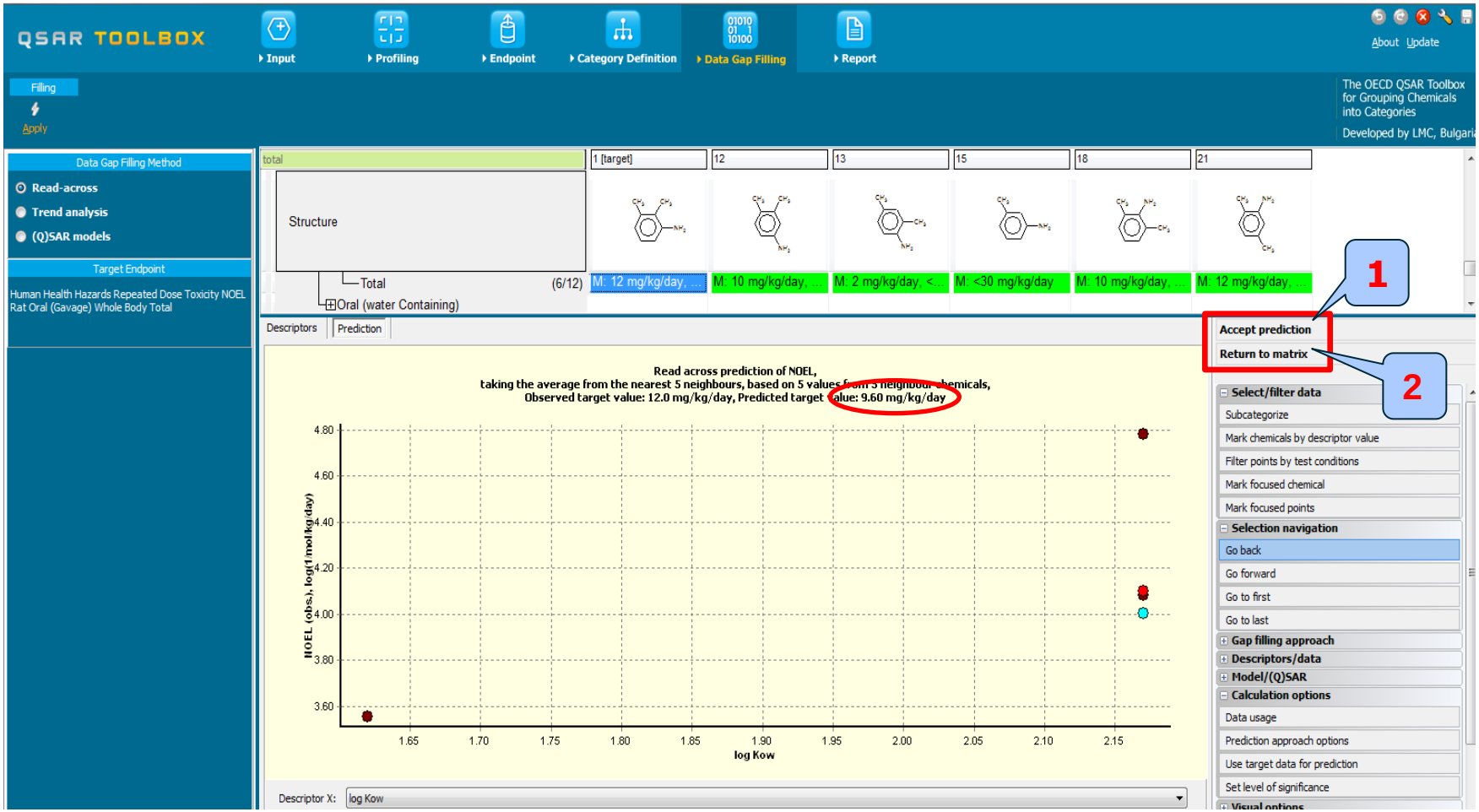
Subcategorization by Structural similarity



1. Hold Control button and select first two bins in order to eliminate them.
2. **Click** Remove to eliminate chemicals with similarity less than 60%

Data Gap Filling

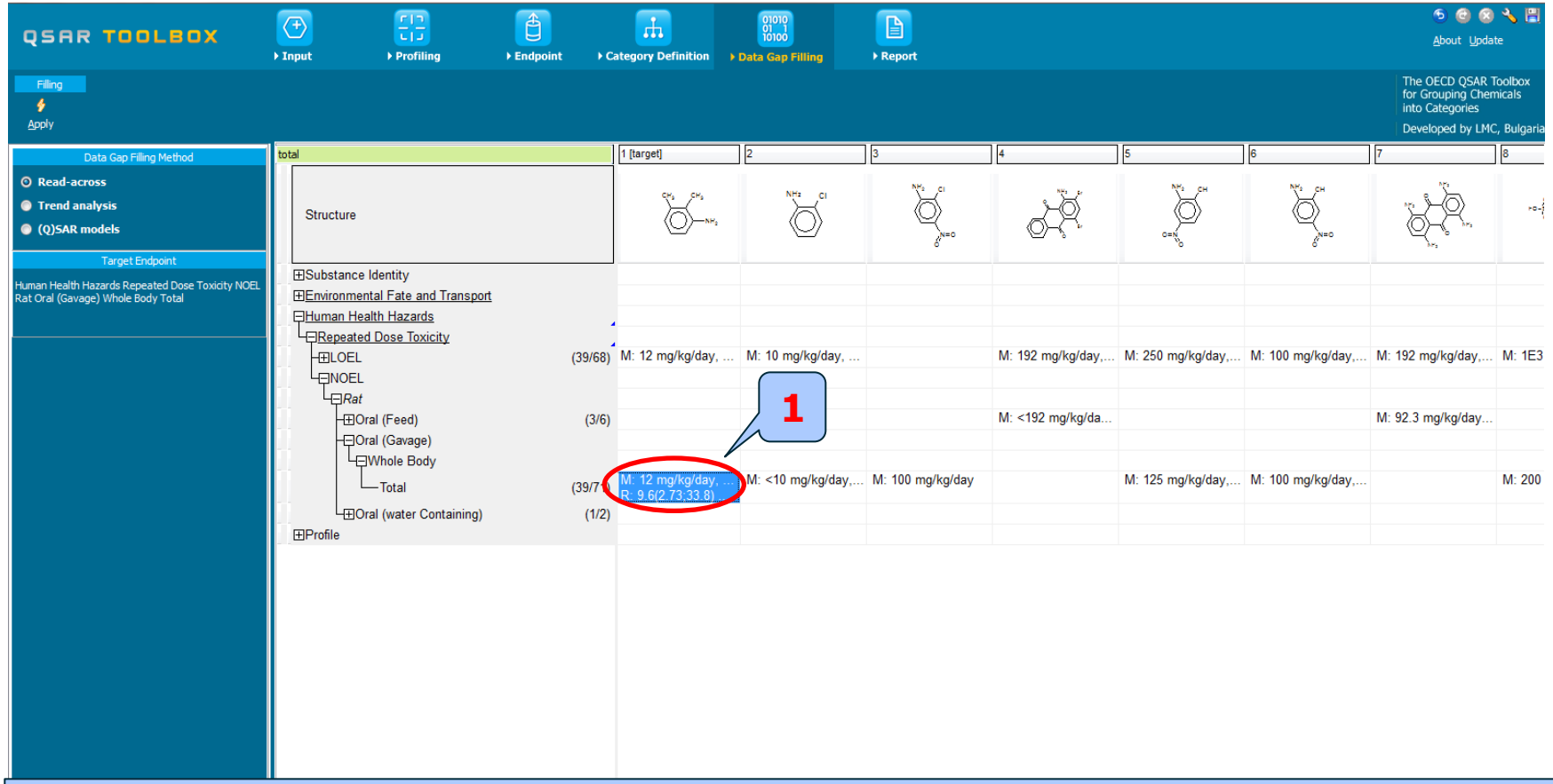
Read across result for Total NOEL



1. **Click** Accept prediction 2. **Click** Return to matrix

Data Gap Filling

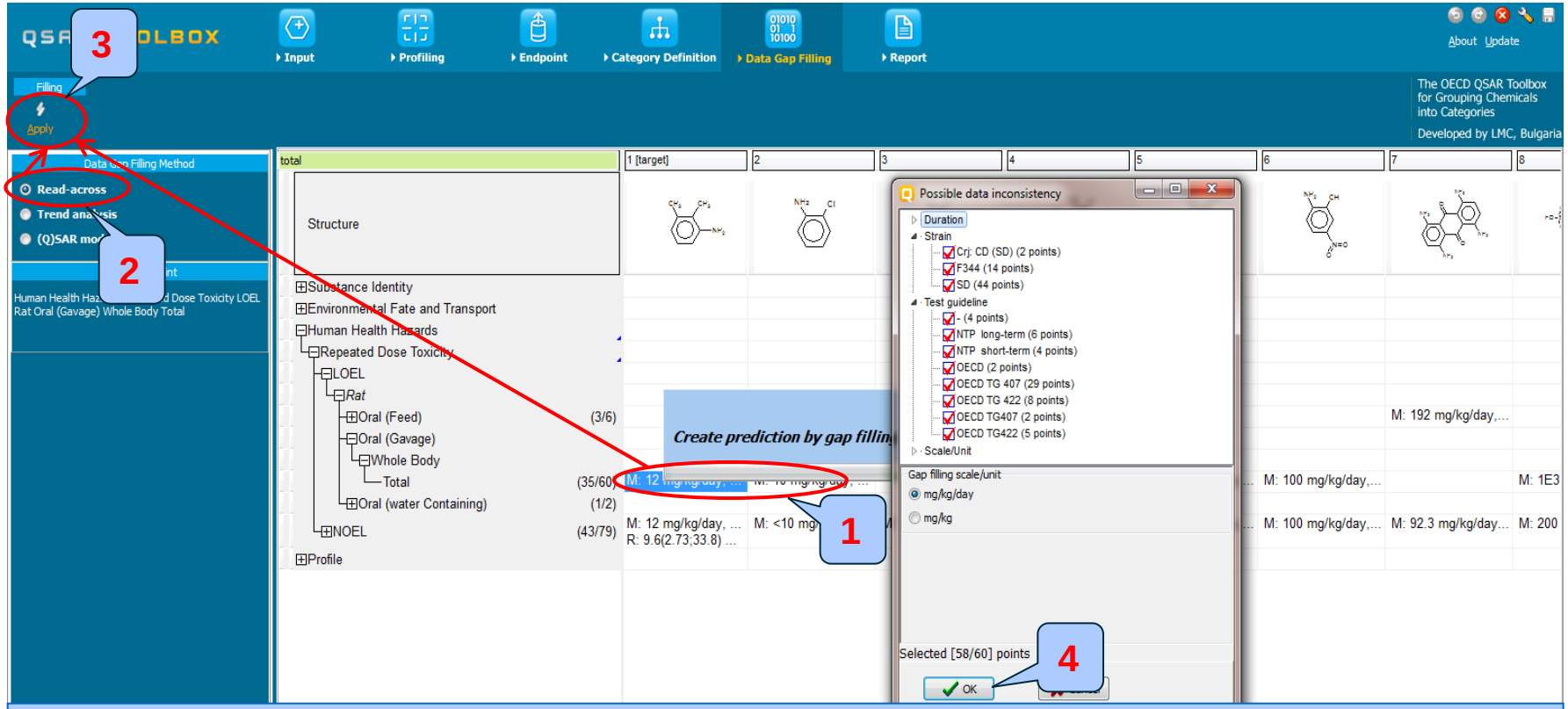
Result of read-across prediction



1. Read across prediction 9.6 mg/kg/day coincide with experimental data (12 mg/kg/day)

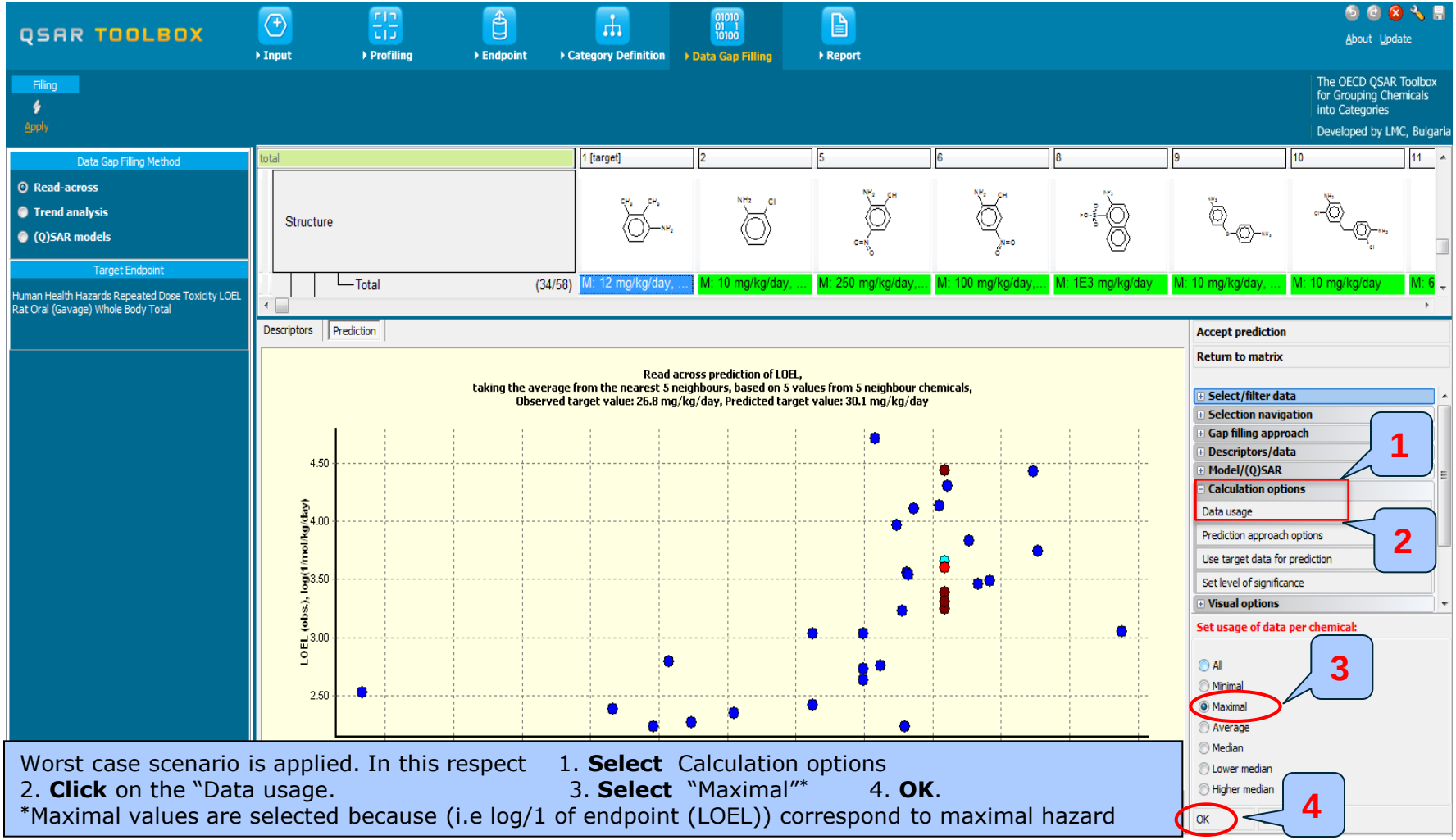
Data Gap Filling

Apply read-across for Total LOEL



1. **Click** on the cell corresponding to "LOEL" total value for the target chemical.
2. **Select** Read-across
3. **Click** Apply
4. **OK**

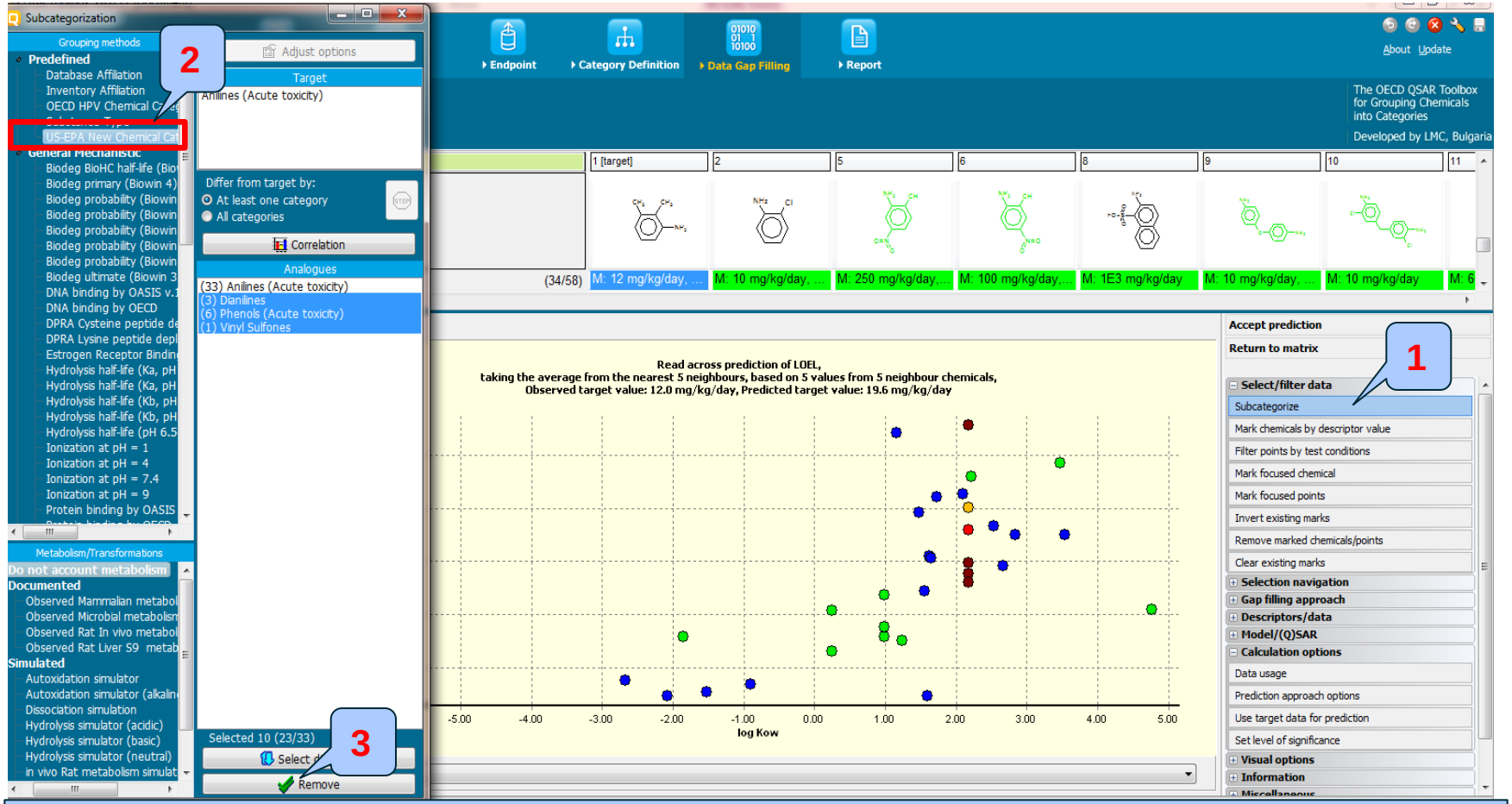
Data Gap Filling
Read-across for LOEL



Worst case scenario is applied. In this respect 1. **Select** Calculation options
2. **Click** on the "Data usage. 3. **Select** "Maximal"* 4. **OK**.
*Maximal values are selected because (i.e log/1 of endpoint (LOEL)) correspond to maximal hazard

Data Gap Filling

Subcategorization by US-EPA New Chemical Categories



1. Click Subcategorize

2. Select US-EPA New Chemical Categories

3. **Click** Remove to eliminate dissimilar chemicals.

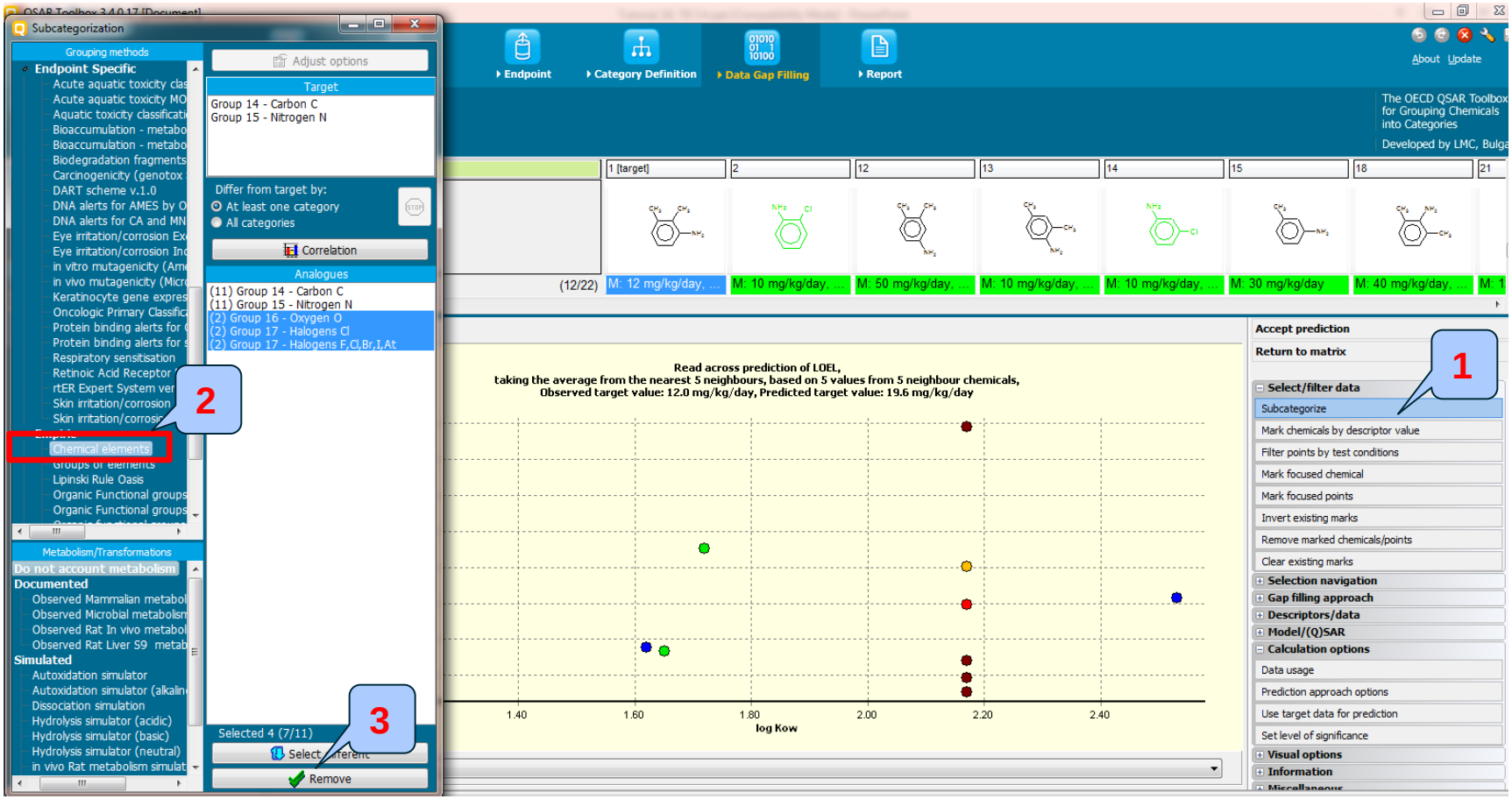
Data Gap Filling

Subcategorization by Repeated dose (HESS)

1. Click Subcategorize. **2. Select** Repeated dose (HESS)
3. Click Remove to eliminate dissimilar chemicals.

Data Gap Filling

Subcategorization by Chemical elements

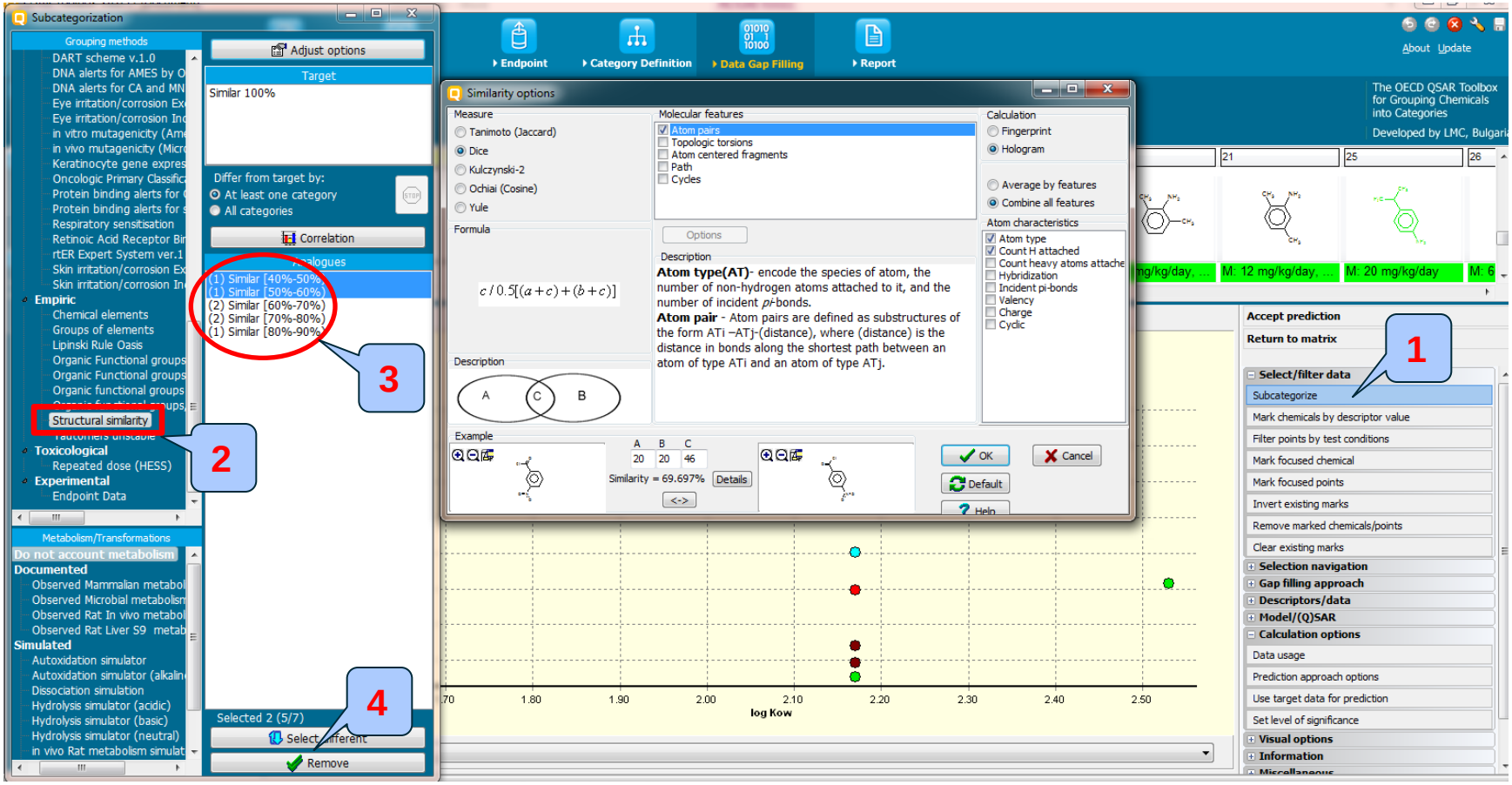


1. Click Subcategorize
2. **Select** Chemical elements

3. **Click** Remove to eliminate dissimilar chemicals.

Data Gap Filling

Subcategorization by Structural similarity



Structural similarity is applied in order to refine the category to the most similar analogues

1. Click Subcategorize.

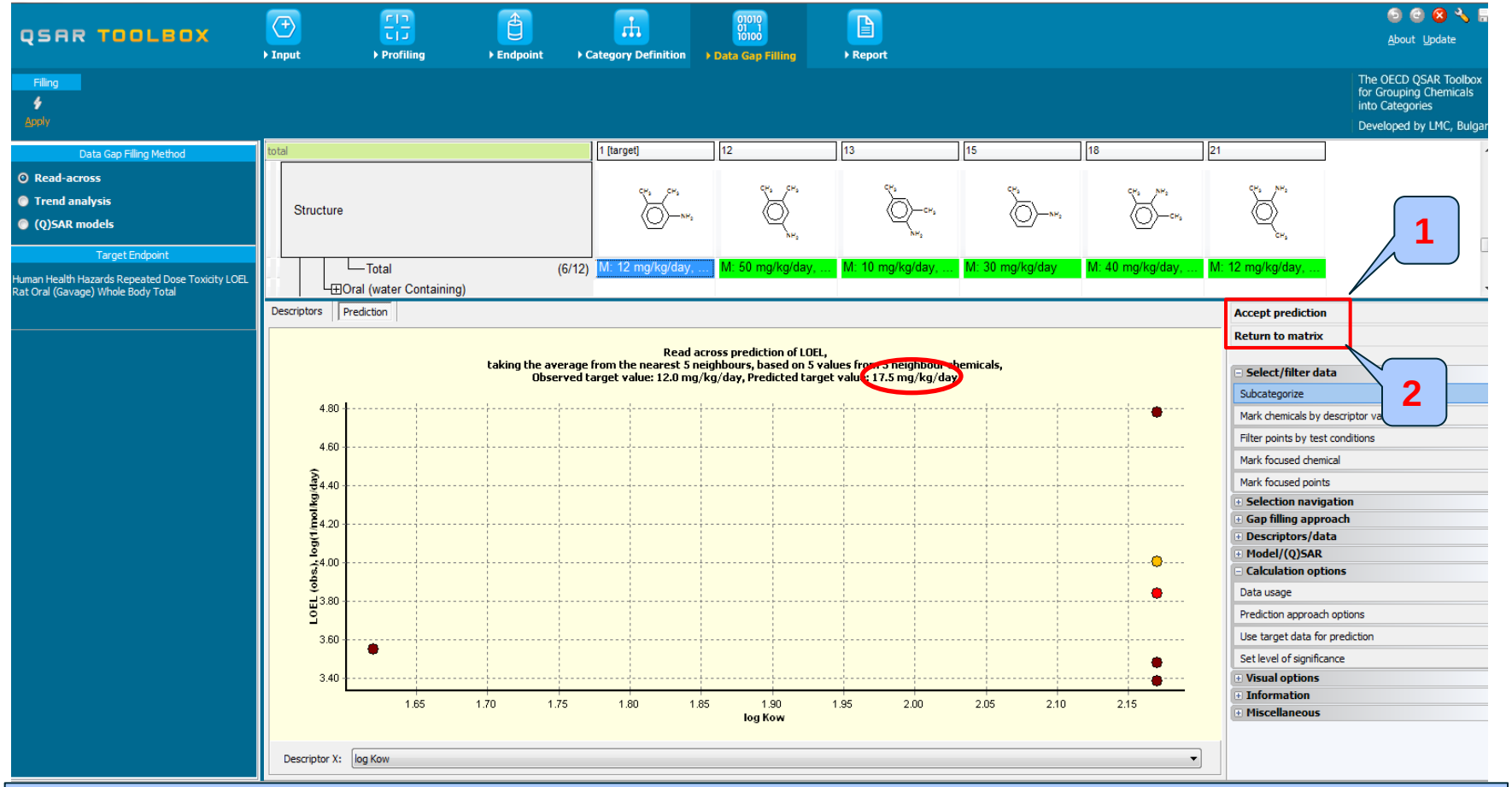
2. **Select** Structural similarity and apply the similarity options as with Total NOEL: Dice, Atom pairs, Atom type; Count H attached

3. **Select** first two categories (hold Ctrl button)

4. **Click** Remove to eliminate chemicals with similarity less than 60%

Data Gap Filling

Read across result for Total LOEL

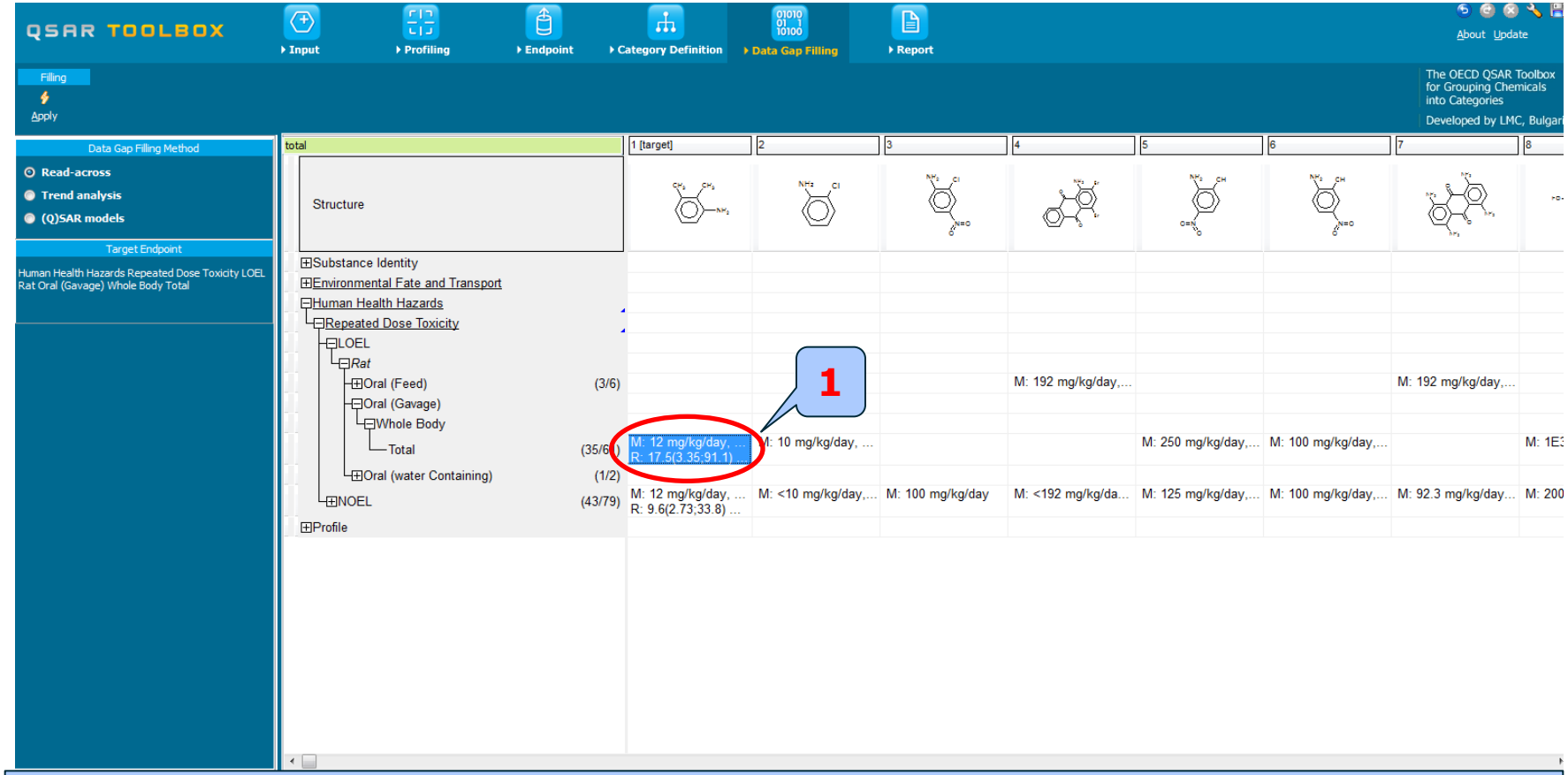


1. **Click** Accept prediction

2. **Click** Return to matrix

Data Gap Filling

Result of read-across prediction



1. Read across prediction for LOEL: 17.5 mg/kg/day coincide with experimental LOEL: 12 mg/kg/day

Read across predictions for 2,3 dimethylaniline (CAS 87-59-2)

Result

Ultimate prediction:

Total NOEL – 9.6 mg/kg/day
Total LOEL – 17.5 mg/kg/day

Based on obtained results (for total LOEL and total NOEL) the target chemical is classified as *Category 2* regarding GHS classification [1]

Table 3.9.2: Guidance values to assist in Category 2 classification

Route of exposure	Units	Guidance value range (dose/concentration)
Oral (rat)	mg/kg bw/d	10 - 100
Dermal (rat or rabbit)	mg/kg bw/d	20 - 200
Inhalation (rat) gas	ppm/6h/d	50 - 250
Inhalation (rat) vapour	mg/litre/6h/d	0.2 - 1.0
Inhalation (rat) dust/mist/fume	mg/litre/6h/d	0.02 - 0.2

[1] Globally Harmonized System of Classification and Labeling of Chemicals (GHS): http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev04/English/ST-SG-AC10-30-Rev4e.pdf

Outlook

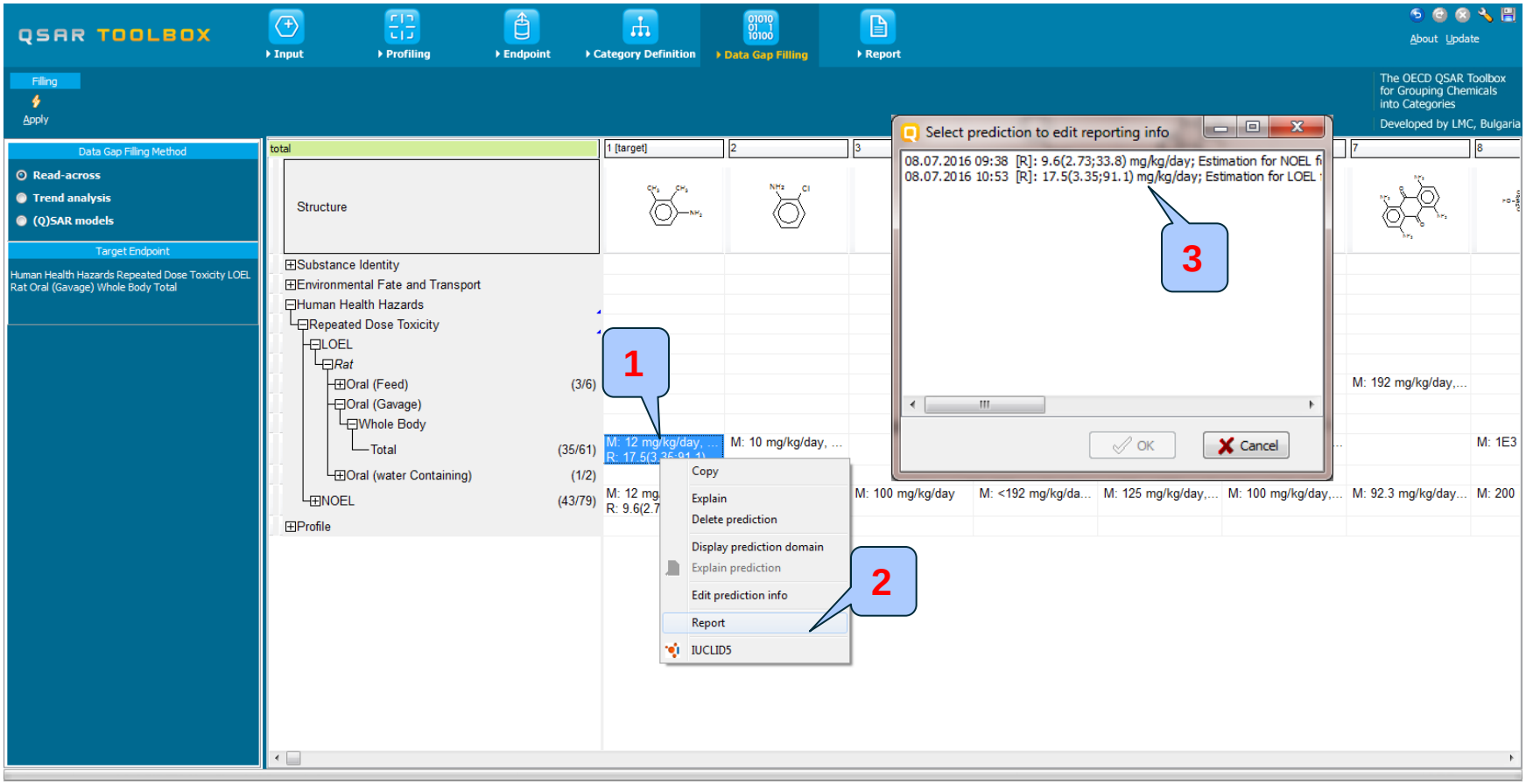
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Category definition
  - Data gap filling
  - **Report**

Report

- Remember the report module allows you to generate a report on the predictions performed with the Toolbox. This module contains predefined report templates as well as a template editor with which users can define their own user defined templates. The report can then be printed or saved in different formats.

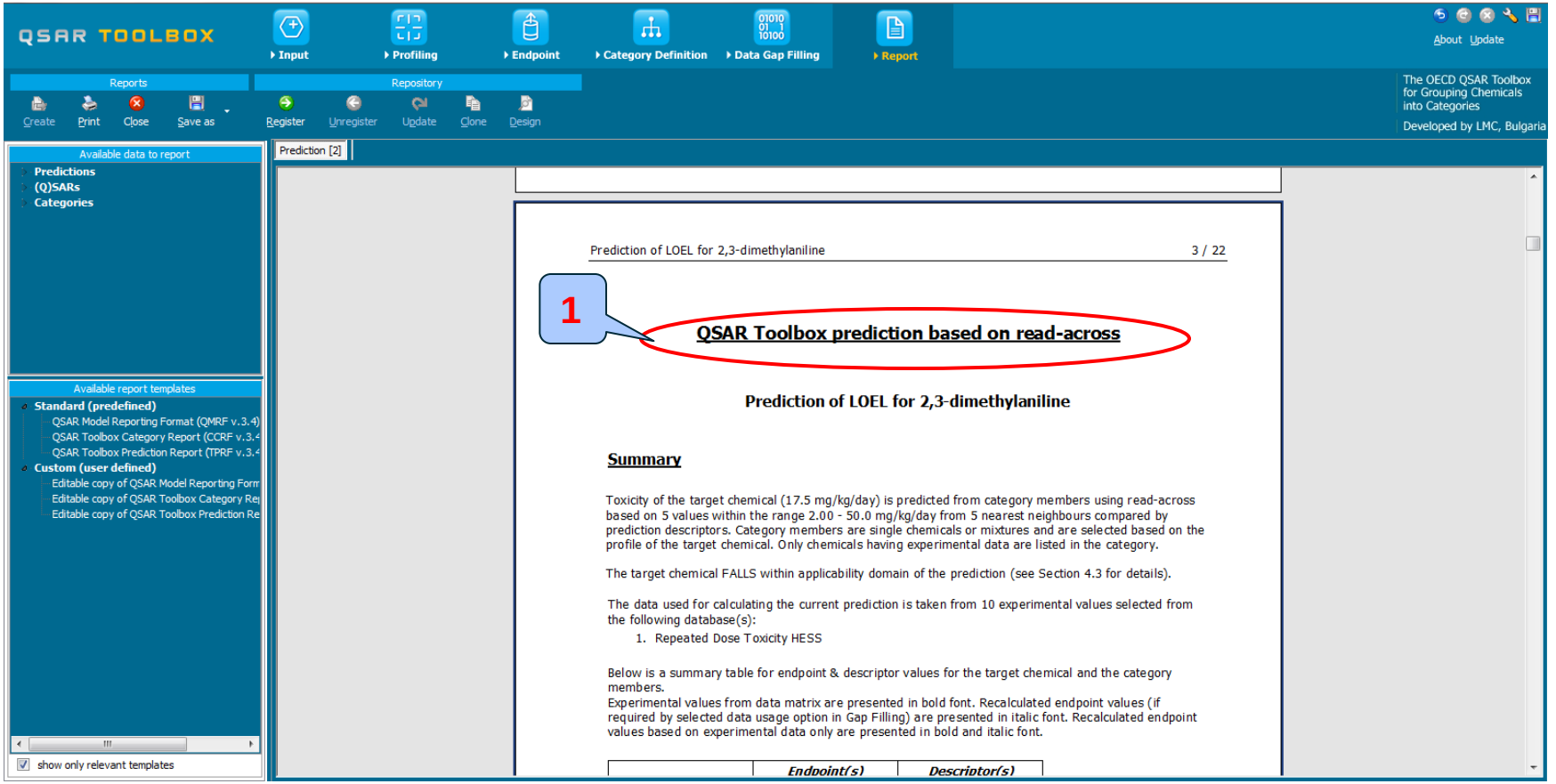
- Generating the report is shown on next screenshots

Report



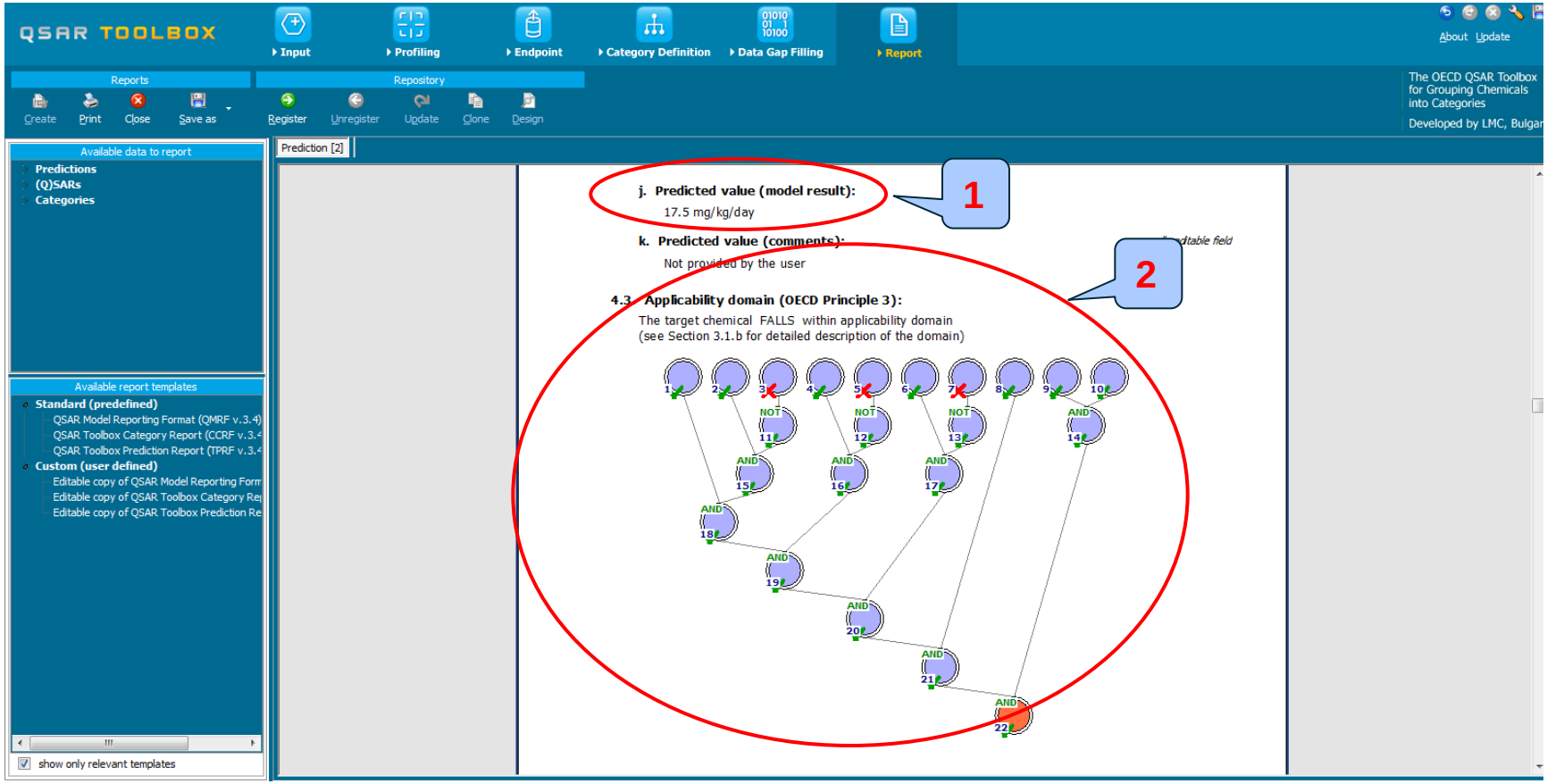
1. **Select** prediction
2. **Right Click** and **Select** Report
3. **Select** the prediction for which you want to generate the report

Report



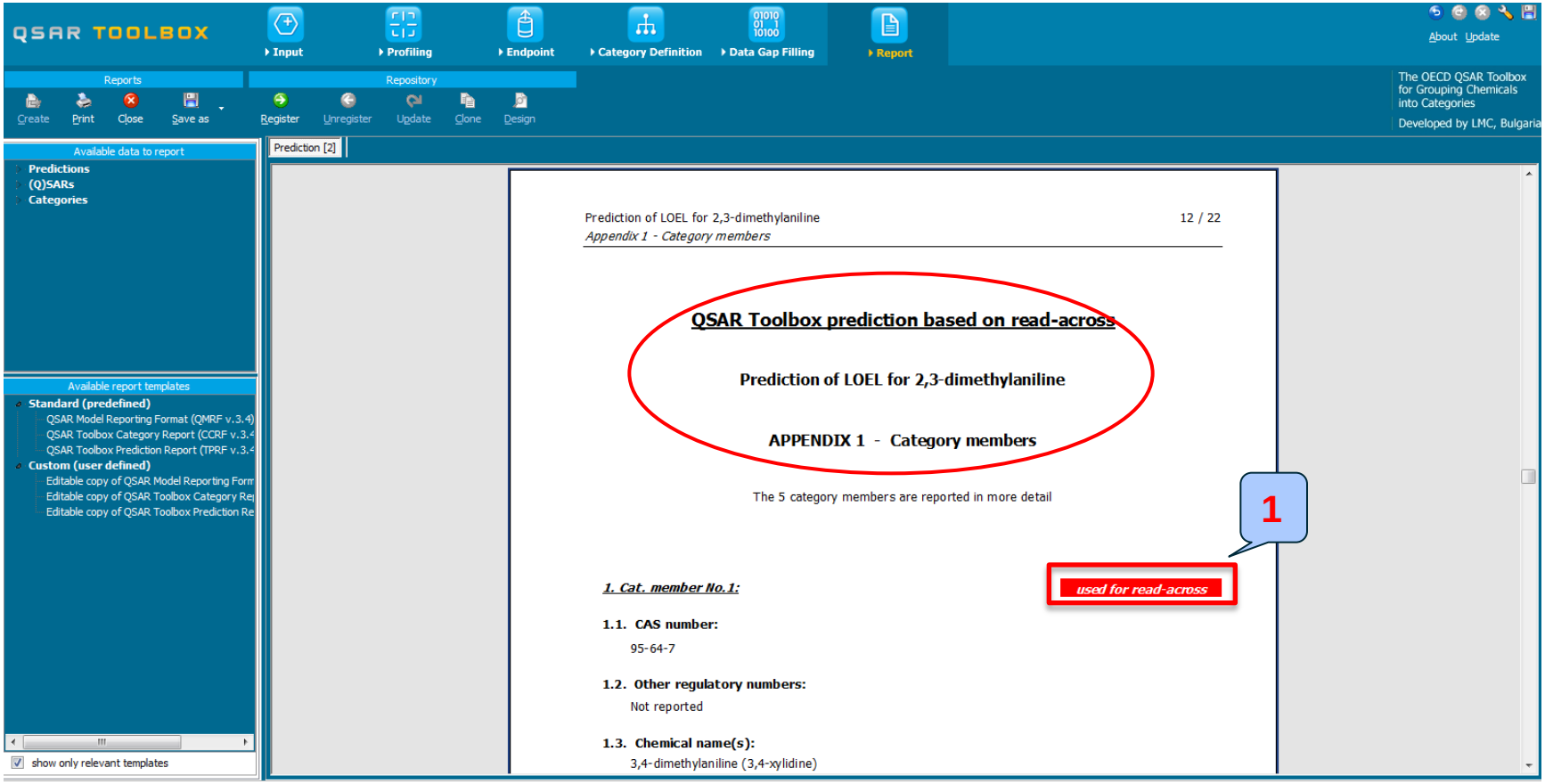
1. Report for LOEL

Report



1. Predicted value 2. Applicability domain

Report



1. Additional information for category members

Outlook

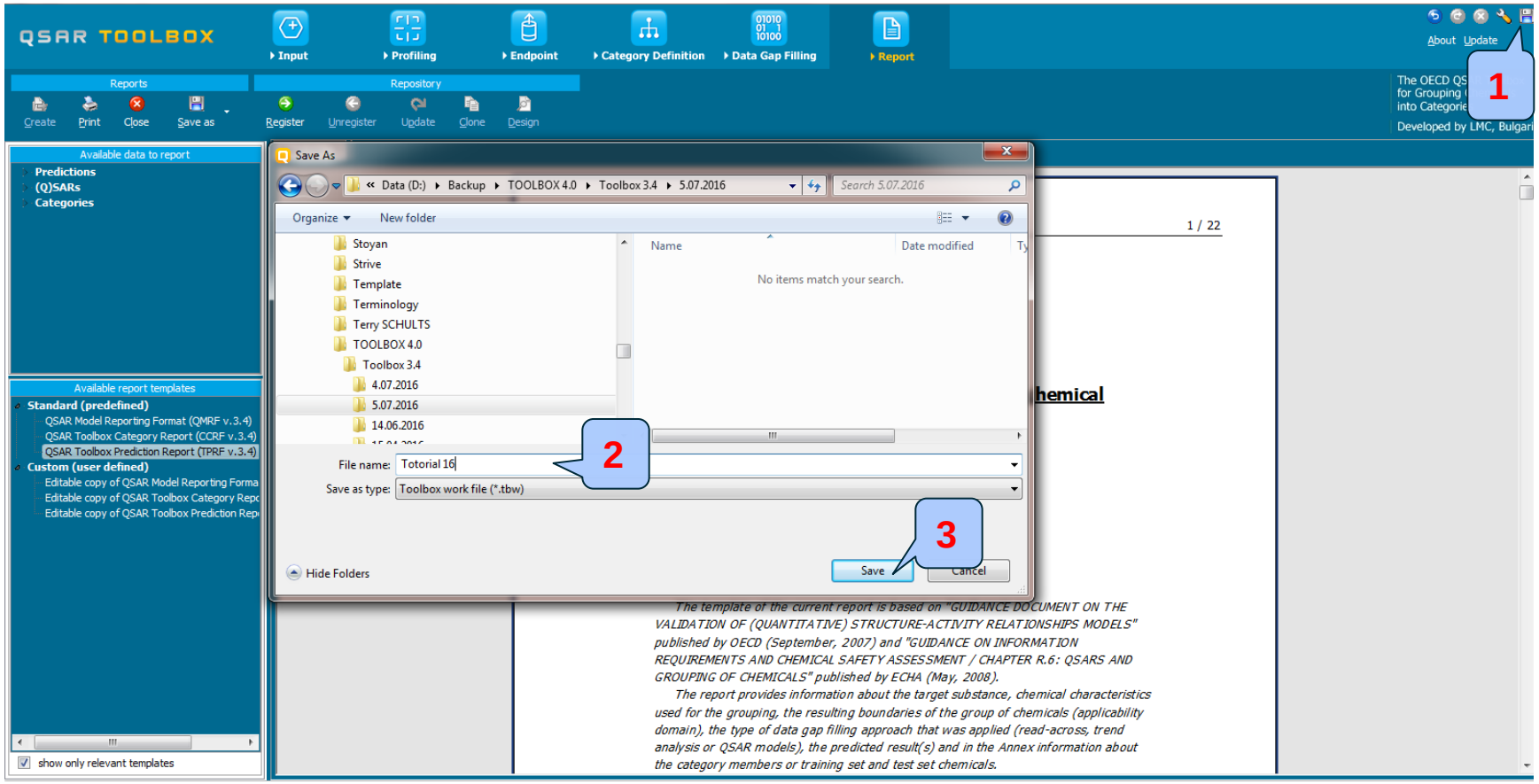
- Background
- Objectives
- The exercise
- Workflow
- **Save prediction**

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).

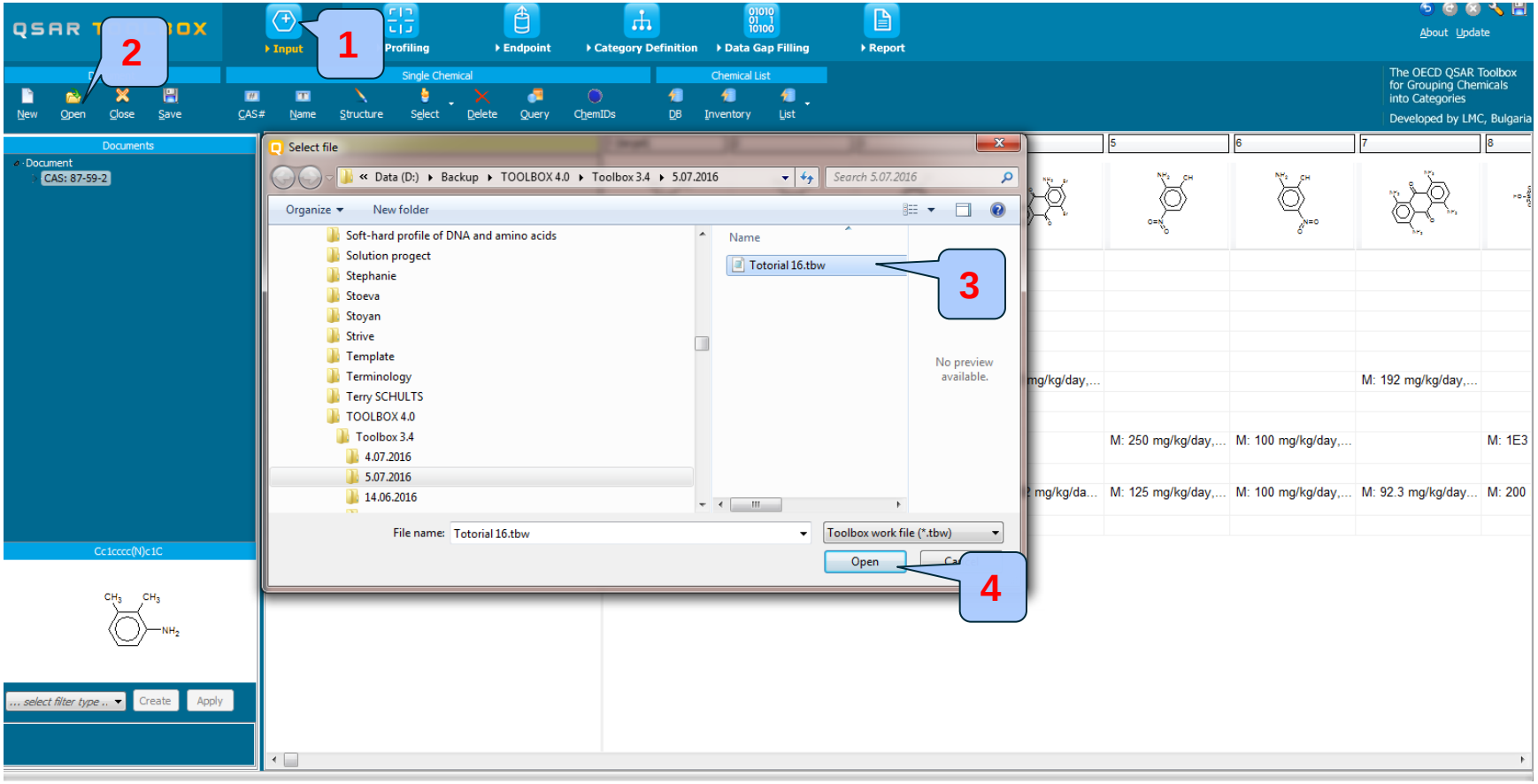
- Saving/Loading the file with TB prediction is shown on next screenshots

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