QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.1

Predicting acute aquatic toxicity to fish of 4-methyl-2-nitroaniline taking into account tautomerism

Outlook

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

Background

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

Outlook

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

Objectives

- **This presentation reviews 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
  - Filling data gap by trend-analysis

Outlook

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

The Exercise

- In this exercise we will predict *LC50* for fish: *P.promelas* for target chemical
- Set of simulated tautomers for the target chemical will be provided
- This prediction will be accomplished by collecting similar analogues presented with their tautomeric set
- The category will be defined using US-EPA New Chemical Categories
- Data gap will be filled by trend-analysis

Outlook

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

Workflow

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

Outlook

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

Chemical Input

There are two ways for simulating tautomers of chemicals

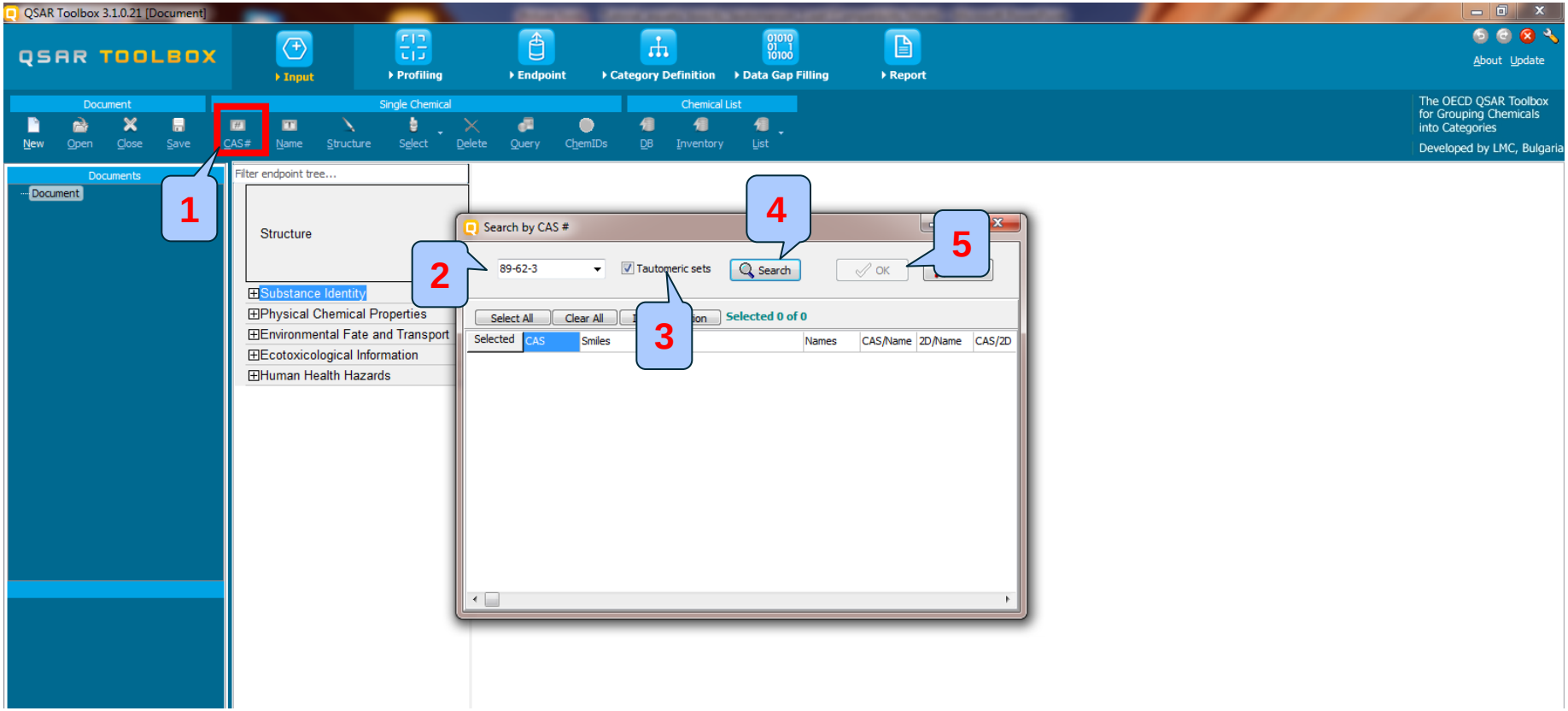
- During the process of entering the structure into the system
- Simulating tautomersim of already entered structure

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

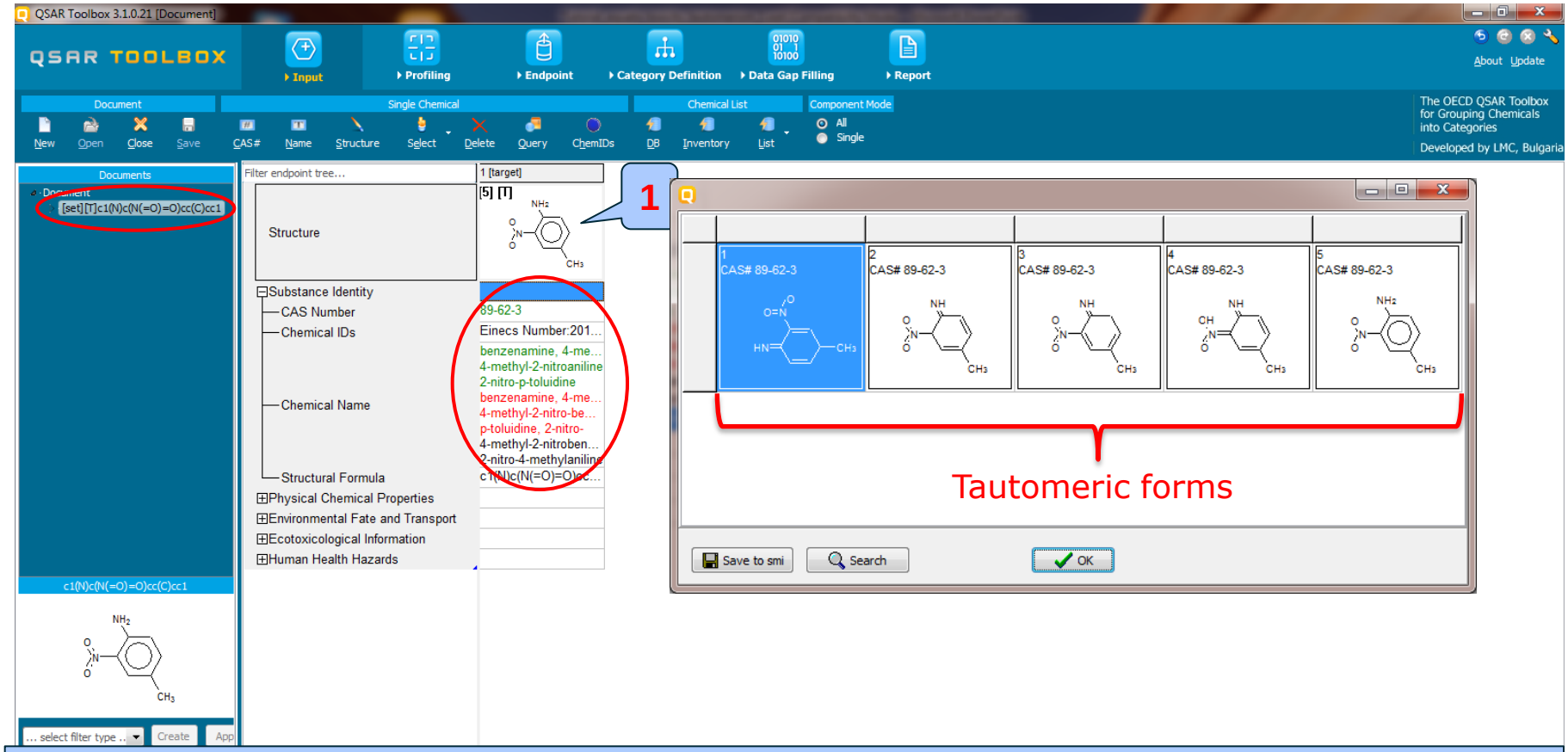
Input target chemical by CAS#



1. **Click** on CAS#; 2. **Enter** 89-62-3; 3. **Select** Tautomeric sets; 4. **Click** Search; 5. **OK**

Note: Tautomeric set functionality search tautomeric forms of entered chemical in previously tautomerized databases

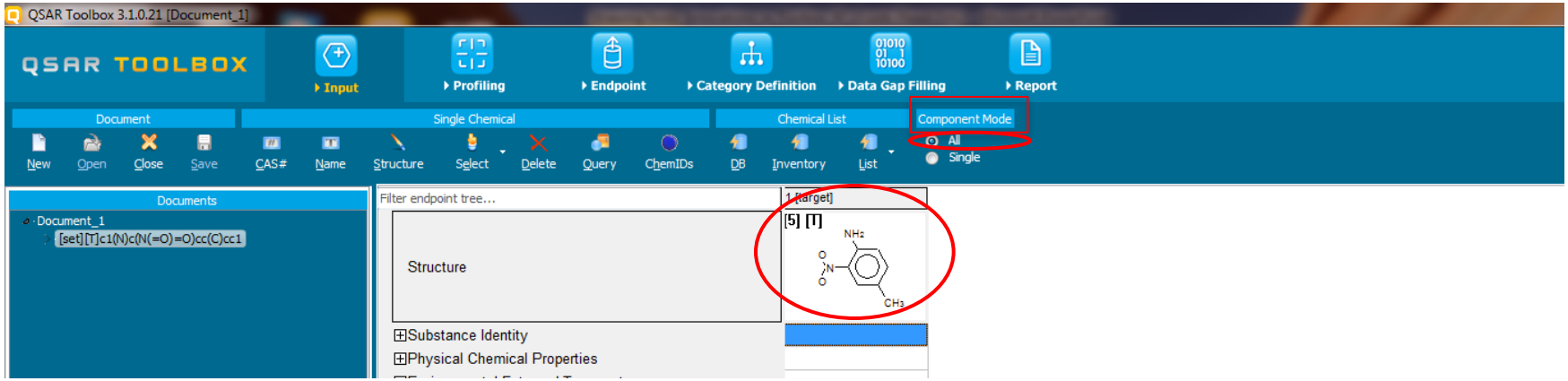
Chemical Input
Target chemical identity



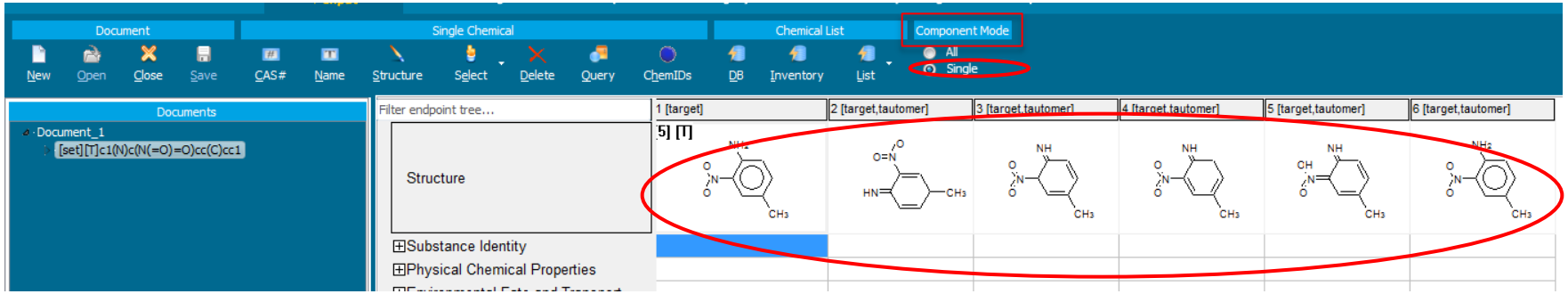
Target with its tautomeric forms are identified and loaded on data matrix.
1. **Double click** over the target structure displays target and its tautomeric forms

Chemical Input
Implementation of Modeling modes

- **Component Mode All –** all tautomeric forms are analyzed in a package



- **Component Mode Single** – each tautomeric form is analyzed individually



Different modes for visualization for the set of target and its tautomeric forms is implemented. A package of target and its tautomeric forms are used in further trend analysis.

Outlook

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

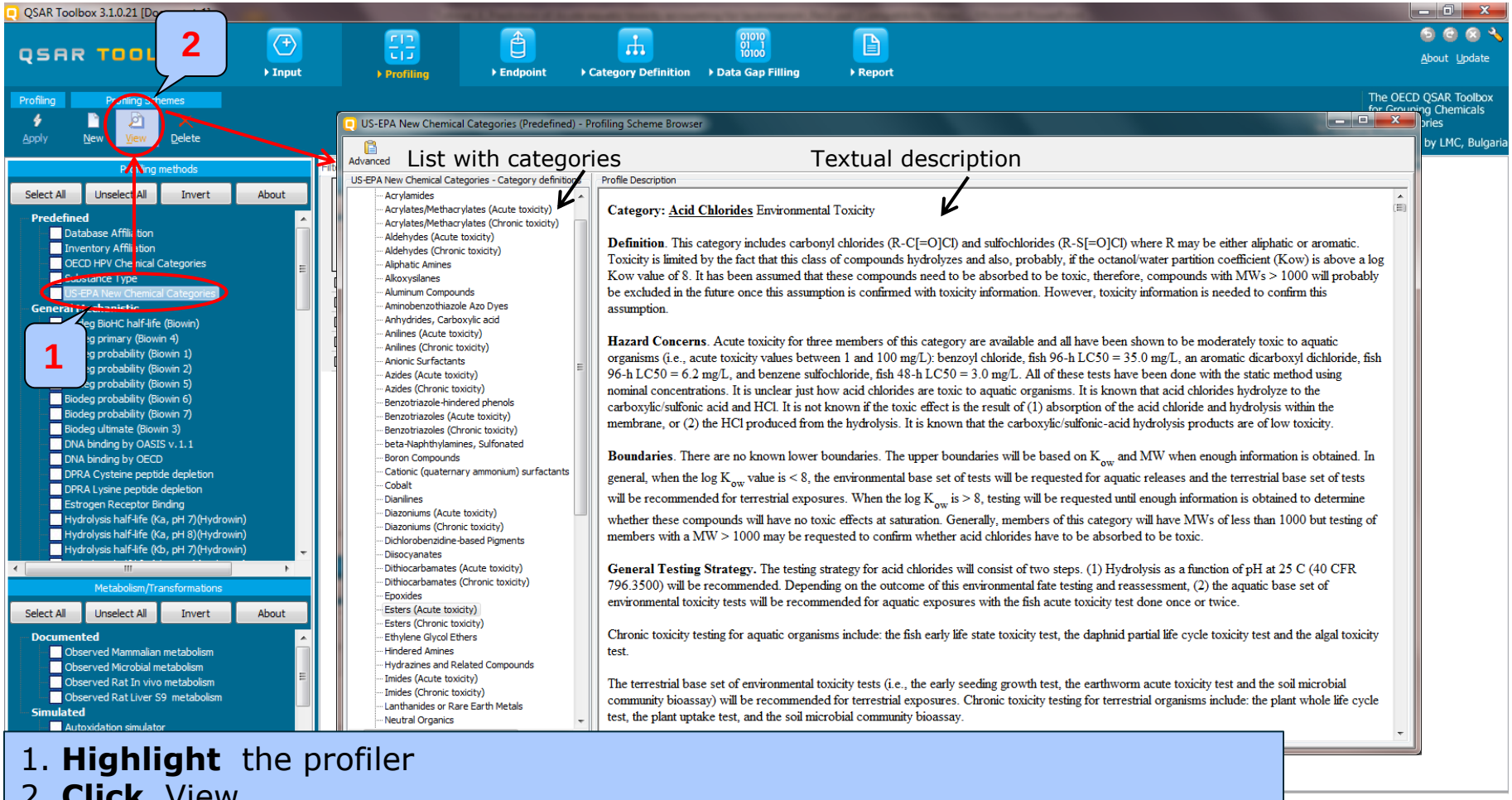
Profiling
Overview

- "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

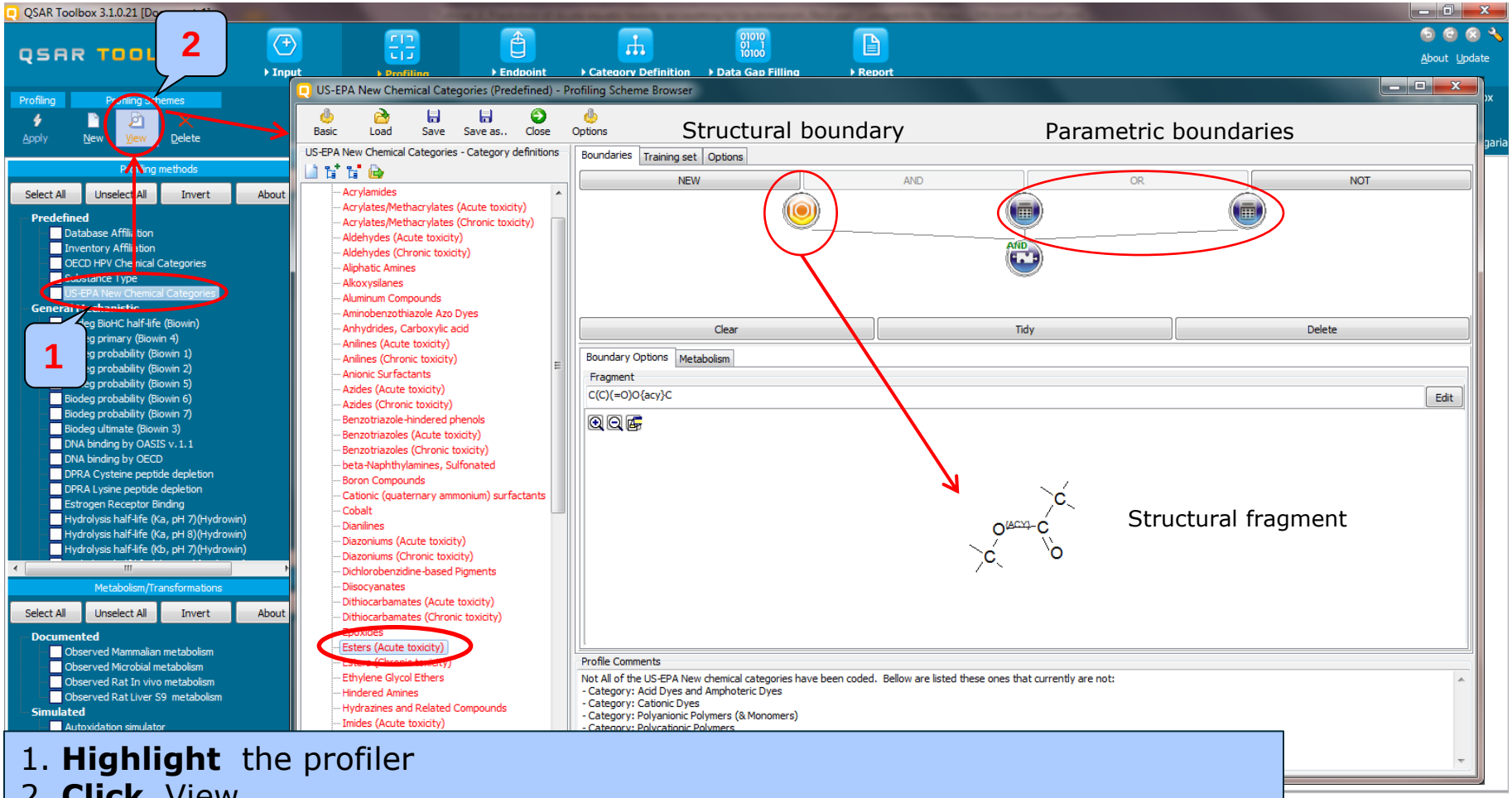
- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, US-EPA New chemical categories and clicking on “View” (see next screen shot).

Profiling
Side-Bar to Profiling



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

Profiling
Side-Bar to Profiling



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

Profiling

Profiling the set of target and tautomers

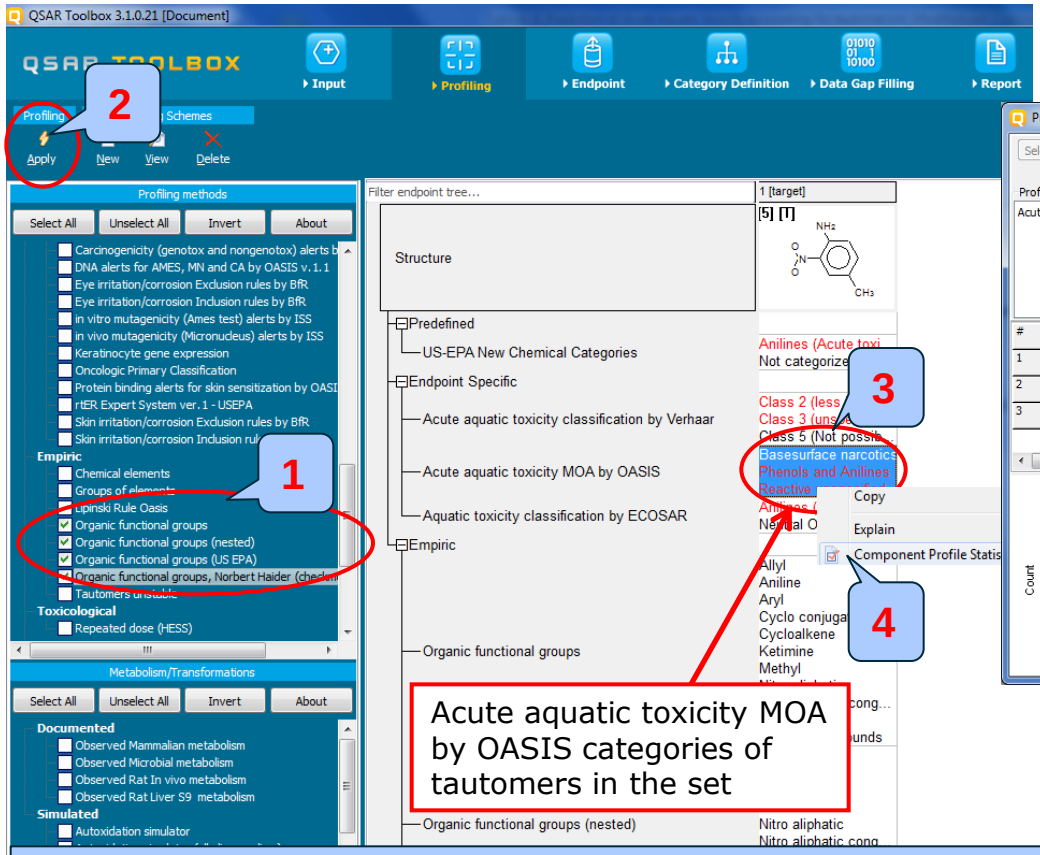
- The following primary profilers relevant to the aquatic toxicity are used in this example(see next screenshot):
  - OECD
  - US-EPA New chemical category
  - Aquatic toxicity classification by ECOSAR
  - Acute aquatic toxicity MOA by OASIS
  - Acute aquatic toxicity classification by Verhaar
  - Organic function groups – all four profilers are used in the assessment

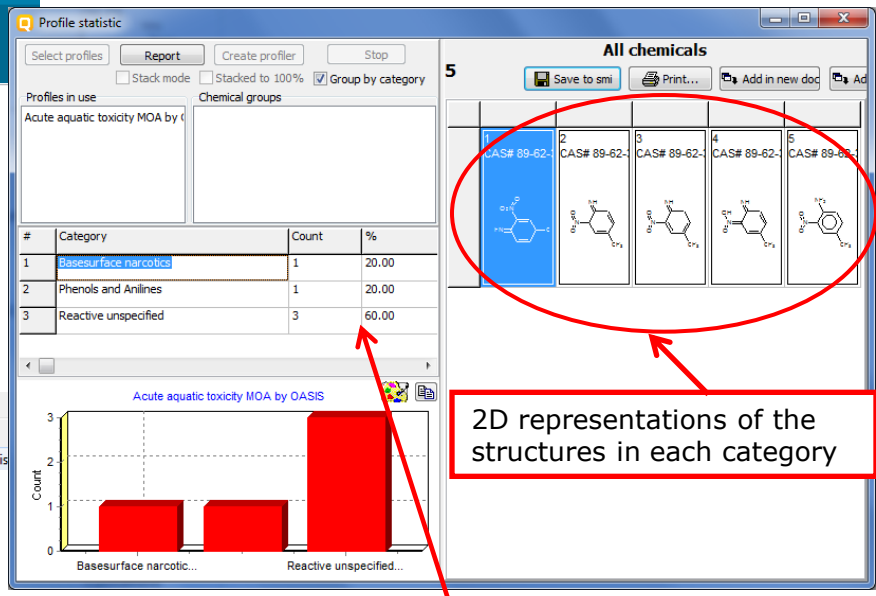
- Select the "Profiling methods" related to the target endpoint by clicking on the box next to the profilers name.

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

Profiling

Profiling the set of target and tautomers





2D representations of the structures in each category

Acute aquatic toxicity MOA by OASIS categories of tautomers in the set

Number of tautomers in a category bin

1. **Check** the profilers related to acute aquatic toxicity as mention on slide #20
2. **Click** Apply
3. Perform **Right click** over the cell with profiling results and
4. **Select** Component profile statistic

Outlook

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

Endpoint

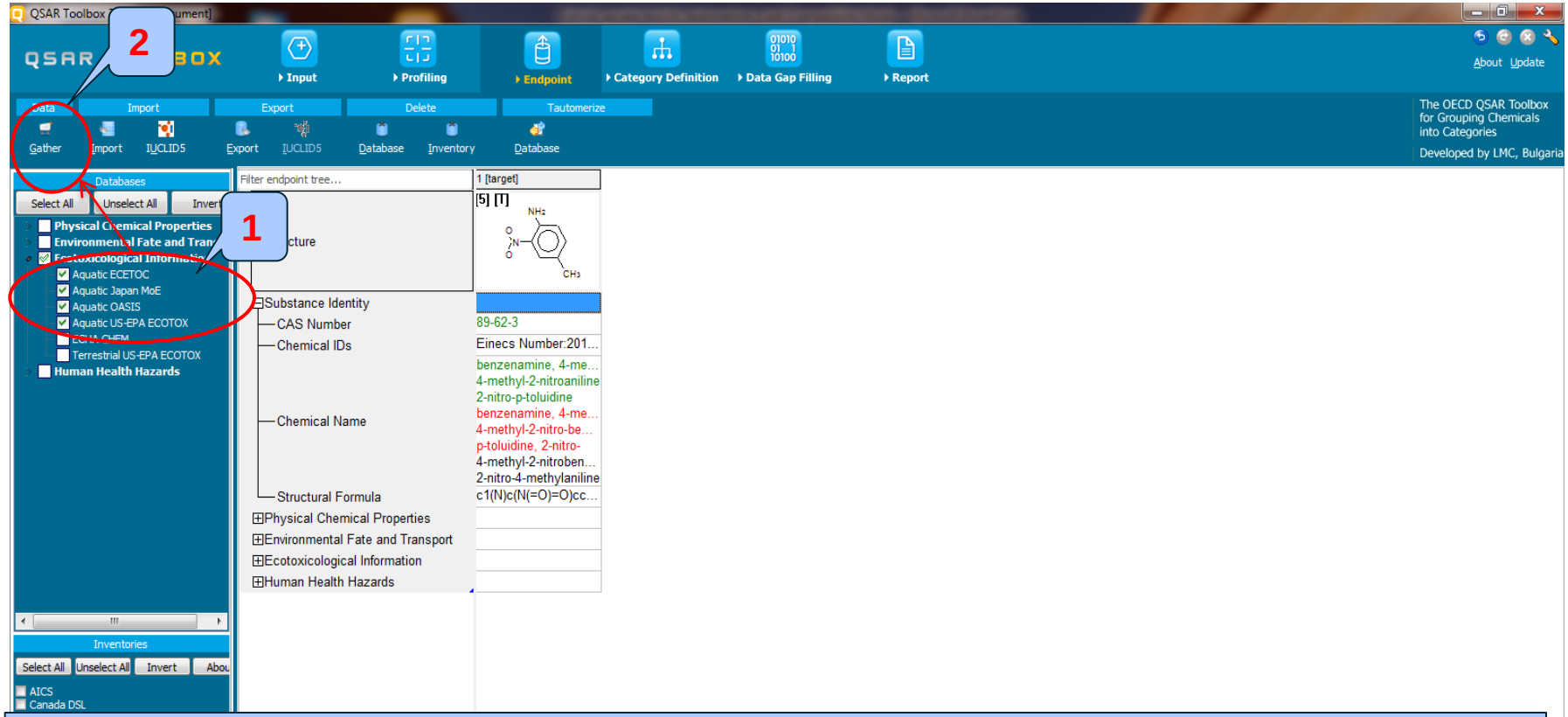
- "Endpoint" refer to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox database.
- Data gathering can be executed in a global fashion (i.e., collecting all data of all endpoints) or on a more narrowly defined basis (e.g., collecting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

Endpoint

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

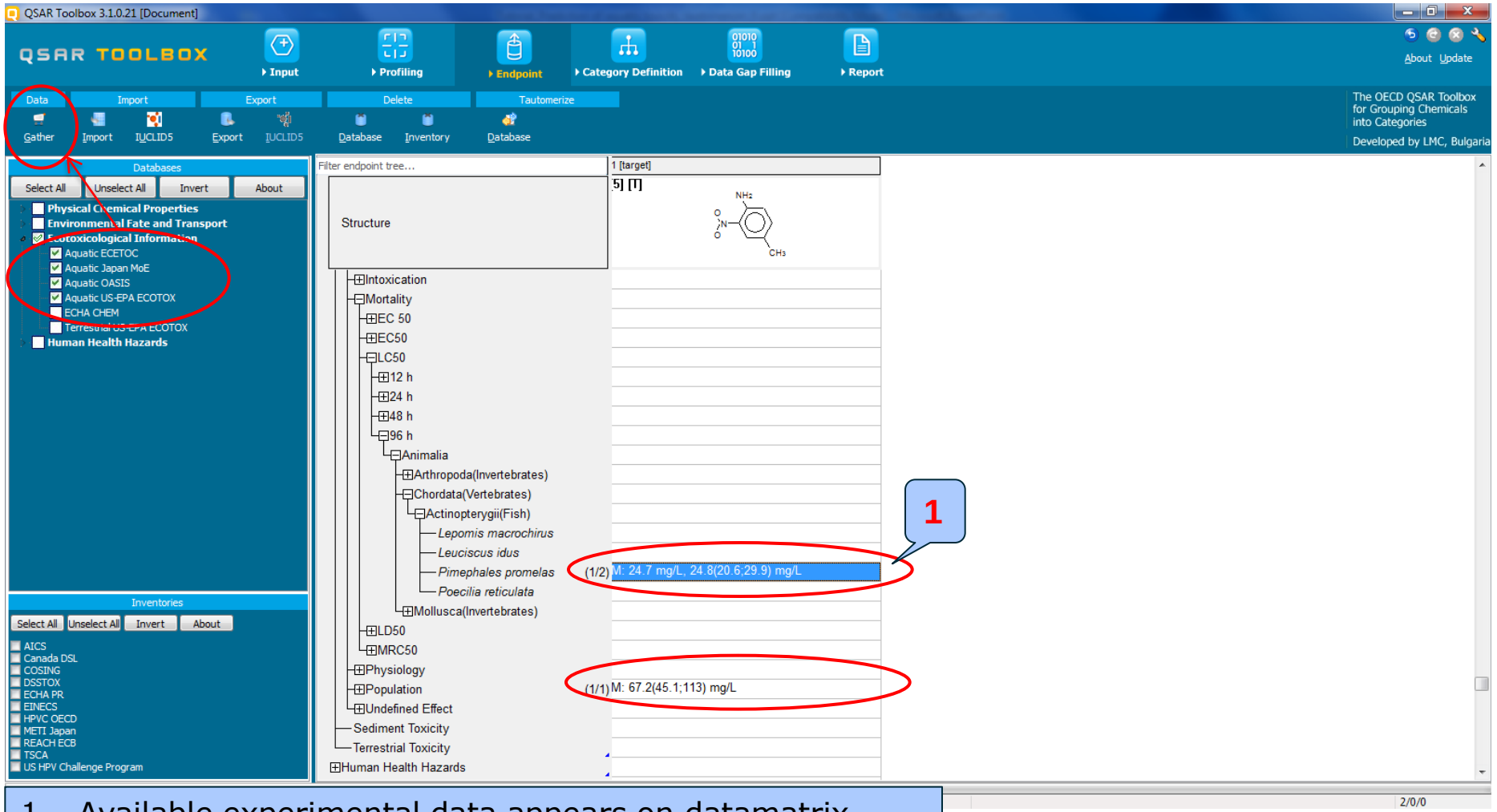
  - ✓ Aquatic ECETOC
  - ✓ Aquatic Japan MoE
  - ✓ Aquatic OASIS
  - ✓ Aquatic US-EPA ECOTOX

Endpoint
Gather data



1. **Select** databases related to the target endpoint by adding a green check in the box before the database name.
2. **Click** Gather

Endpoint



1. Available experimental data appears on datamatrix.

Recap

- You have entered the target chemical
- You have simulate the tautomeric forms of target chemical
- You have gather data if available and found experimental data for one of tautomeric forms (in our case for entered structure).
- 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

*Another general approach for development of categories for acute aquatic tox endpoints are summarized in document "Strategies for grouping chemicals for data gap filling for acute aquatic toxicity endpoints" posted on OECD Website

Category Definition
Grouping methods

- For this example, specific classifications of target and its tautomers are identified by the following profilers: US-EPA, MOA of action and EcoSAR (phase I)
- For this example analogues identified by US-EPA New chemicals category are used for further data gap filling
- Subsequent search of analogues is applied over the set of tautomers having same categories as those of the target tautomeric set

Category definition is a tool for grouping chemicals, which allows to group chemicals based on different measures of "similarity". 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

Also we strongly recommend training exercises. For more details see:

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

Category Definition
Side-bar of US-EPA New chemical categories

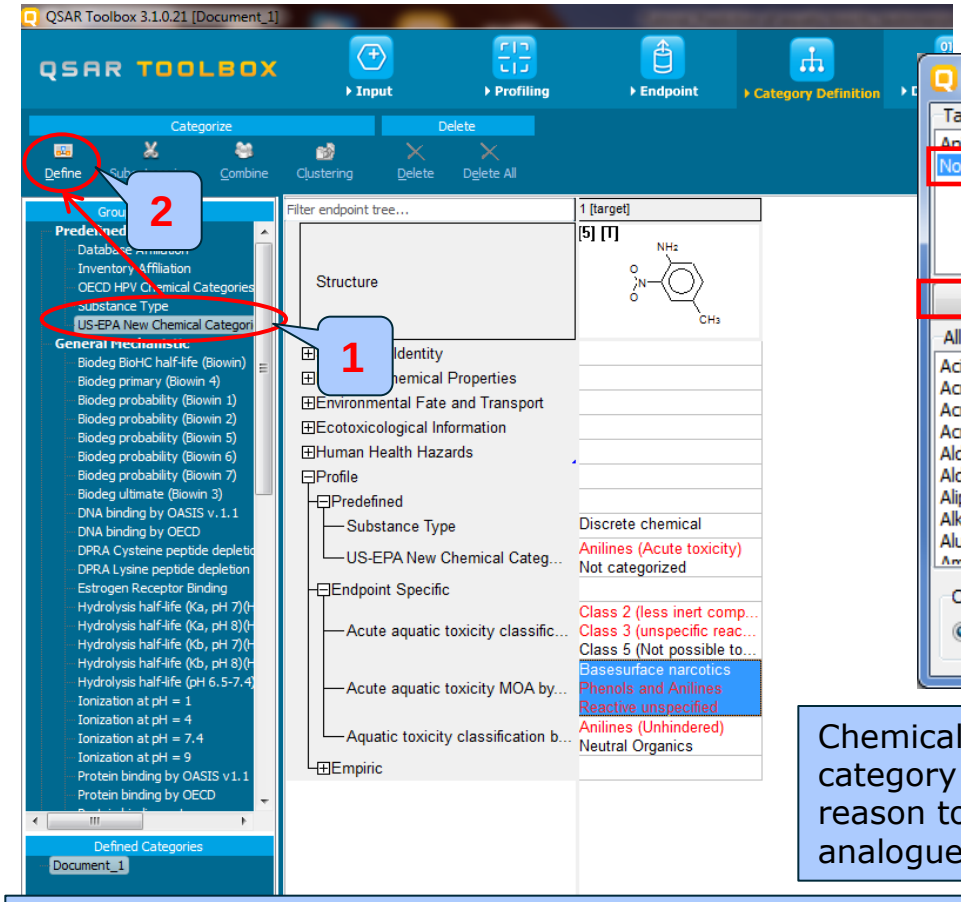
- US-EPA has been used by the U.S. Environmental Protection Agency to predict the aquatic toxicity of new industrial chemicals in the absence of test data

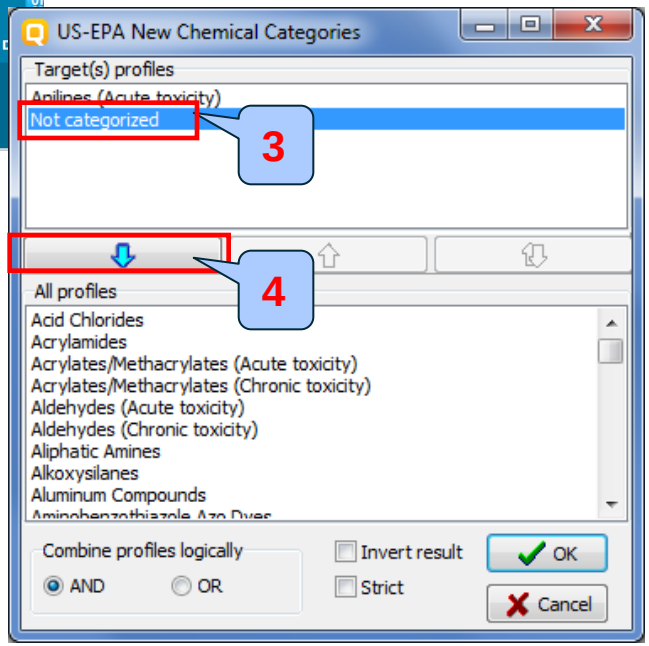
- US-EPA include classes of chemicals for which sufficient regulatory history has been accumulated

- "Classification by US-EPA" in the Toolbox is used for grouping of chemicals by structural similarity which may have mechanistic meaning. Experience has shown US-EPA to be a robust profiler which makes it a logical choice in an initial profiling scheme.

Category Definition

Defining US-EPA New Chemical category



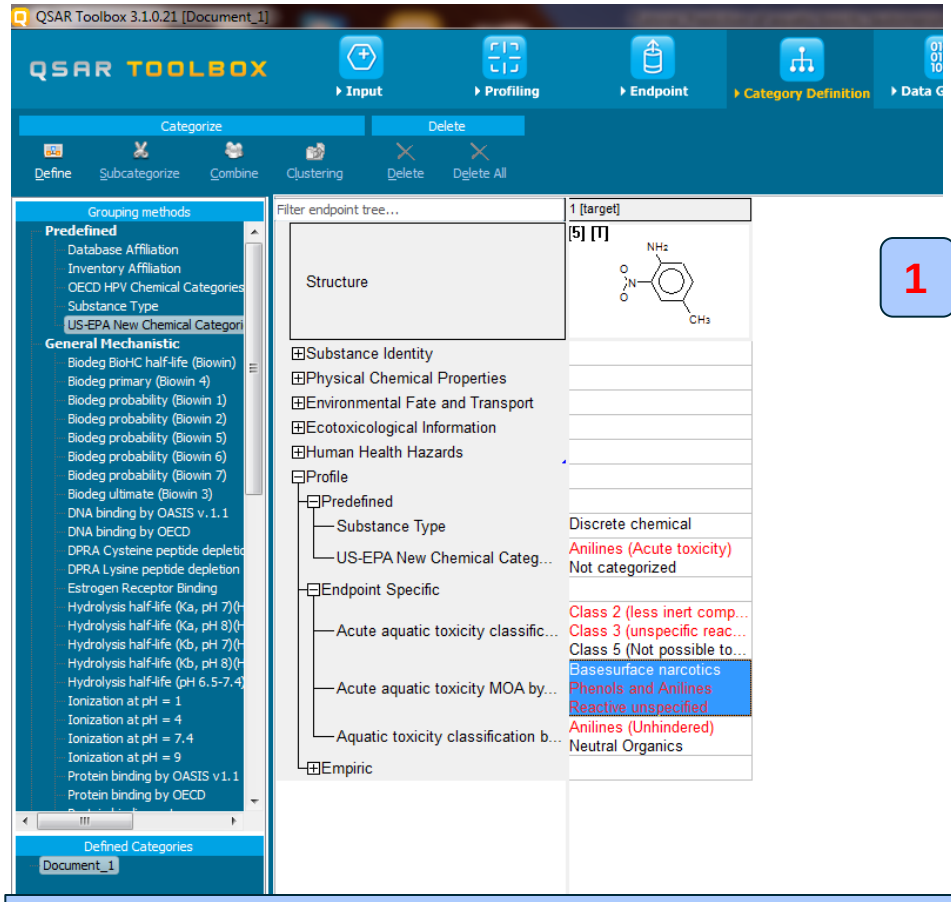


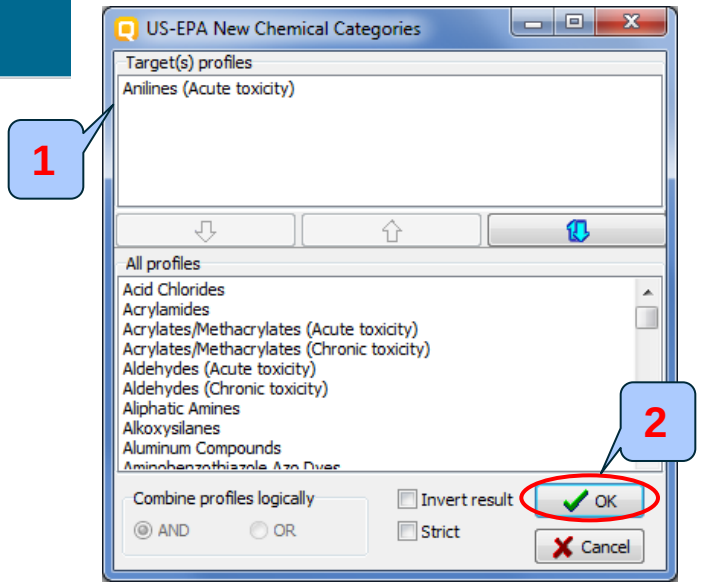
Chemicals (in this case tautomers) without identified toxic category are classified as "Not categorized" This is the reason to exclude this category from the set of searched analogues.

1. **Highlight** "US-EPA New Chemical Categories "
2. **Click** Define
3. **Select** "Not categorized"
4. **Click** arrow down to remove selected category

Category Definition

Defining US-EPA New Chemical category

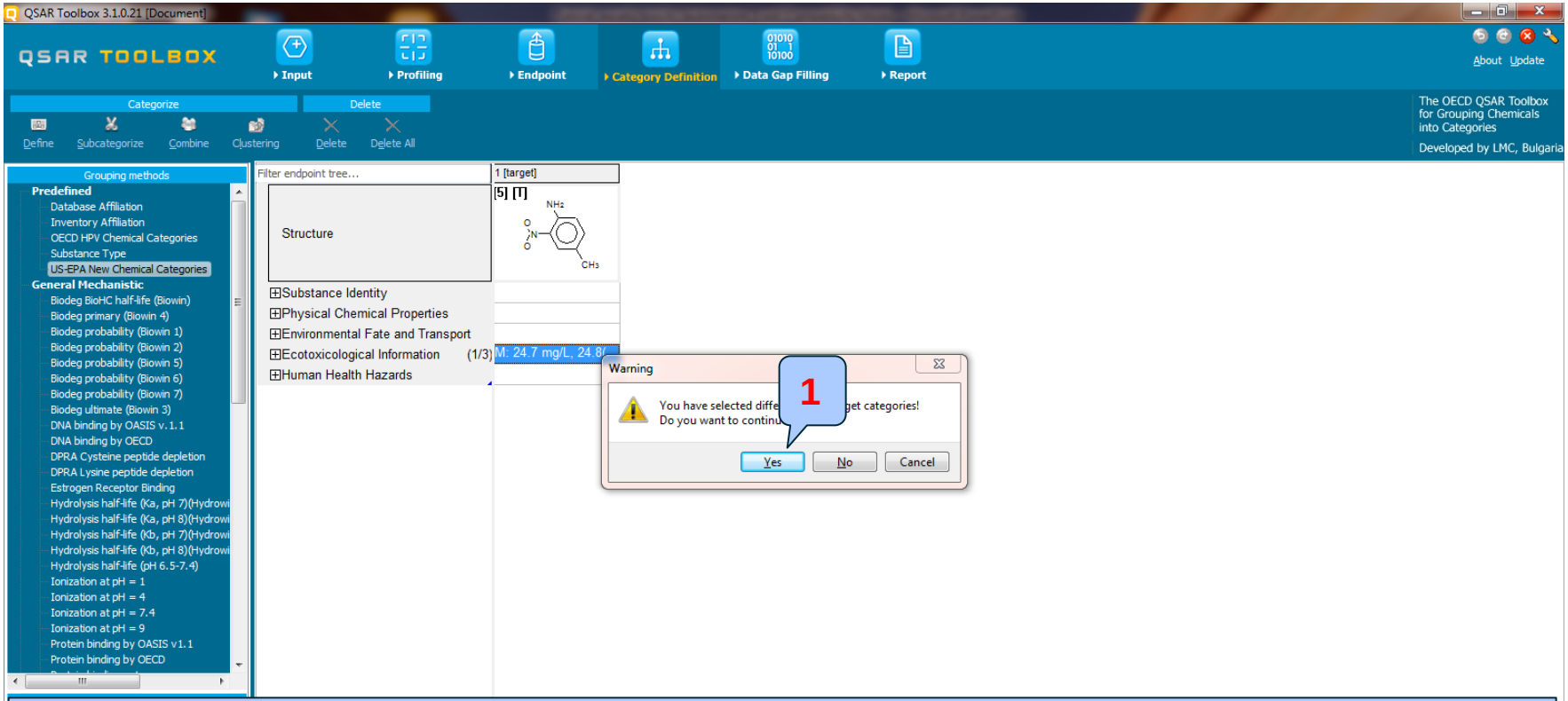




1. Category **"Anilines"** is used for further categorization
2. **Click** OK

Category Definition

Defining US-EPA New Chemical category

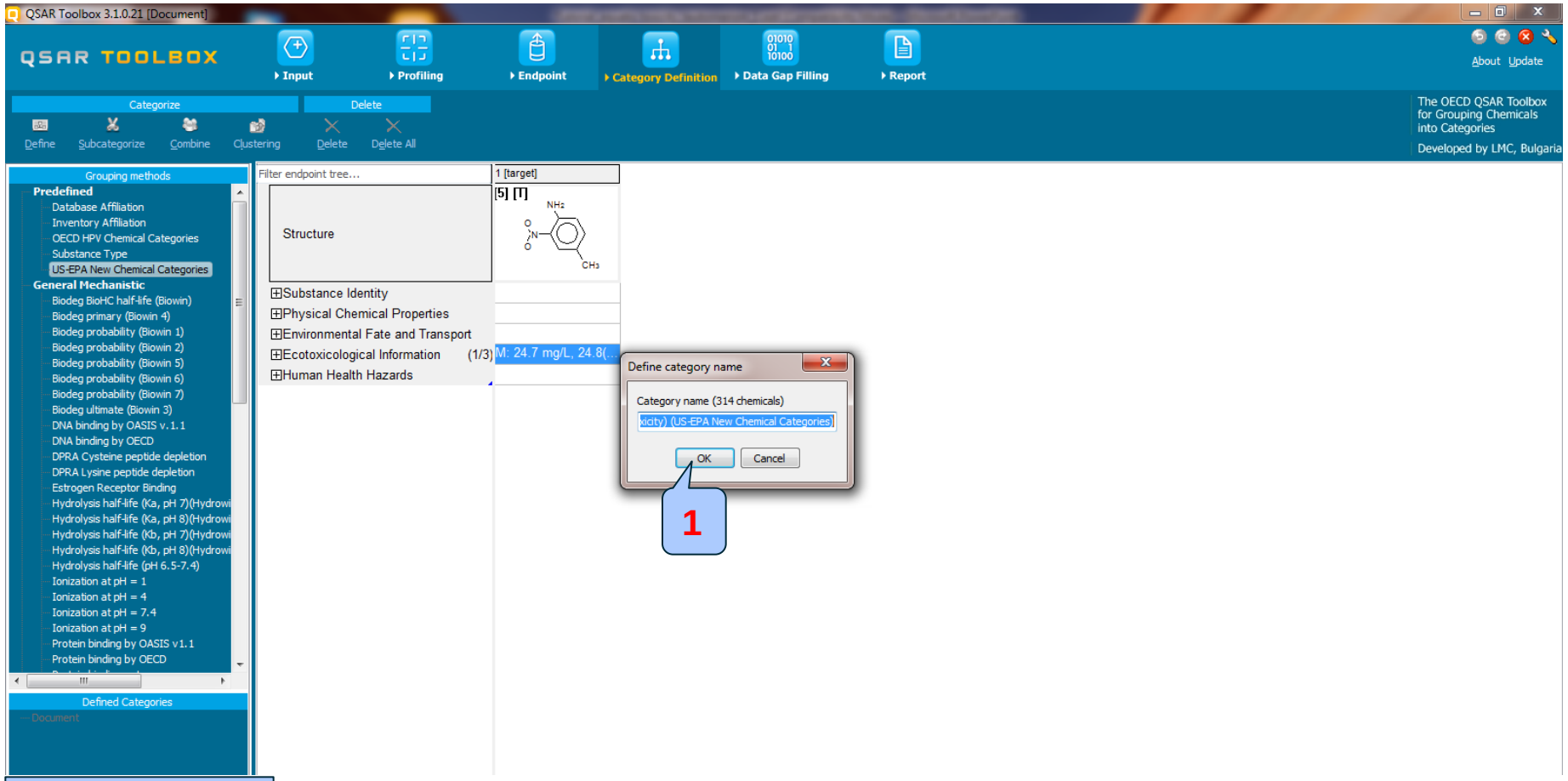


A notification message informs you that you have selected different categories from those of the target.

1. **Select** Yes

Category Definition

Defining US-EPA New Chemical category



1. **Click** OK

Category Definition
Analogues

- The Toolbox now identifies all chemicals represented as tautomeric sets corresponding to the US-EPA classification of "Anilines" which are listed in the databases selected under "Endpoint".
- 314 analogues(tautomeric sets) are identified. Along with the target they form a category (Anilines) which can be used for data gap filling.
- The name of the category appear in the "Defined Categories" window, along with the number of substances belonging to the category.

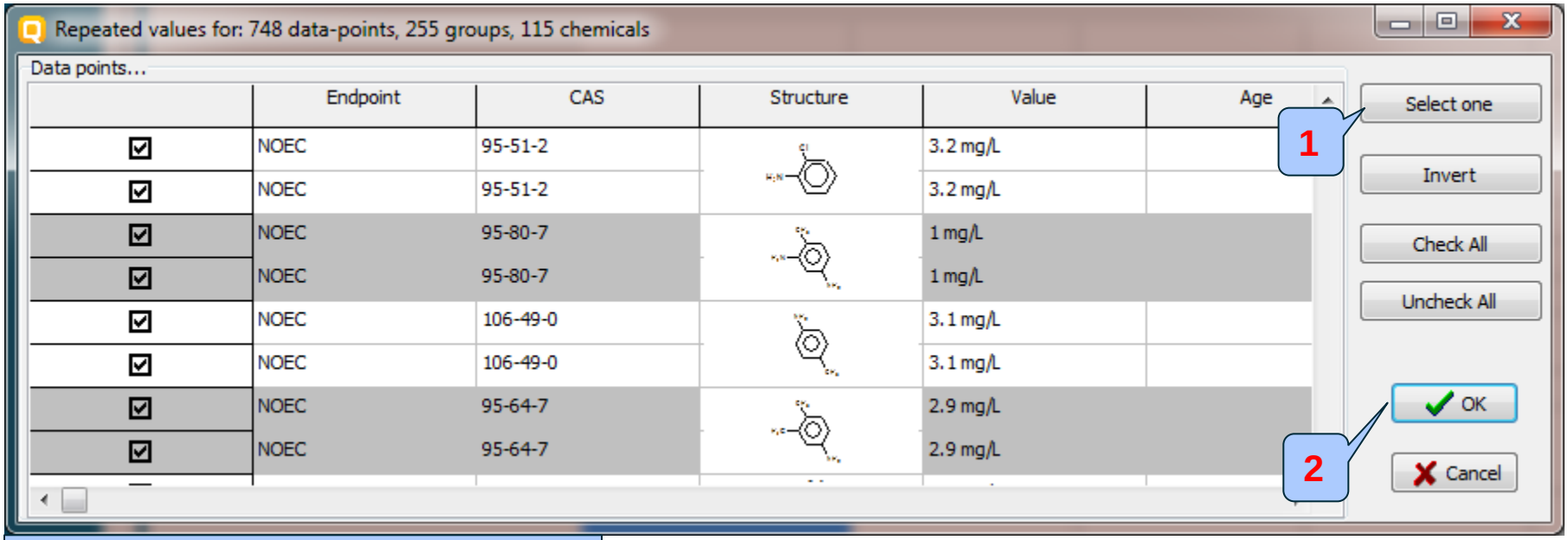
Category Definition
Read data for Analogues

- 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 ecotoxicological endpoints are selected, both options give the same results.
- As the Toolbox must search the database, this may take some time.

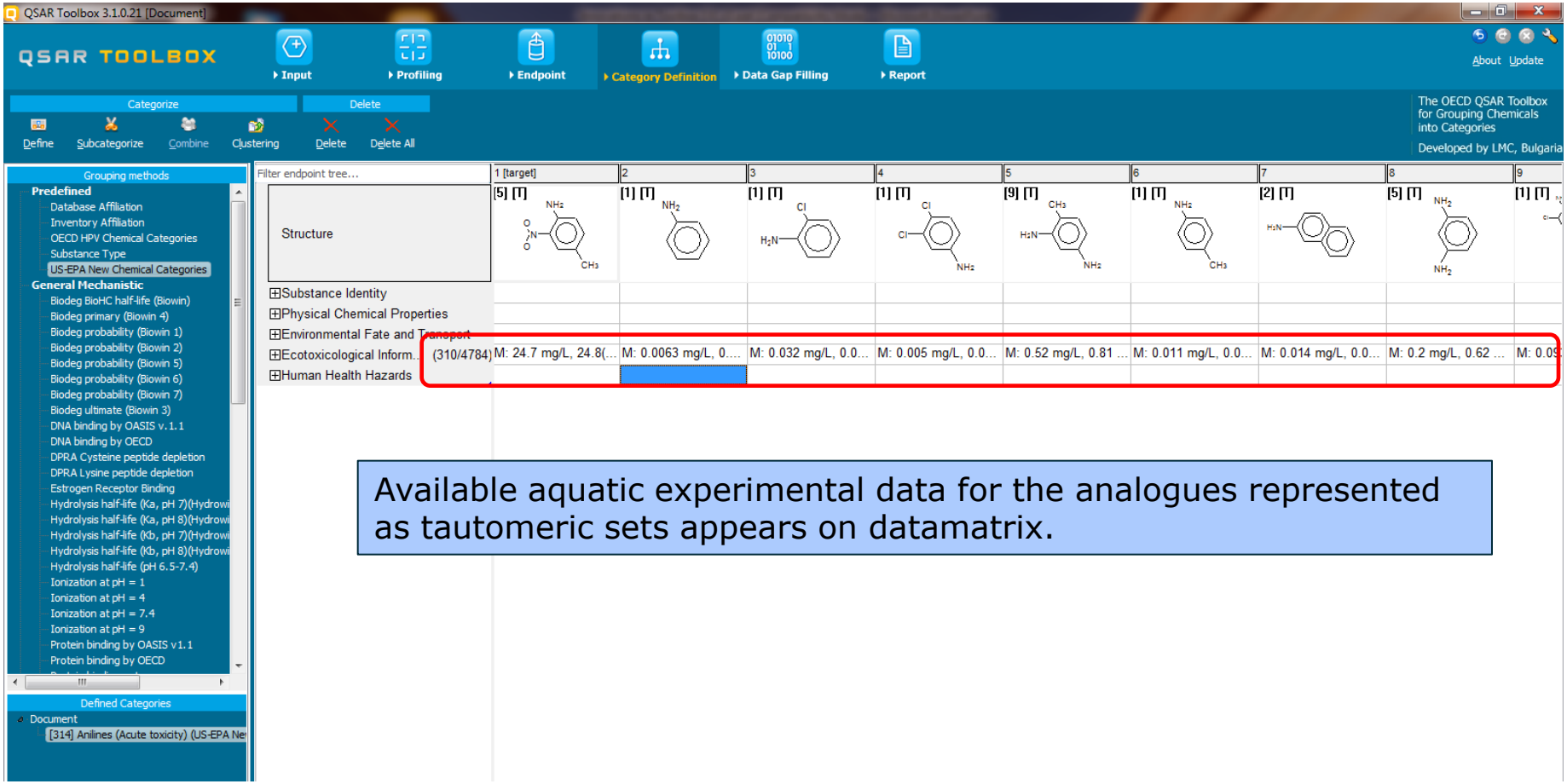
Category Definition
Read data for Analogues

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 and then
2. **Click** OK

Category Definition
Summary information for Analogues



Available aquatic experimental data for the analogues represented as tautomeric sets appears on datamatrix.

Recap

- You have identified a category (“anilines”) with the “US-EPA New Chemical Categories” profiler for the target chemical 4-methyl-2-nitroaniline and its tautomeric forms

- The available experimental results for these 314 analogues represented as tautomeric sets have been collected from the selected databases (Aquatic ECETOC, Aquatic Japan MoE, Aquatic USEPA ECOTOX, and Aquatic OASIS).

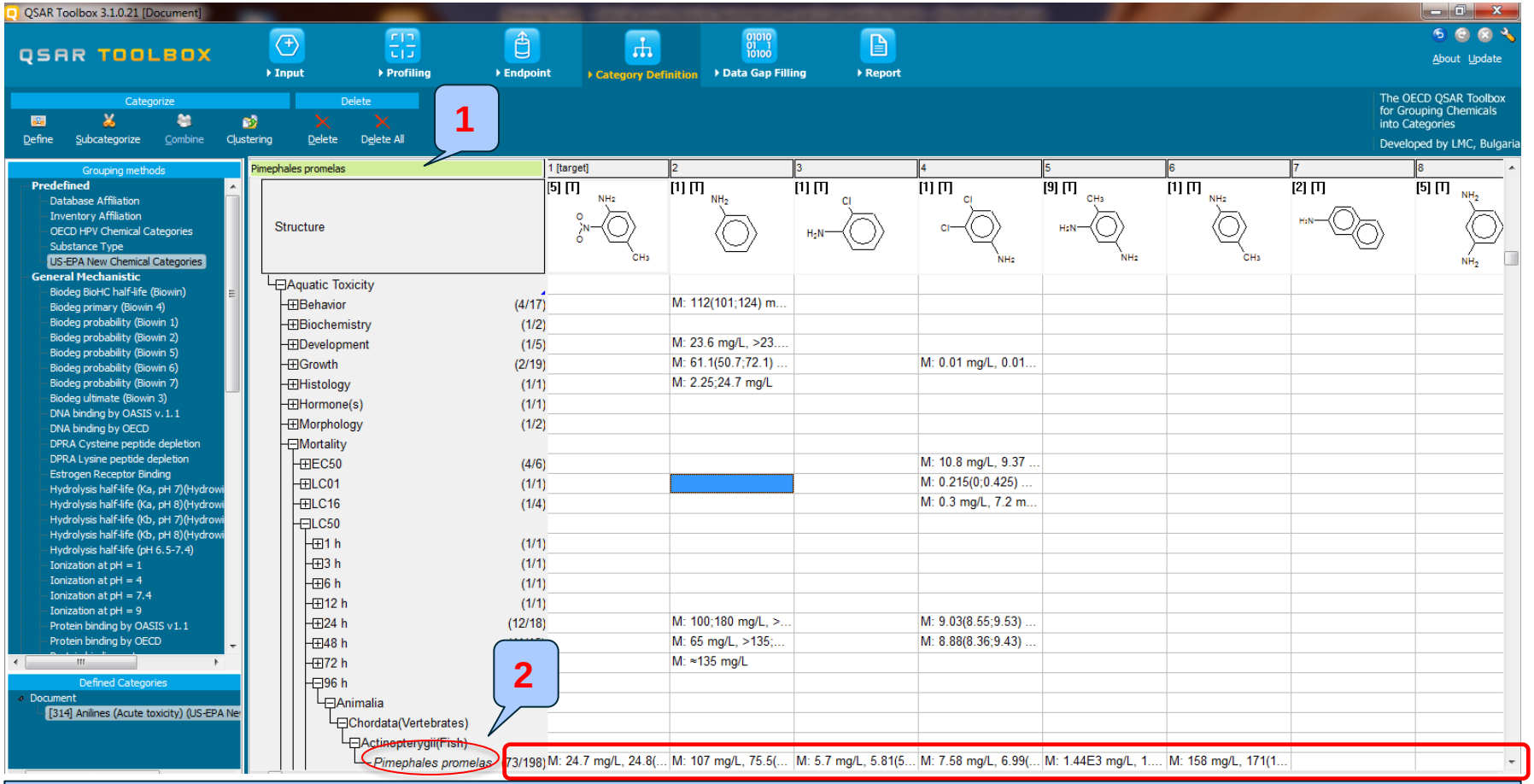
- But before the user can proceed with the “Filling Data Gap” module, he/she should navigate through the endpoint tree and find the specific gap that will be filled.

Category Definition
Navigation through the endpoint tree

- The user can navigate through the data tree by opening (or closing) the nodes of the tree.
- The data tree is extensive but logically constructed; it can be mastered with a practice.
- In this example, the "96 h LC50 Mortality for *Pimephales promelas*" is the target endpoint.
- You can navigate through the endpoint tree by typing the species "*Pimephales promelas*" in the "Filter endpoint tree…" box and double click on Aquatic Toxicity, Mortality, LC50, 96 h, Animalia, etc to *Pimephales promelas* - the specific endpoint (see next screenshot)

Category Definition

Navigation through the endpoint tree



1. **Type** "Pimephales promelas" in the filter box, then **press** Enter
2. **Open** the tree to the target endpoint by **single left** click on the ⊞ sign

Recap

- You have now retrieved the available experimental data on aquatic toxicity for 314 analogue chemicals of target and its tautomeric forms classified as “anilines” by the “US-EPA New Chemical Categories” profiler.

- You have identified the target endpoint of “96 h LC50 Mortality for *Pimephales promelas*”.

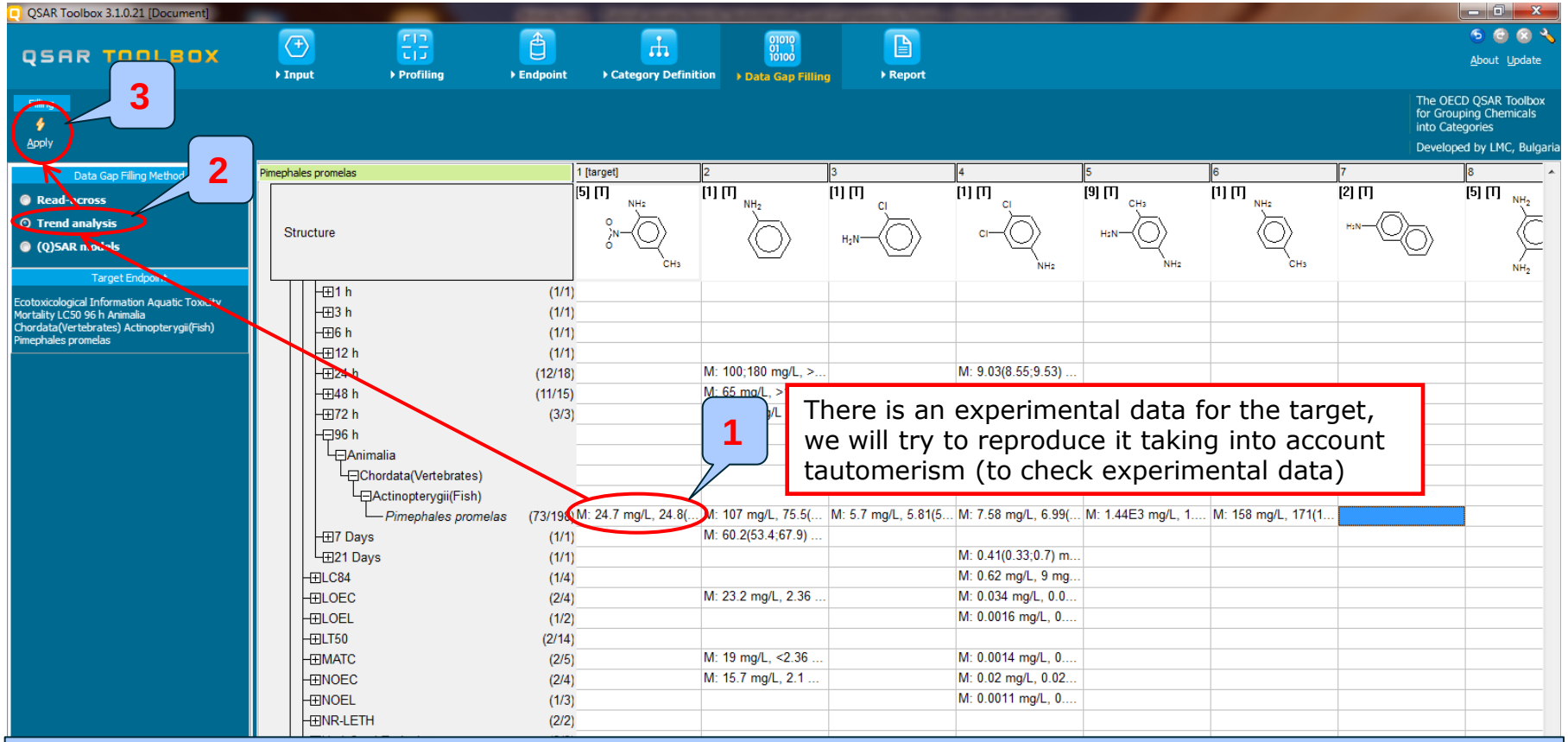
- There is an experimental data for the investigated endpoint, in our exercise we will try to reproduce the experimental data taking into account tautomeric forms of the target

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

Outlook

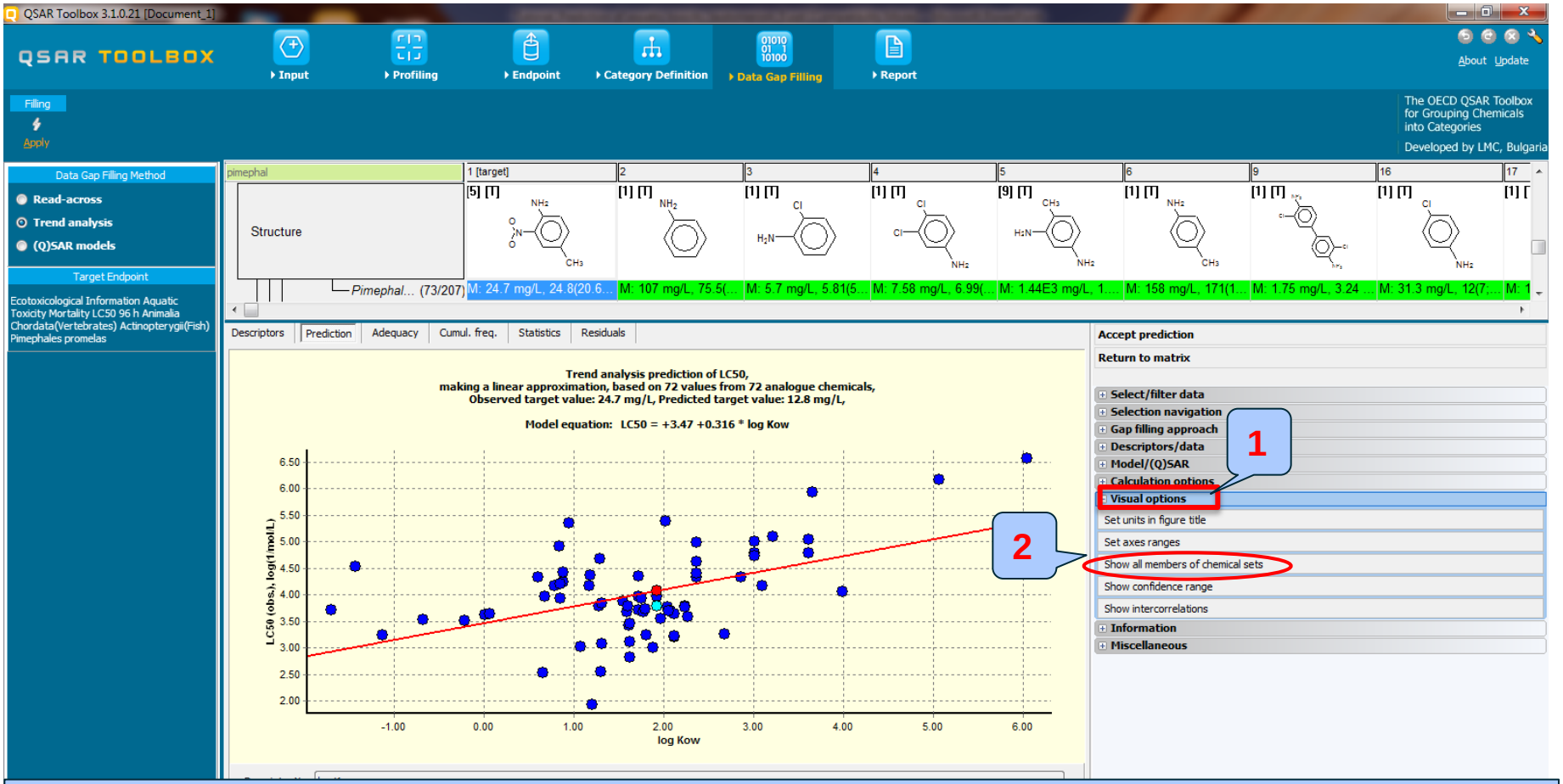
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Category definition
  - **Data Gap Filling**

Data Gap Filling
Apply Trend analysis



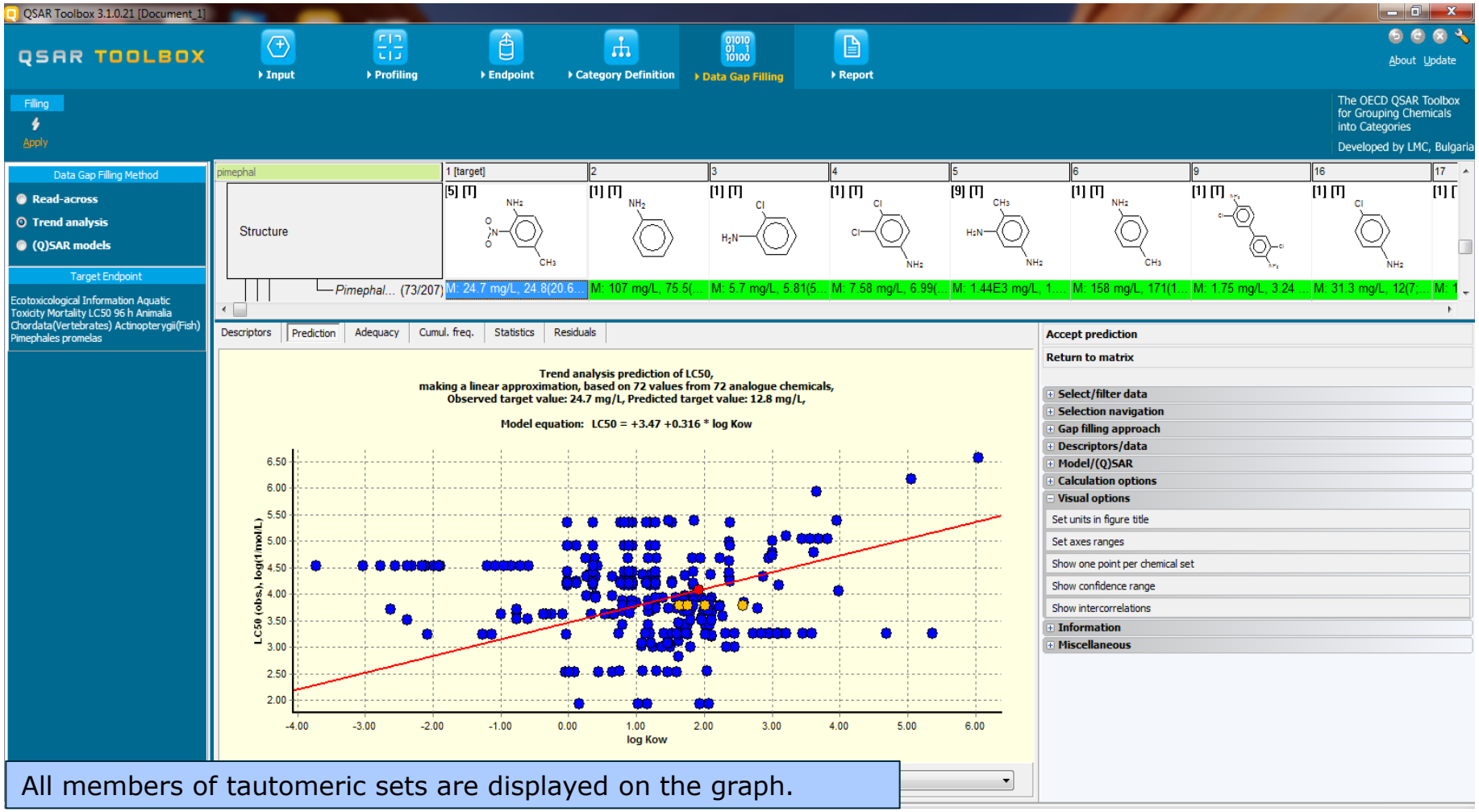
1. **Highlight** the endpoint box corresponding to *Pimephales promelas*/LC50/96h under the target chemical.
2. **Select** Trend analysis
3. **Click** Apply

Data Gap Filling
Results of Trend analysis



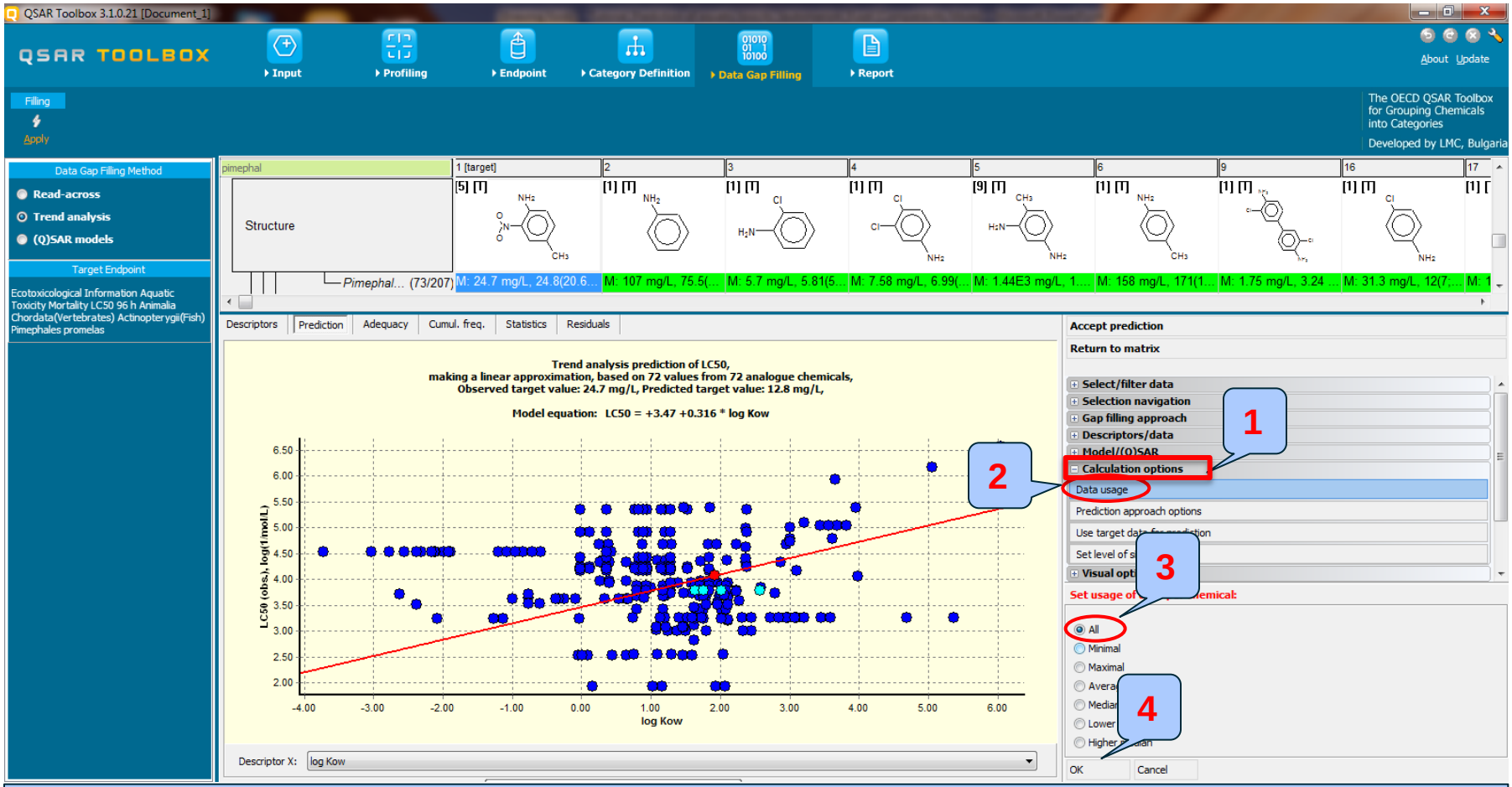
Visualization of members of chemical sets is possible when **click** on 1. Visual options, then **2. Select** Show all members of chemical sets

Data Gap Filling
Results of Trend analysis



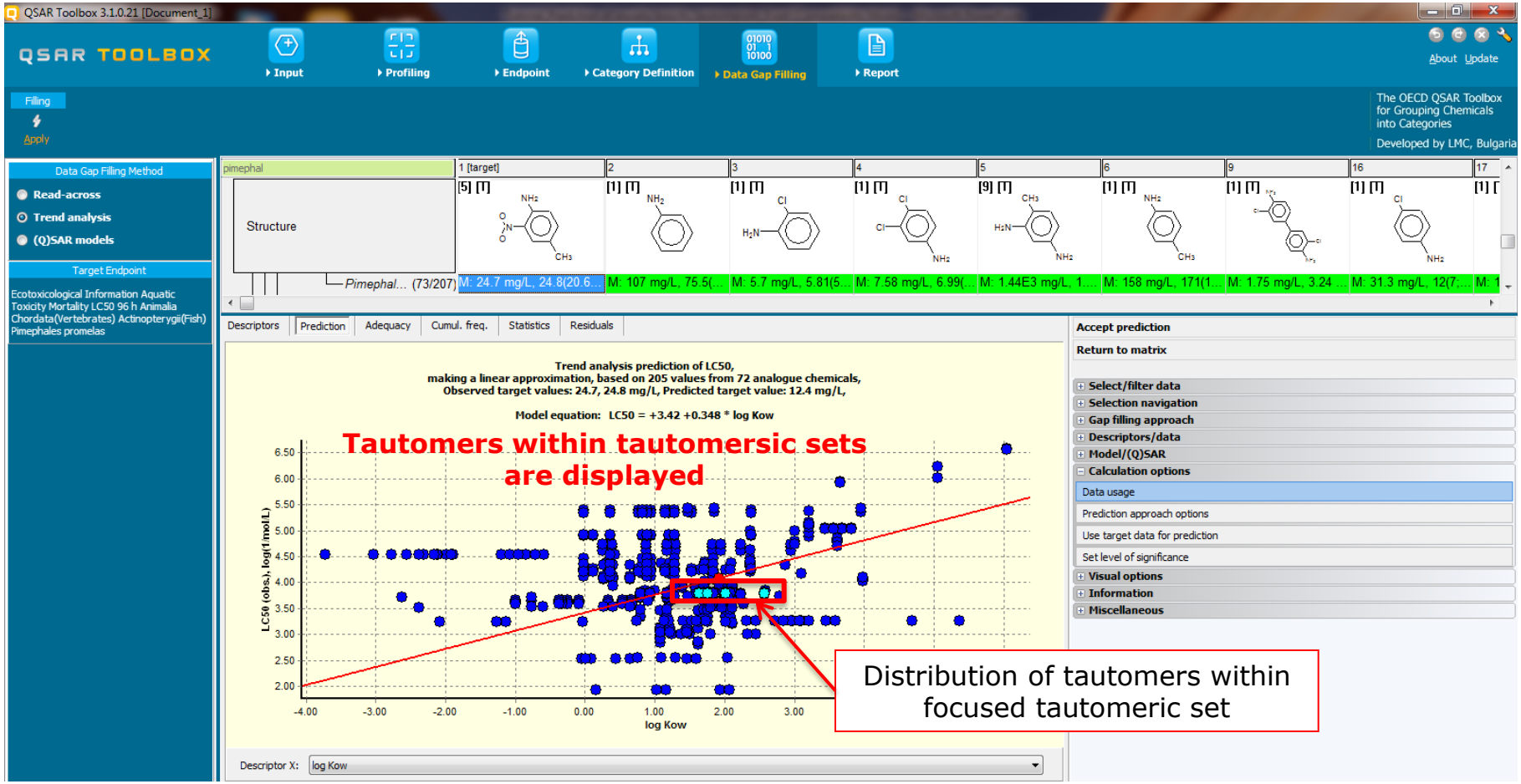
All members of tautomeric sets are displayed on the graph.

Data Gap Filling
Results of Trend analysis



All observed data for chemicals in tautomeric sets could be used in trend analysis when 1. **Open** Calculation options, then 2. **Select** Data usage and 3. **Select** All. Finally 4. **Click** OK

Data Gap Filling
Results of Trend analysis



All members of tautomeric sets are displayed on the graph and all experimental data is taken into account.

Data Gap Filling
Side-Bar of Subcategorisation

- Remember in the Toolbox, a category retrieved for tautomeric set refers to a group of chemicals with its tautomeric forms which have same profiling result according to one of the profilers listed in the module "Profiling".
- Subcategorisation refers to the process of applying additional profilers to the previously defined category; subcategorization procedure can be applied on:
  - **Single chemical** – eliminate chemicals having different categories than those of the target
  - **Set of tautomers** – eliminate tautomeric sets as a whole having different categories than those of the target tautomeric set
  - **Tautomers within tautomeric set** – eliminate specific tautomers within tautomeric set, which have categories different than those of the target.
- Graphical illustration of subcategorization procedure is given on next three slides

Data Gap Filling

Side-Bar of Subcategorisation

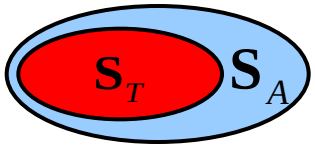
Elimination of single chemical

Features of the target: S_T

Features of analogue: S_A

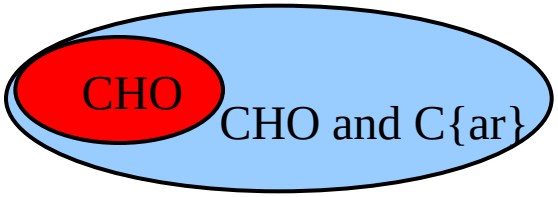
Refinement of category (Sub-categorization)

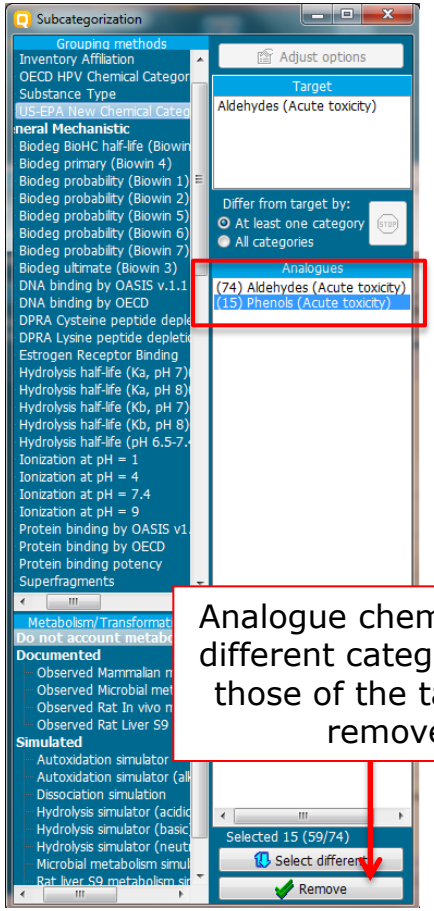
Analogue has different features of the target:



$$S_A \neq S_T$$

Example





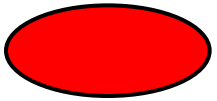
Analogue chemicals with different categories than those of the target are removed

Data Gap Filling

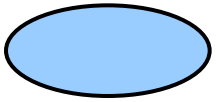
Side-Bar of Subcategorisation

Eliminating set of tautomers

Features of the target tautomeric set: S_T

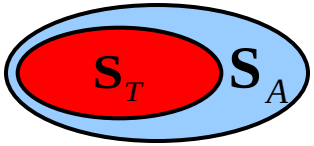


Features of analogues (as tautomeric sets): S_A



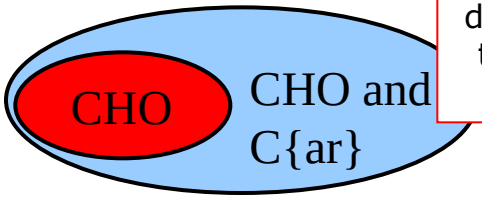
Refinement of category (Sub-categorization)

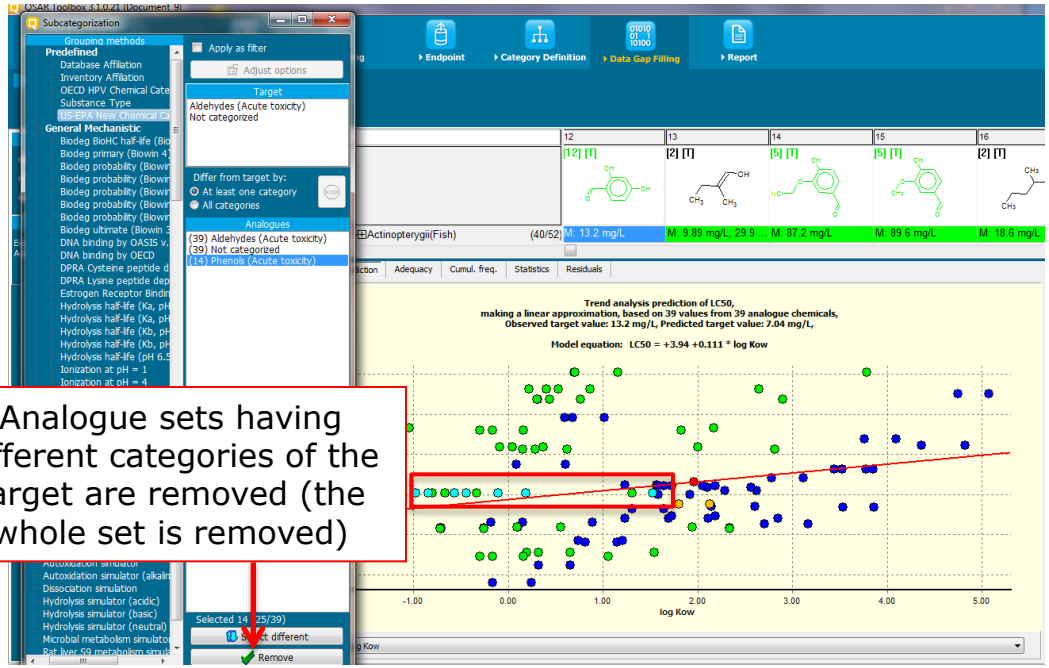
Analogue as tautomeric set has different features of the target:



$$S_A \neq S_T$$

Example



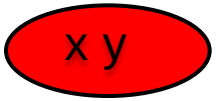


Data Gap Filling

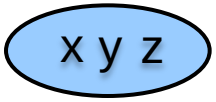
Side-Bar of Subcategorisation

Eliminating tautomers from a tautomeric sets

Features of the target tautomeric set: S_T



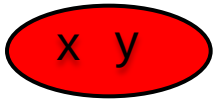
Features of analogues (as tautomeric sets): S_A



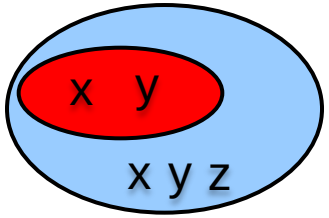
Refinement of category (Sub-categorization)

Tautomers within tautomeric sets have different features of the target:

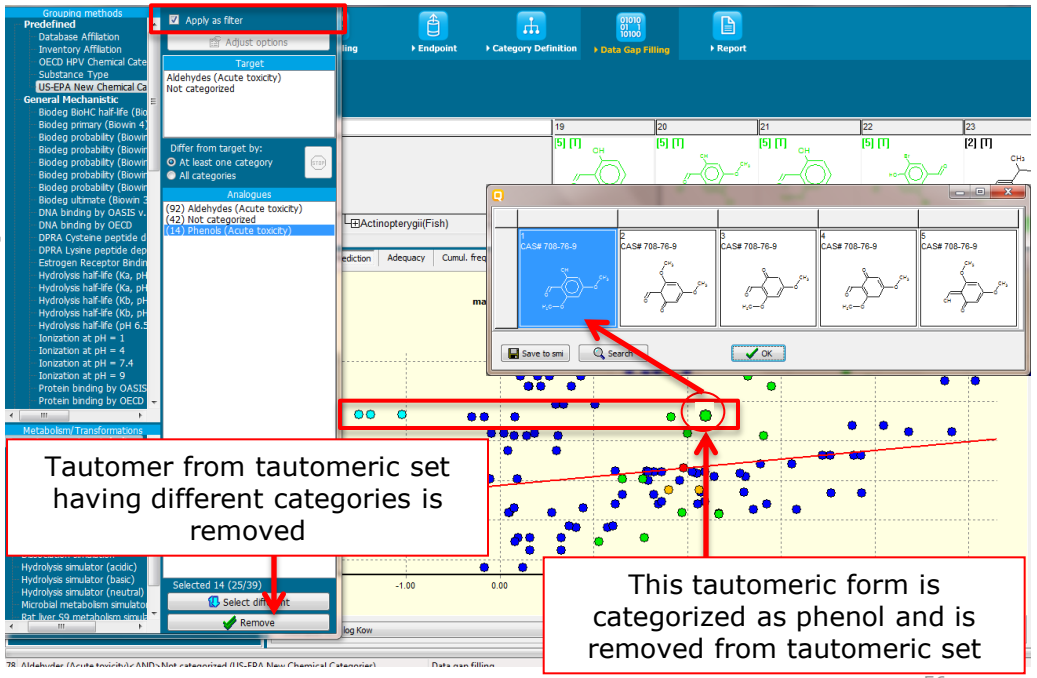
Tautomeric set of target



Tautomeric set of analogue



$$S_A \neq S_T$$



Data Gap Filling
Side-Bar of Subcategorisation

- In this example, the following subcategorizations are applied in order to eliminate dissimilar tautomeric sets:

  - Substance type

    The categorisation based on substance type allows keeping among the analogues only those that are of the same chemical type: discrete chemicals, mixtures, polymers, inorganics, organometalics. The current target represented as tautomeric set include discrete chemicals only. Hence the analogues (tautomeric sets) should also be discrete chemicals.

  - Aquatic toxicity classification by ECOSAR

    The categorization based on mode of action identifies analogues (in this case tautomeric sets) having the same mode of action as the target (i.e phenols and anilines). The analogues (tautomeric sets) having different categories should be eliminated.

Cont'd on next slide

Data Gap Filling
Side-Bar of Subcategorisation

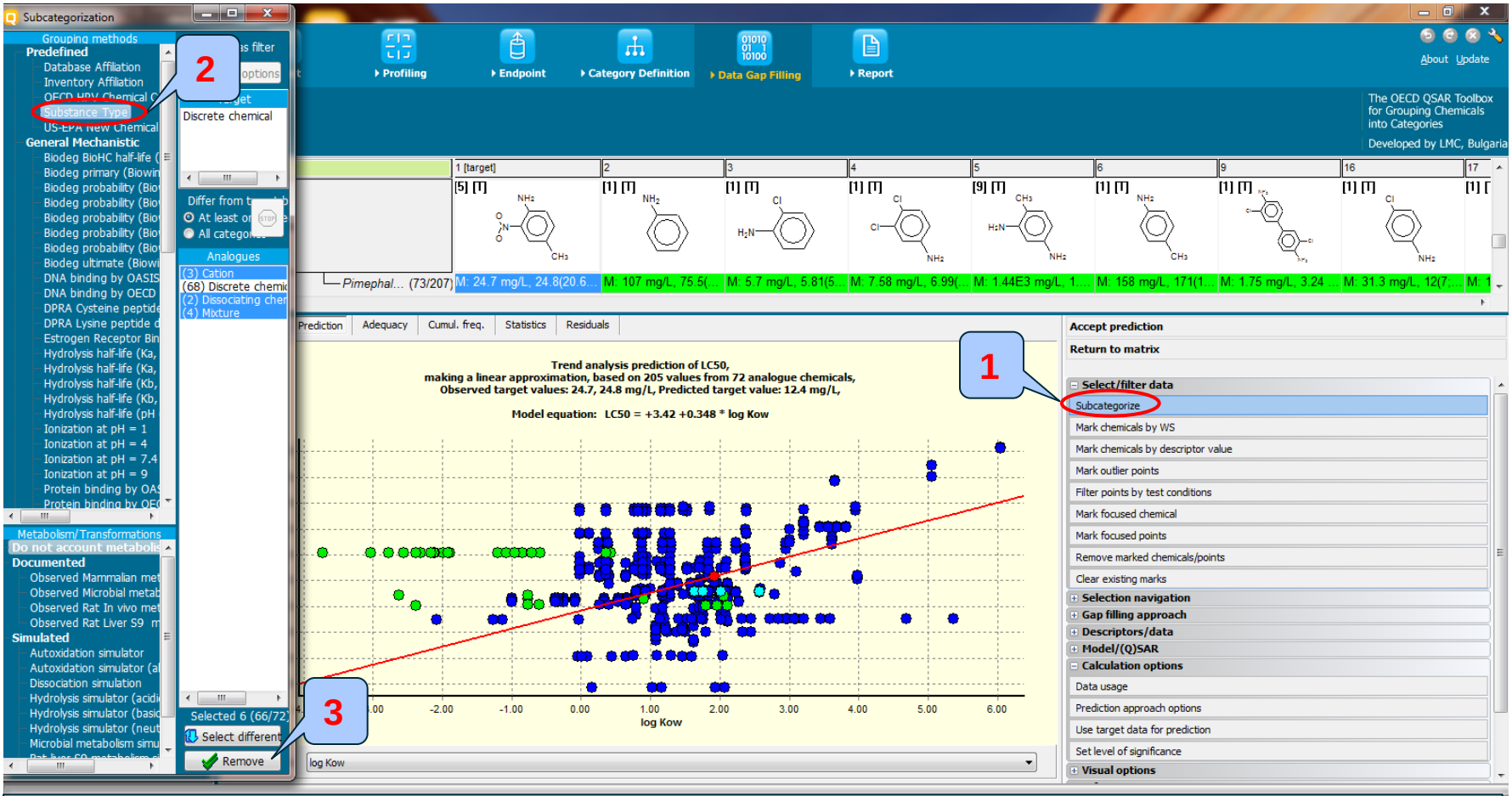
- In this example, the following subcategorizations is applied in order to eliminate dissimilar tautomeric sets
  - Chemical elements

    The categorisation based on Chemical elements allows keeping among the analogues (tautomeric sets) only those that have same chemical elements as the target tautomeric sets.

Subcategorisation steps are demonstrated on the next 4 screen shots.

Data Gap Filling

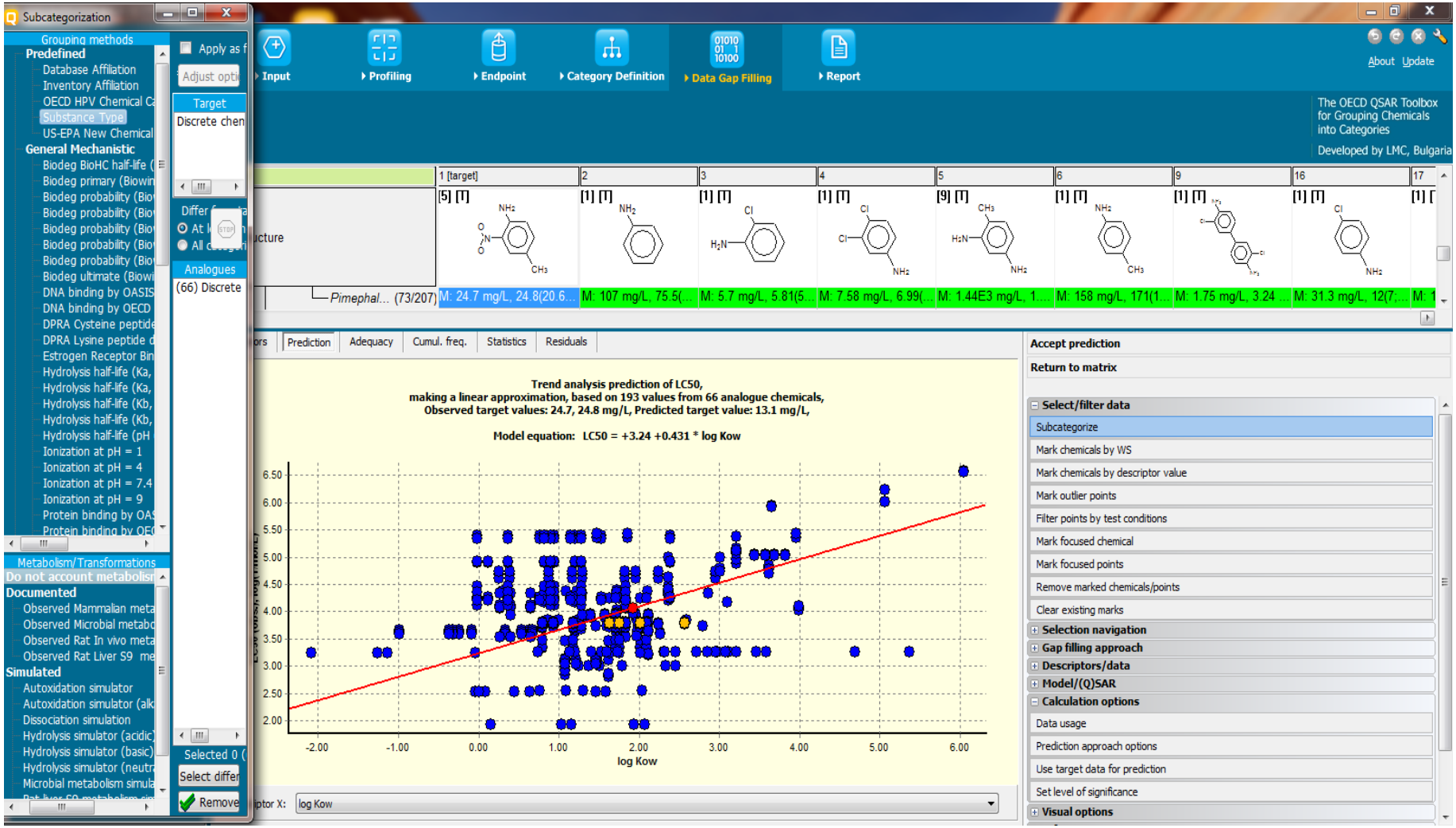
Subcategorisation by Substance type



1.**Click** Subcategorize 2. **Select** Substance type 3. **Click** Remove to eliminate dissimilar tautomeric sets (the whole set is removed)

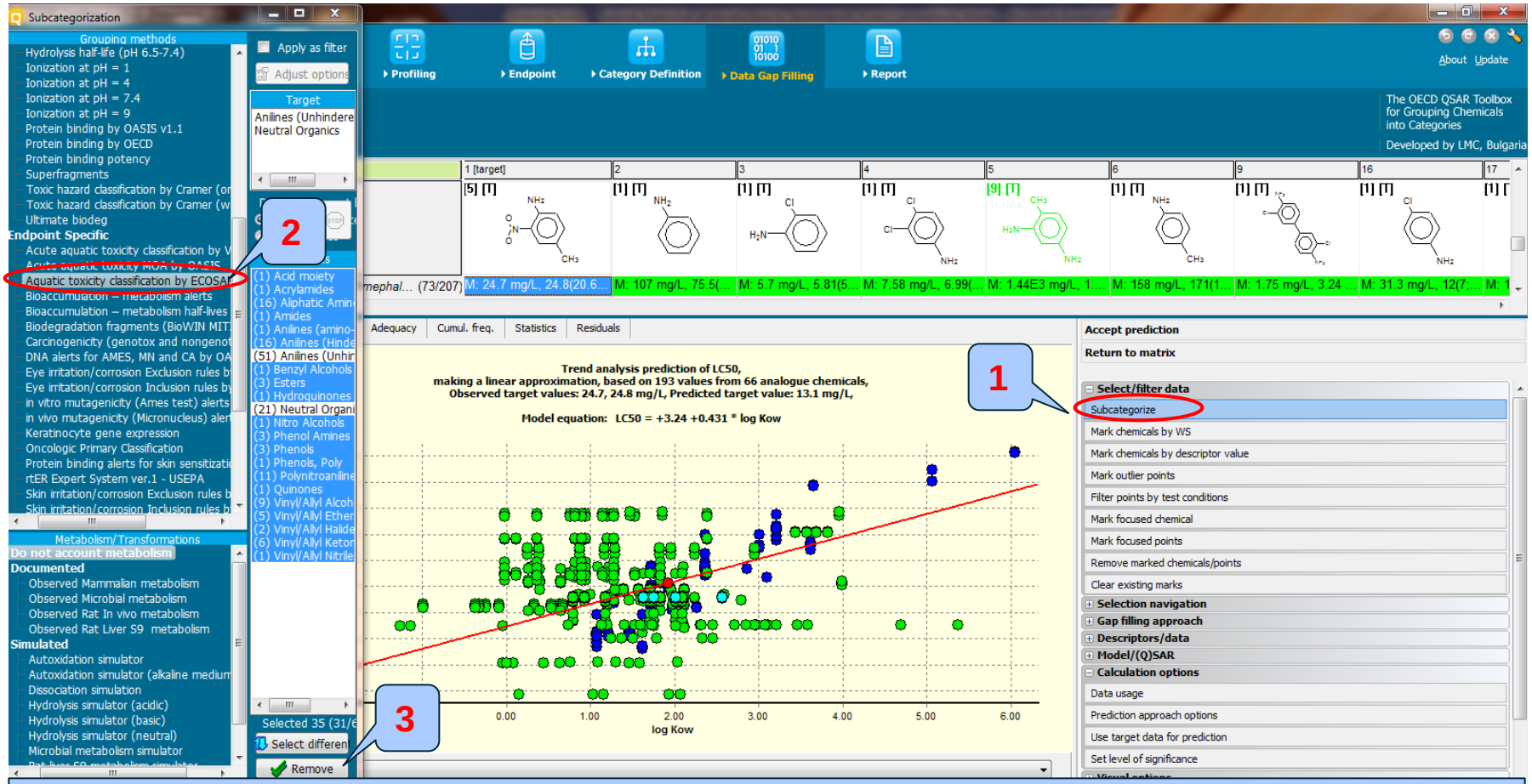
Data Gap Filling

Result of Subcategorisation by Substance type



Data Gap Filling

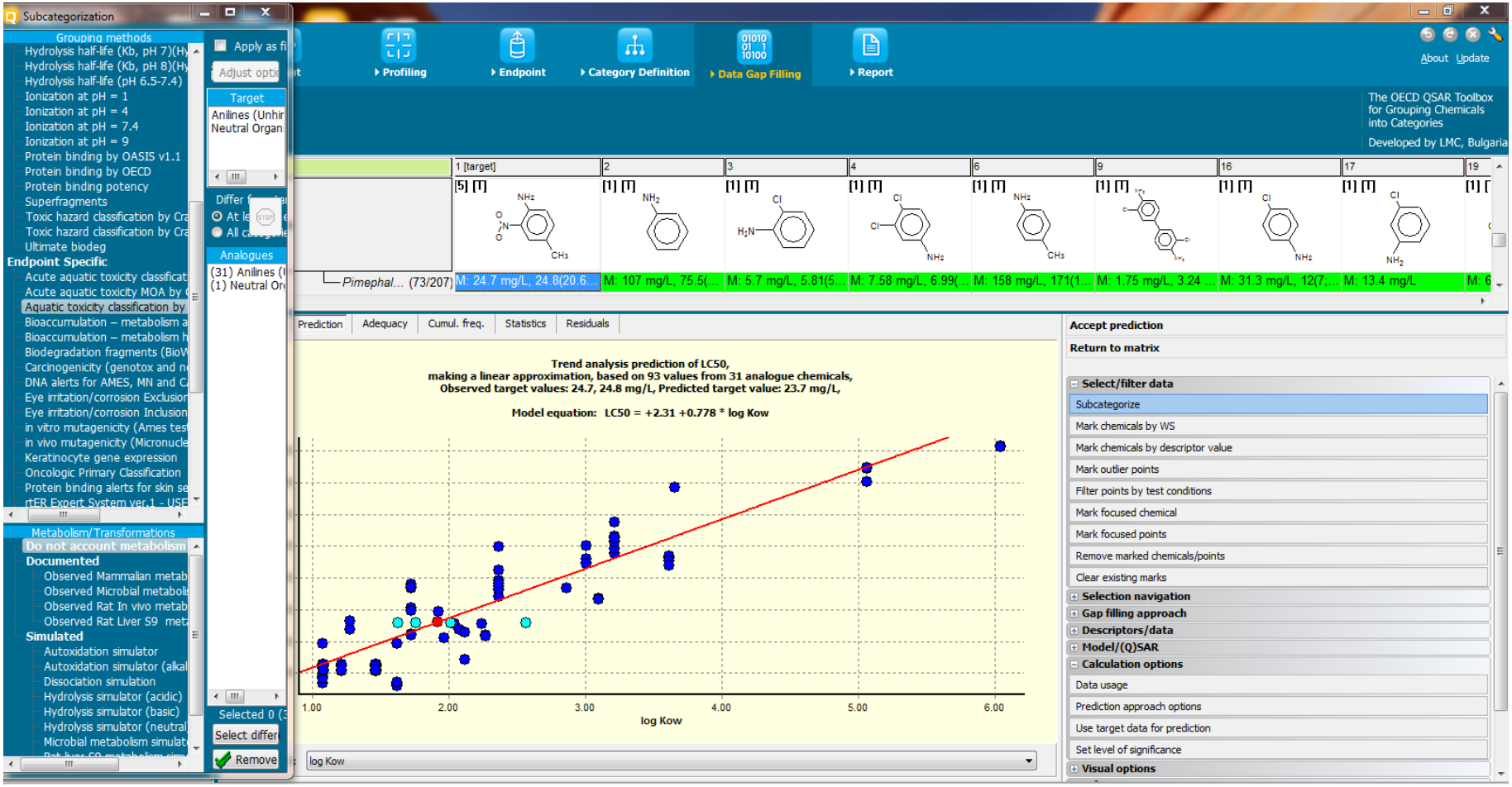
Subcategorisation by Aquatic toxicity classification by ECOSAR



1. **Click** Subcategorize 2. **Select** Aquatic toxicity classification by ECOSAR
3. **Click** Remove to eliminate dissimilar tautomeric sets (the whole set is removed)

Data Gap Filling

Result of Subcategorisation by Aquatic toxicity classification by ECOSAR



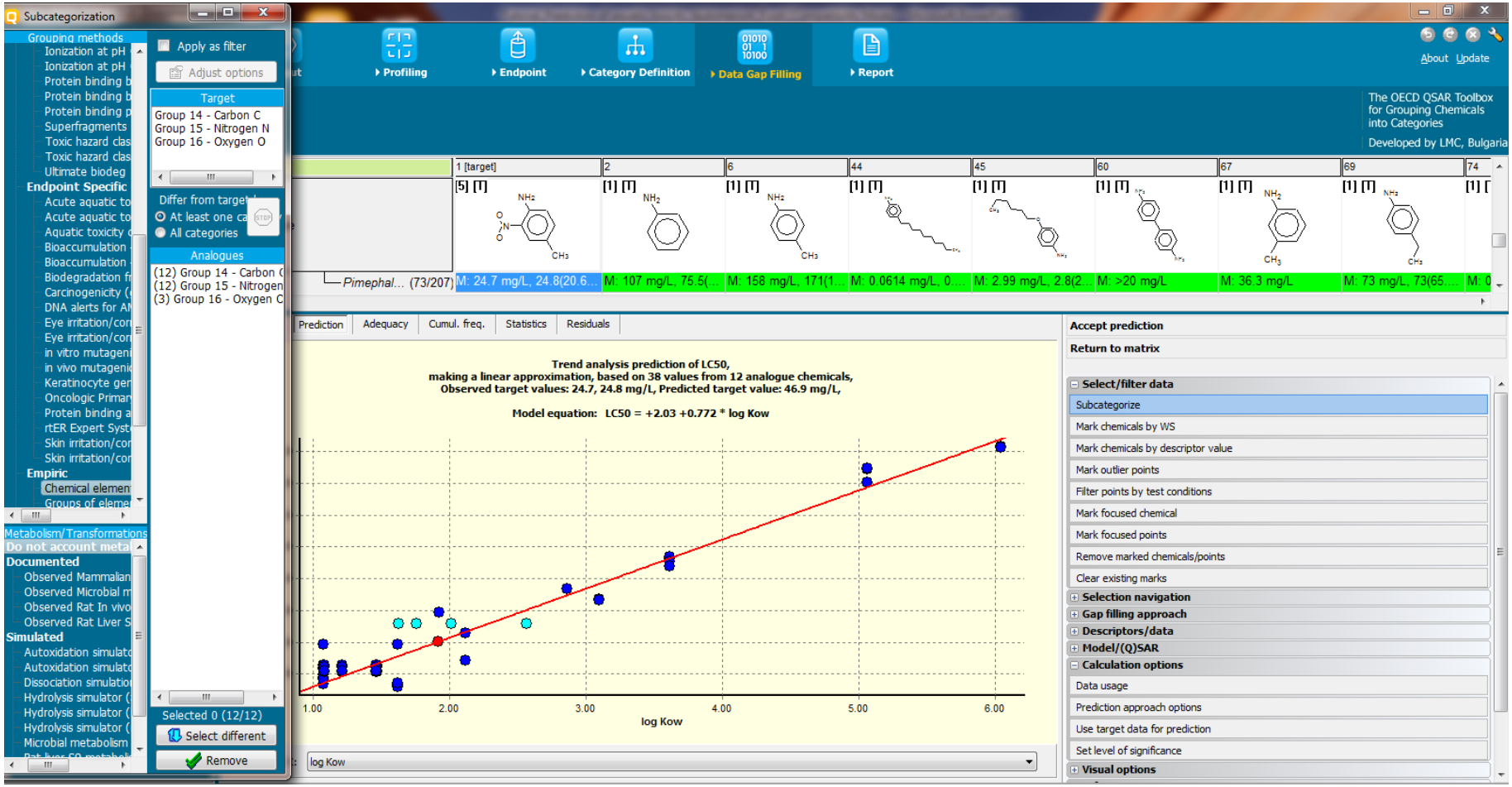
Data Gap Filling

Subcategorisation by Chemical elements

1. **Select** Chemical elements
2. **Click** Remove to eliminate dissimilar tautomeric sets (the whole set is removed)

Data Gap Filling

Result of Subcategorisation by Chemical elements



Data Gap Filling
Side-Bar of Subcategorisation

The last subcategorisation procedure eliminates unstable tautomeric forms from given tautomeric sets. This elimination is possible with respect to

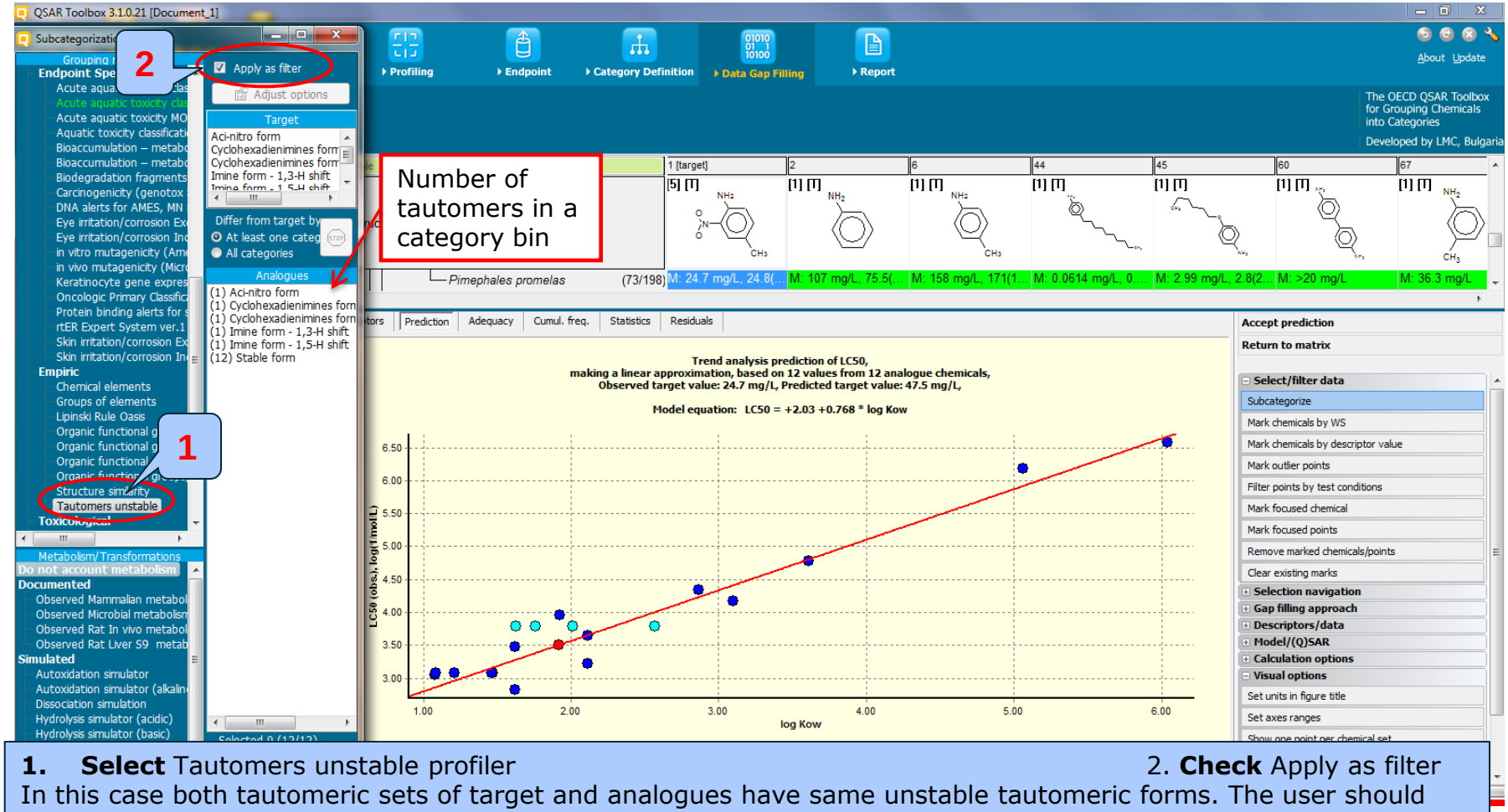
- Tautomers unstable profiler

  The categorisation based on Tautomers unstable allows keeping among the set of analogues only those thautomeric forms that are stable. For toxic effects conditioned by less specific interactions (such as mortality, growth inhibition, immobilization, etc.) the stable tautomeric forms appear to be the dominant toxicants. The tautomeric sets of target chemical and analogues include stable and unstable tautomeric forms. Based on the above recommendation the set of analogues and the target should contain only stable tautomeric forms. In this respect filtering the tautomeric sets should be applied. ("Apply filter option" should be selected)

See next two slides

Data Gap Filling

Subcategorization by Tautomers unstable

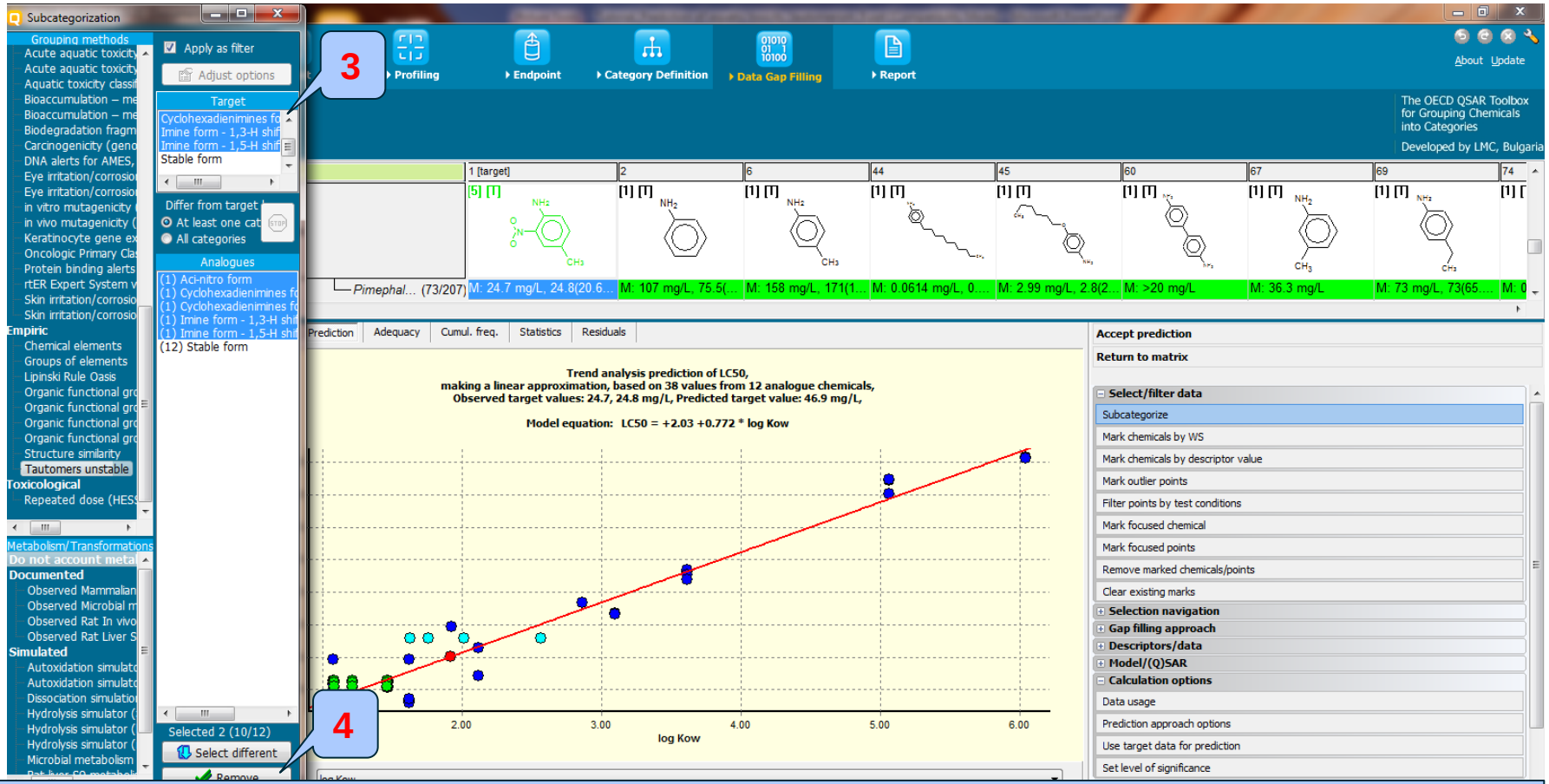


1. Select Tautomers unstable profiler 2. **Check** Apply as filter

In this case both tautomeric sets of target and analogues have same unstable tautomeric forms. The user should manually select unstable tautomeric forms in order to remove them, because from the system's point of view all labels are equal and the system cannot prefer the label "stable" to other (unstable ones). (see next slide)

Data Gap Filling

Subcategorization by Tautomers unstable

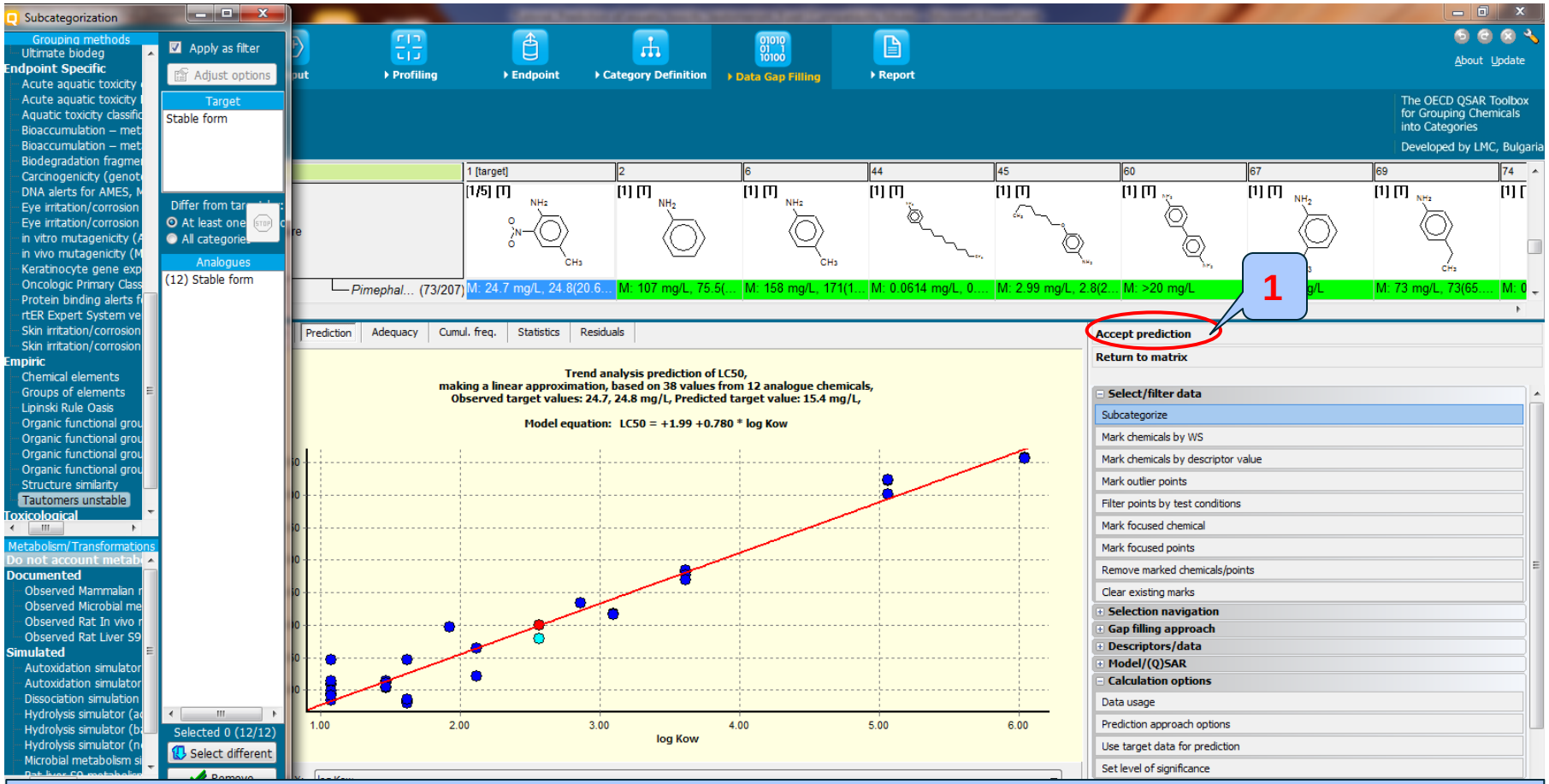


1. **Hold** Ctrl button and select unstable forms from the target tautomeric set, then the system automatically will select unstable tautomeric forms from analogues sets

4. **Click** Remove to eliminate dissimilar tautomers from tautomeric sets

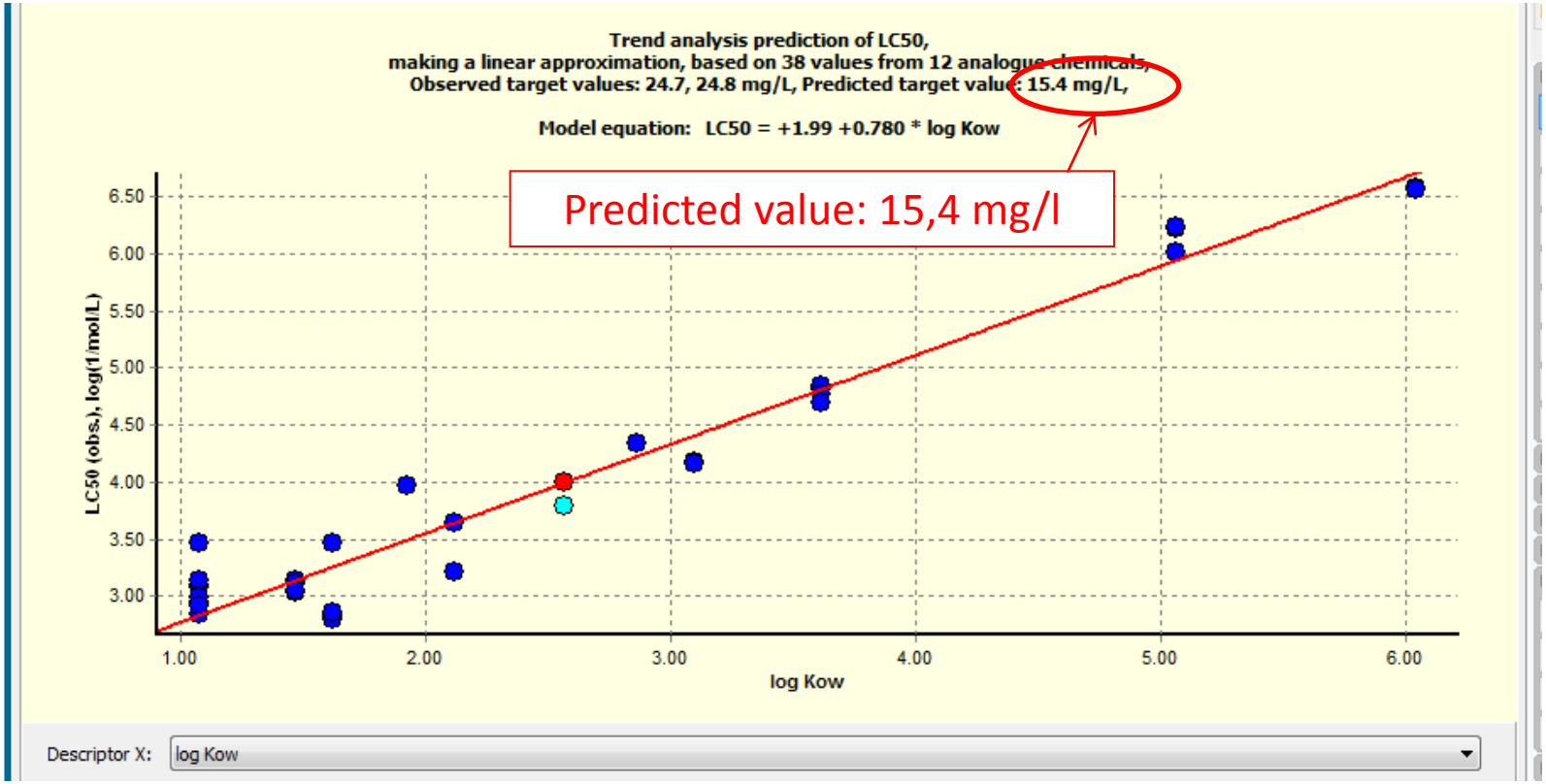
Data Gap Filling

Result of Subcategorisation by Tautomers unstable

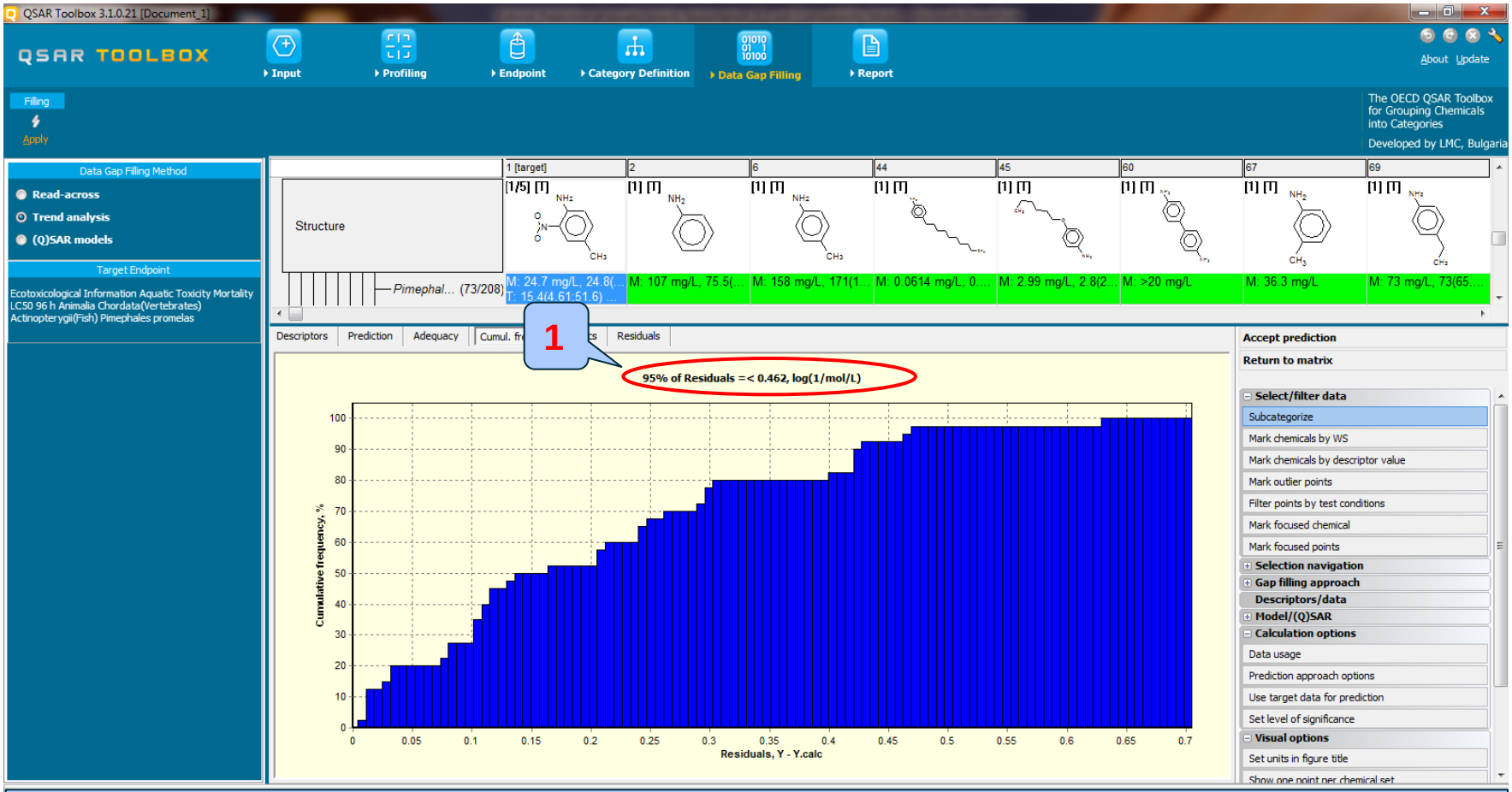


1. Accept prediction

Data Gap Filling
Result

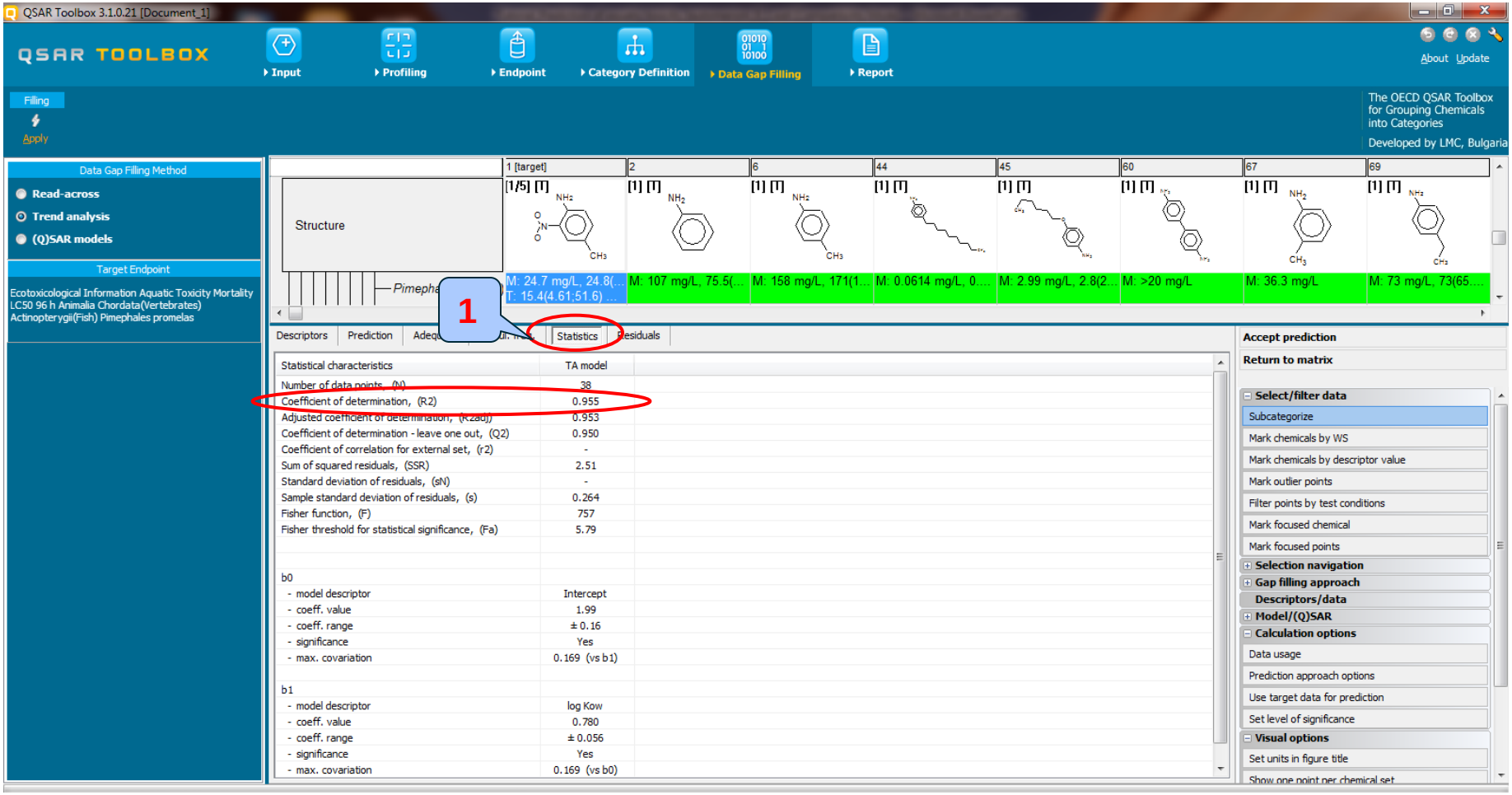


Data Gap Filling
Cumulated frequency



1. 95% of residuals are in the range of experimental error

Data Gap Filling
Statistics



1. Coefficient of determination is high

Data Gap Filling

Summary on implementation of tautomers in trend analysis

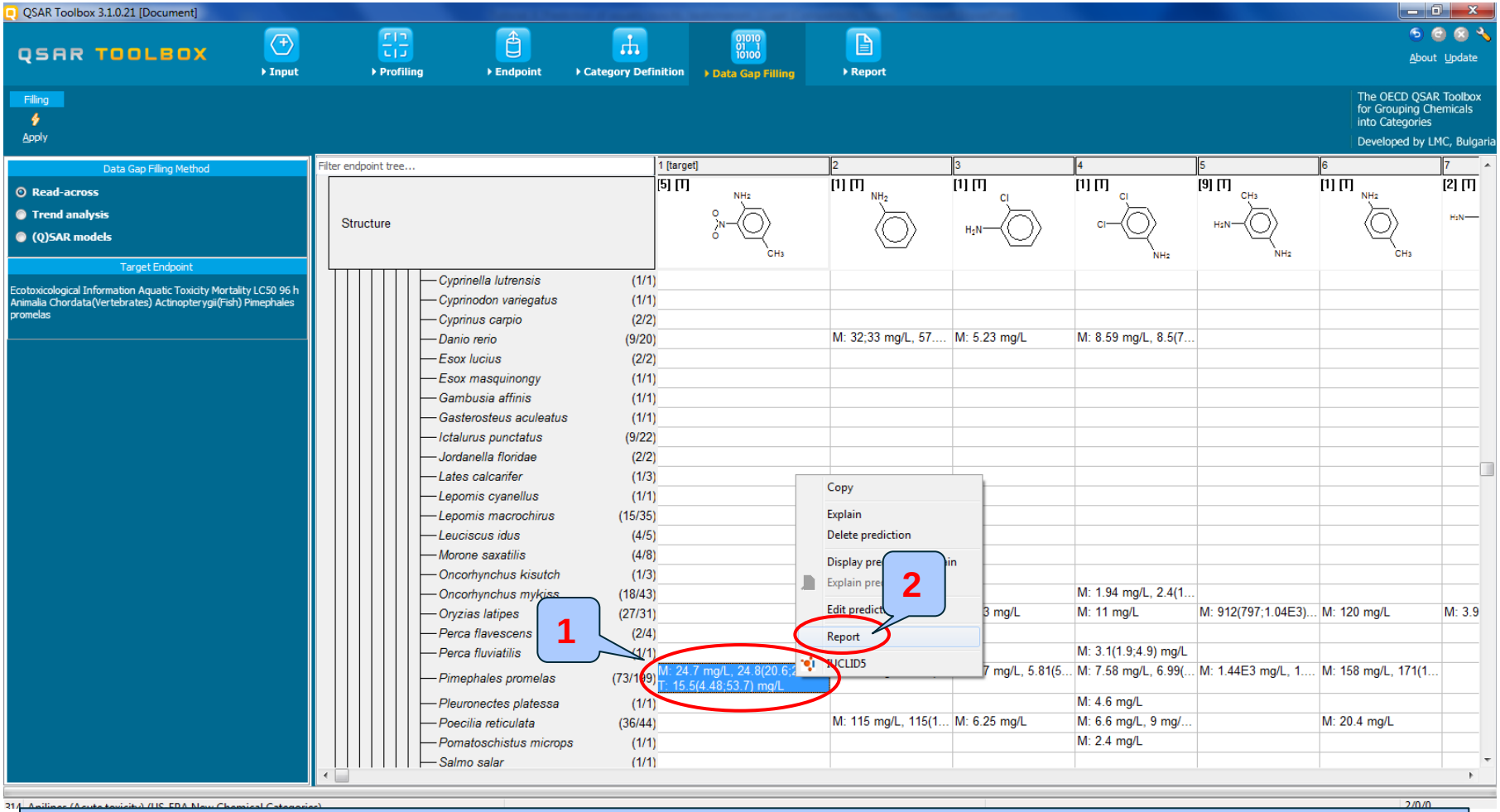
- For toxic effects conditioned by less specific interactions (such as mortality, growth inhibition, immobilization, etc.) the stable tautomeric forms appear to be the dominant toxicants.

- **Recommendation:** to use the most stable tautomers for representation of the chemicals

Outlook

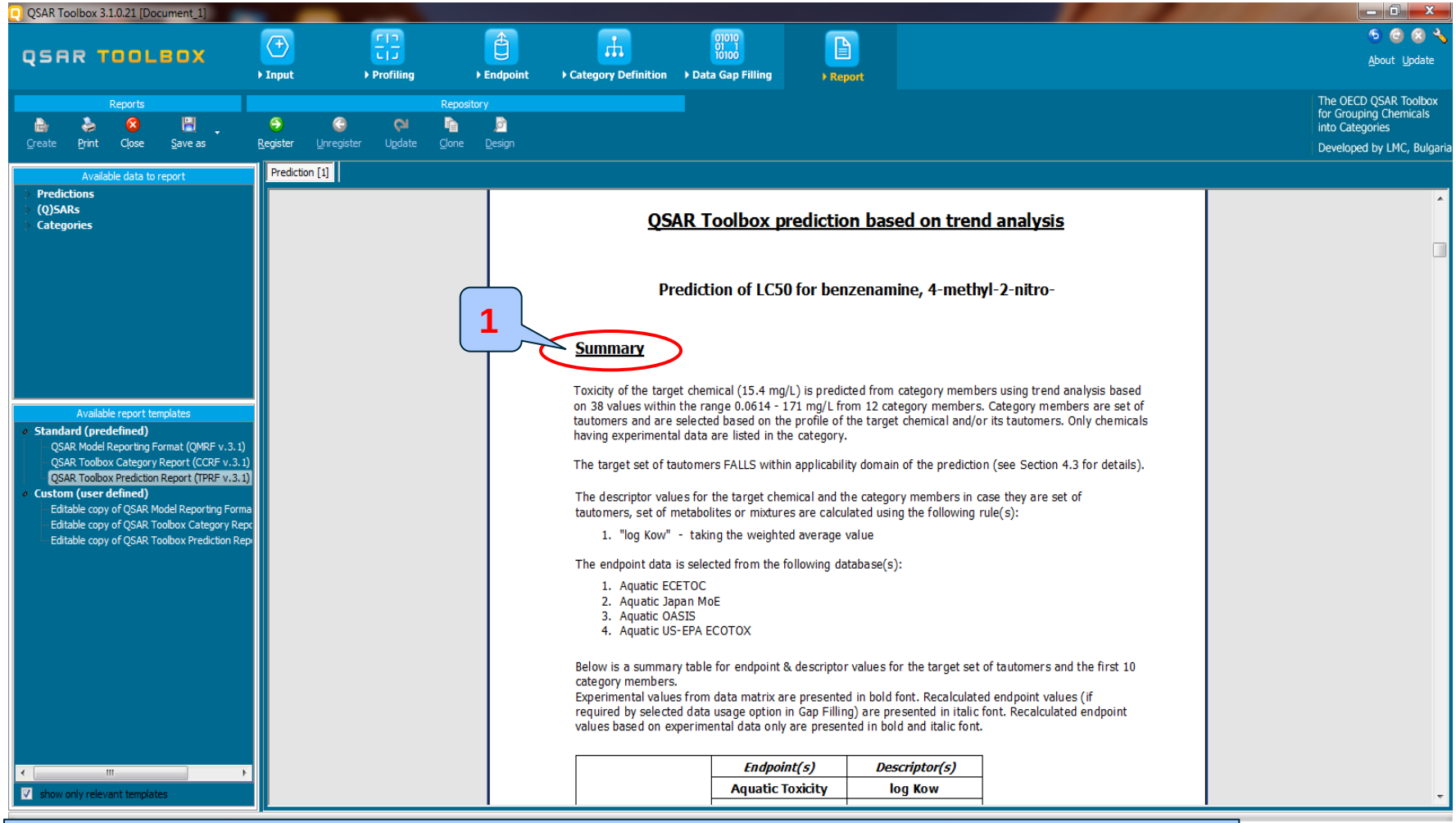
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Category definition
  - Data Gap Filling
  - **Report**

Report



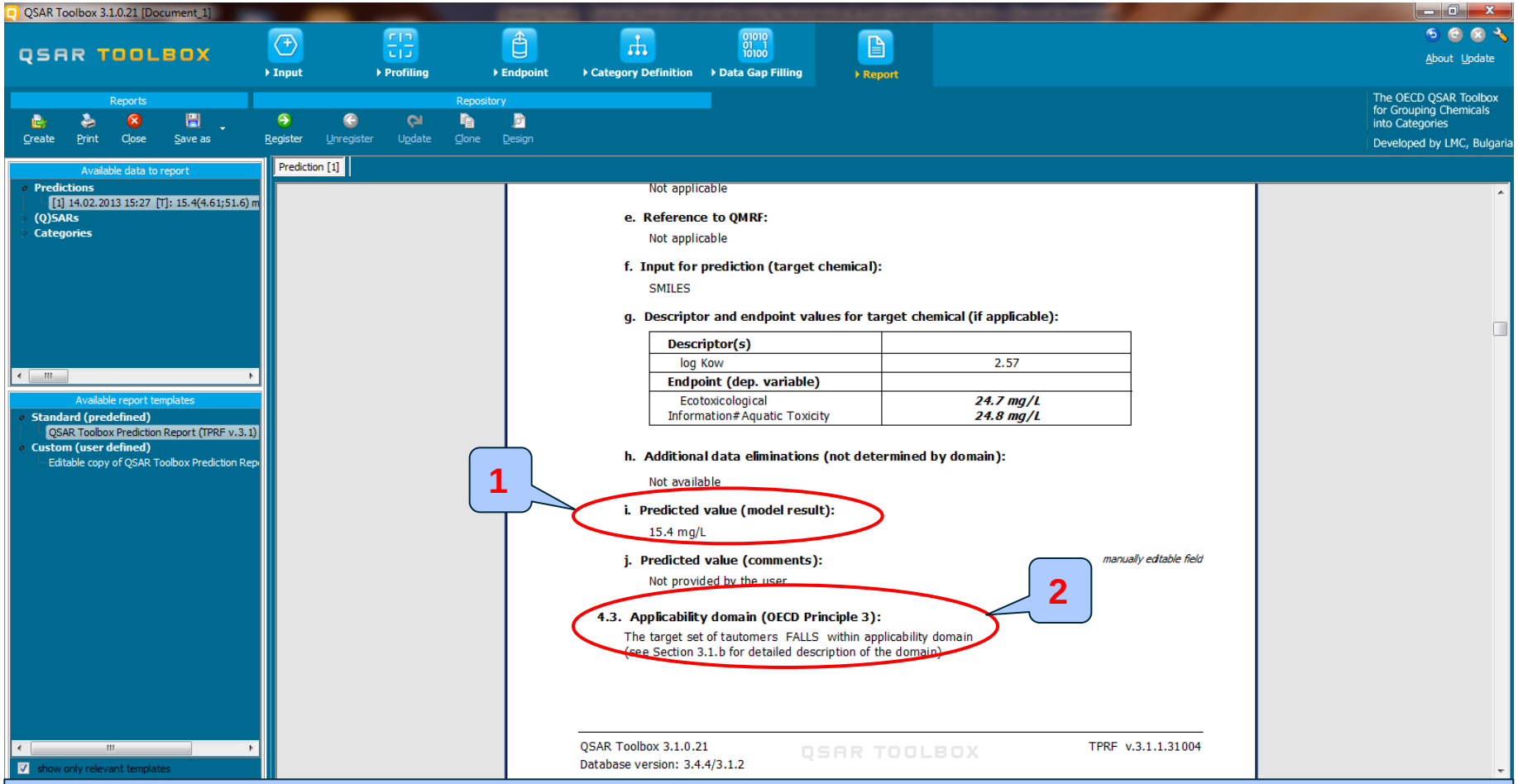
1. **Click** on the cell with prediction
2. Perform **Right click** and **Select** Report

Report



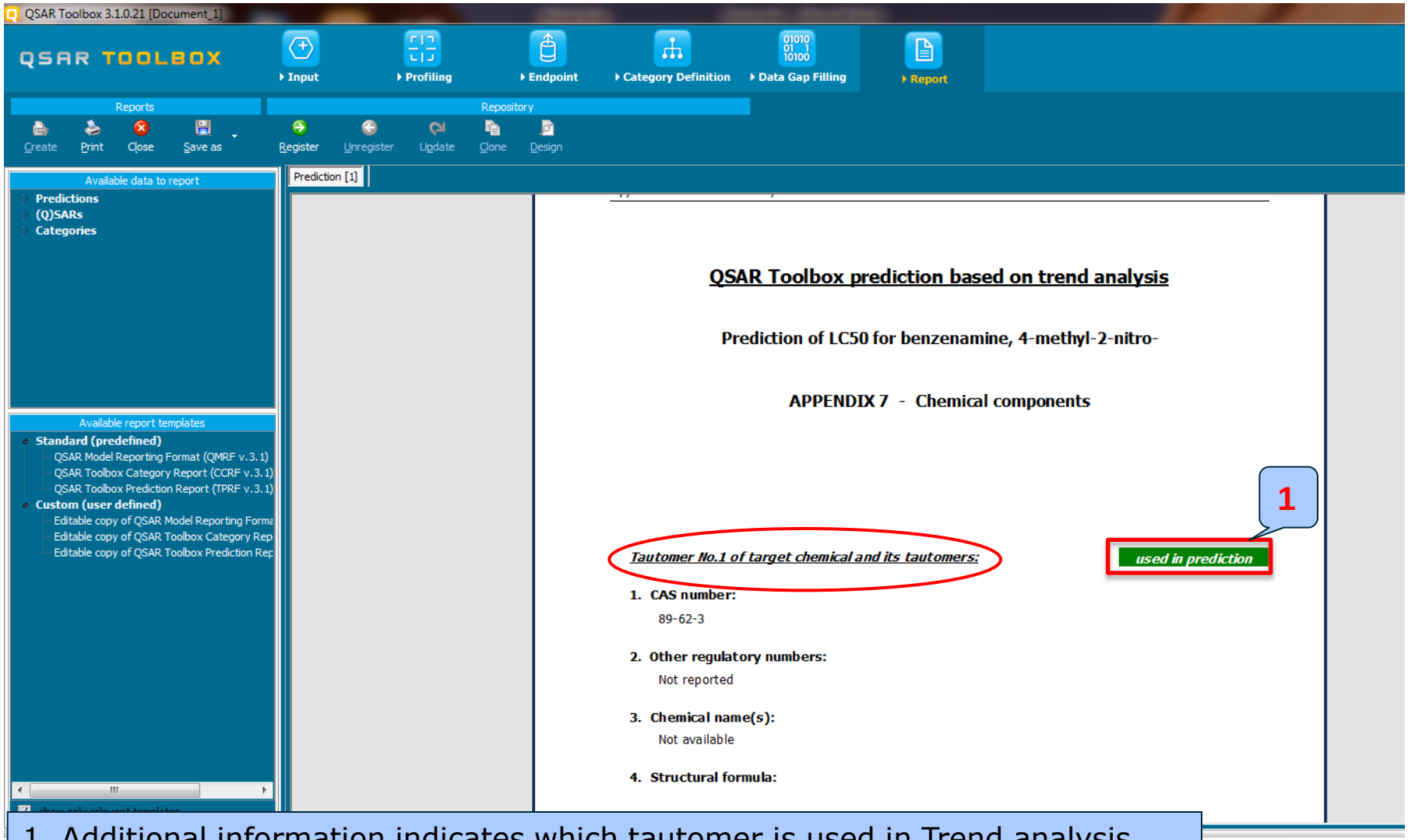
1. Summary information for tautomer prediction

Report



1. Predicted value 2. Applicability domain

Report



1. Additional information indicates which tautomer is used in Trend analysis