QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Step-by-step example of how to evaluate an ad-hoc category of aliphatic amines and to predict an ecotoxicological endpoint

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for evaluating of an ad-hoc category.
- You will learn several new functionalities which we be repeated to assure a consistent category is defined.
- It is assumed that you now have some experience in using the Toolbox so there will be multiple key strokes between screen shots.

Outlook

- Background
- **Category evaluation - overview**

Category Evaluation
Two phases of evaluation process

Phase I. **Evaluation of Category & Data Gap Filling:**

- Investigating the structural consistency of an ad hoc category (e.g., a category submitted by an industry consortium to a regulatory assessment program).
- Implementation of an ad hoc category building and data gap filling.

Phase II. **Extension of Category & Data Gap Filling:**

- Search for other analogues which are consistent with the submitted category.
- Data gap filling using new data matrix.

Outlook

- Background
- Category evaluation – overview
  - **<u>Phase I.</u> Evaluation of Category & Data Gap Filling**
    - *<u>Step 1:</u> Investigating the structural consistency of an ad hoc category.*
      - ***Case study***

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Case Study

- The submission consists of a category with 19 aliphatic amines.
- The predicted ecotoxicological endpoint EC50, 48h, *D.magna* of 2-Butanamine(CAS 13952-84-6) will be reviewed.

Outlook

- Background
- Category evaluation – overview
  - **<u>Phase I.</u> Evaluation of Category & Data Gap Filling**
    - *<u>Step 1:</u> Investigating the structural consistency of an ad hoc category.*
      - *Case study*
      - ***Workflow***

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow

The following input workflow is used:

- Input the file for submission as an user list.(In this case input file Aliphatic amines.smi*)
- Evaluate the category applying the following profiling schemes:
  - US-EPA New Chemical Categories
  - Aquatic toxicity classification by ECOSAR
  - Aquatic toxicity MOA of action
  - Organic functional groups (nested)

Aliphatic amines.smi* - file is packed in the zip file with all power point tutorials posted on the LMC website.

Also it is placed in the example folder of the Toolbox installation.

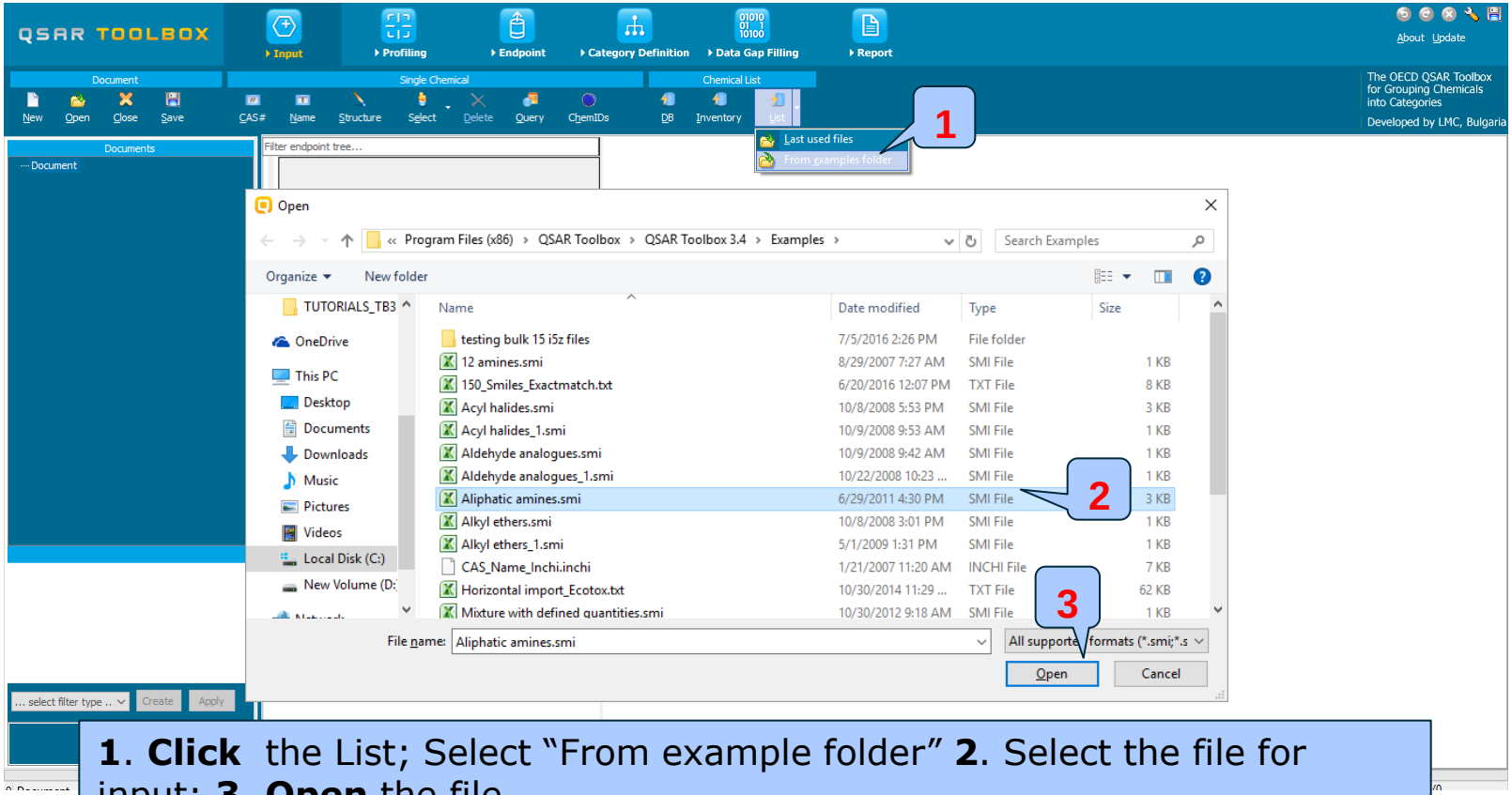
Outlook

- Background
- Category evaluation – overview
  - **Phase I. Evaluation of Category & Data Gap Filling**
    - *Step 1: Investigating the structural consistency of an ad hoc category.*
      - *Case study*
      - ***Workflow***
        - ***Input***

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Input



1. Click the List; Select "From example folder" **2.** Select the file for input; **3. Open** the file.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

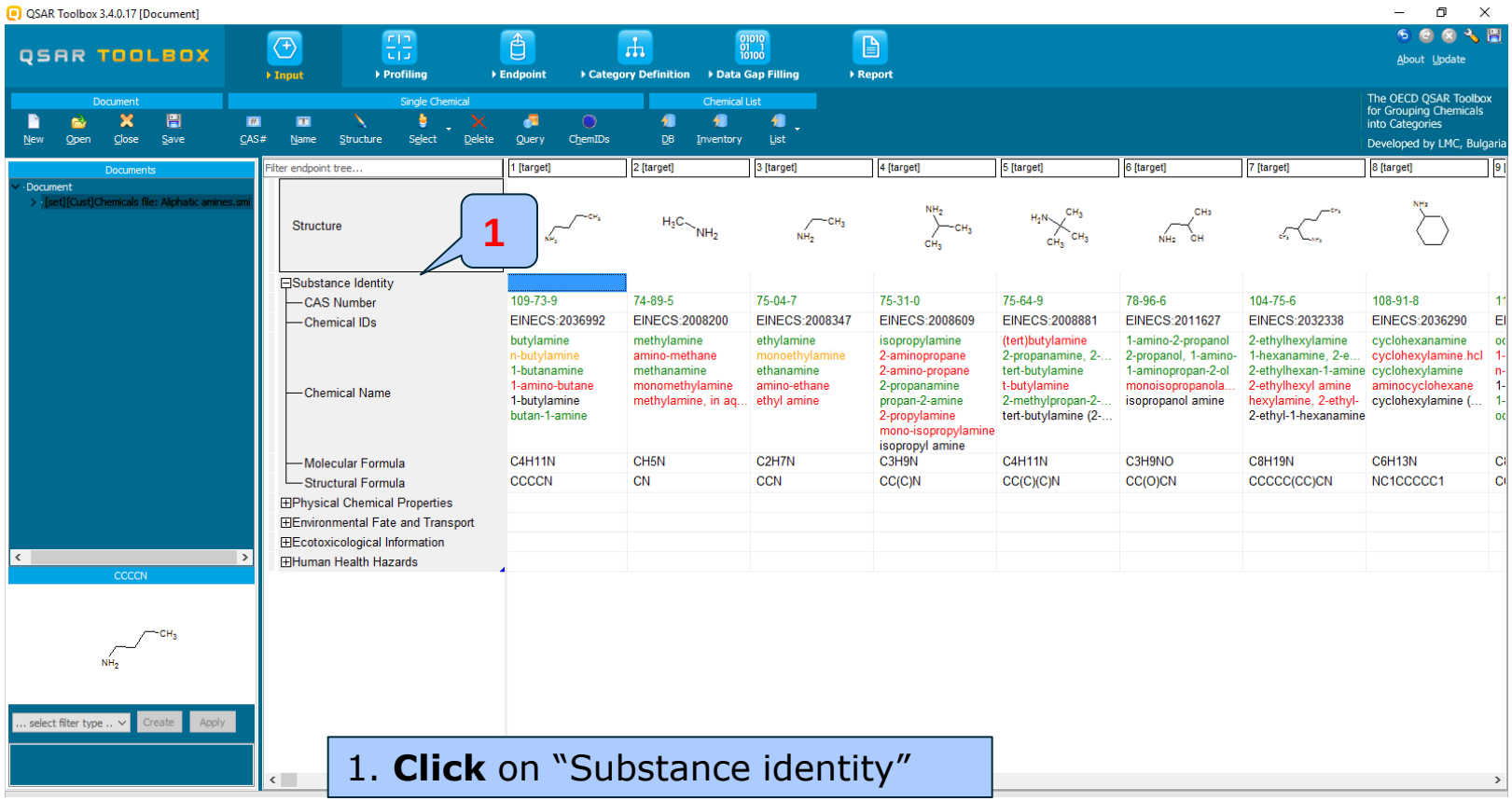
Workflow/Input

- You have now inserted your chemical list into the system.
- Click on the box next to "Substance Identity"; this displays the chemical identification information (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Input



Outlook

- Background
- Category evaluation – overview
  - **Phase I. Evaluation of Category & Data Gap Filling**
    - *Step 1: Investigating the structural consistency of an ad hoc category.*
      - *Case study*
      - ***Workflow***
        - *Input*
        - ***Profiling***

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling

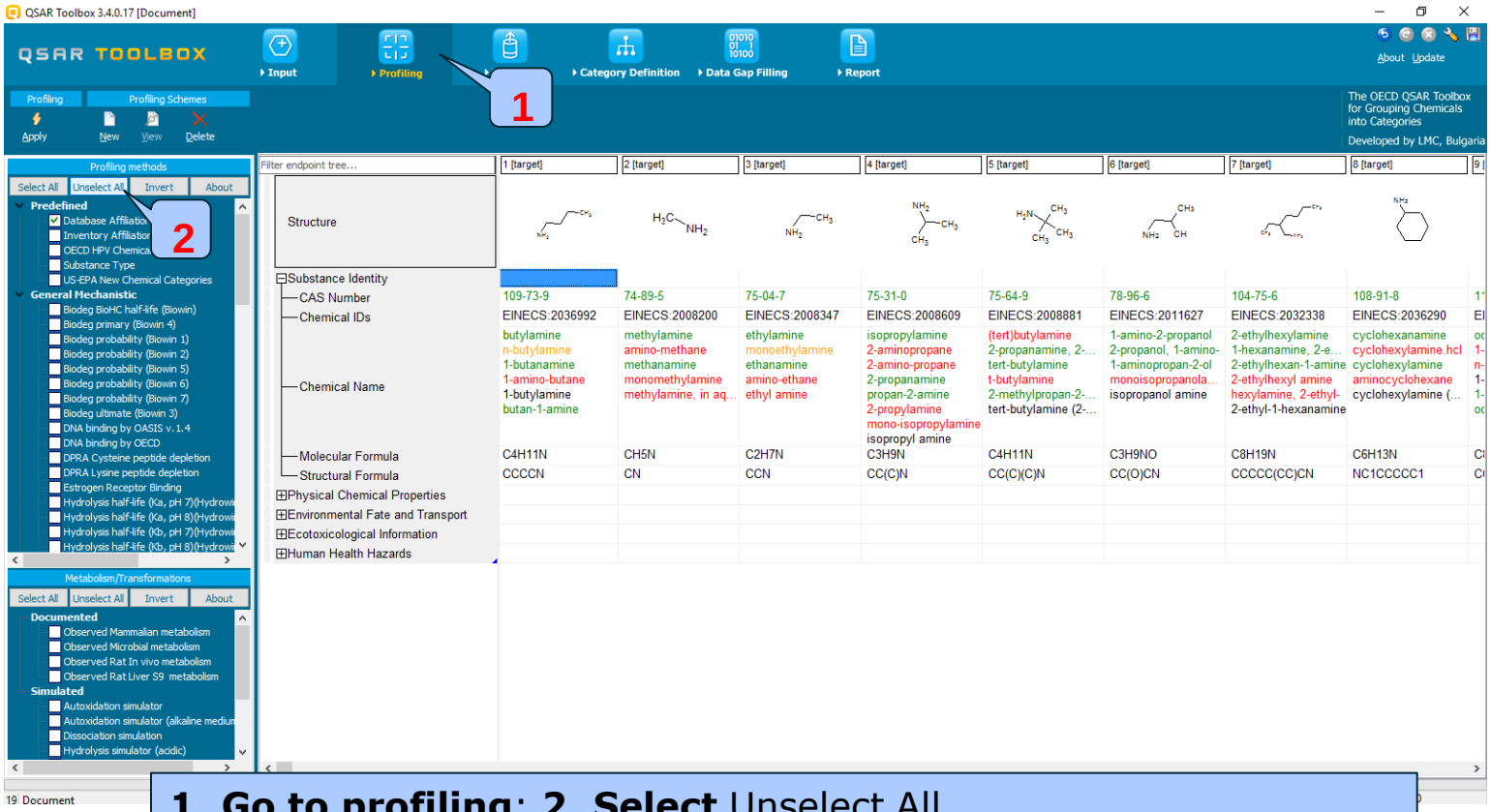
- The first step of the category evaluation according to Phase I is the consistency check of the category which starts with evaluation of the robustness with respect to structural functionalities. The following schemes could be used for this purpose:
  - US-EPA New Chemical Categories
  - Aquatic toxicity classification by ECOSAR
  - Aquatic toxicity MOA of action
  - Organic functional groups (nested)

- Select "profiling methods" by clicking on the boxes before the names of the profilers and Click "Apply". Before selecting the profiling methods unselect all (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling

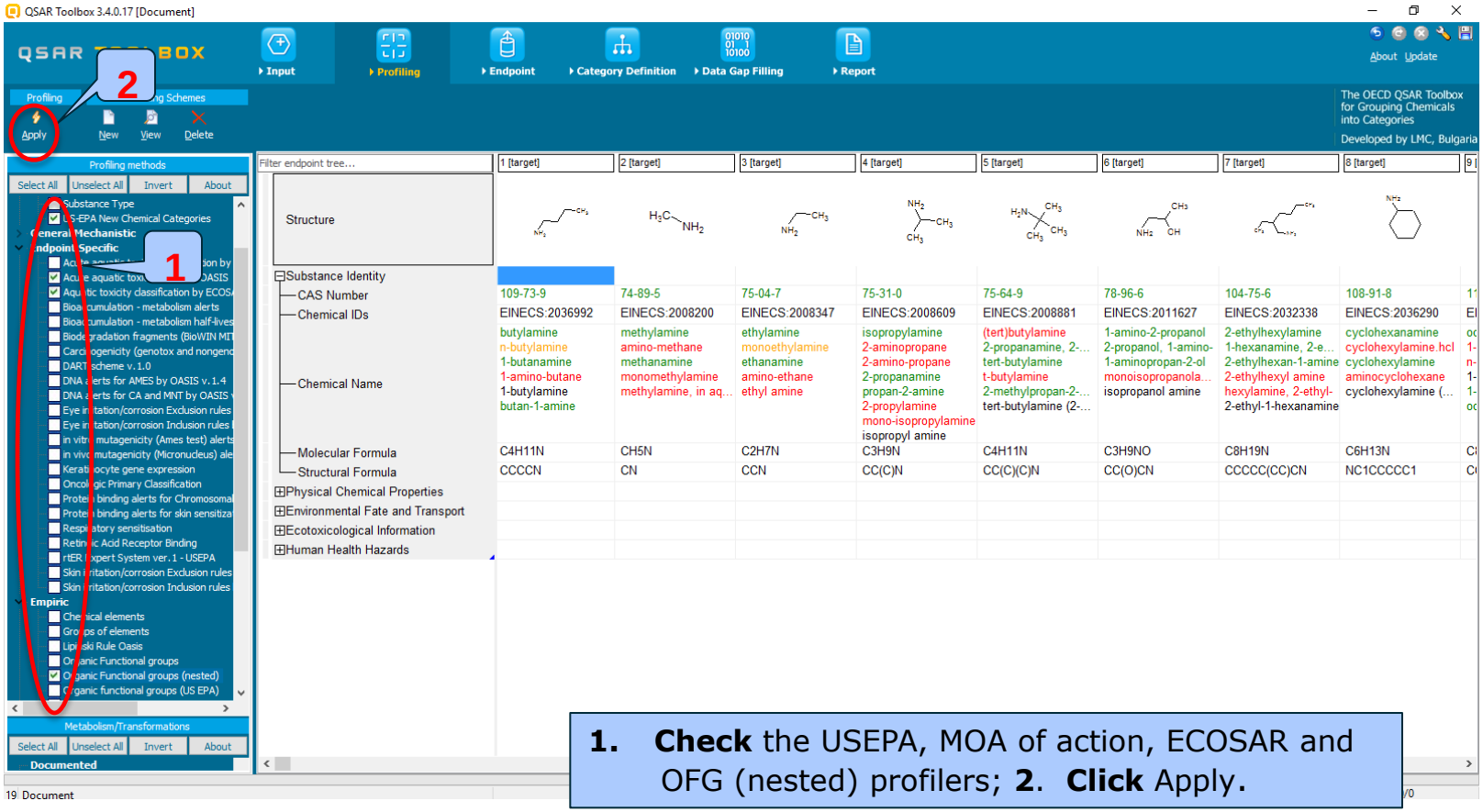


1. Go to profiling; 2. Select Unselect All.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling



1. **Check** the USEPA, MOA of action, ECOSAR and OFG (nested) profilers; **2.** **Click** Apply.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

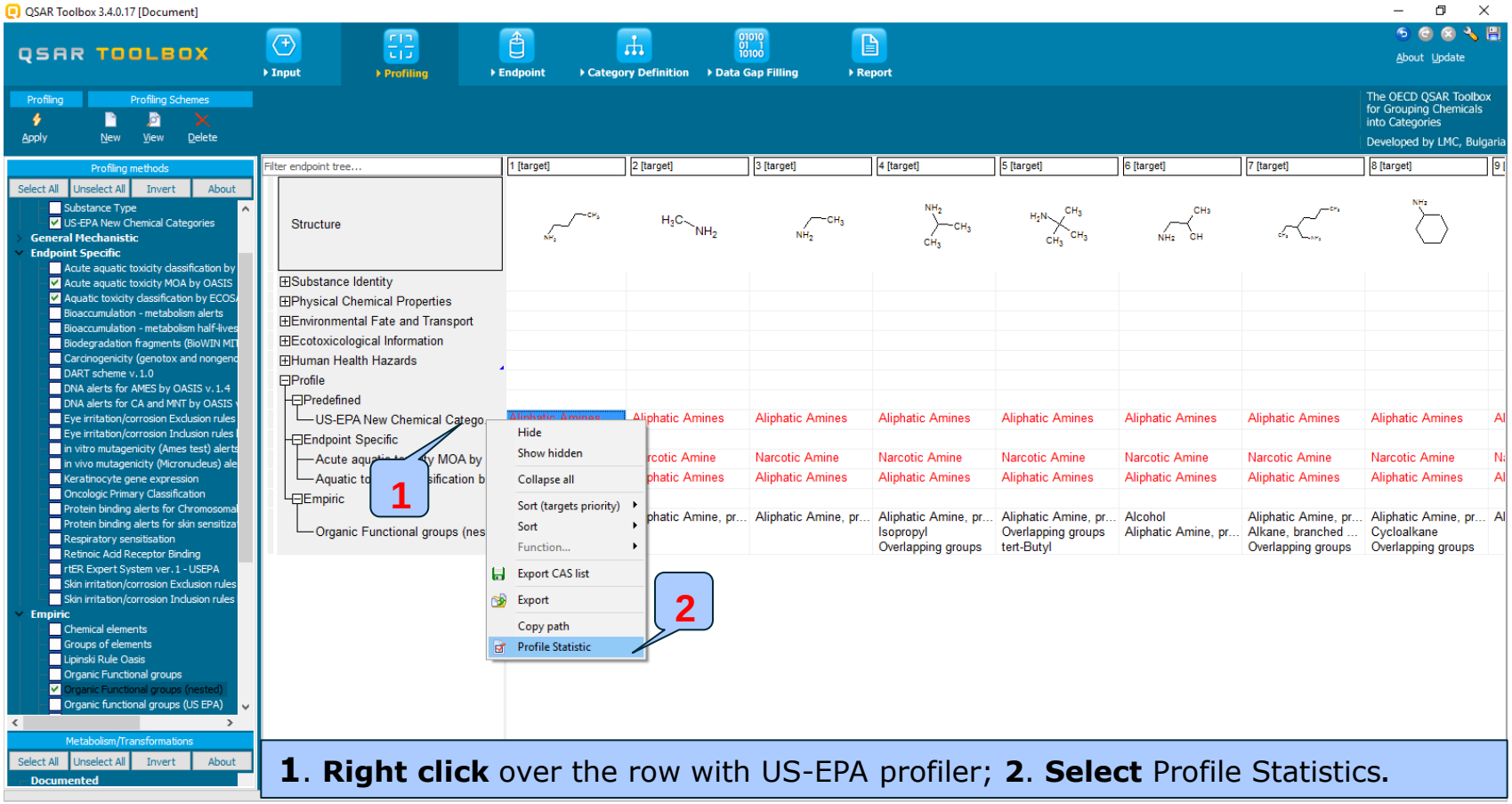
Workflow/Profiling

- The actual profiling will take several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appeared as a dropdown box under the target chemical.
- The result from profile statistics, can be seen by right clicking in the space above the profiler in the endpoint tree and select Profile statistics from the dropdown menu (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling

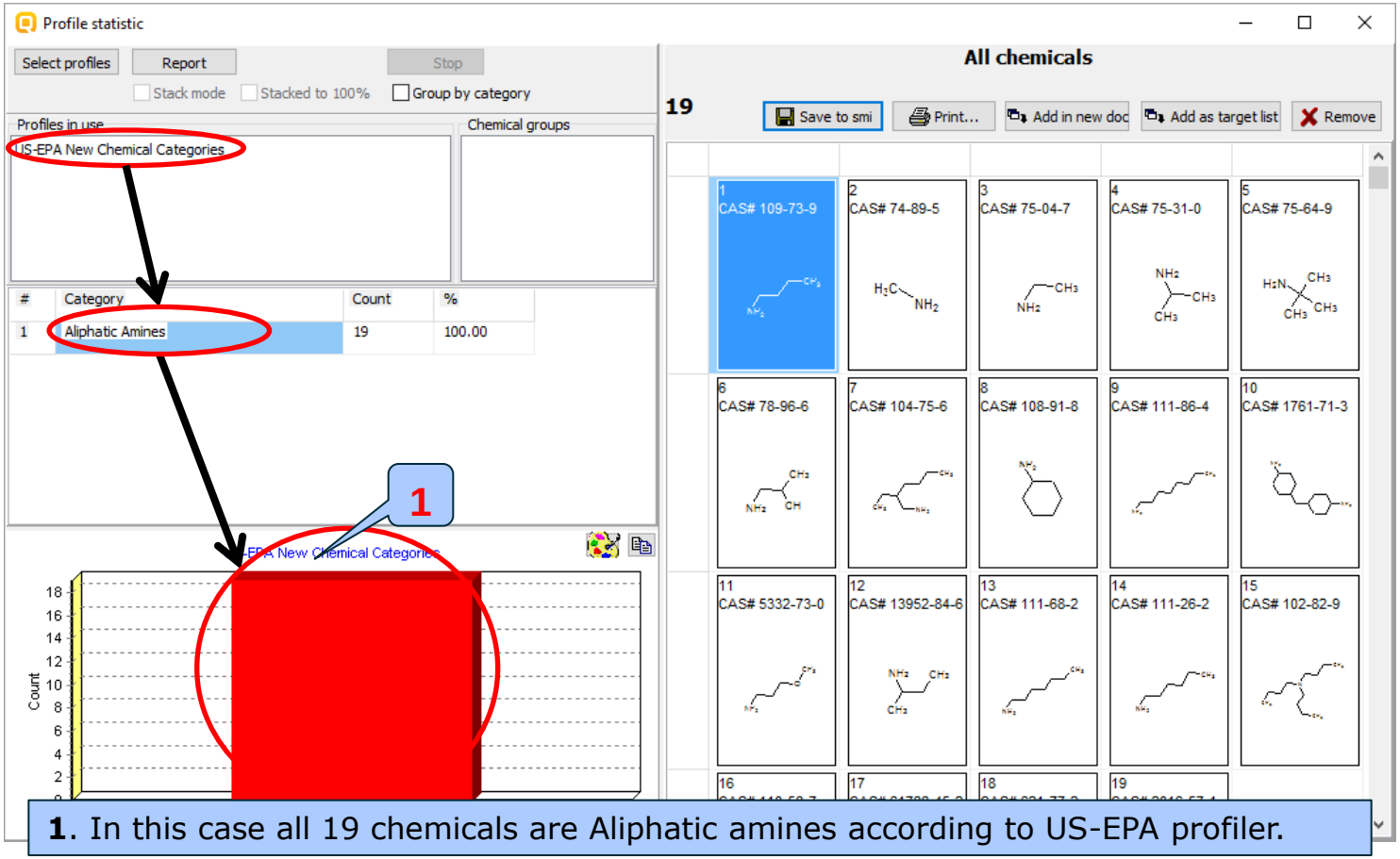


1. **Right click** over the row with US-EPA profiler; **2**. **Select** Profile Statistics.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling/Statistics according to US-EPA

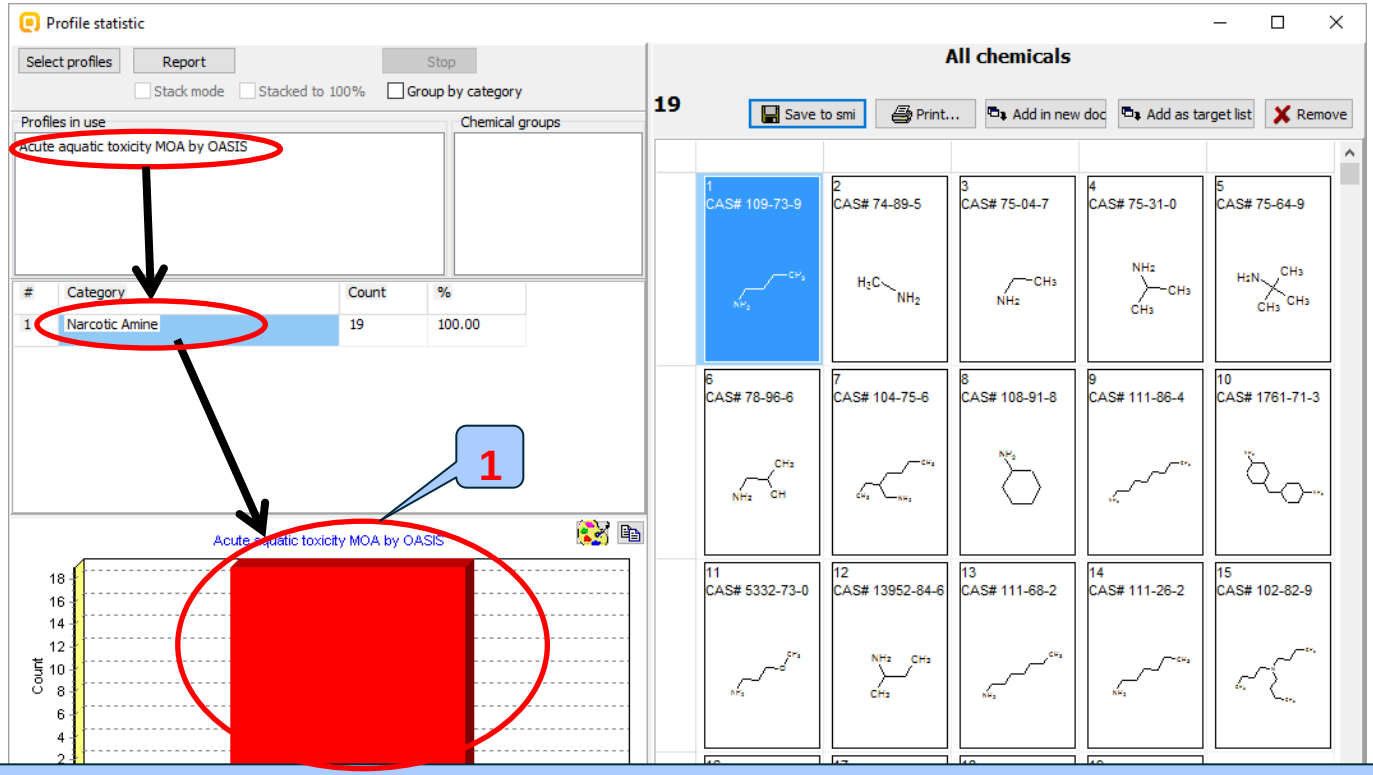


1. In this case all 19 chemicals are Aliphatic amines according to US-EPA profiler.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling/Statistics according to Aquatic toxicity MOA by OASIS

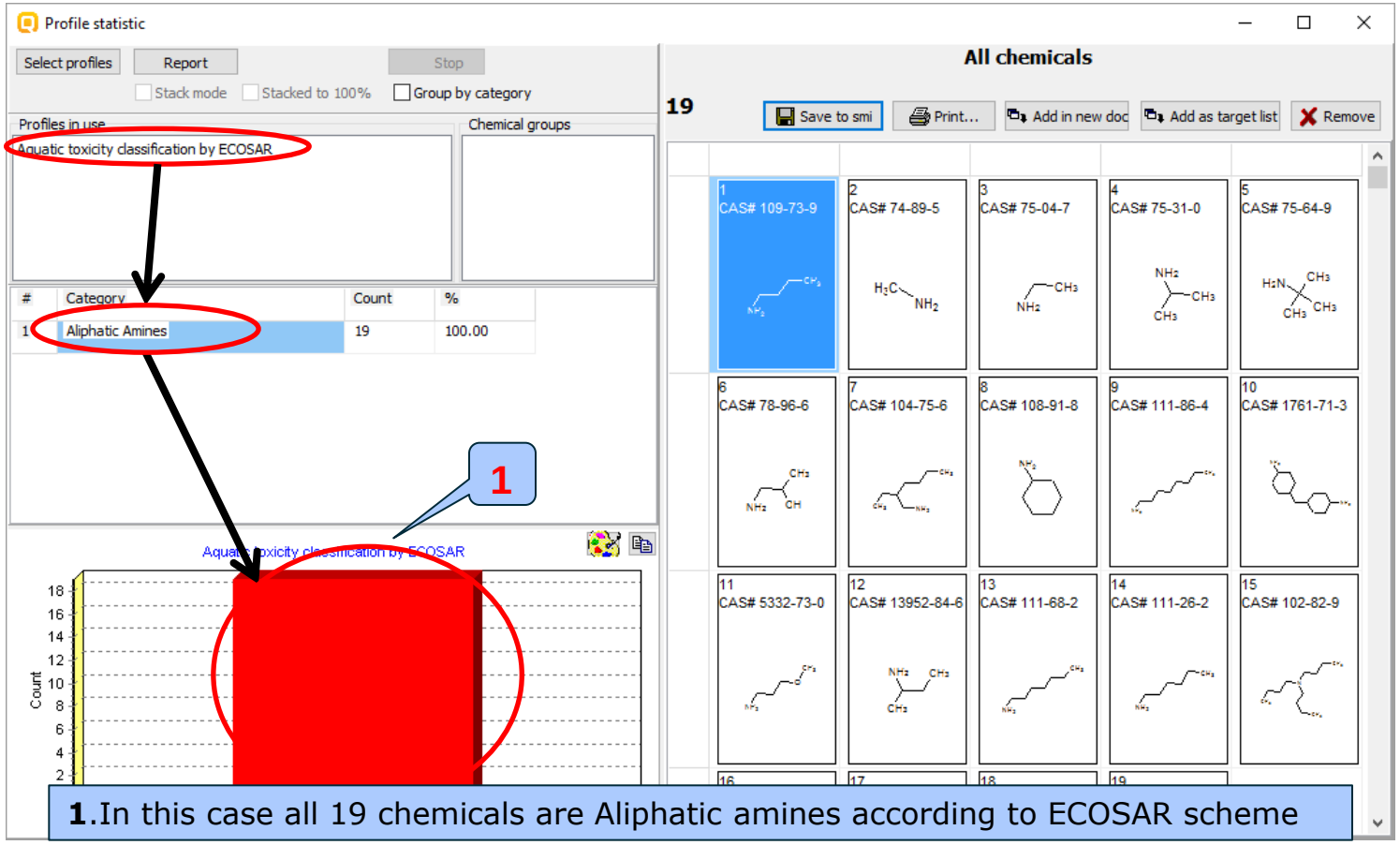


1. In this case all 19 chemicals are Narcotic amines according to MOA by OASIS profiling scheme

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling/Statistics according to ECOSAR

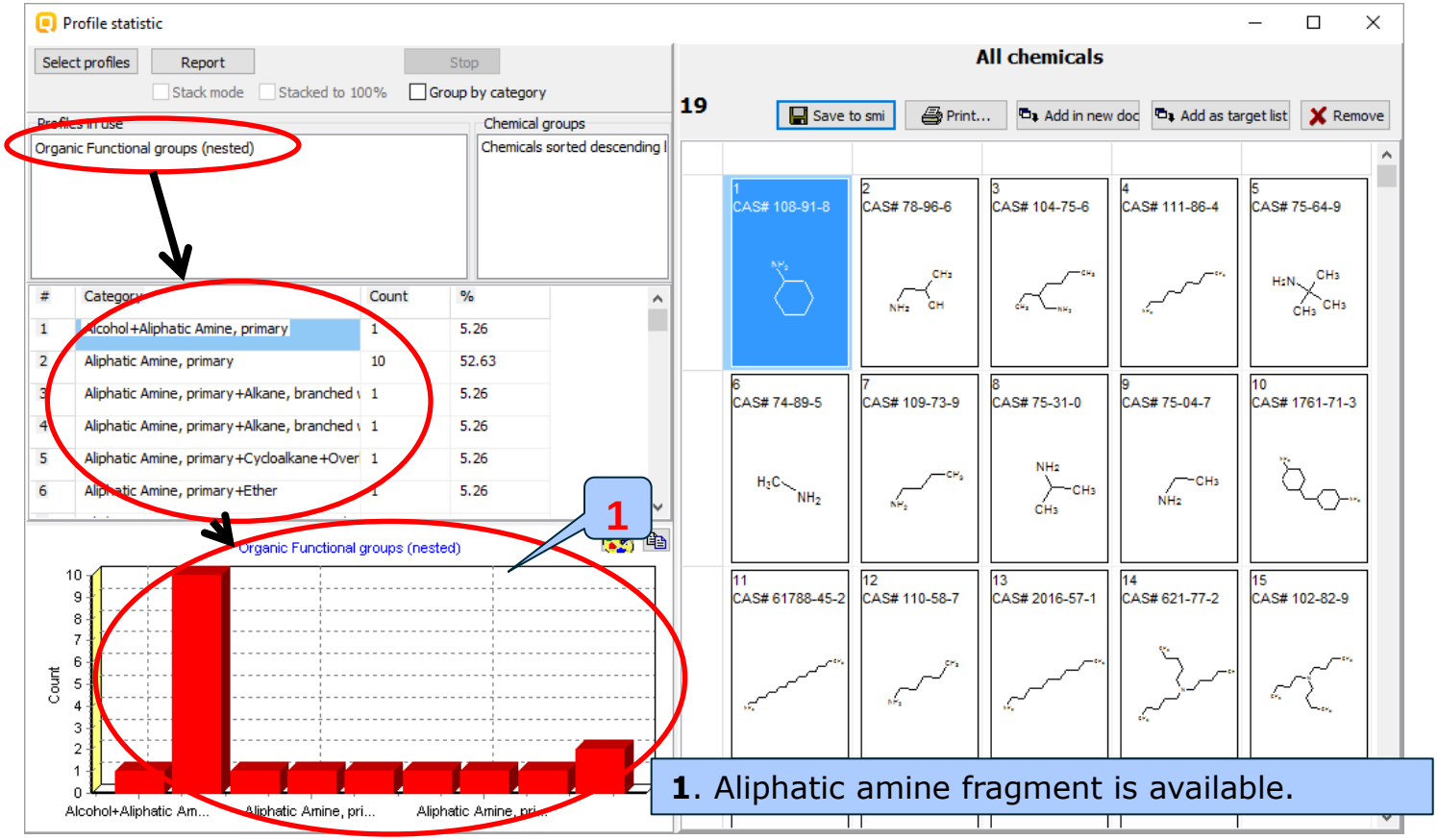


1.In this case all 19 chemicals are Aliphatic amines according to ECOSAR scheme

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling/Statistics according to OFG(nested)



1. Aliphatic amine fragment is available.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Recap

- **Chemicals are defined as:**
  - Aliphatic amines (broader category than primary amines) – US-EPA categories, ECOSAR classification
  - The statistics of organic functional groups provide detailed alert description of all 19 structures. However, all 19 chemicals have aliphatic amines fragment.
- **It could be concluded that the category is consistent with respect to structural functionalities (chemicals are empirically similar).**

Outlook

- Background
- Category evaluation – overview
  - **<u>Phase I.</u> Evaluation of Category & Data Gap Filling**
    - *<u>Step 1:</u> Investigating the structural consistency of an ad hoc category.*
    - *<u>Step 2:</u> Investigating the a*pplicability domain of an ad hoc category.

Phase I: Evaluation of Category & Data Gap Filling

Step 2. Investigating the applicability domain of an ad hoc category

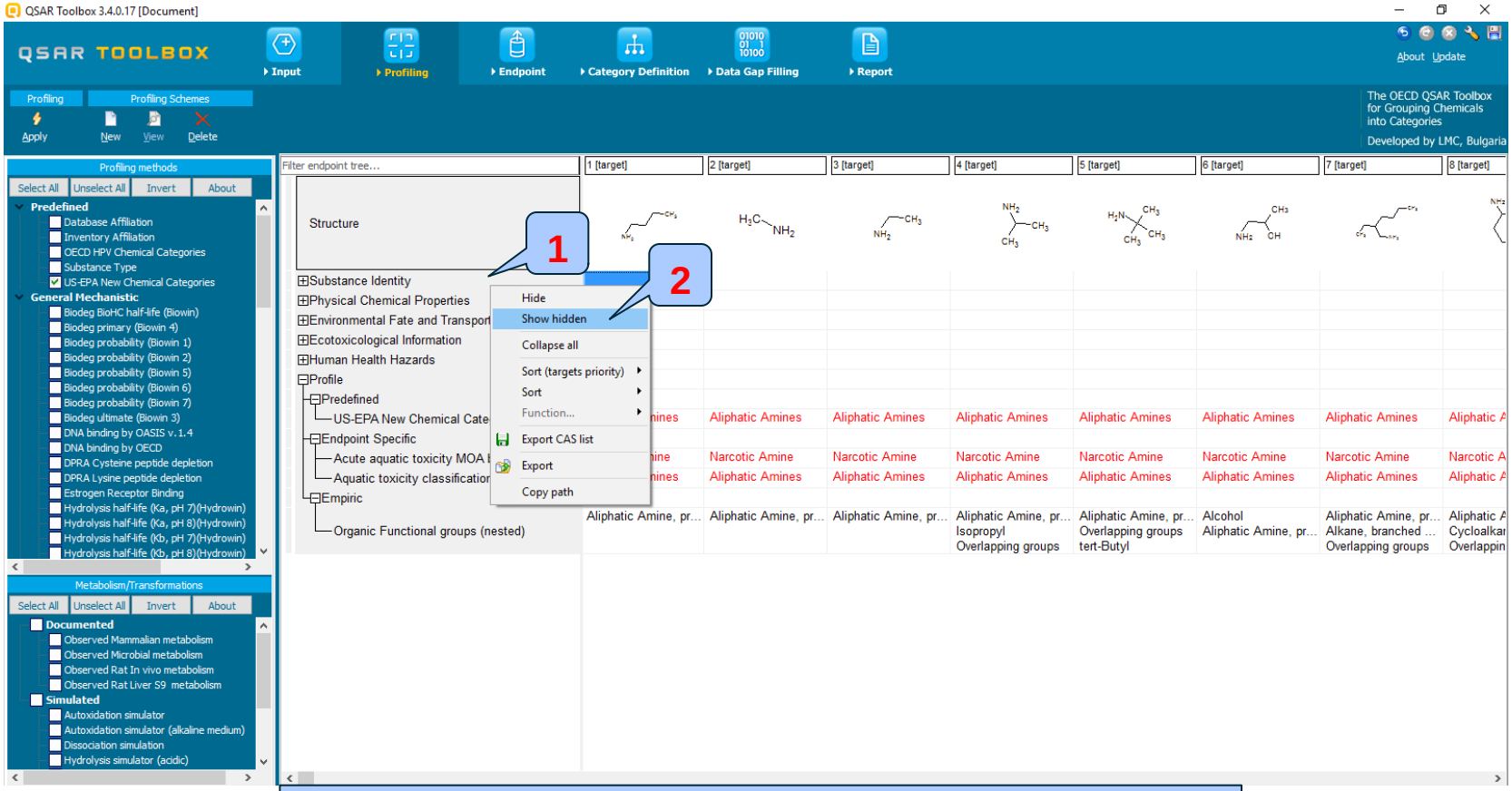
- Boundaries of structural functionalities
  - Aliphatic amines
- Parametric boundaries
  - log Kow (from 0.64 to 7.71)
  - Molecular weight (from 31 to 269 Da)
  - Water solubility (from $0.48x10^{-1}$ to $1x10^{6}$ mg/l)

You are now ready to extract the 2D and/or 3D parameters (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 2. Investigating the applicability domain of an ad hoc category

Extracting 2D and 3D parametrs

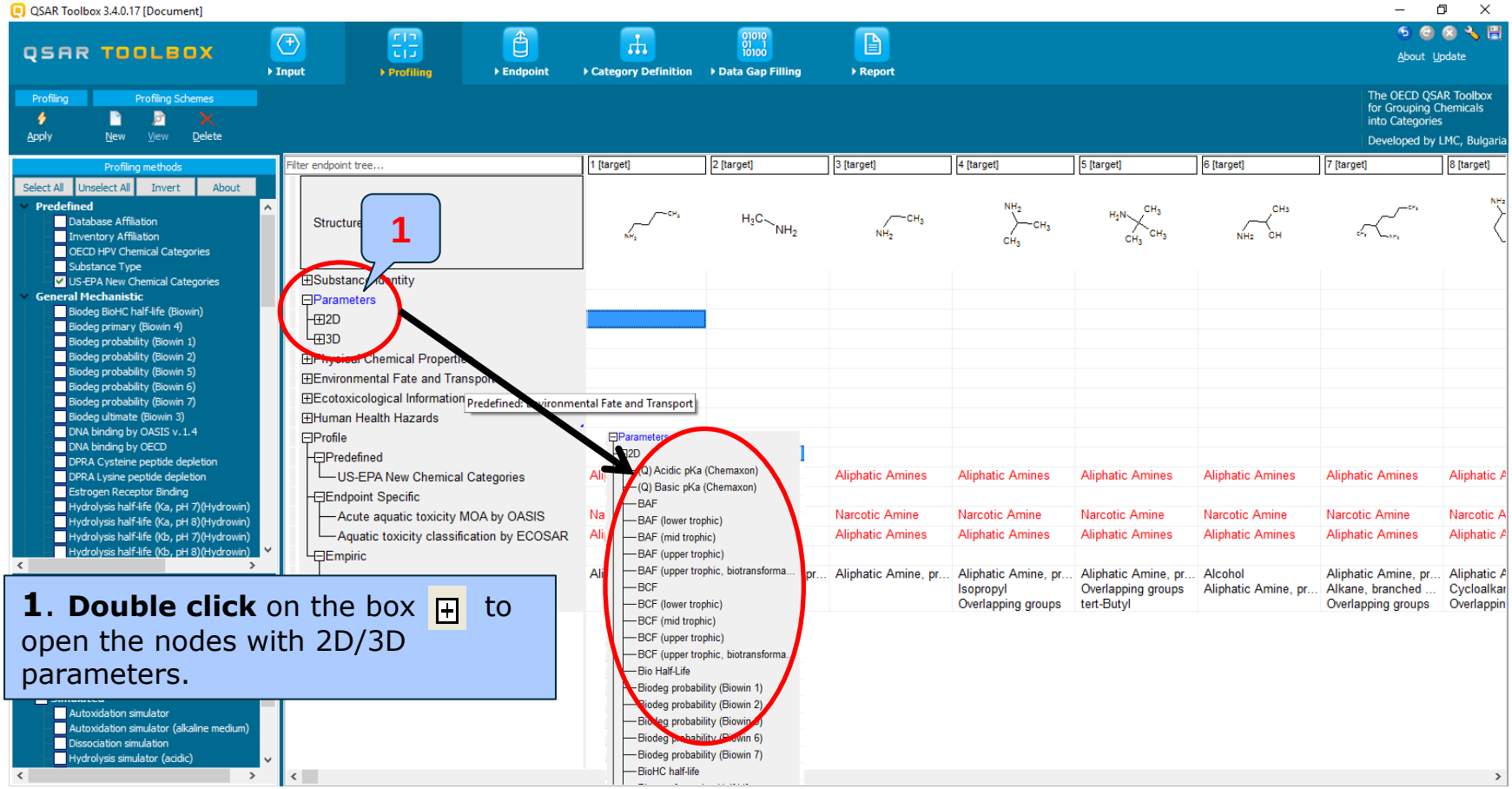


1. Right click above endpoint tree; **2. Select** Show hidden.

Phase I: Evaluation of Category & Data Gap Filling

Step 2. Investigating the applicability domain of an ad hoc category

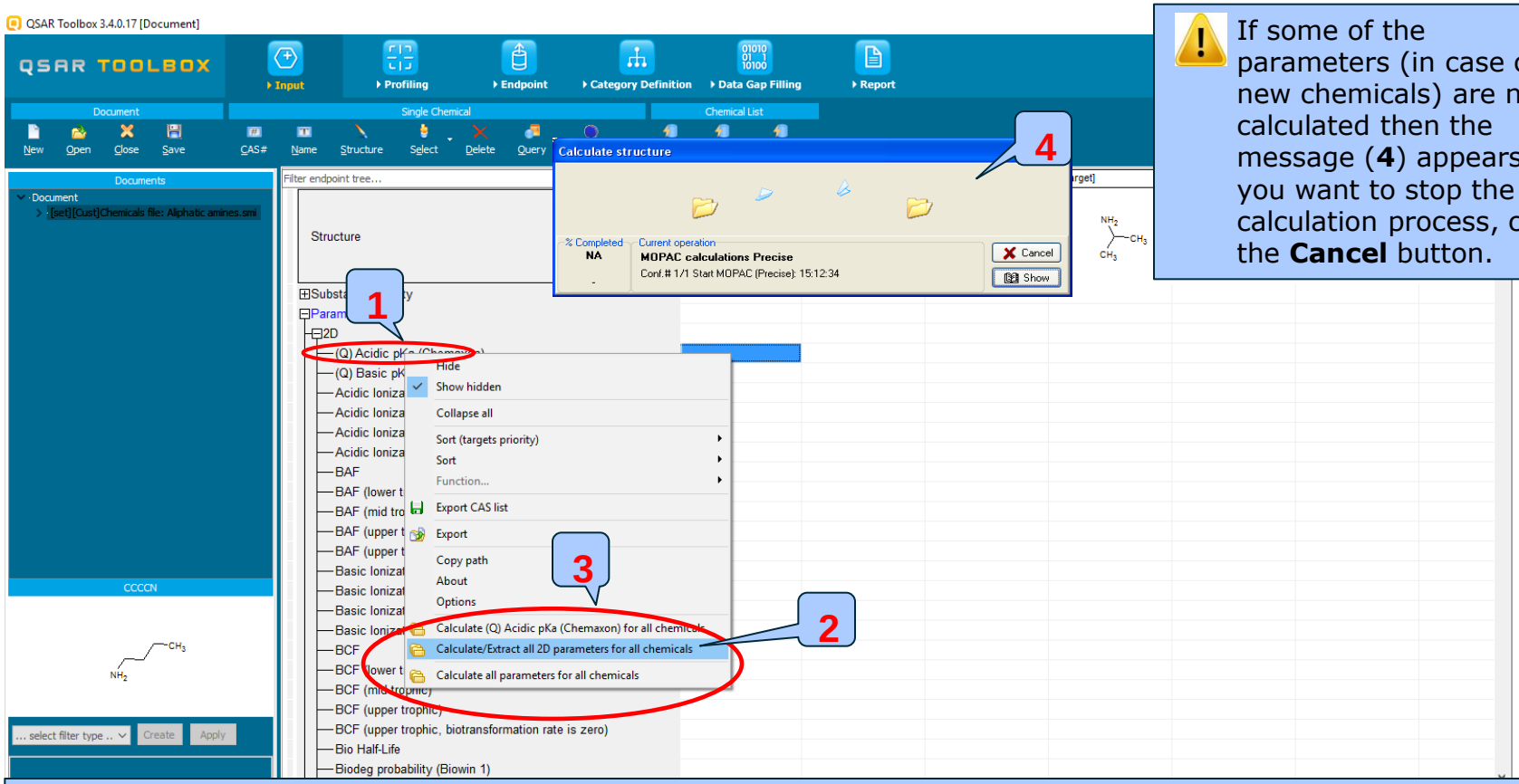
Extracting 2D and 3D parametrs



Phase I: Evaluation of Category & Data Gap Filling

Step 2. Investigating the applicability domain of an ad hoc category

Extracting 2D and 3D parametrs



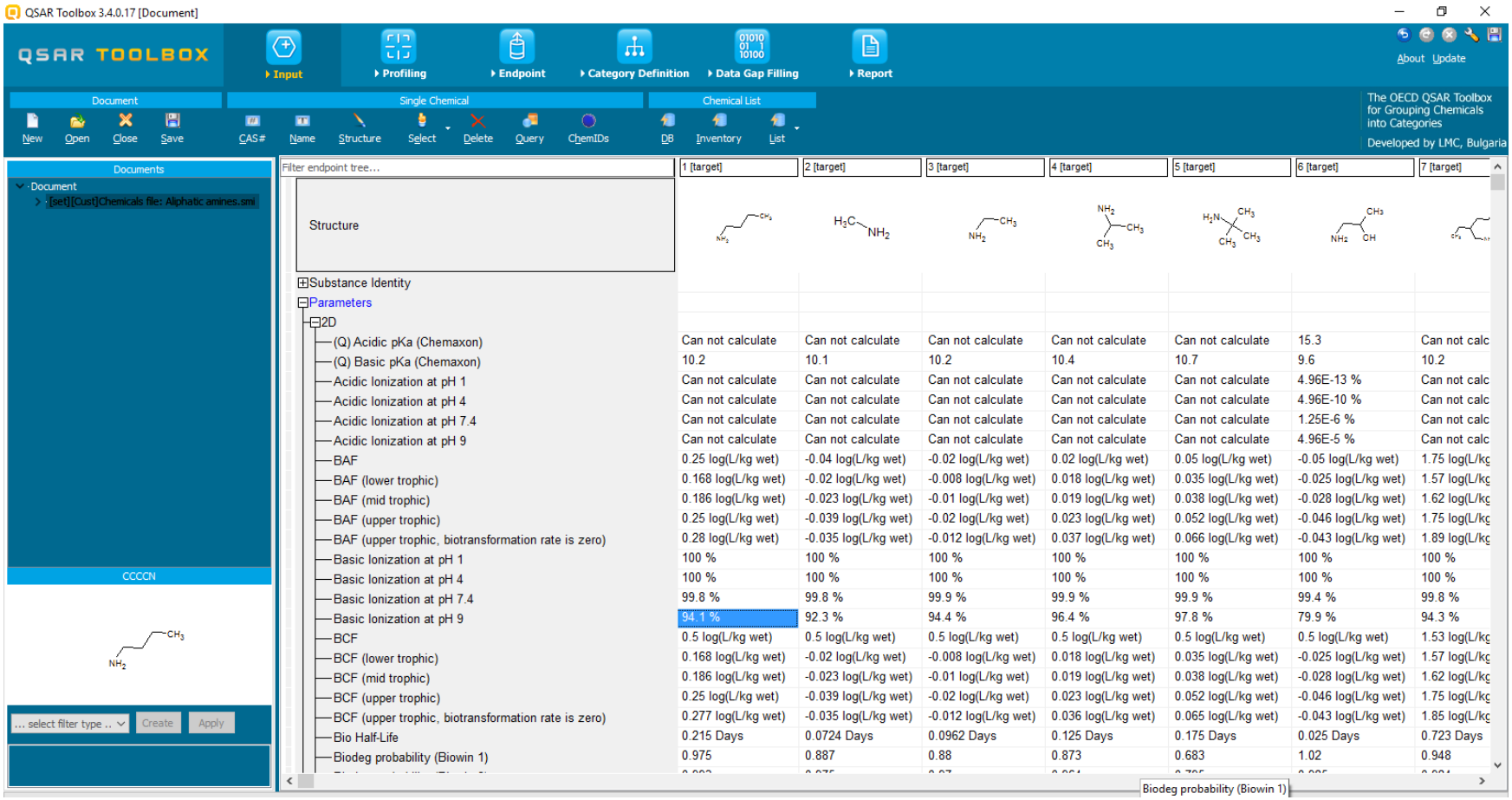
If some of the parameters (in case of new chemicals) are not calculated then the message (**4**) appears If you want to stop the calculation process, click the **Cancel** button.

1. **Right click** above the parameter in the endpoint tree; **2**. The user can calculate all parameters or extract current parameter for all chemicals in the current row. **3**. In this exercise calculate all (2D) parameters.

Phase I: Evaluation of Category & Data Gap Filling

Step 2. Investigating the applicability domain of an ad hoc category

Extracting 2D and 3D parameters - results



Outlook

- Background
- Category evaluation – overview
  - **<u>Phase I.</u> Evaluation of Category & Data Gap Filling**
    - *<u>Step 1:</u> Investigating the structural consistency of an ad hoc category.*
    - *<u>Step 2:</u> Investigating the a*pplicability domain of an ad hoc category.
    - *<u>Step 3:</u> Reading data for the analogues*

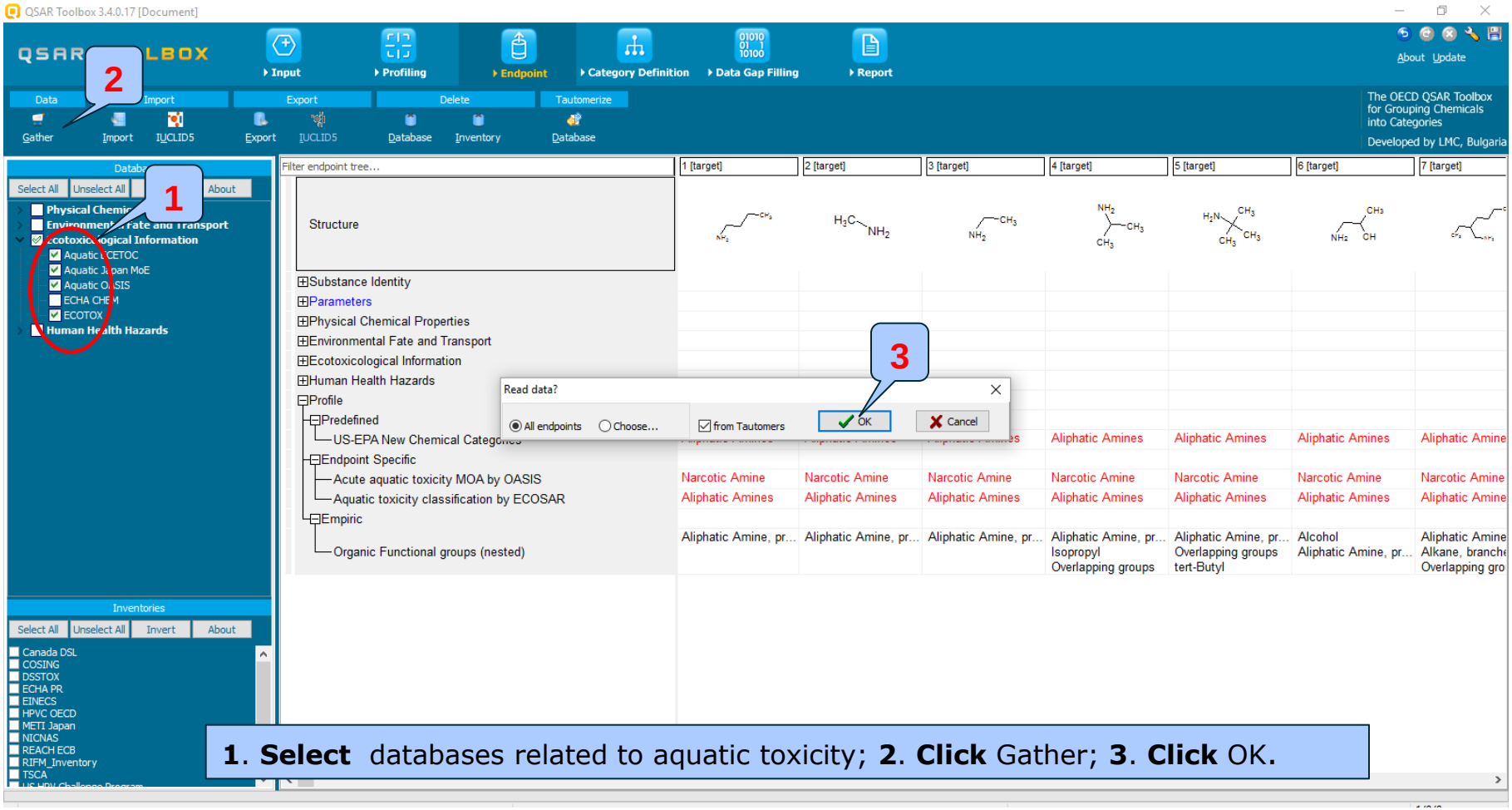
Phase I: Evaluation of Category & Data Gap Filling

Step 3. Reading data

- Next, the Data matrix is constructed by extracting available experimental results for all 19 members of the category.
- Based on preceding category evaluation, no outliers have been identified violating the structural and mechanistic consistency of the category.

Phase I: Evaluation of Category & Data Gap Filling

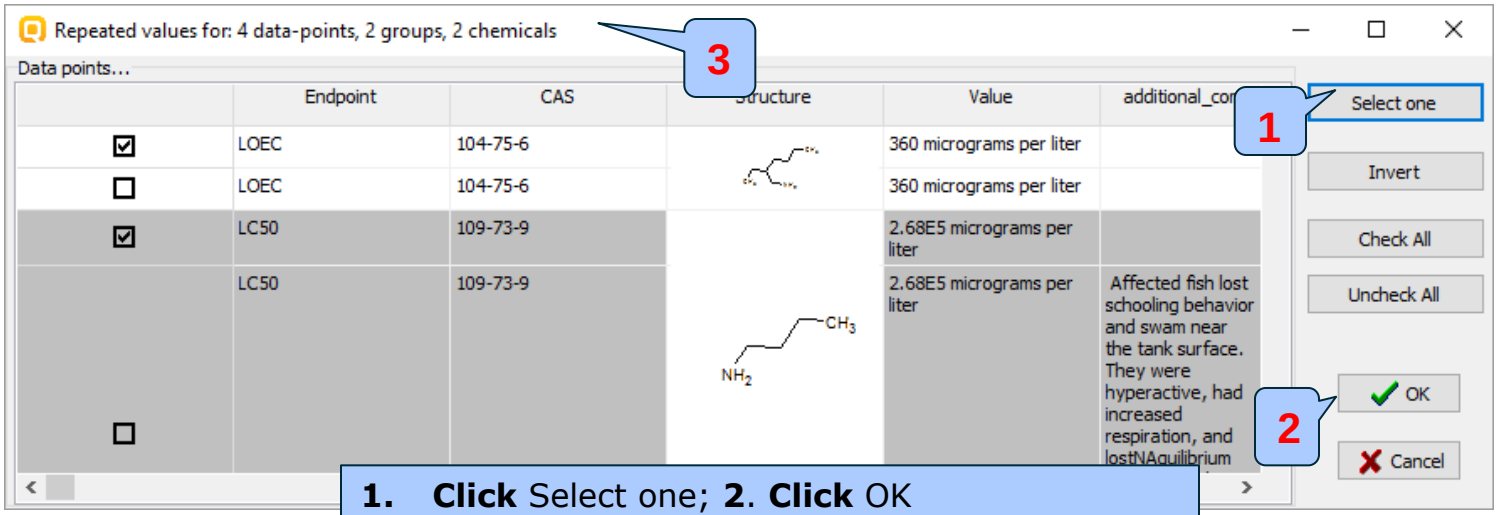
Step 3. Reading data



Phase I: Evaluation of Category & Data Gap Filling

Step 3. Reading data

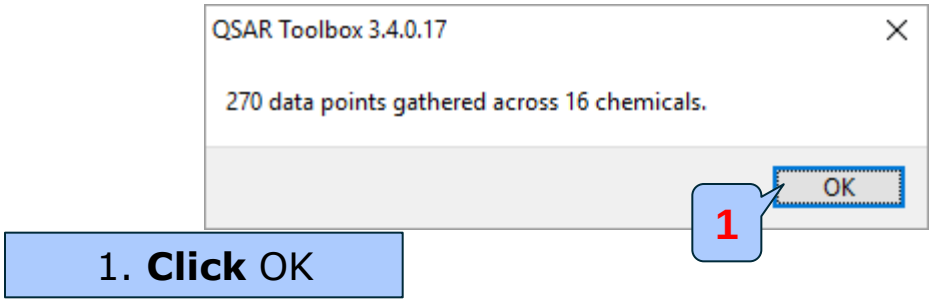
Due to the overlap between the Toolbox databases same data for intersecting chemicals could be found simultaneously in more than one database. The data redundancy is identified and the user has the opportunity to select either a single data value or all data values.



Phase I: Evaluation of Category & Data Gap Filling

Step 3. Reading data

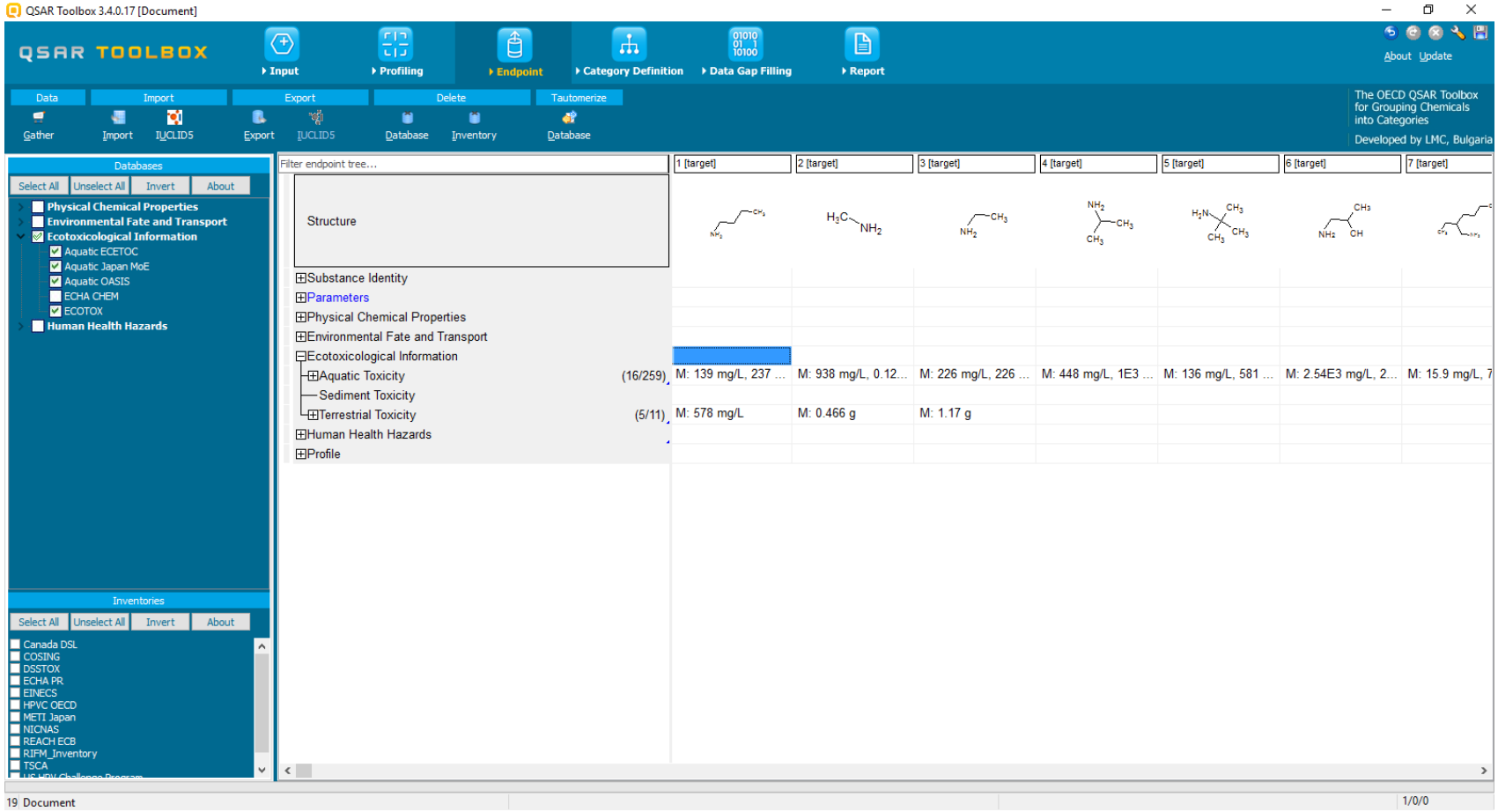
The system automatically gives indication for the number of gather experimental data points



Phase I: Evaluation of Category & Data Gap Filling

Step 3. Reading data.

Inserting data into the data matrix

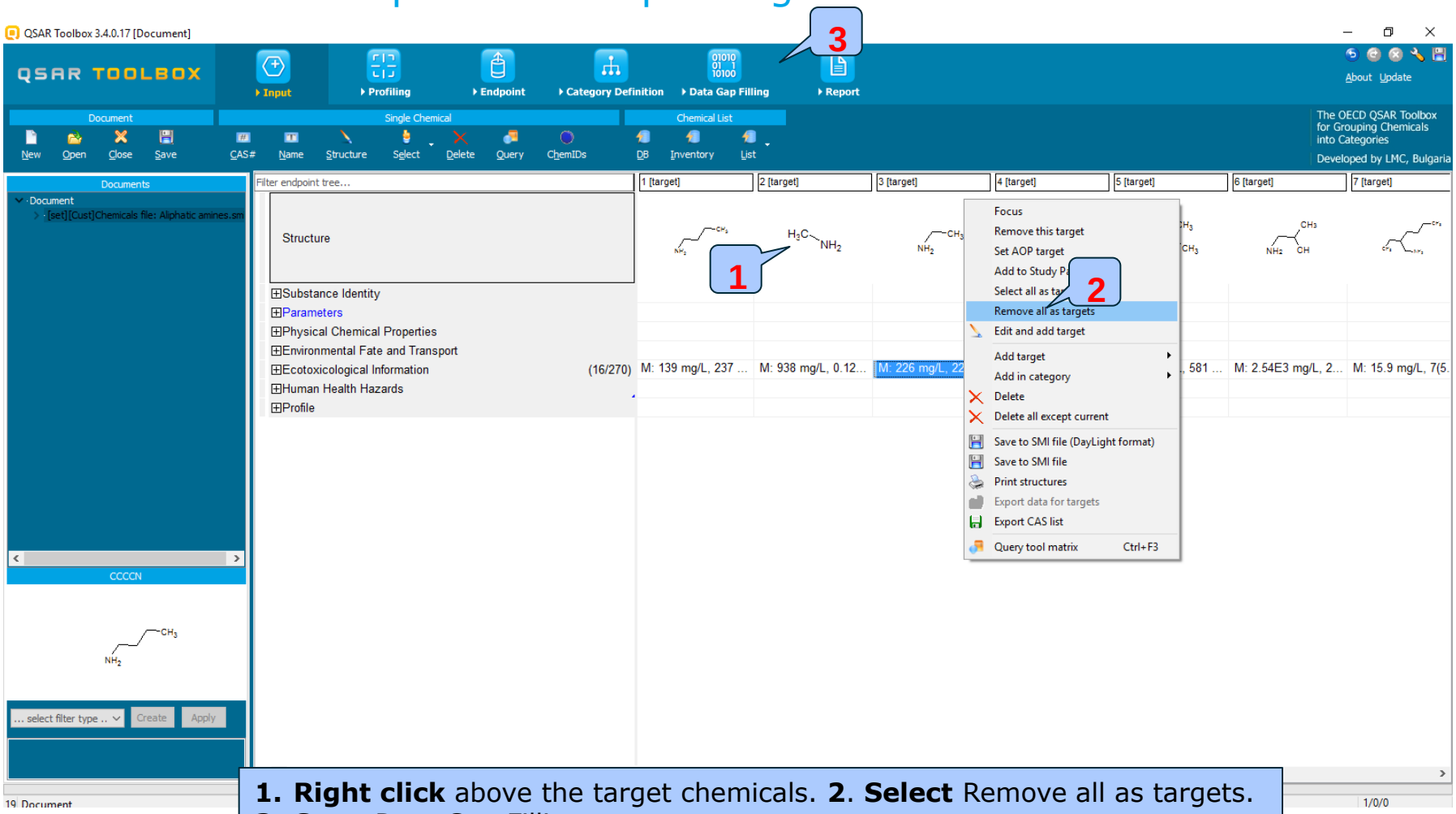


Outlook

- Background
- Category evaluation – overview
  - **Phase I. Evaluation of Category & Data Gap Filling**
    - *Step 1: Investigating the structural consistency of an ad hoc category.*
    - *Step 2: Investigating the a*pplicability domain of an ad hoc category.
    - *Step 3: Reading data for the analogues*
    - *Step 4: Data gap filling for 2-Butanamine*

Phase I: Evaluation of Category & Data Gap Filling

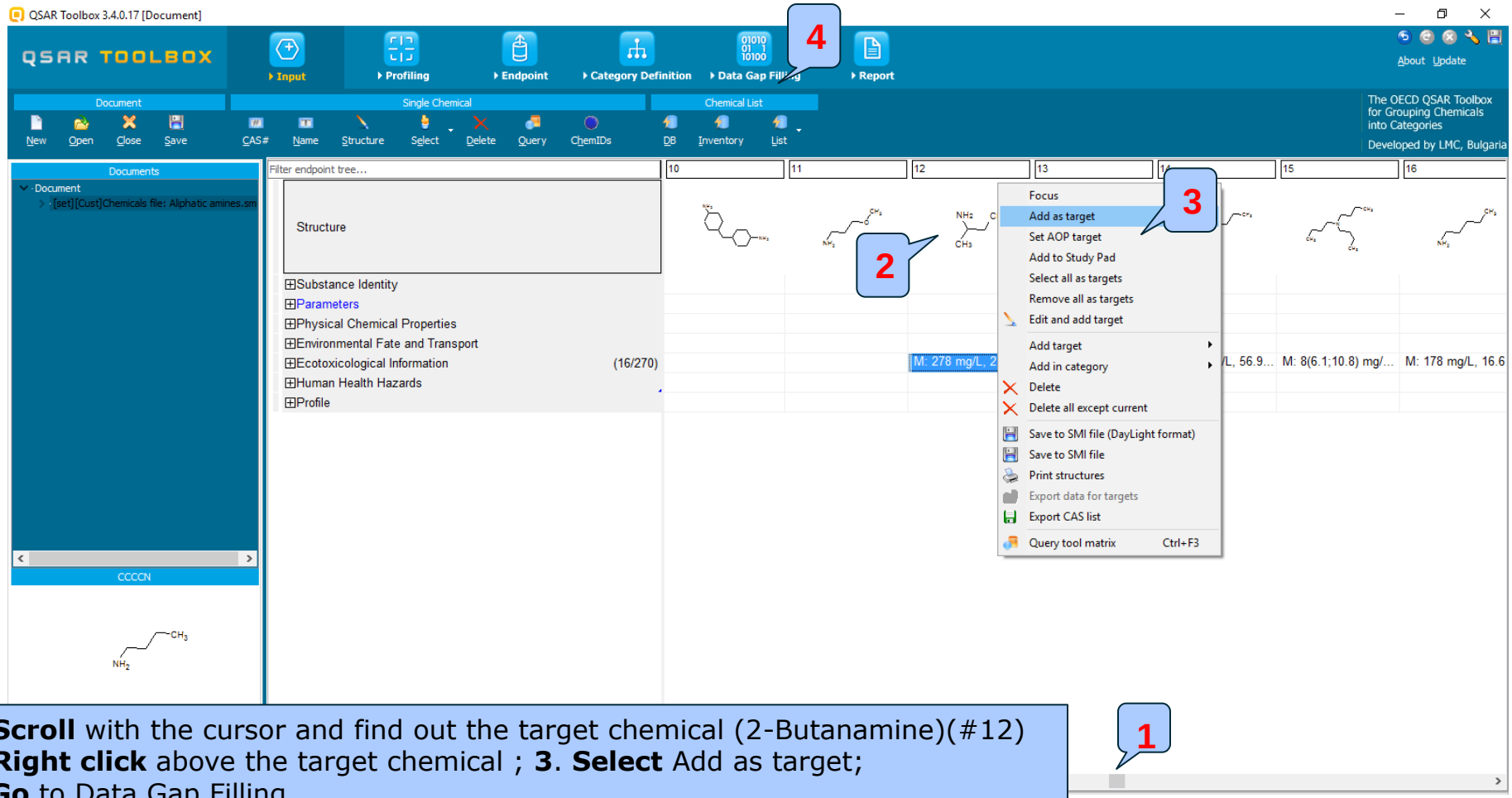
Step 4. Data Gap Filling for 2-Butanamine



1. Right click above the target chemicals. **2. Select** Remove all as targets.
3. Go to Data Gap Filling

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine

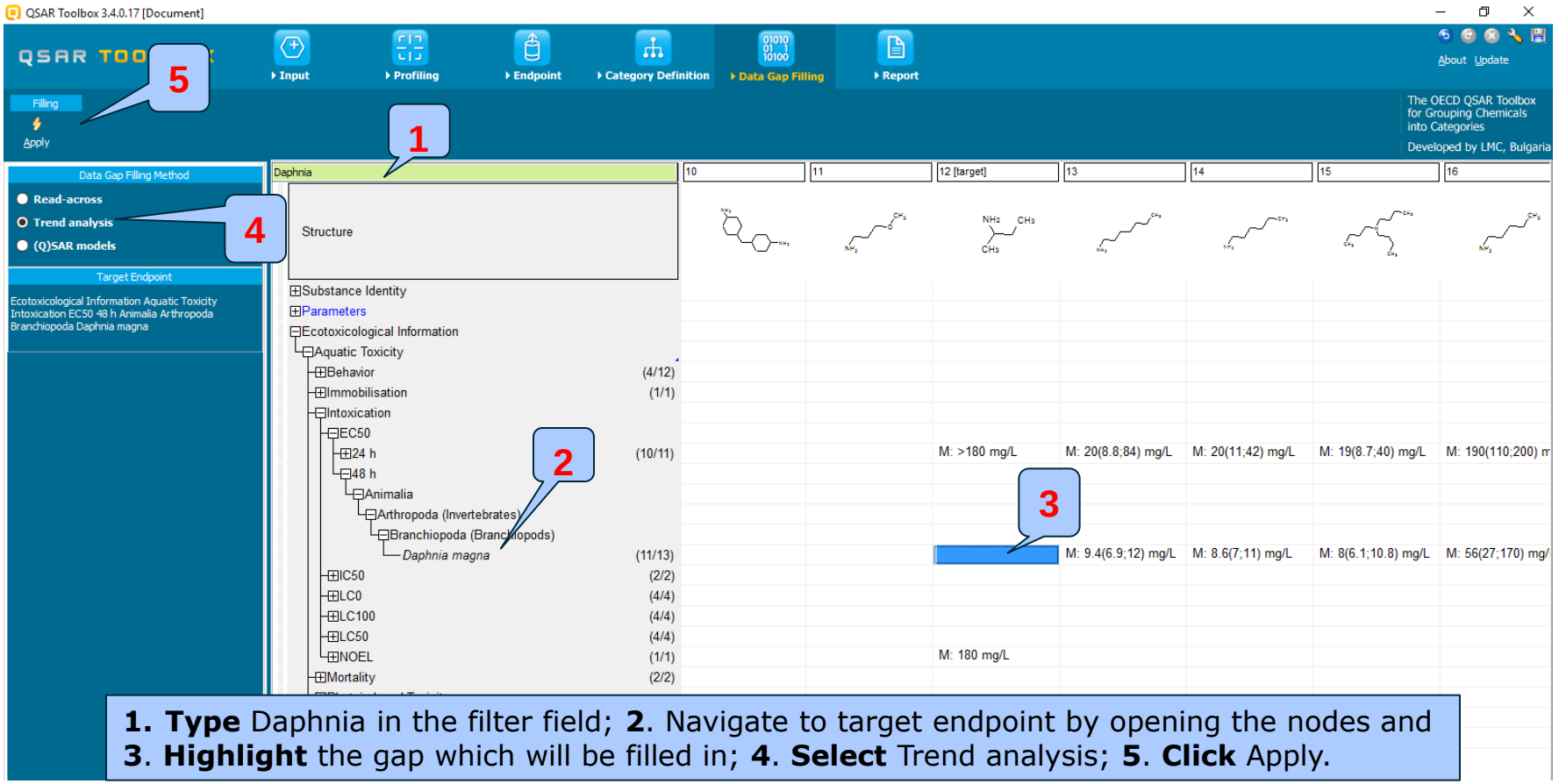


1. Scroll with the cursor and find out the target chemical (2-Butanamine)(#12)
2. Right click above the target chemical ; **3. Select** Add as target;
4. Go to Data Gap Filling

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine

Navigate to the target endpoint: Intoxication, D.magna,EC50, 48h



1. Type Daphnia in the filter field; **2**. Navigate to target endpoint by opening the nodes and **3**. **Highlight** the gap which will be filled in; **4**. **Select** Trend analysis; **5**. **Click** Apply.

Outlook

- Background
- Category evaluation – overview
  - **Phase I. Evaluation of Category & Data Gap Filling**
    - *Step 1: Investigating the structural consistency of an ad hoc category.*
    - *Step 2: Investigating the a*pplicability domain of an ad hoc category.
    - *Step 3: Reading data for the analogues*
    - *Step 4: Data gap filling for 2-Butanamine*
      - *Step 4.1*: How to reorder the endpoint tree*

Phase I: Evaluation of Category & Data Gap Filling

Step 4.1*. How to reorder the endpoint tree

* Detailed information describing how to reproduce the order of endpoint tree shown in the current presentation

Endpoint tree consist of two parts

Predefined

Ecotoxicological Information
Aquatic Toxicity

This part cannot be reordered or changed.

Dynamic part

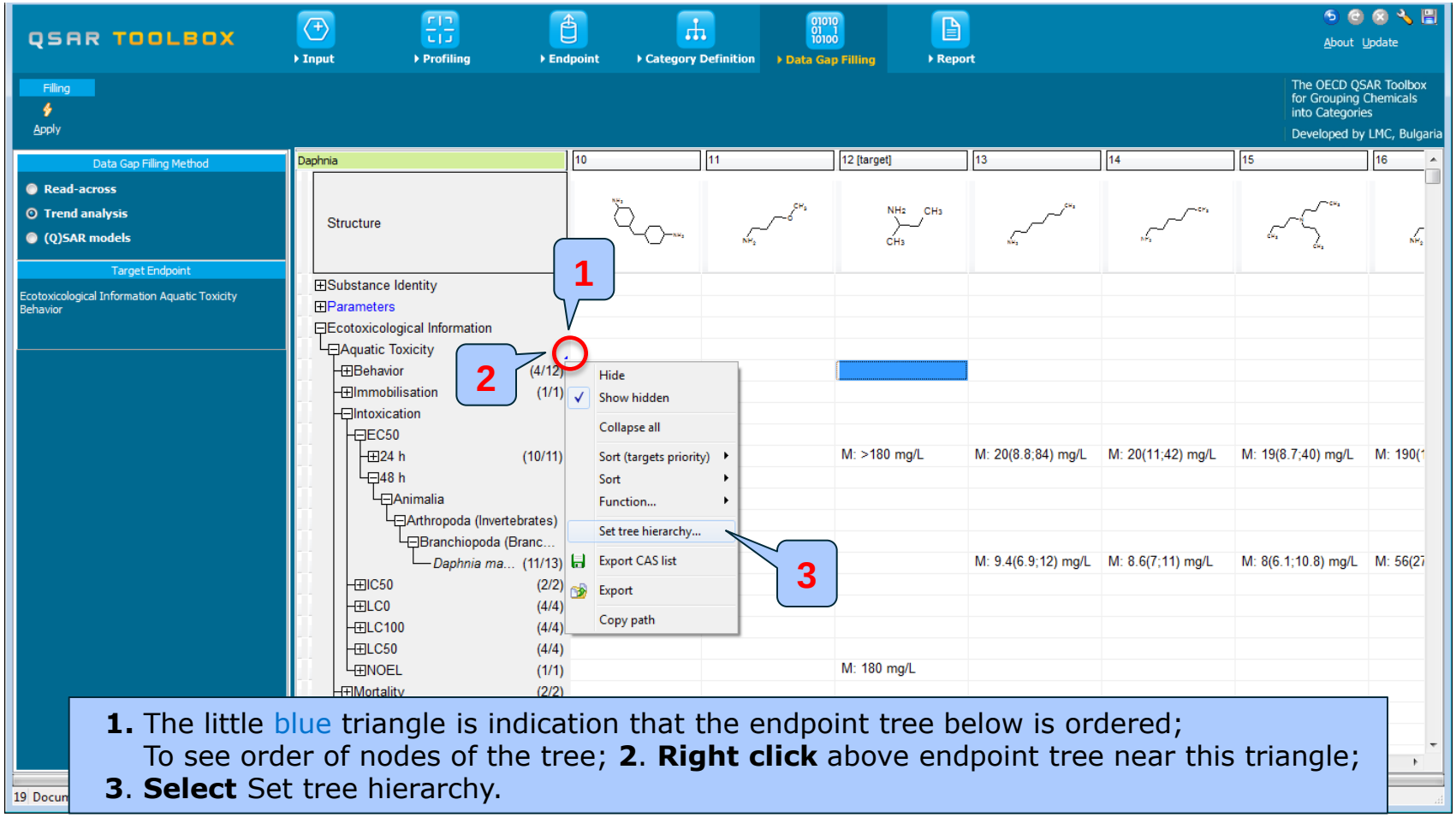
Animalia	Kingdom
Arthropoda(Invertebrates)	Phylum
Branchiopoda(branchiopods)	Class
Intoxication	Effect
Daphnia magna	Test organism
24 h	Duration
EC 50	Endpoint

Scientific information automatically collected from the Taxonomic bank.

Dynamic part, includes fields coming from databases. These fields have different labels like effect for intoxication, or endpoint for EC50. The position of these field in endpoint tree is managing using the **Set tree hierarchy** option (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 4.1. How to reorder the endpoint tree



1. The little blue triangle is indication that the endpoint tree below is ordered;
To see order of nodes of the tree; **2**. **Right click** above endpoint tree near this triangle;
3. **Select** Set tree hierarchy.

Phase I: Evaluation of Category & Data Gap Filling

Step 4.1. How to reorder the endpoint tree

Intoxication — Effect

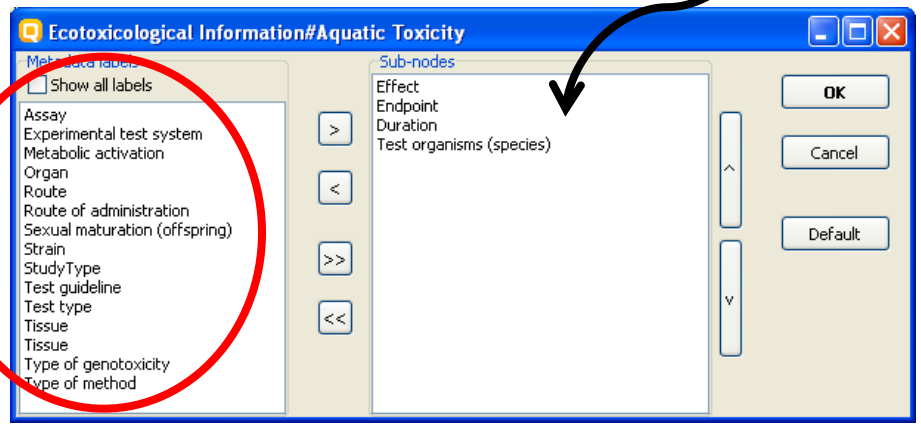
Daphnia magna — Test organism

24 h — Duration

EC50 — Endpoint

Left panel contains most usable labels of fields coming from databases.

The endpoint tree is organized following the order shown in right panel



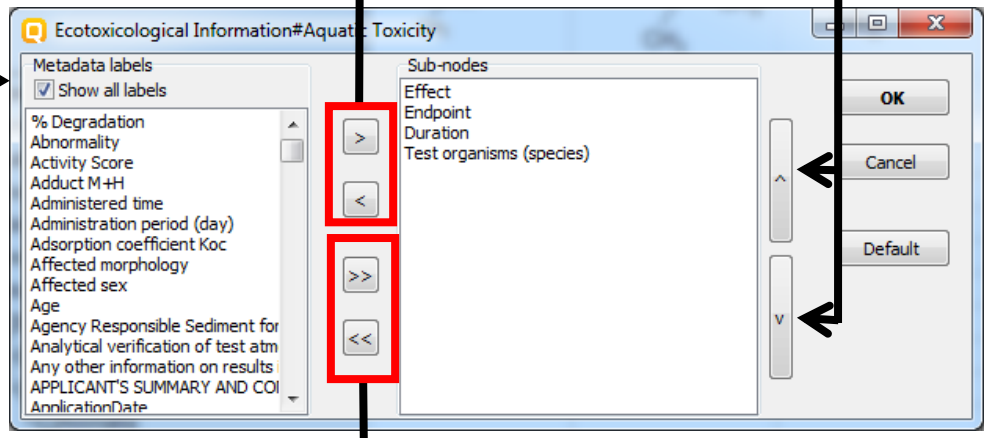
Phase I: Evaluation of Category & Data Gap Filling

Step 4.1. How to reorder the endpoint tree

All available fields coming from databases could be shown when the box **Show all labels** is selected.

To remove or change some of the fields which are visualizing on the tree, highlight the field from left panel and move it to the right panel using the **Right** or **Left** button.

Reordering of the fields is possible using the **Up** and **Down** buttons.



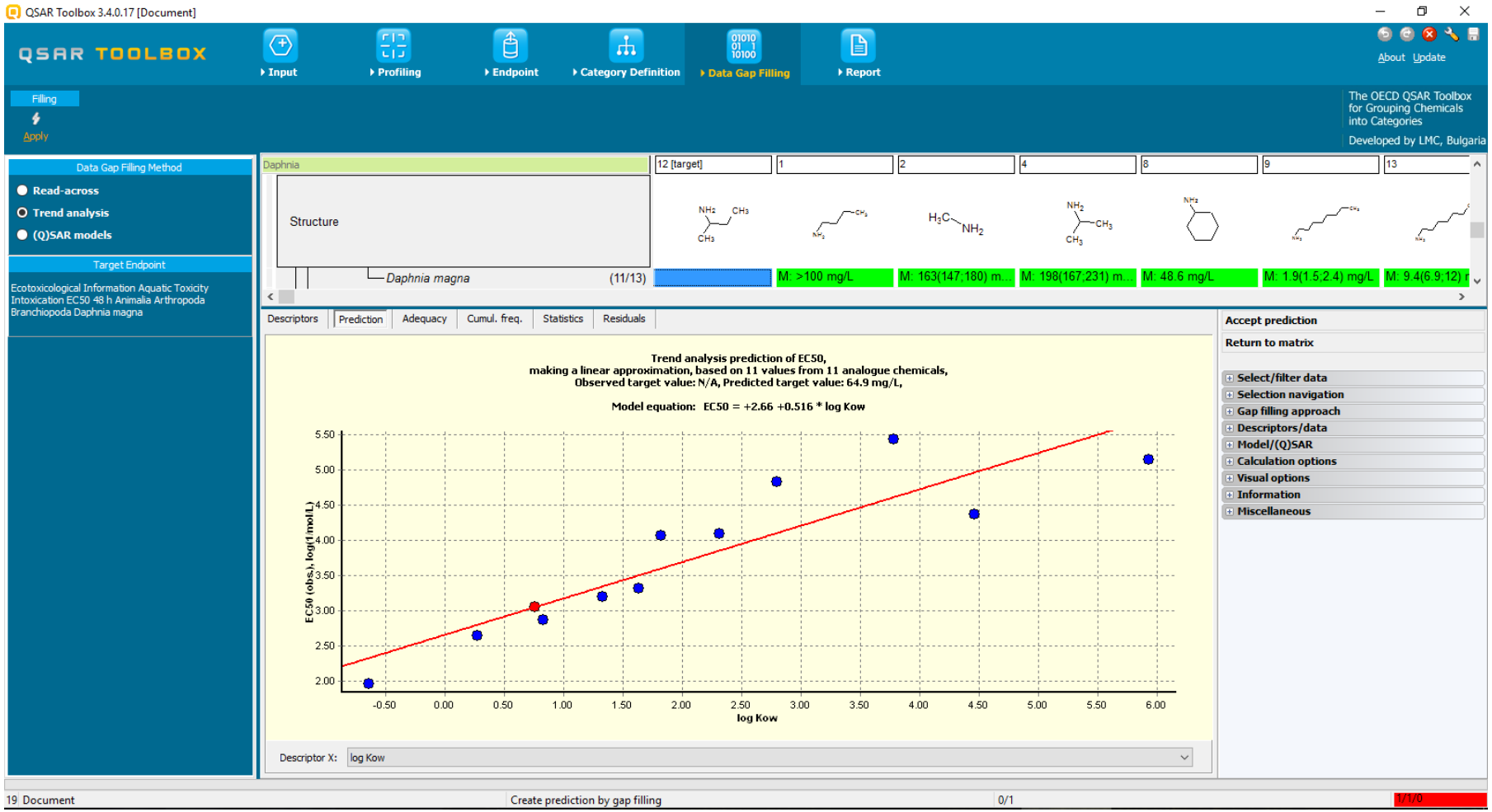
To move all fields form one panel to another use **Right** and **Left** button.

Outlook

- Background
- Category evaluation – overview
  - **Phase I. Evaluation of Category & Data Gap Filling**
    - *Step 1: Investigating the structural consistency of an ad hoc category.*
    - *Step 2: Investigating the a*pplicability domain of an ad hoc category.
    - *Step 3: Reading data for the analogues*
    - *Step 4: Data gap filling for 2-Butanamine*
      - *Back to Trend analysis*

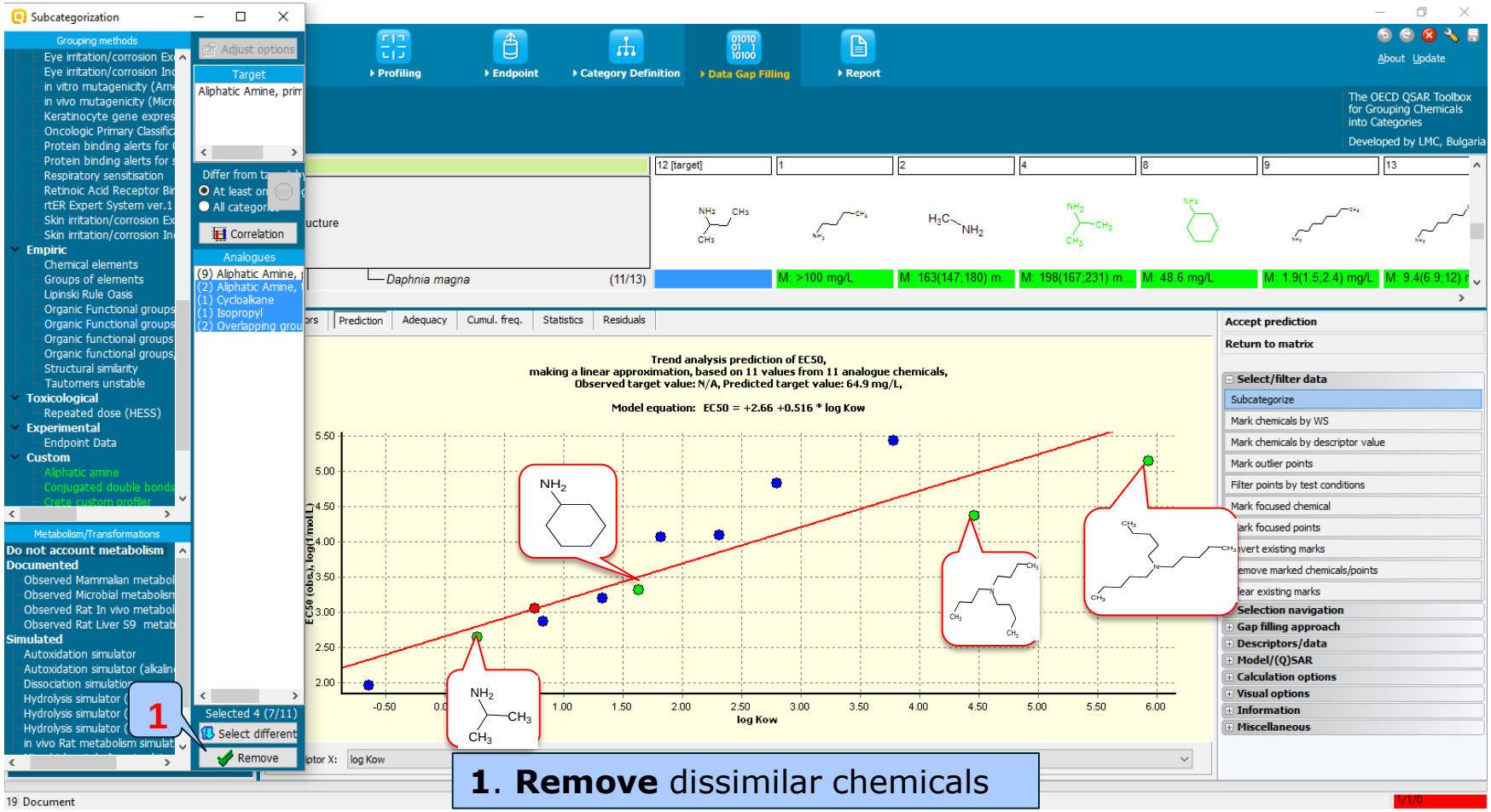
Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine



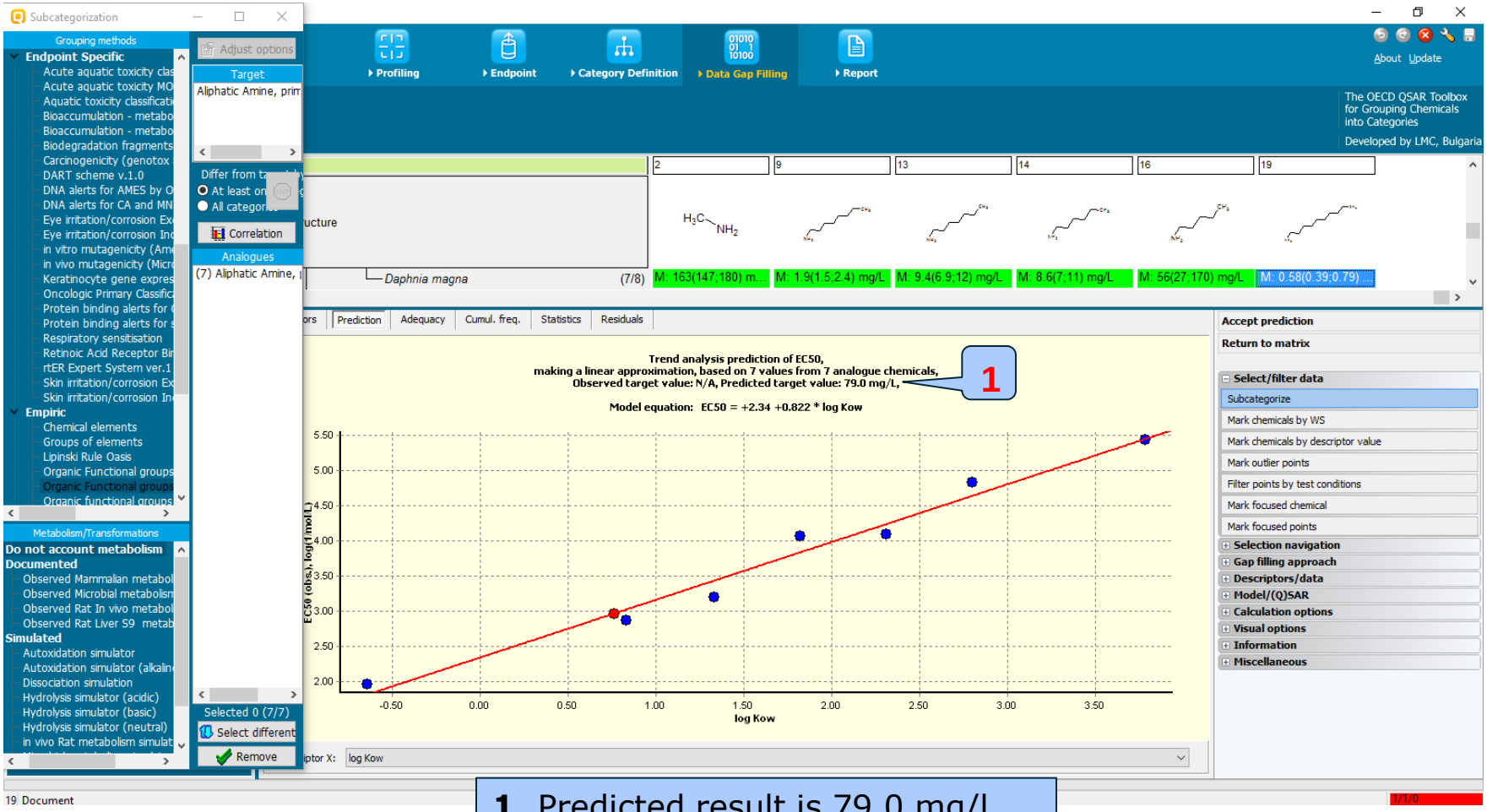
Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine/*Subcategorize by OFG(nested)*



Phase I: Evaluation of Category & Data Gap Filling

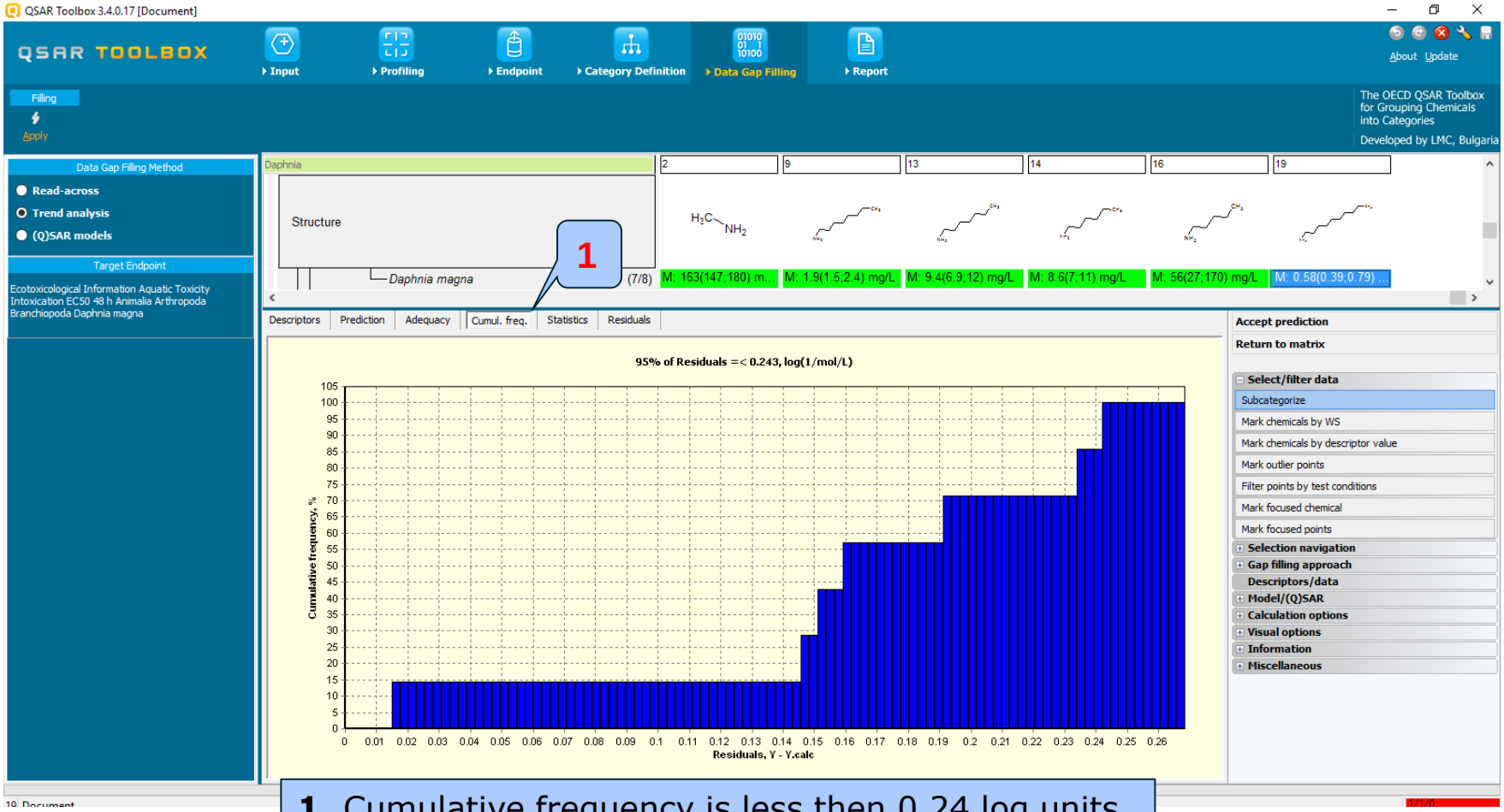
Step 4. Data Gap Filling for 2-Butanamine/*Predicted result*



1. Predicted result is 79.0 mg/l

Phase I: Evaluation of Category & Data Gap Filling

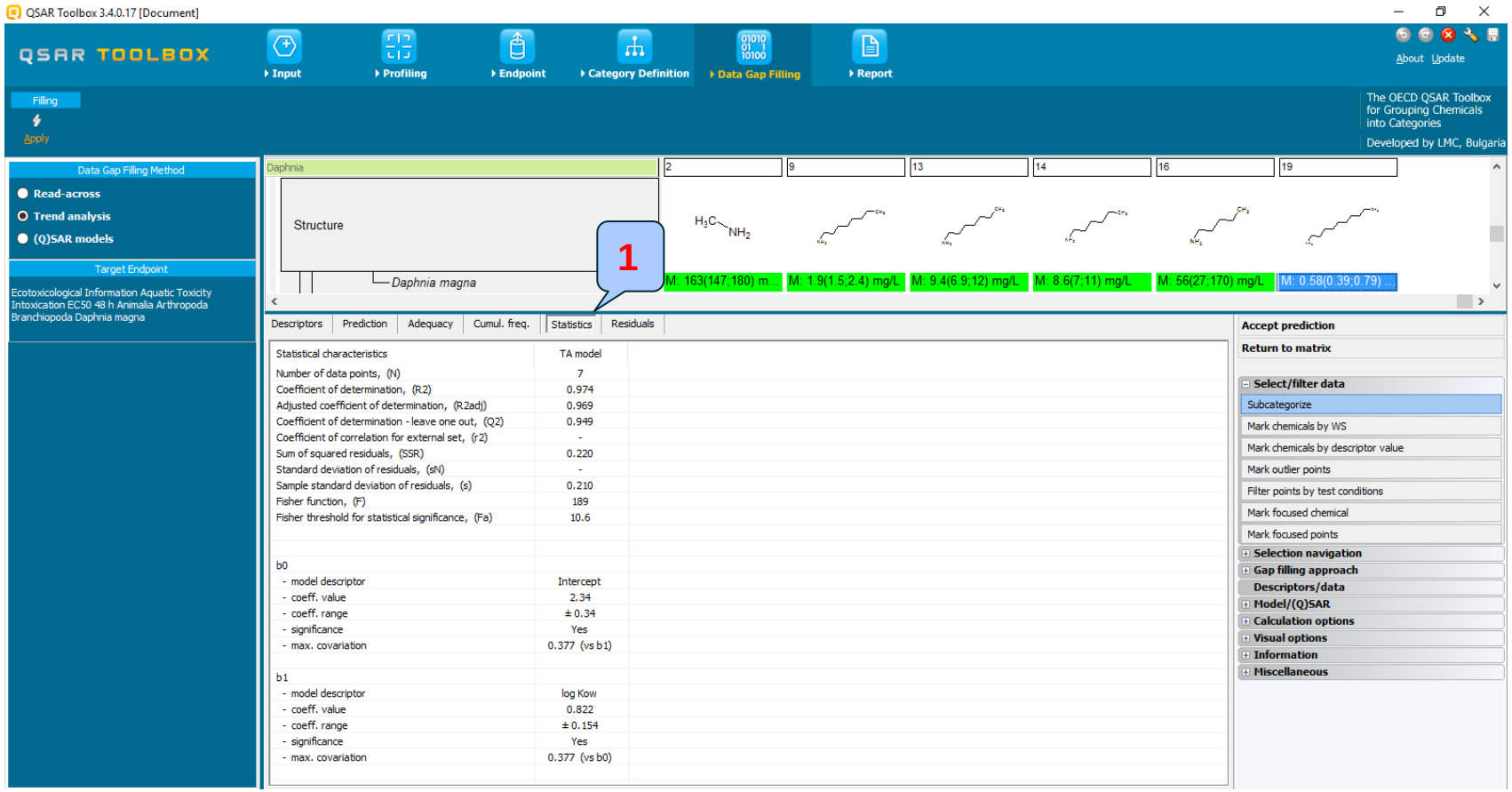
Step 4. Data Gap Filling for 2-Butanamine/*Cumulative frequency*



1. Cumulative frequency is less then 0.24 log units

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine/*Statistics*



Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine/*Interpretation of the result*

- The structurally similar analogs across category of aliphatic amines is used for data gap filling
- Subcategorization by Organic functional groups(nested) is applied
- The prediction based on the defined category is acceptable.
- The predicted value based on predefined category of aliphatic amines is 79.0 mg/l

Outlook

- Background
- Category evaluation – overview
  - Phase I. Evaluation of Category & Data Gap Filling
  - **Phase II. Extension of the Category& Data gap Filling**

Phase II: Extension Category & Data gap filling

- The extension of the category is performed by using Phase II of the category evaluation process (*Extension of Category & Data Gap Filling*). Other analogues are searched in the Toolbox, which are structurally and mechanistically consistent with the predefined category.
- The structural analogues could be defined using ECOSAR grouping

Outlook

- Background
- Category evaluation – overview
  - Phase I. Evaluation of Category & Data Gap Filling
  - **Phase II. Extension of the Category& Data gap Filling**
    - Step 1: Category definition

Phase II: Extension Category & Data gap filling

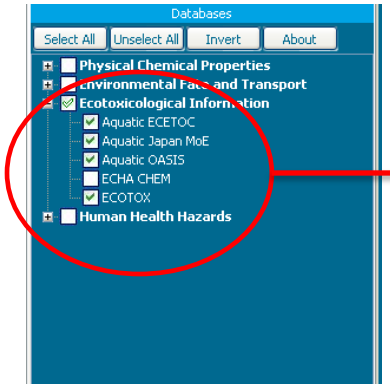
Step 1: Category definition

- The ECOSAR strict category is used to define an broader category used in further analysis.
- According to ECOSAR categorization **a list of 302 aliphatic amines have been selected** for which data are available in the Toolbox (identified in the Toolbox databases).
- The same endpoint: EC 50 48h *D.magna* will be predicted as with the predefined category

Phase II: Extension Category & Data gap filling

Step 1: Category definition

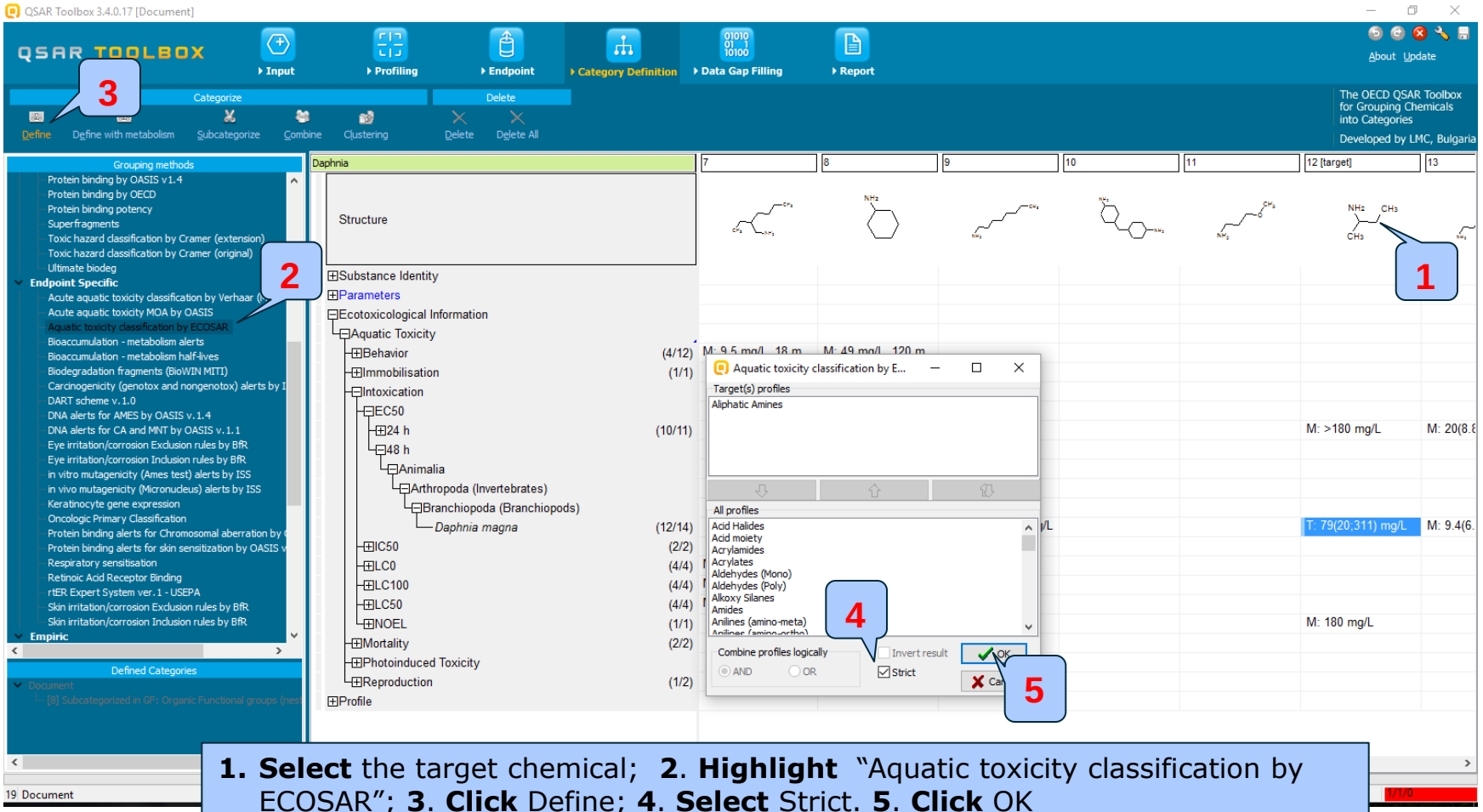
- Define ECOSAR category
- Before defining the category, the following databases related to the predicted endpoint are selected:



Aquatic ECETOC
Aquatic Japan MoE
Aquatic OASIS
ECOTOX

Phase II: Extension Category & Data gap filling

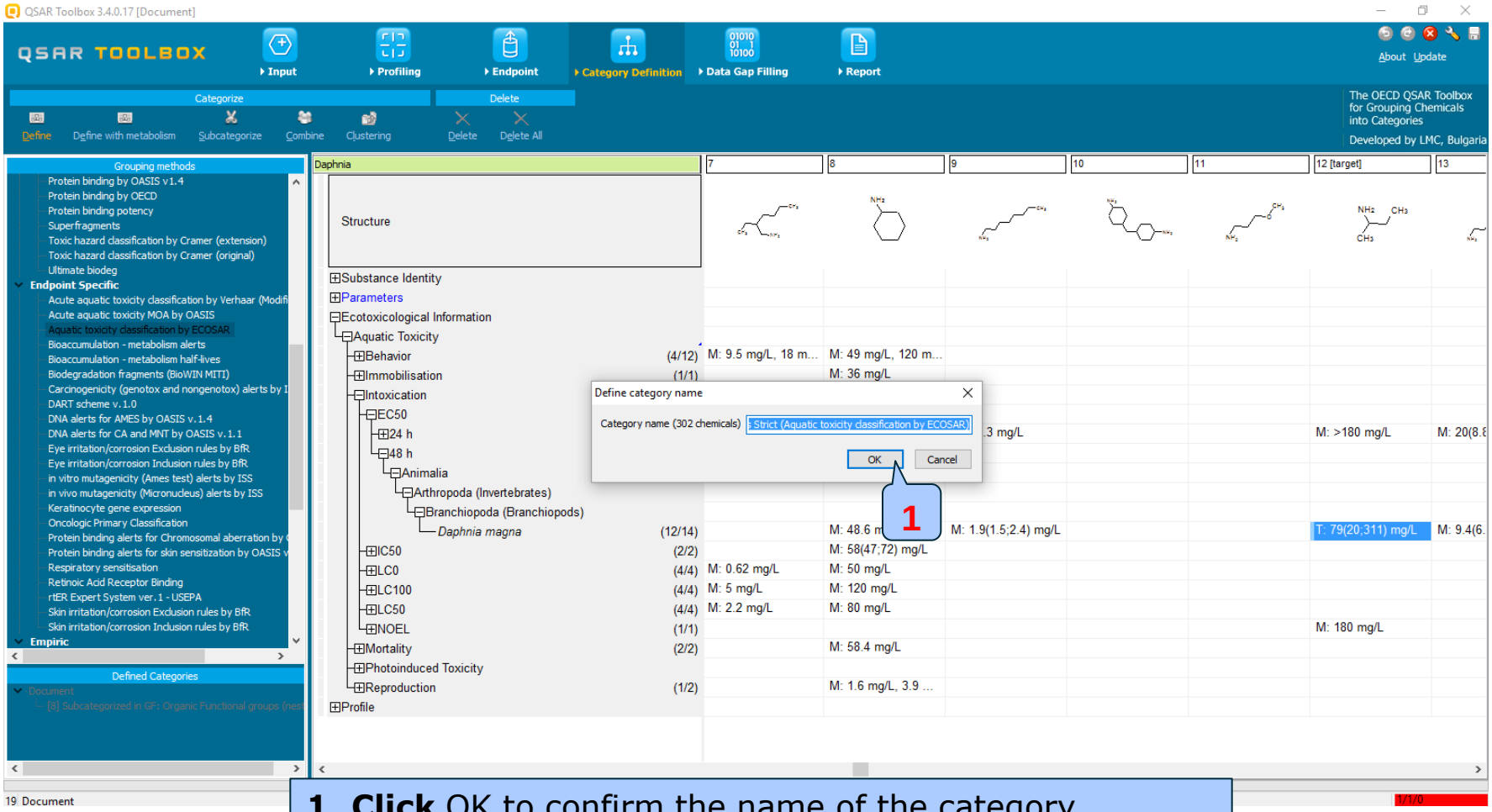
Step 1: Category definition/*Defining ECOSAR (strict)*



1. Select the target chemical; **2. Highlight** "Aquatic toxicity classification by ECOSAR"; **3. Click** Define; **4. Select** Strict. **5. Click** OK

Phase II: Extension Category & Data gap filling

Step 1: Category definition/*Defining ECOSAR (strict)*



1. Click OK to confirm the name of the category.

Phase II: Extension Category & Data gap filling

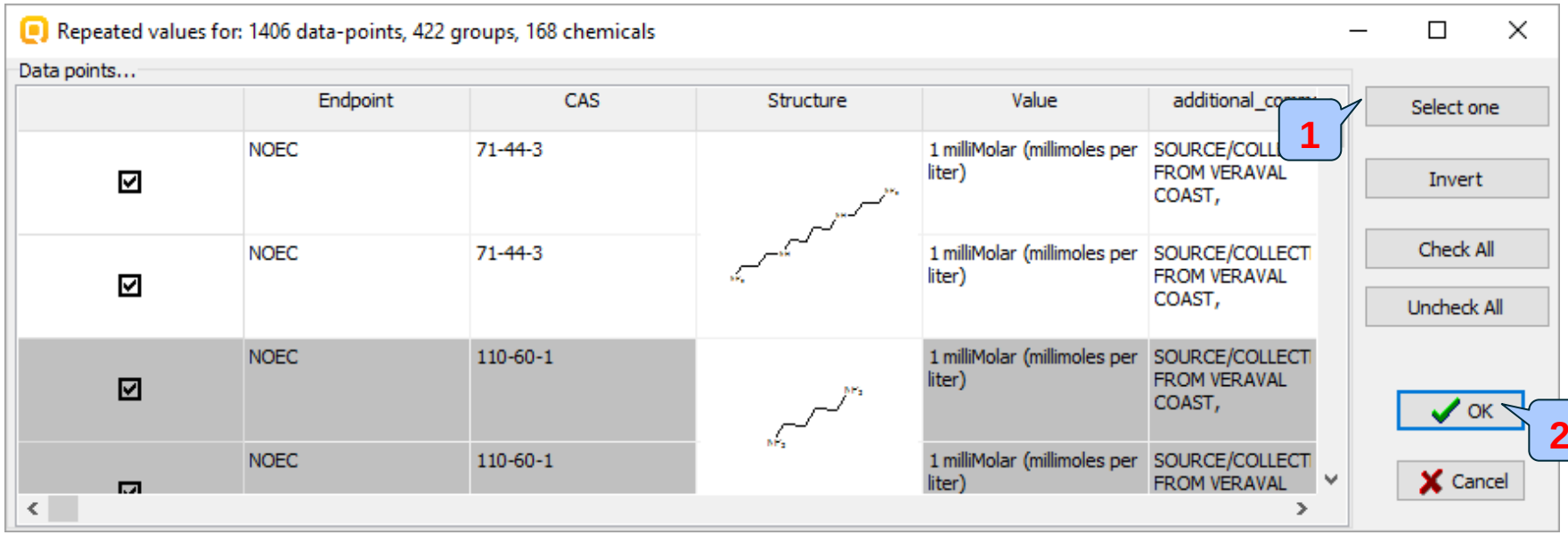
Step 1: Category definition/*Reading data*

- The Toolbox will now retrieve those chemicals that have the same ECOSAR functionality as the target compound.
- 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, as only databases are selected that contain information for aquatic toxicity endpoint, both options give the same results.

Phase II: Extension Category & Data gap filling

Step 1: Category definition/*Reading data*

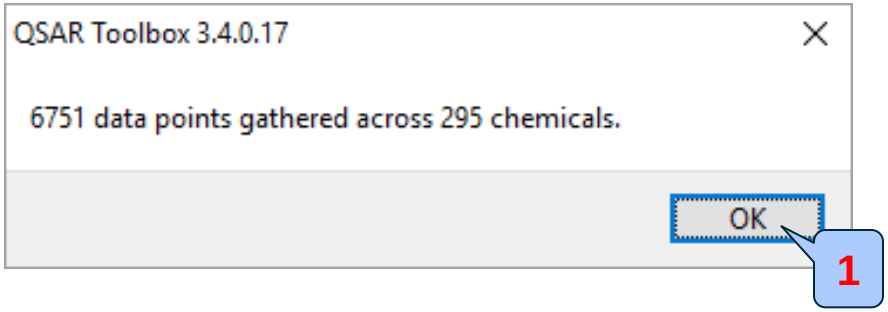


1. Click Select one and then; **2. Click** OK**.**

Phase II: Extension Category & Data gap filling

Step 1: Category definition/*Reading data*

The system automatically gives indication for the number of gather experimental data points



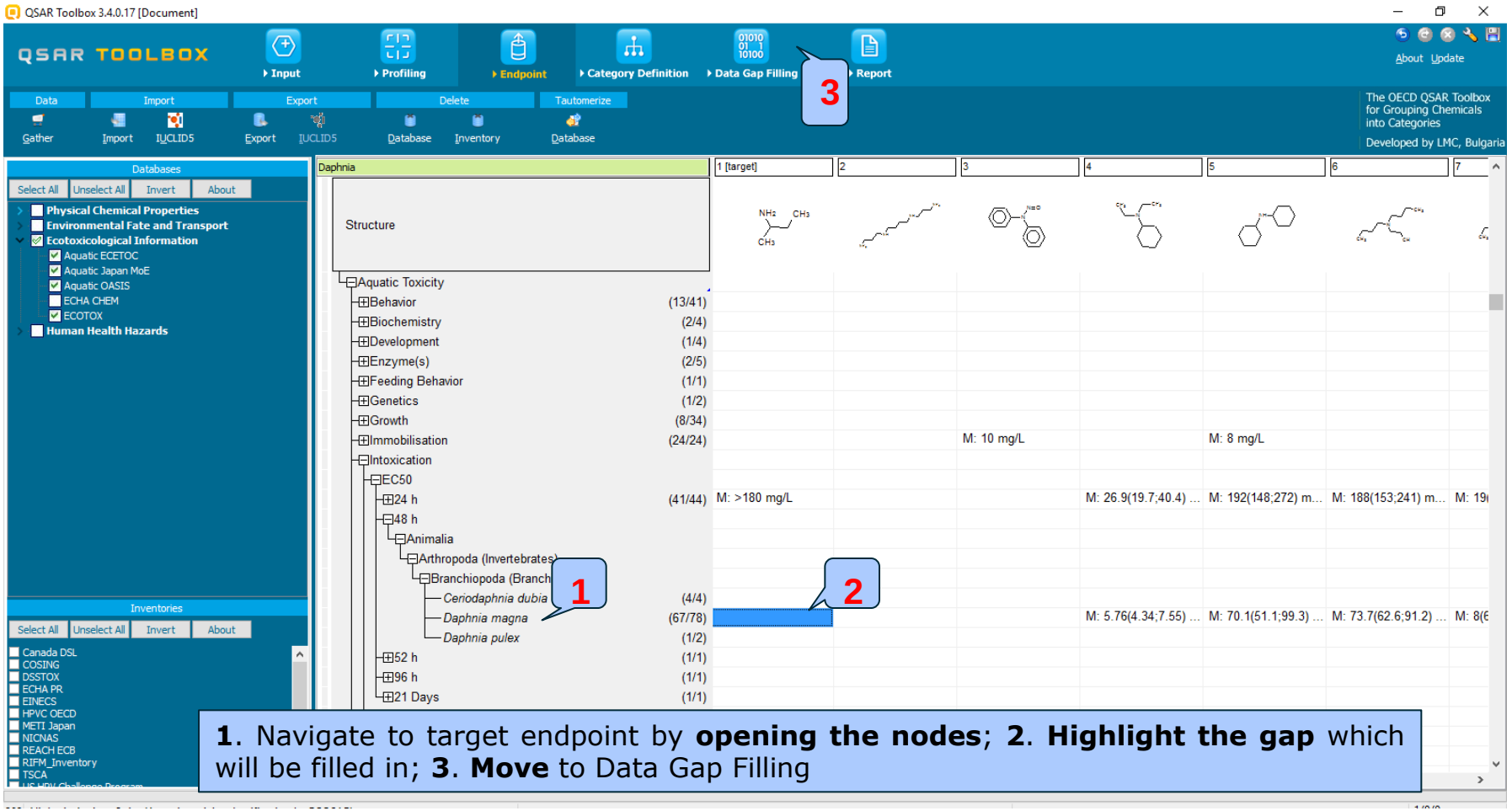
1. **Click** OK

Outlook

- Background
- Category evaluation – overview
  - Phase I. Evaluation of Category & Data Gap Filling
  - **Phase II. Extension of the Category& Data gap Filling**
    - Step 1: Category definition
    - Step 2: Navigate to the target endpoint

Phase II: Extension Category & Data gap filling

Step 2: Navigate to the target endpoint/*EC 50, 48h, D.magna*



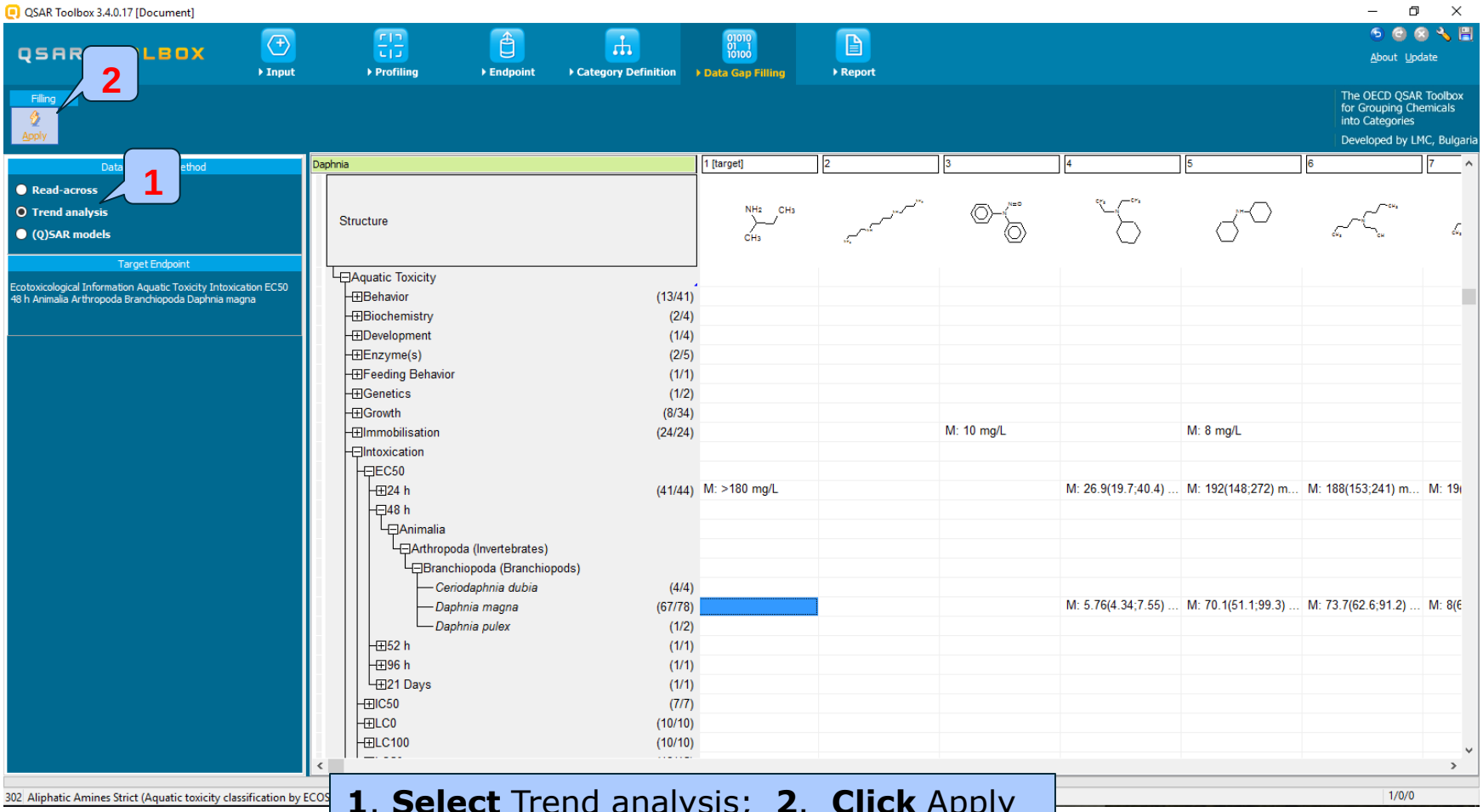
1. Navigate to target endpoint by **opening the nodes**; **2**. **Highlight the gap** which will be filled in; **3**. **Move** to Data Gap Filling

Outlook

- Background
- Category evaluation – overview
  - Phase I. Evaluation of Category & Data Gap Filling
  - **Phase II. Extension of the Category& Data gap Filling**
    - Step 1: Category definition
    - Step 2: Navigate to the target endpoint
    - Step 3: Data Gap Filling

Phase II: Extension Category & Data gap filling

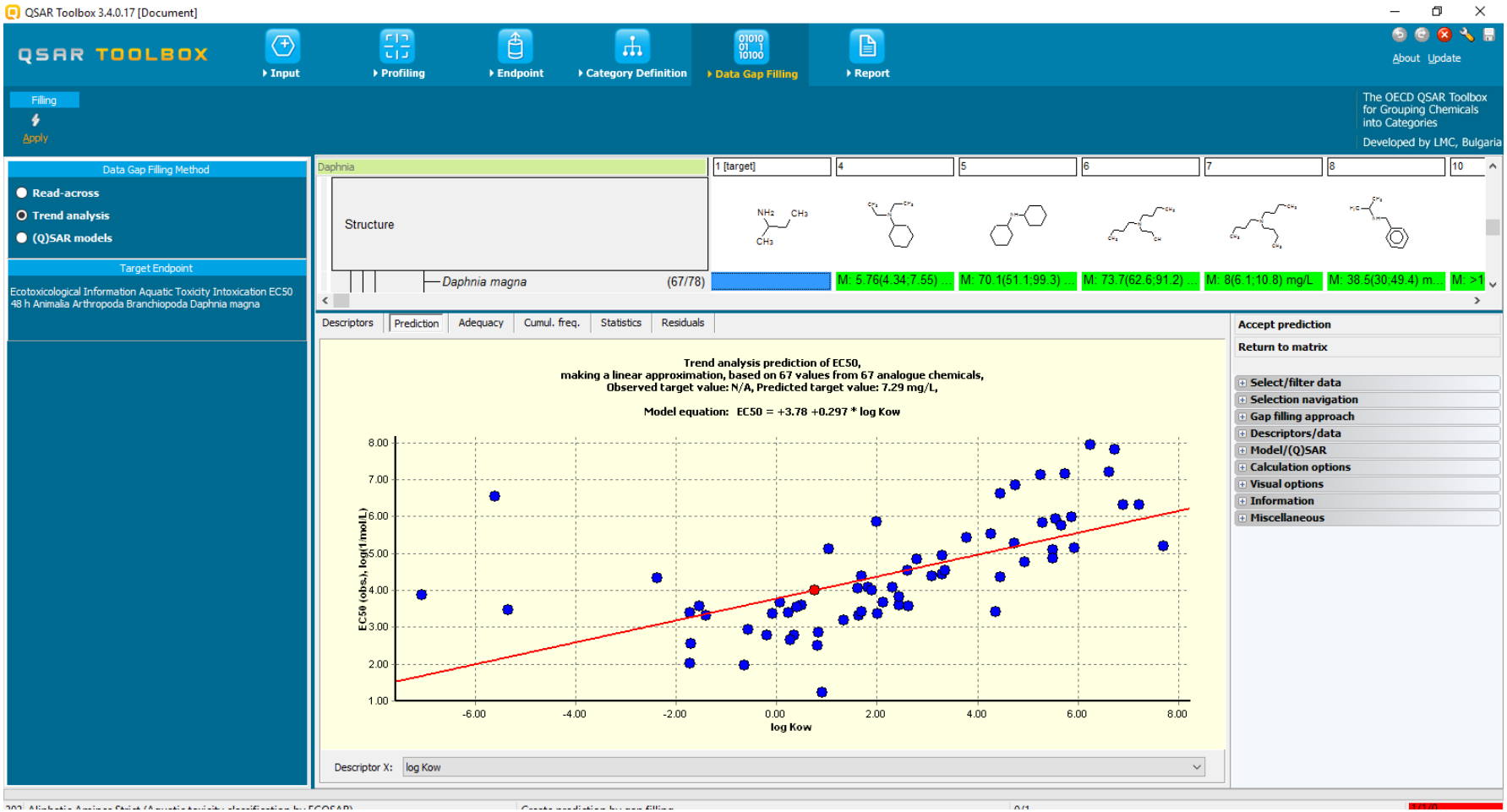
Step 3: Data Gap Filling



1. **Select** Trend analysis; **2**. **Click** Apply

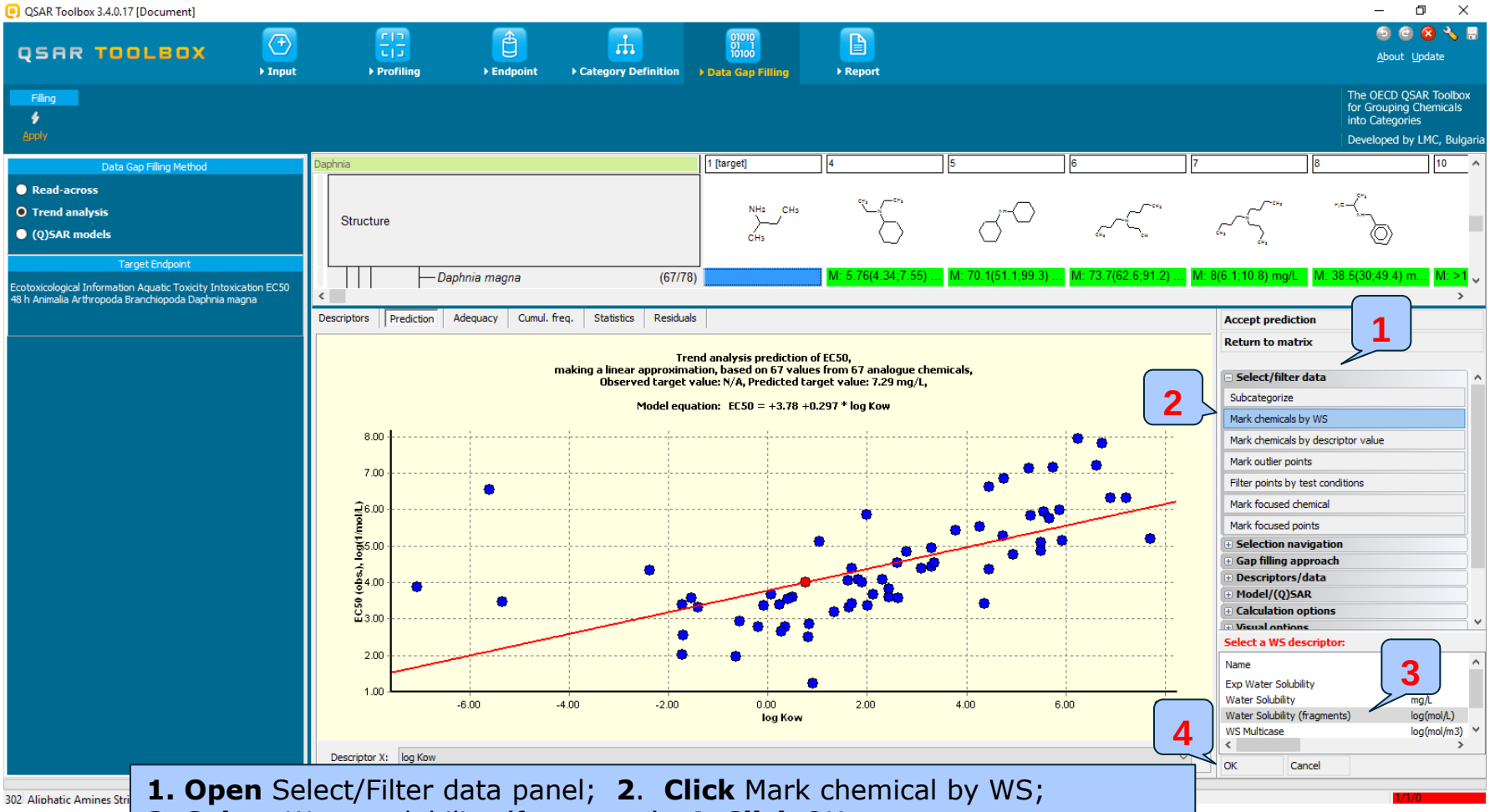
Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling



Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/*Subcategorize by Water solubility*



1. Open Select/Filter data panel; **2. Click** Mark chemical by WS;
3. Select Water solubility (fragments); **4. Click** OK

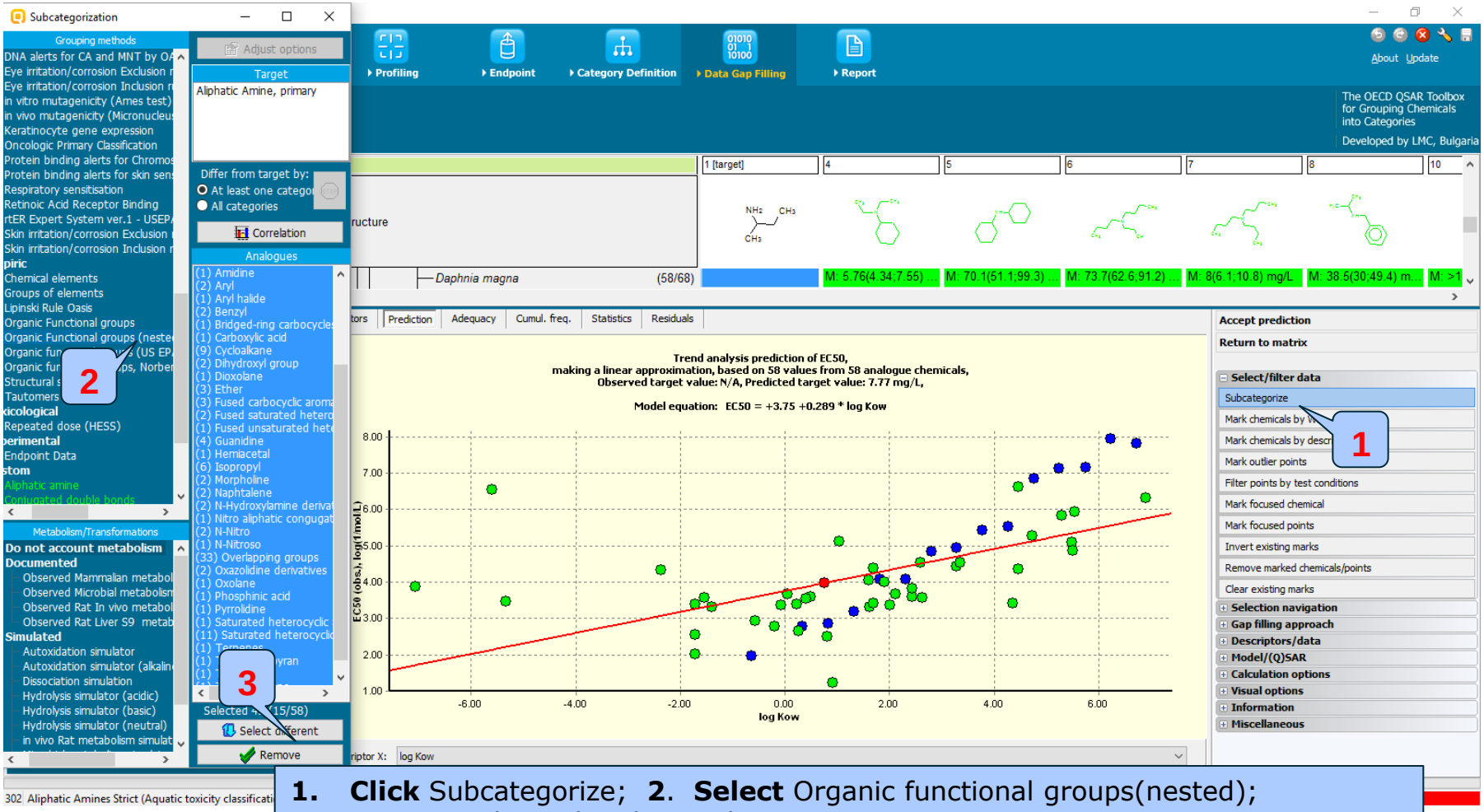
Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/*Subcategorize by Water solubility*

1. Nine chemicals are marked in green; **2**.**Click** OK; **3**. **Remove** marked chemicals

Phase II: Extension Category & Data gap filling

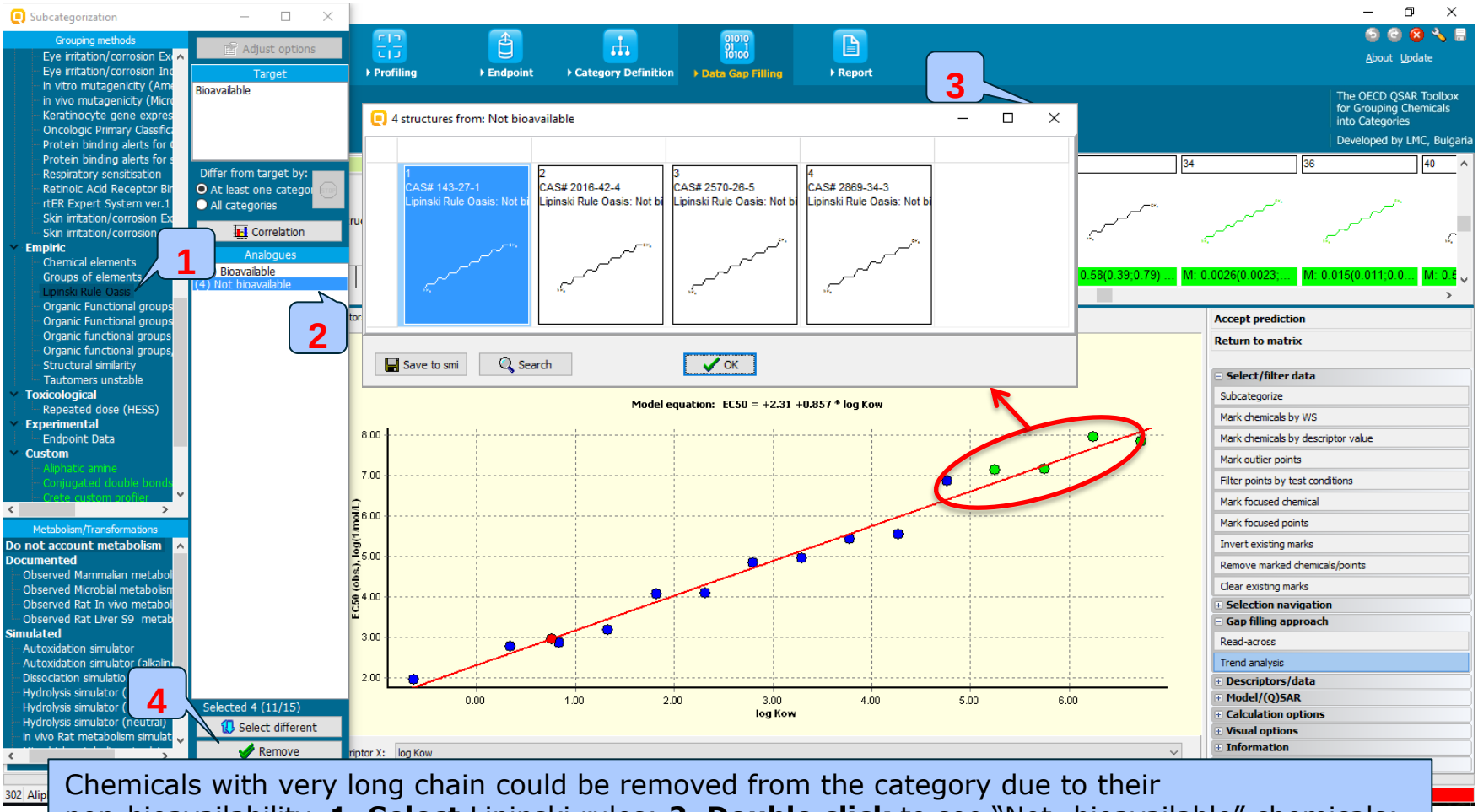
Step 3: Data Gap Filling/*Subcategorize by OFG(nested)*



1. **Click** Subcategorize; **2.** **Select** Organic functional groups(nested);
3. **Remove** dissimilar chemicals

Phase II: Extension Category & Data gap filling

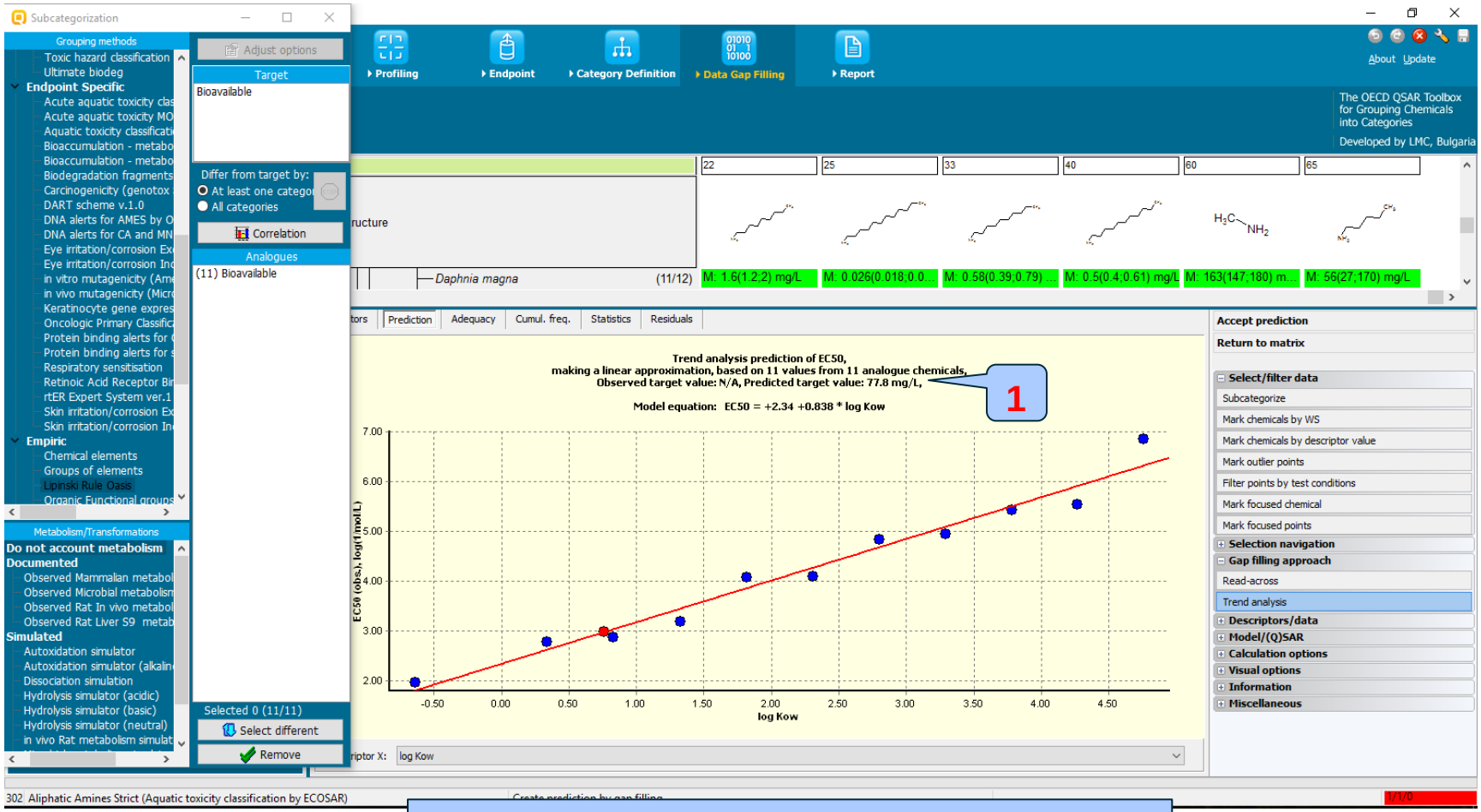
Step 3: Data Gap Filling/*Subcategorize by Lipinski rules*



Chemicals with very long chain could be removed from the category due to their non-bioavailability. **1**. **Select** Lipinski rules; **2**. **Double click** to see "Not- bioavailable" chemicals; **3**. **Close** the appeared window; **4**. **Remove** dissimilar chemicals

Phase II: Extension Category & Data gap filling

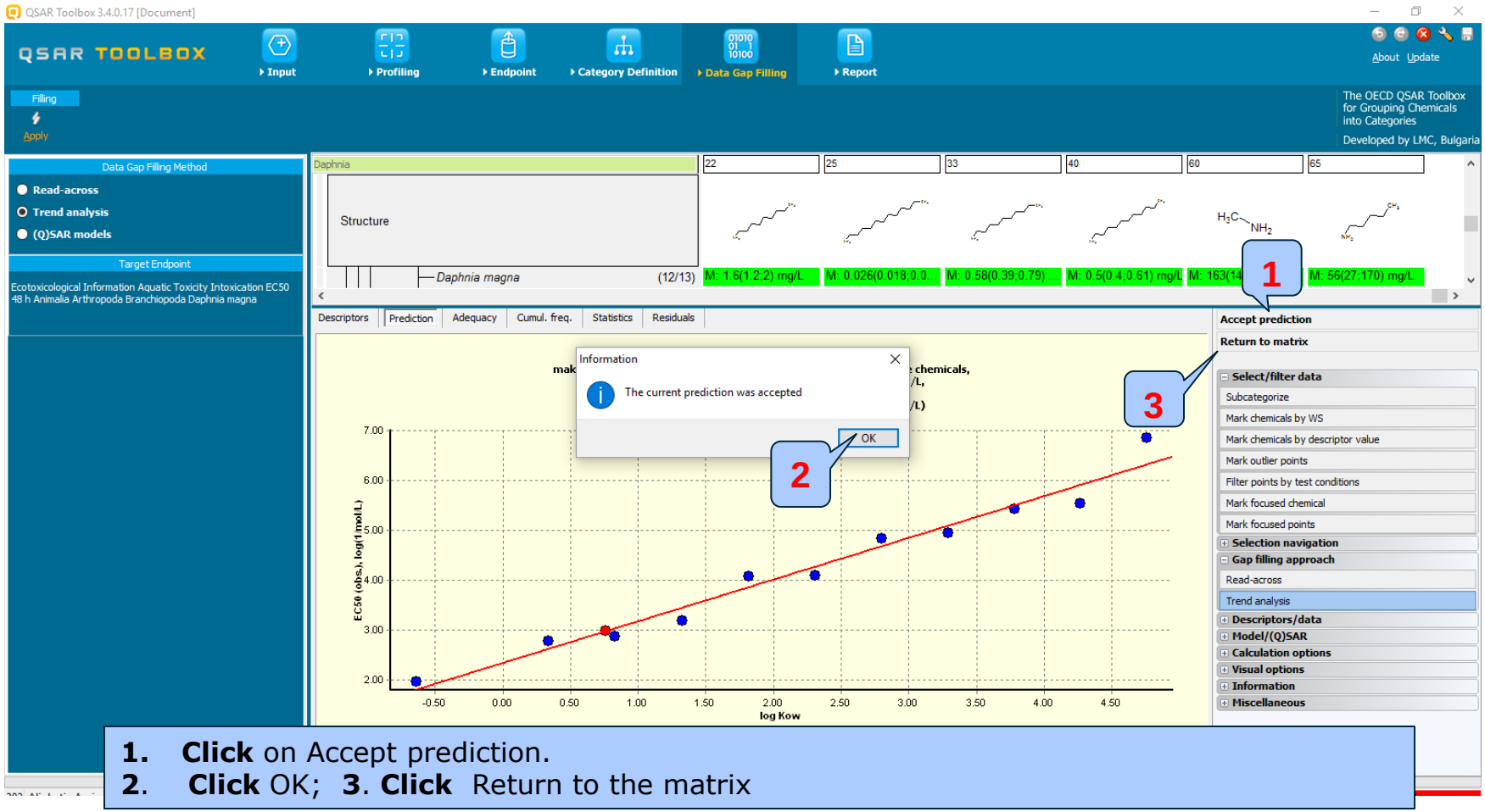
Step 3: Data Gap Filling/*Prediction result*



Prediction result is 77.8 mg/l.

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/*Accept the Prediction result*



1. **Click** on Accept prediction.
2. **Click** OK; **3. Click** Return to the matrix

Outlook

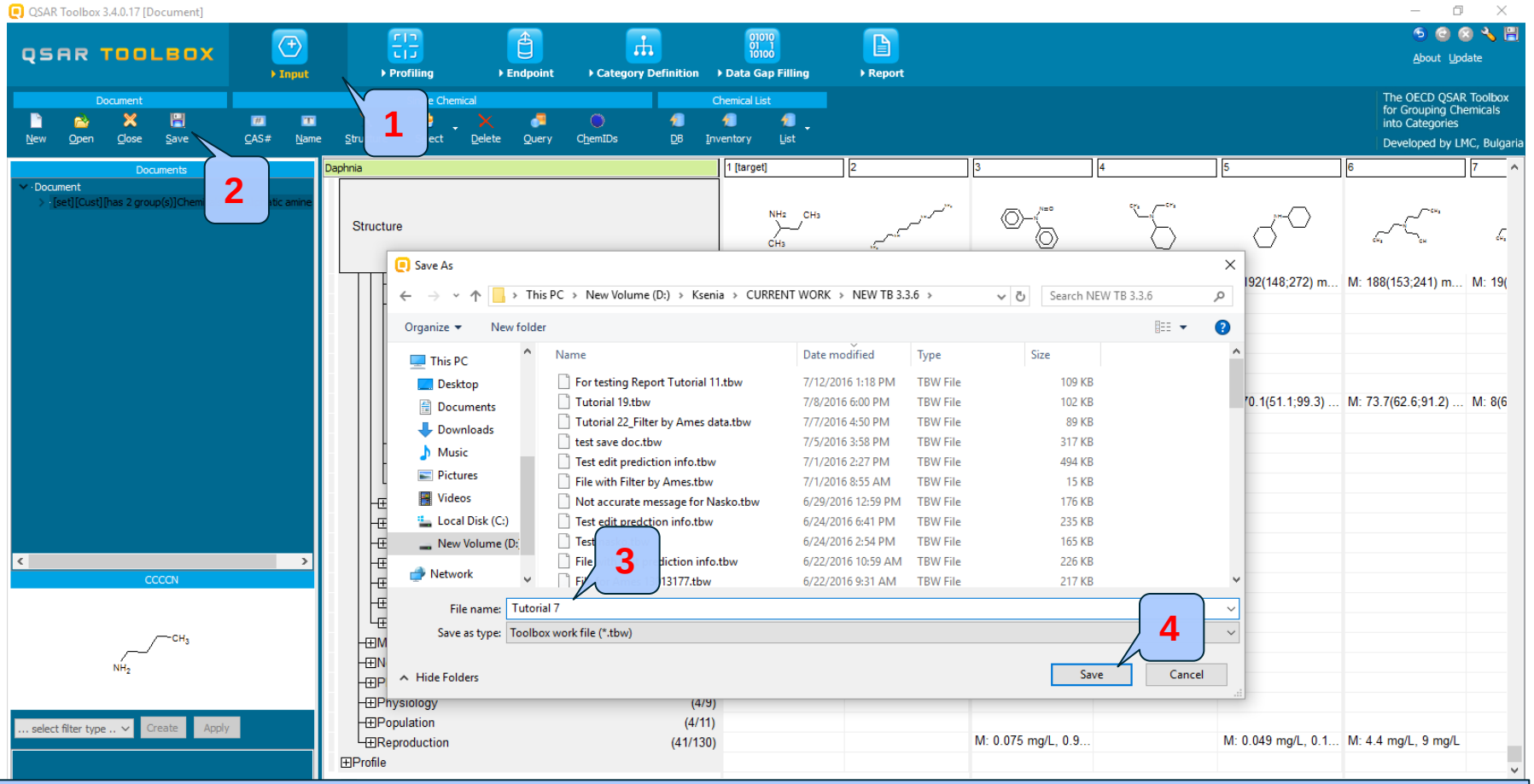
- Background
- Category evaluation – overview
  - Phase I. Evaluation of Category & Data Gap Filling
  - Phase II. Extension of the Category& Data gap Filling
- Save the prediction result

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).

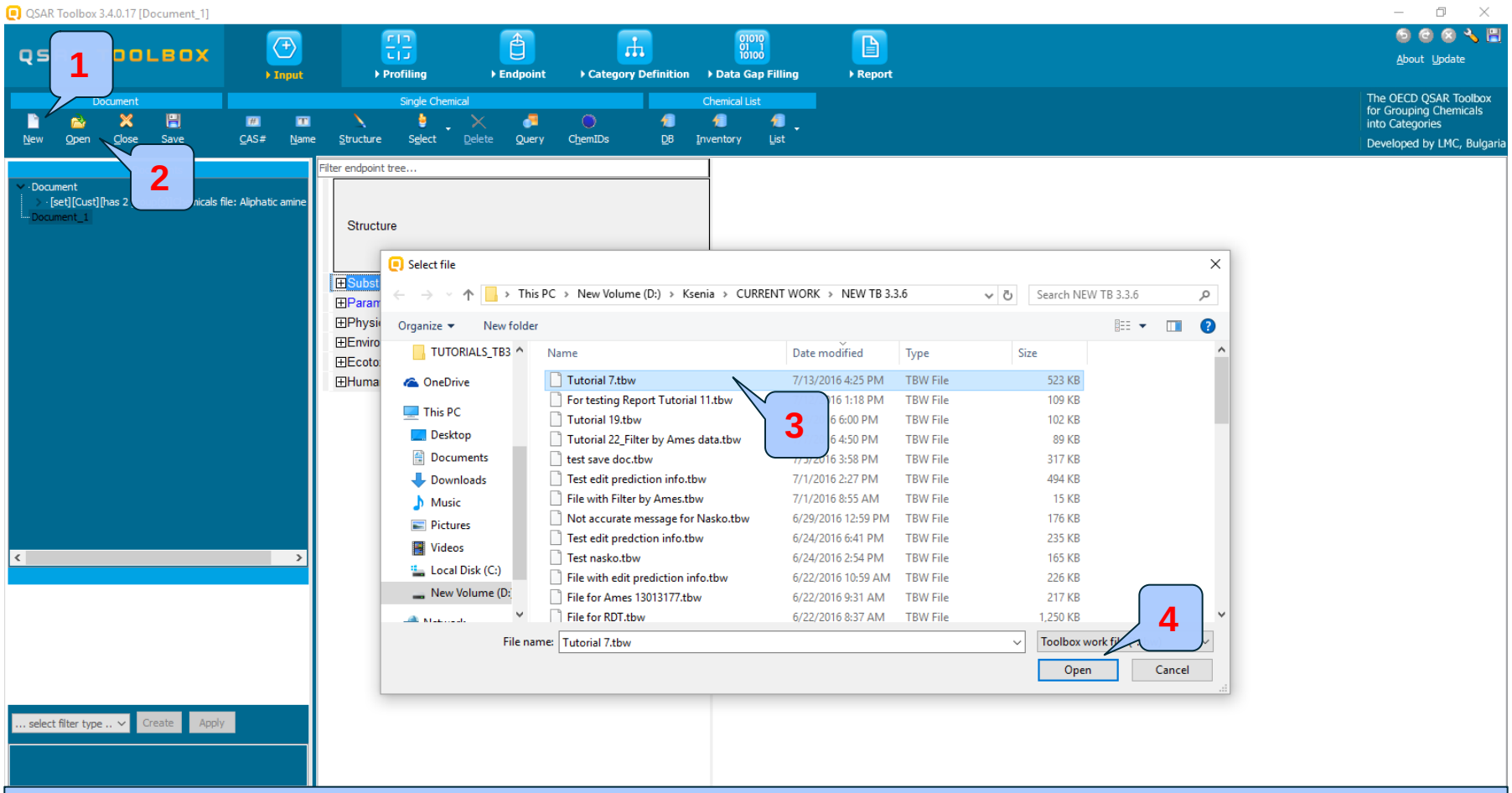
- Saving/Loading the file with TB prediction is shown on next screenshots

Saving the prediction result



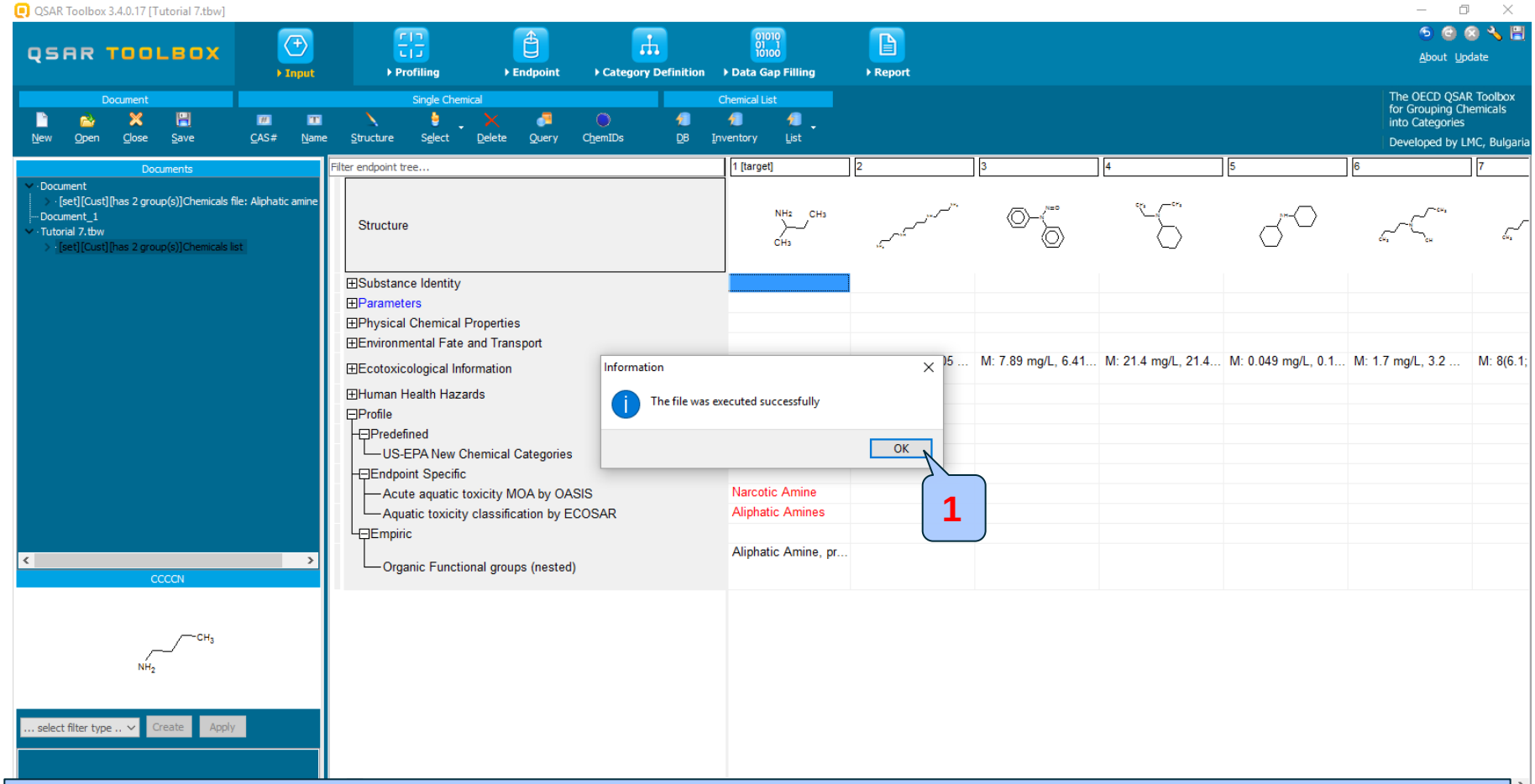
1. Go to Input section 2.**Click** on Save button 3. **Define** name of the file; 4. **Click** Save button

Open saved file



1. **Create** new document 2. **Click** Open; 3. **Find** and **select file**; 4. **Click** Open

Open saved file



1. The file is opened successfully 1. **Click** OK