QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.4.1

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

Outlook

- **Background**
- Keywords
- Category evaluation - overview

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for evaluating an ad-hoc category.
- You will learn several new functionalities which will 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
- **Keywords**
- Category evaluation - overview

Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options (e.g. Profiling)

WORKFLOW – the use, in combination, of the different modules (e.g. prediction workflow: from input to report)

PROFILER - algorithm (rule set) for the identification of specific features of the chemicals. Several types of profilers are available, such as structural (e.g. Organic functional groups), mechanistic (e.g. Protein binding by OECD) and endpoint-specific (e.g. in vitro in vitro mutagenicity (Ames test) alerts by ISS) profilers.

ALERT - the profilers consist of sets of rules or alerts. Each of the rules consists of a set of queries. The queries could be related to the chemical structure, physicochemical properties, experimental data, comparison with the target or list with substances and external queries from other predefined profilers (reference queries).

CATEGORY – "group" of substances sharing same characteristics (e.g. the same functional groups or mode of action). In a typical Toolbox workflow, it consists of the target chemical and its analogues gathered according to the selected profilers

ENDPOINT TREE – Endpoints are structured in a branched scheme, from a broader level (Phys-Chem properties, Environmental Fate and transport, Ecotoxicology, Human health hazard) to a more detailed one (e.g. EC3 in LLNA test under Human health hazard-Skin sensitization)

DATA MATRIX – Table reporting the chemical(s) and data (experimental results, profilers outcomes, predictions). Each chemical is in a different column and each data in a different row

Outlook

- Background
- Keywords
- **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 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
- Keywords
- Category evaluation – overview
  - **Phase I. Evaluation of Category & Data Gap Filling**
    - *Step 1: 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
- Keywords
- 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***

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* from example directory).
- 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 available with TB installation, located at C:\Program Files (x86)\Common Files\QSAR Toolbox 4.4\Config\Examples

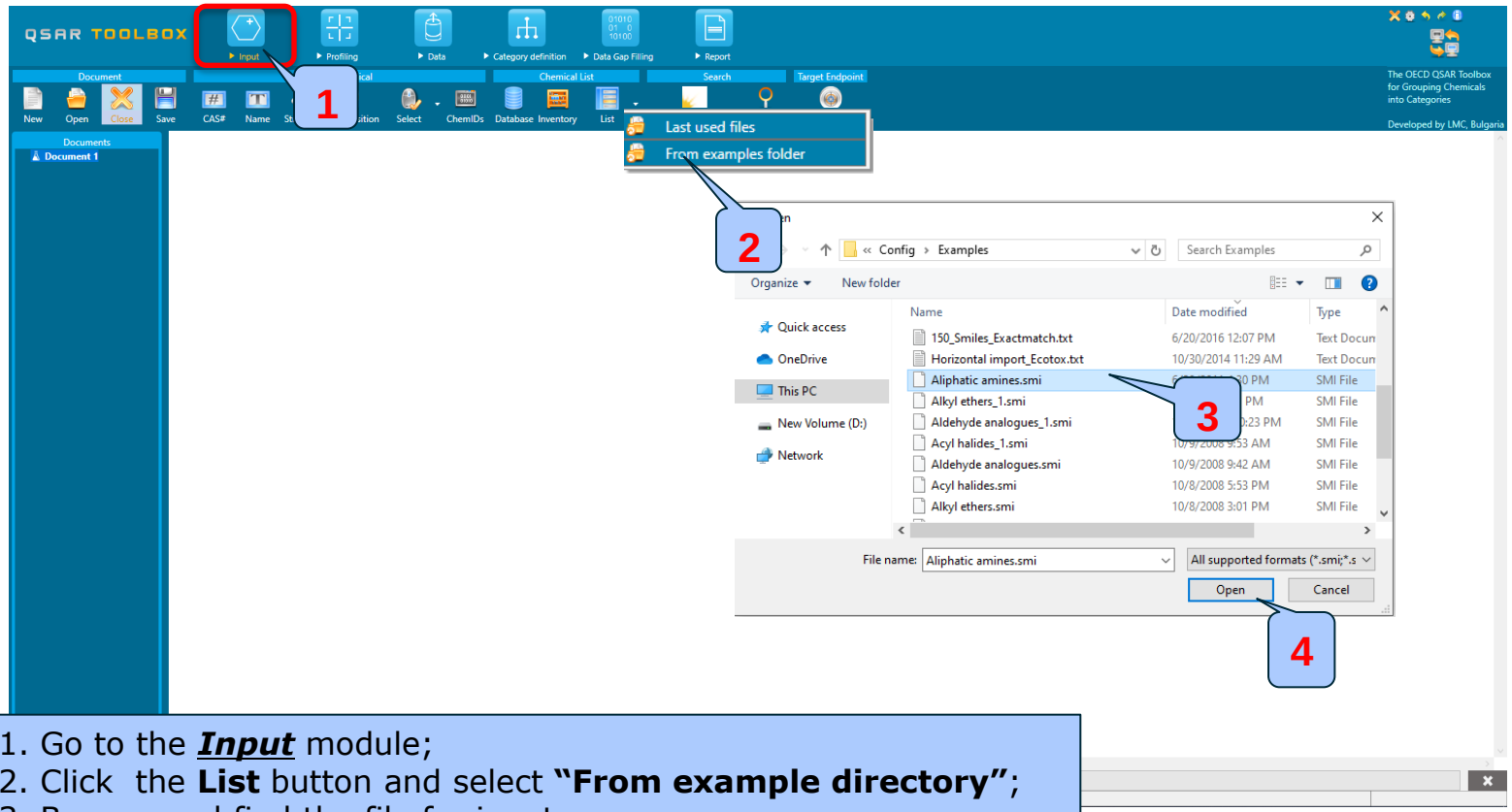
Outlook

- Background
- Keywords
- 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. Go to the ***Input*** module;
2. Click the **List** button and select **"From example directory"**;
3. Browse and find the file for input;
4. Click **Open**.

Phase I: Evaluation of Category & Data Gap Filling

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

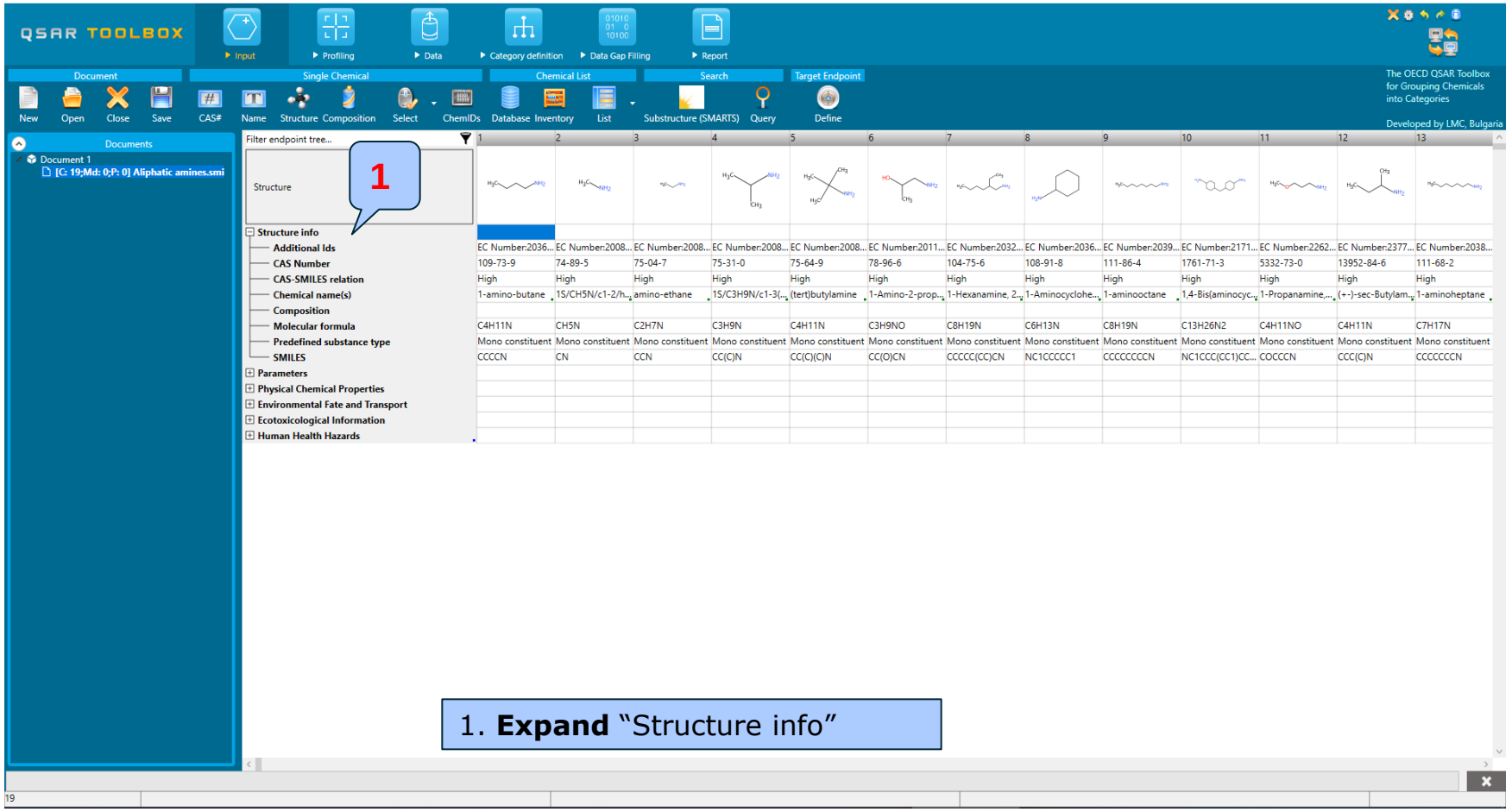
Workflow/Input

- You have now inserted your chemical list into the system.
- Click on the box next to "Structure info"; 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

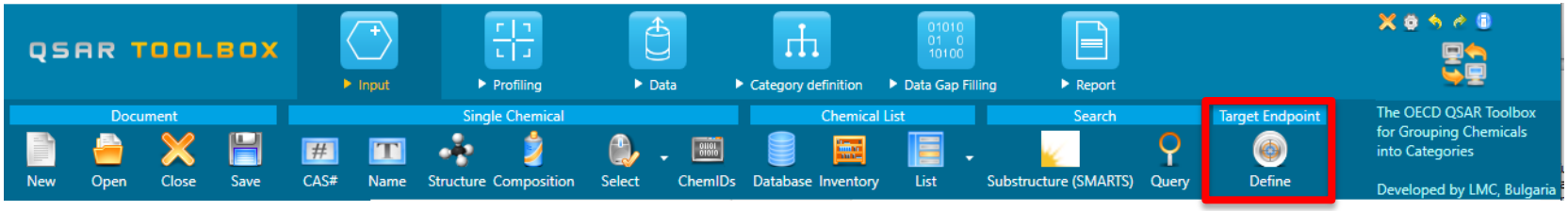


Outlook

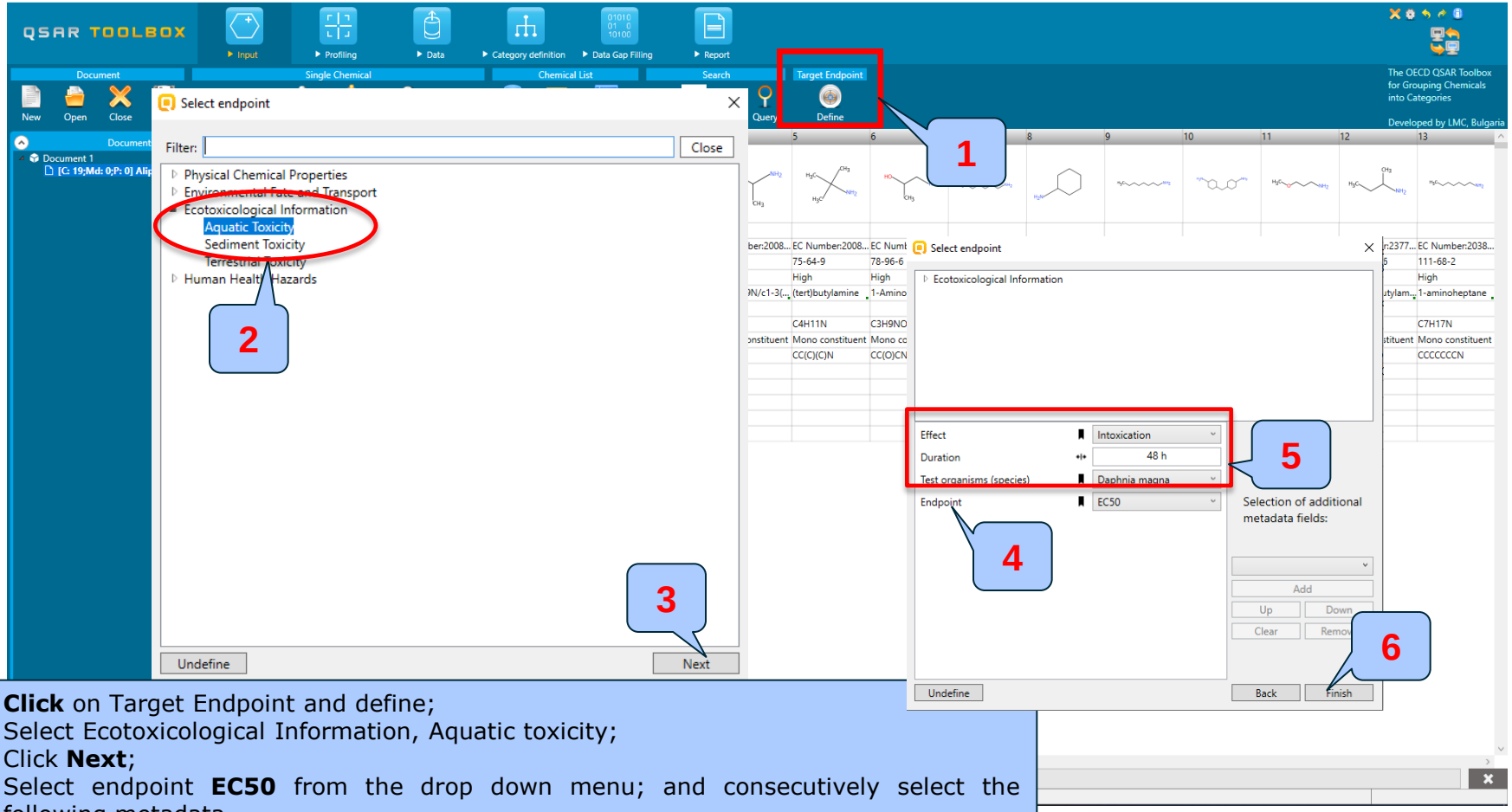
- Background
- Keywords
- Category evaluation – overview
  - **Phase I. Evaluation of Category & Data Gap Filling**
    - *Step 1: Investigating the structural consistency of an ad hoc category.*
      - *Case study*
      - ***Workflow***
        - *Input*
        - ***Define target endpoint***

Input
Define target endpoint

- Defining of the endpoint allows entering the endpoint of interest e.g. EC3, LC50, gene mutation etc., along with specific metadata information. Based on the metadata, relevancy of the profiles and databases is provided expressed in different highlighting:
  - In green are highlighted the most suitable profilers related to the endpoint and databases including data for the defined target endpoint, while
  - In orange are colored profilers which are plausible with respect to the defined target endpoint.



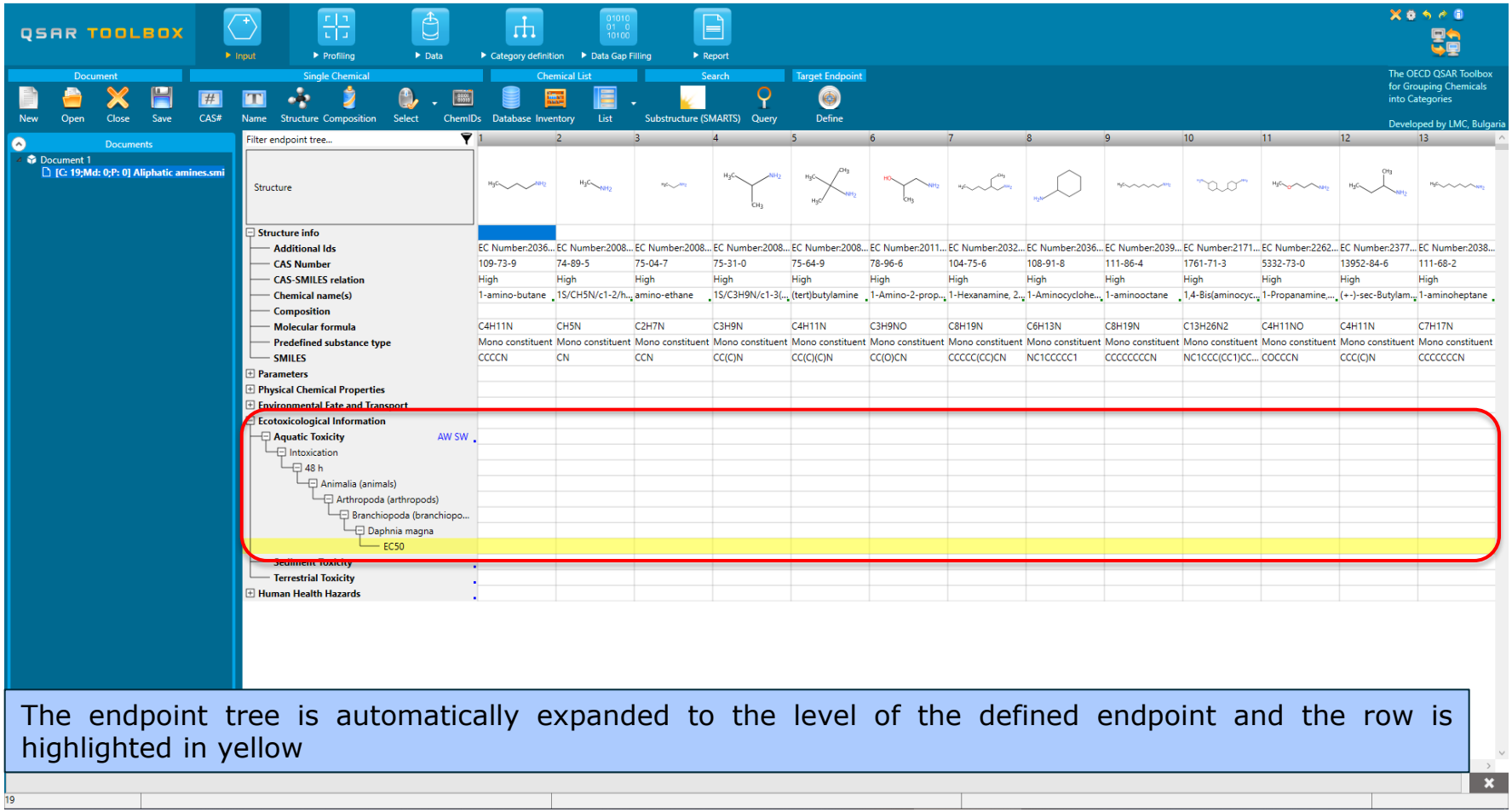
Define target endpoint
Overview



1. **Click** on Target Endpoint and define;
2. Select Ecotoxicological Information, Aquatic toxicity;
3. Click **Next**;
4. Select endpoint **EC50** from the drop down menu; and consecutively select the following metadata
5. Test organism(spices) – **Daphnia magna**; Duration – **48 h**; Effect – **Intoxication**;
6. Click **Finish.**

Define target endpoint

Overview



The endpoint tree is automatically expanded to the level of the defined endpoint and the row is highlighted in yellow

Phase I: Evaluation of Category & Data Gap Filling

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

Workflow/Profiling

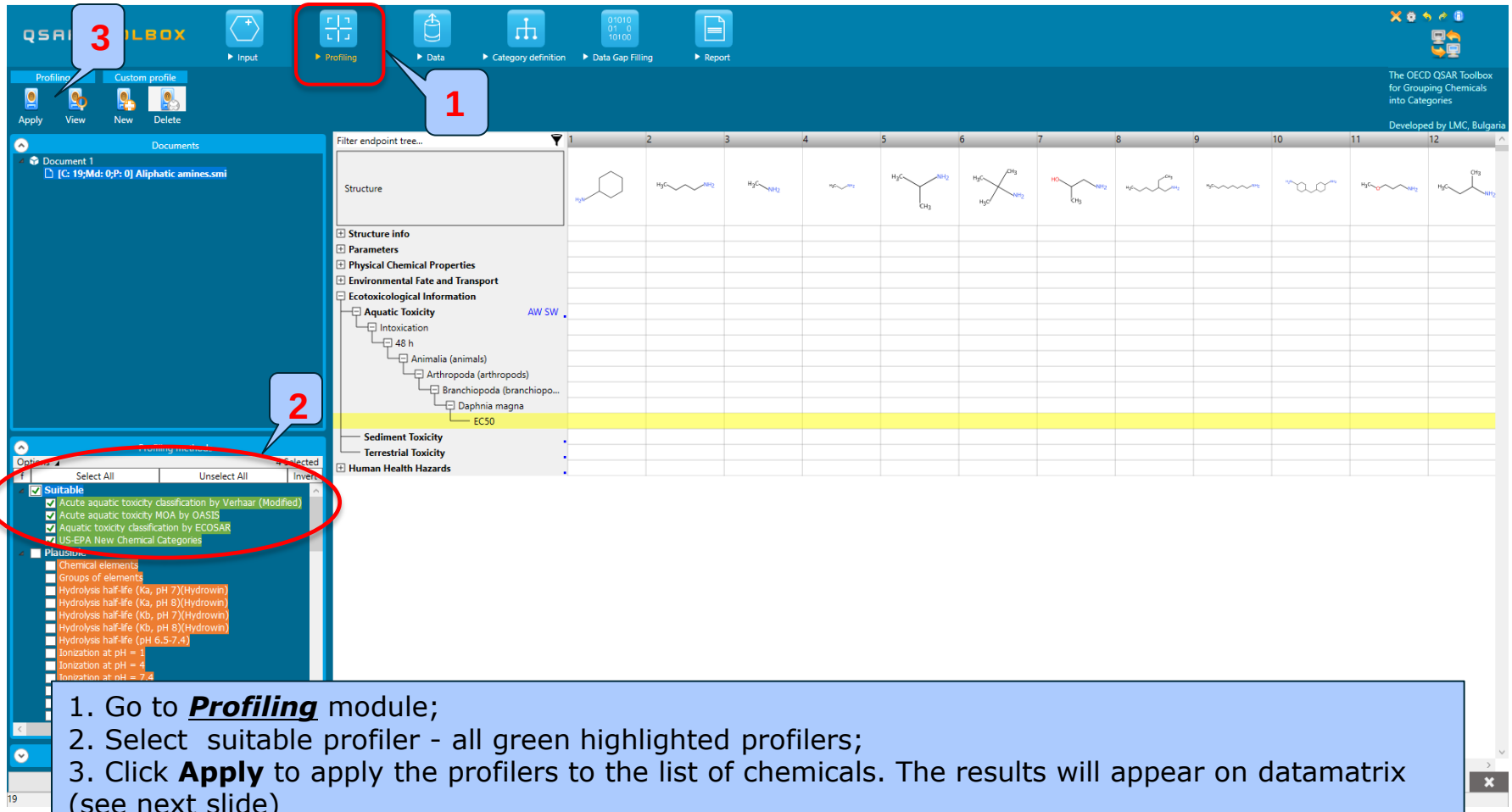
- The first step of the category evaluation according to Phase I is the consistency check of the category which starts with evaluation of the robustness with respect to structural functionalities. The following schemes could be used for this purpose:
  - US-EPA New Chemical Categories (hereafter cited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

Profiling

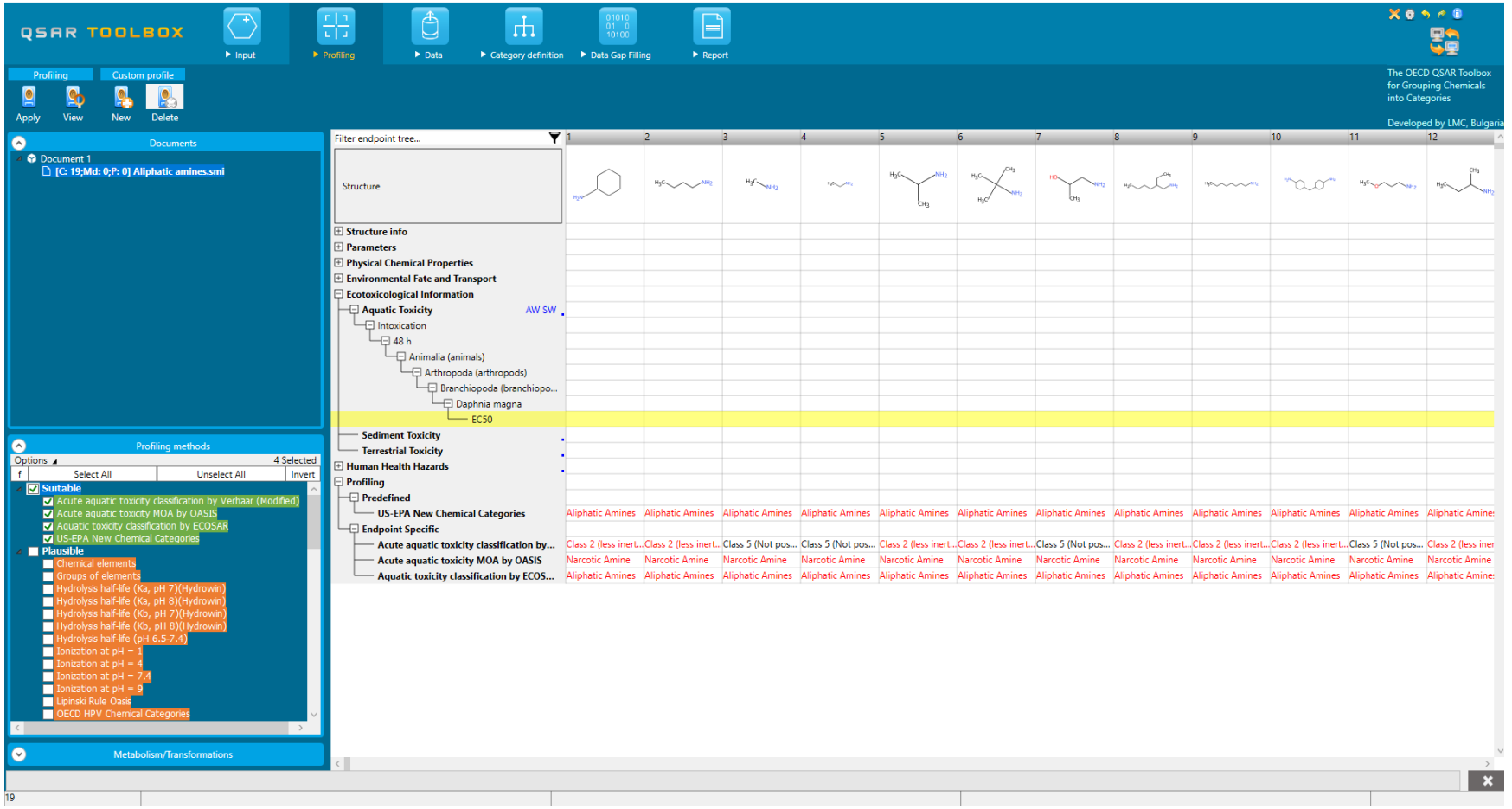


1. Go to ***Profiling*** module;
2. Select suitable profiler - all green highlighted profilers;
3. Click **Apply** to apply the profilers to the list of chemicals. The results will appear on datamatrix (see next slide)

Phase I: Evaluation of Category & Data Gap Filling

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

Profiling



Phase I: Evaluation of Category & Data Gap Filling

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

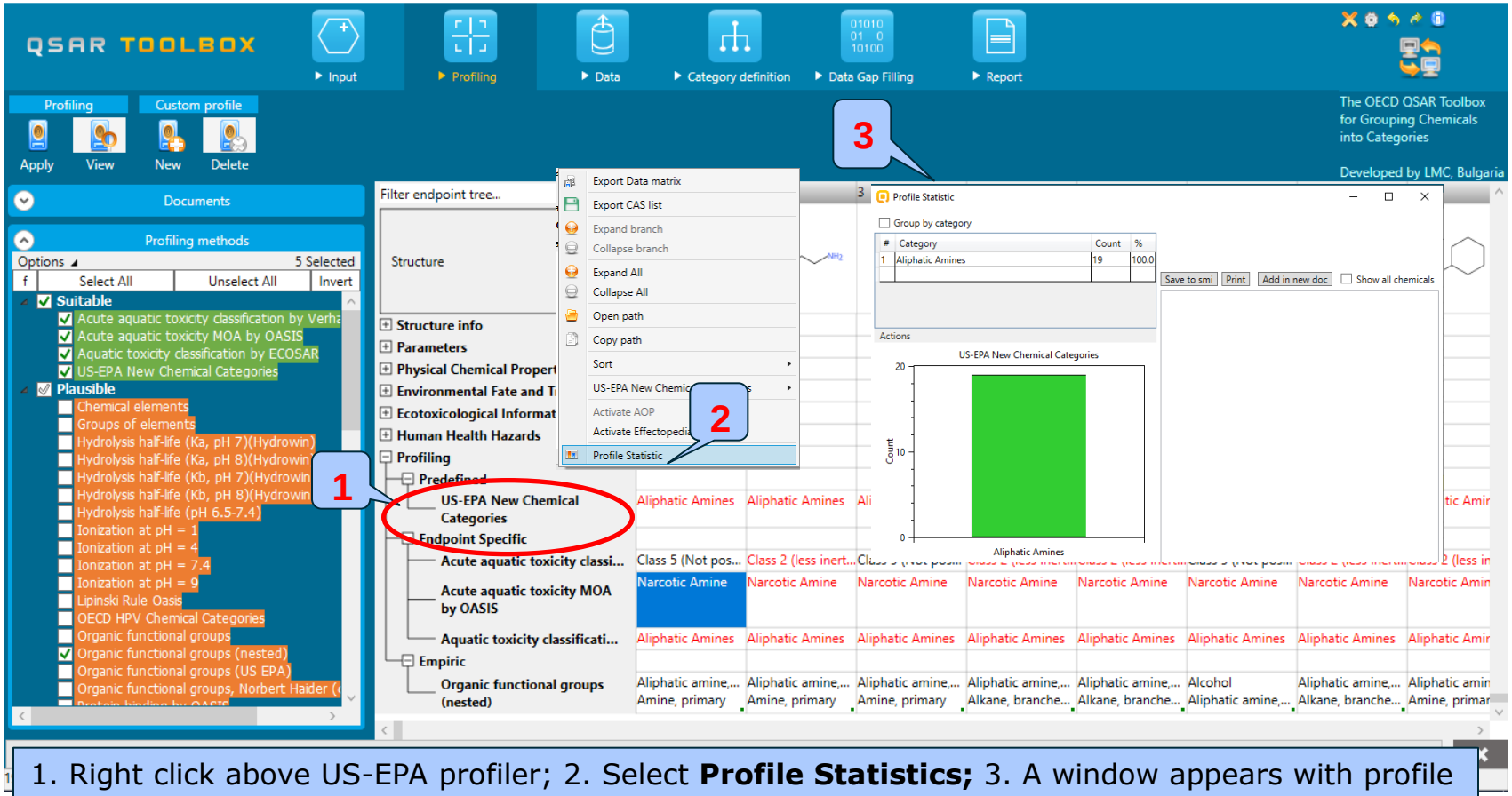
Workflow/Profiling

- 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 overall result for the list of chemicals can be seen by right clicking in the space near 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

Profiling

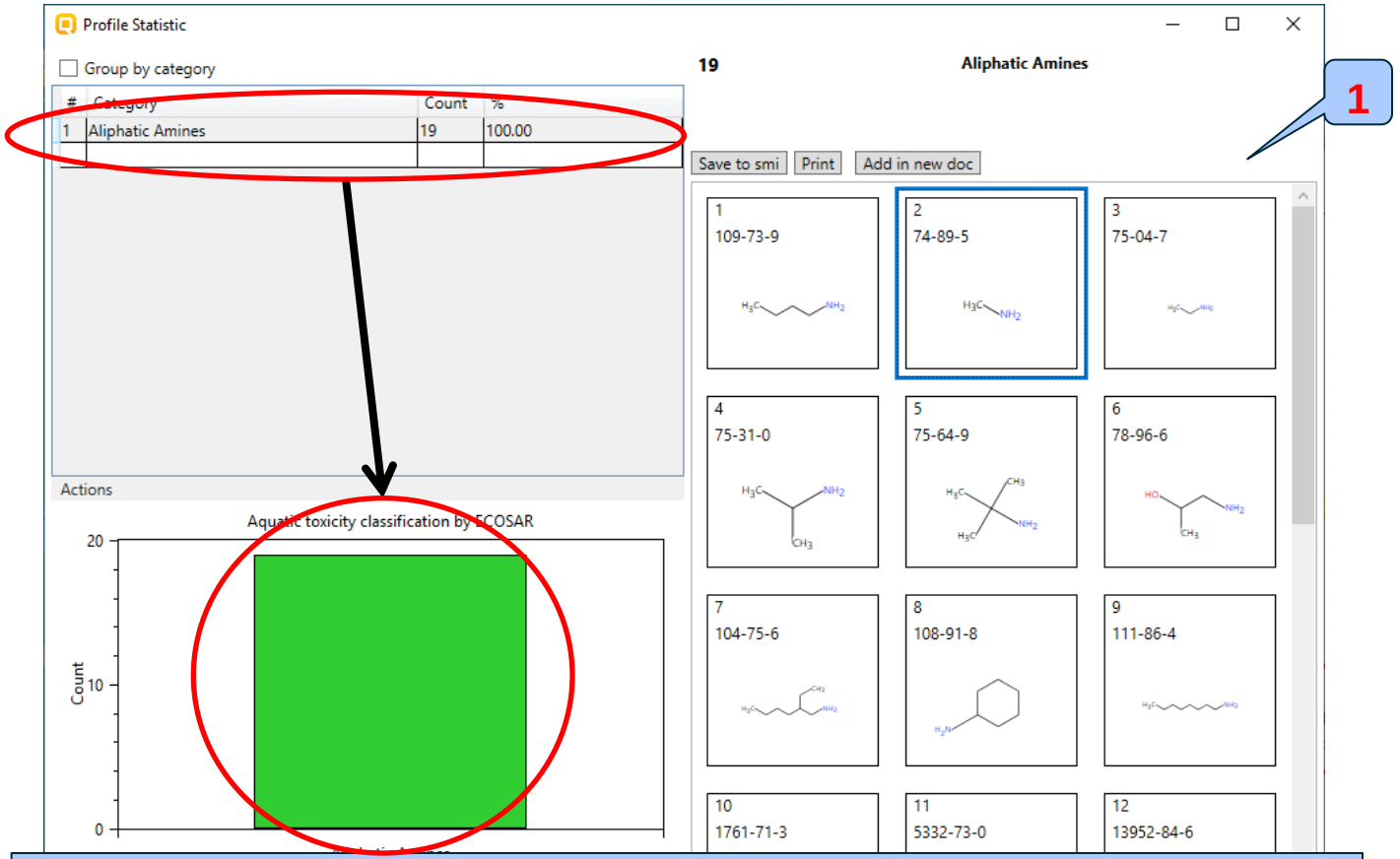


1. Right click above US-EPA profiler; 2. Select **Profile Statistics;** 3. A window appears with profile statistics for the list of chemicals across "US-EPA" profiler. All the chemicals are "Aliphatic amine"

Phase I: Evaluation of Category & Data Gap Filling

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

Profiling/Statistics according to ECOSAR

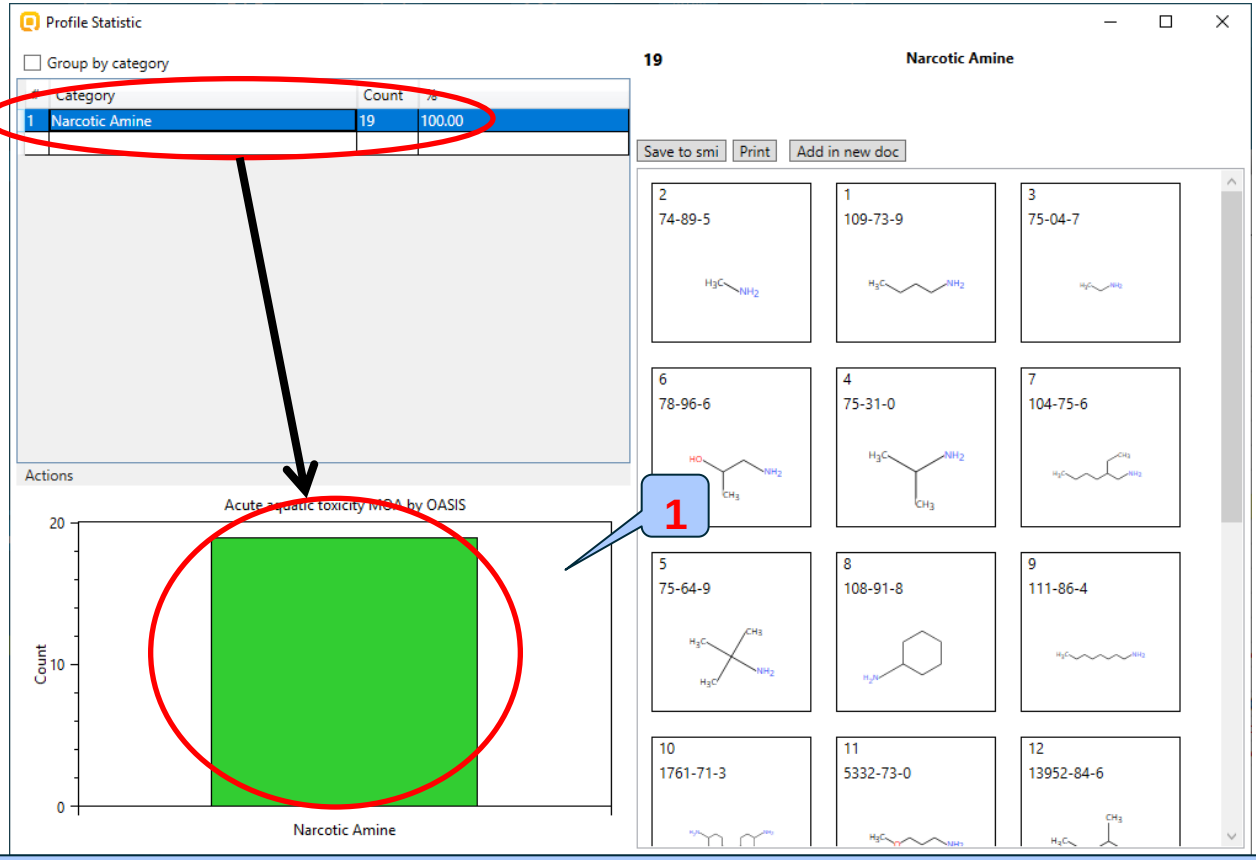


Repeat steps 1 and 2 from slide 24.
1. In this case all 19 chemicals are Aliphatic amines according to ECOSAR profiler.

Phase I: Evaluation of Category & Data Gap Filling

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

Profiling/Statistics according to Aquatic toxicity MOA by OASIS

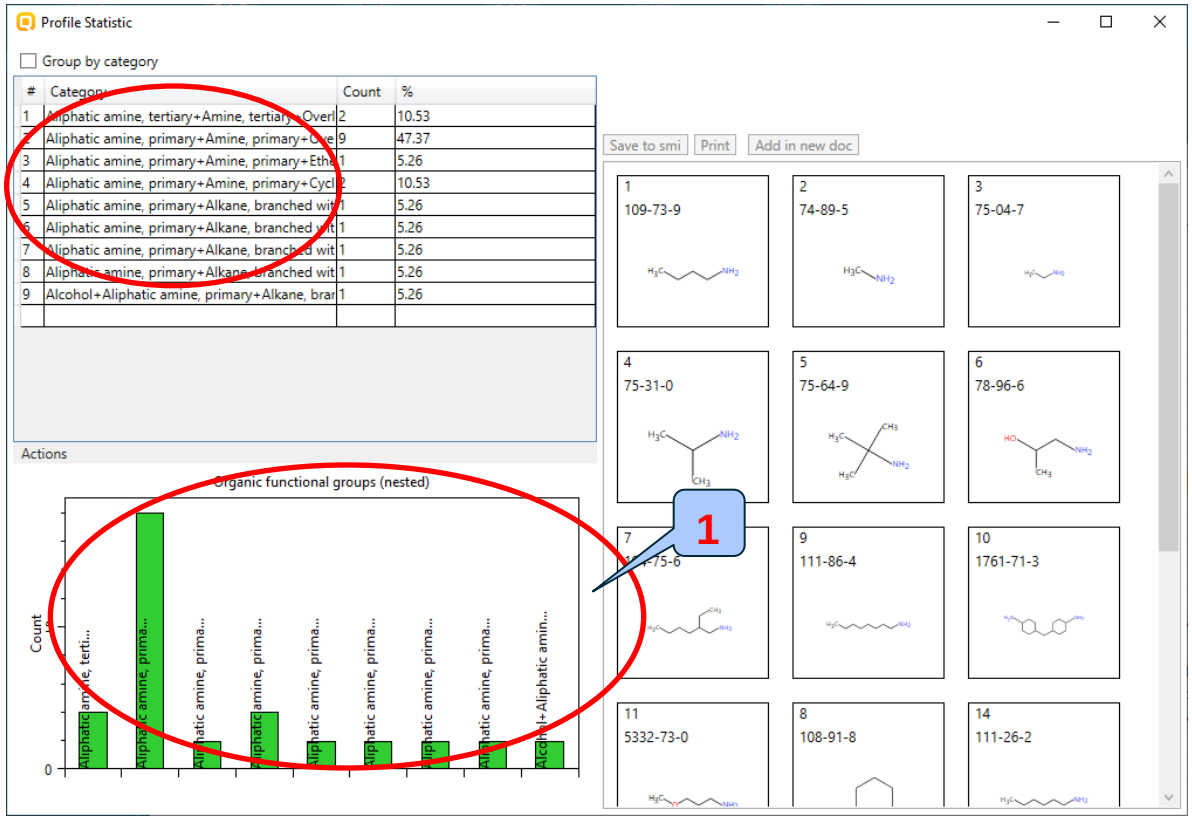


1. In this case all 19 chemicals are Narcotic amines according to MOA by OASIS profiling scheme

Phase I: Evaluation of Category & Data Gap Filling

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

Profiling/Statistics according to OFG (nested)



1.In this case all the chemicals in the category are categorized as "Aliphatic amine," according to OFG (nested) scheme

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- they all are aliphatic amines).**

Outlook

- Background
- Keywords
- 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.**

Phase I: Evaluation of Category & Data Gap Filling

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

In step 2, the applicability domain of the category will be investigated. The domain consist of two layers:

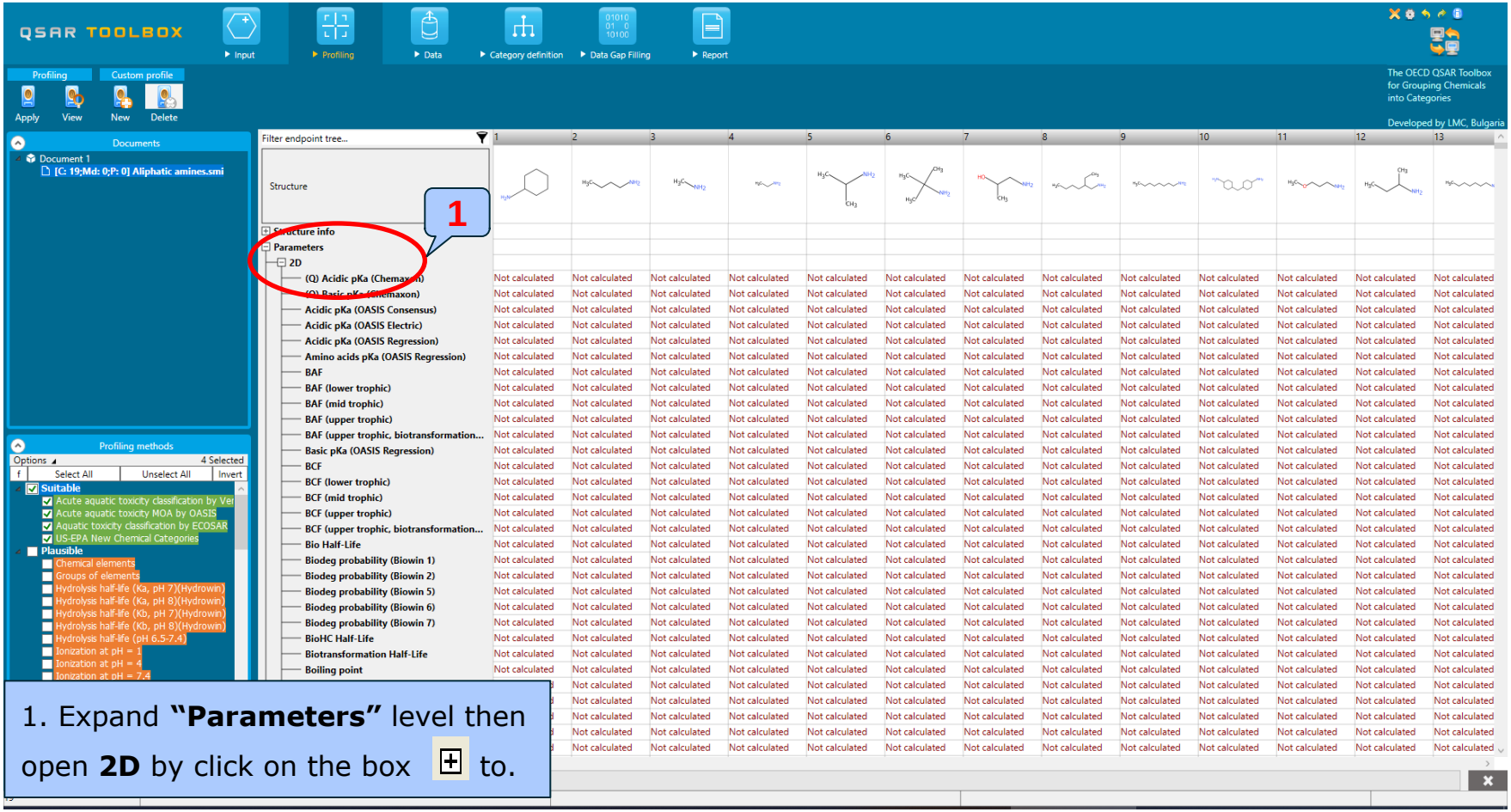
- Boundaries of structural functionalities
  - Aliphatic amines
- Parametric boundaries
  - log Kow (from 0.64 to 7.71)
  - Molecular weight (from 31 to 269 Da)
  - Water solubility (from $0.48x10^{-1}$ to $1x10^{6}$ mg/l)

You are now ready to extract the 2D and/or 3D parameters in order to check the parametric boundaries (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 parameters

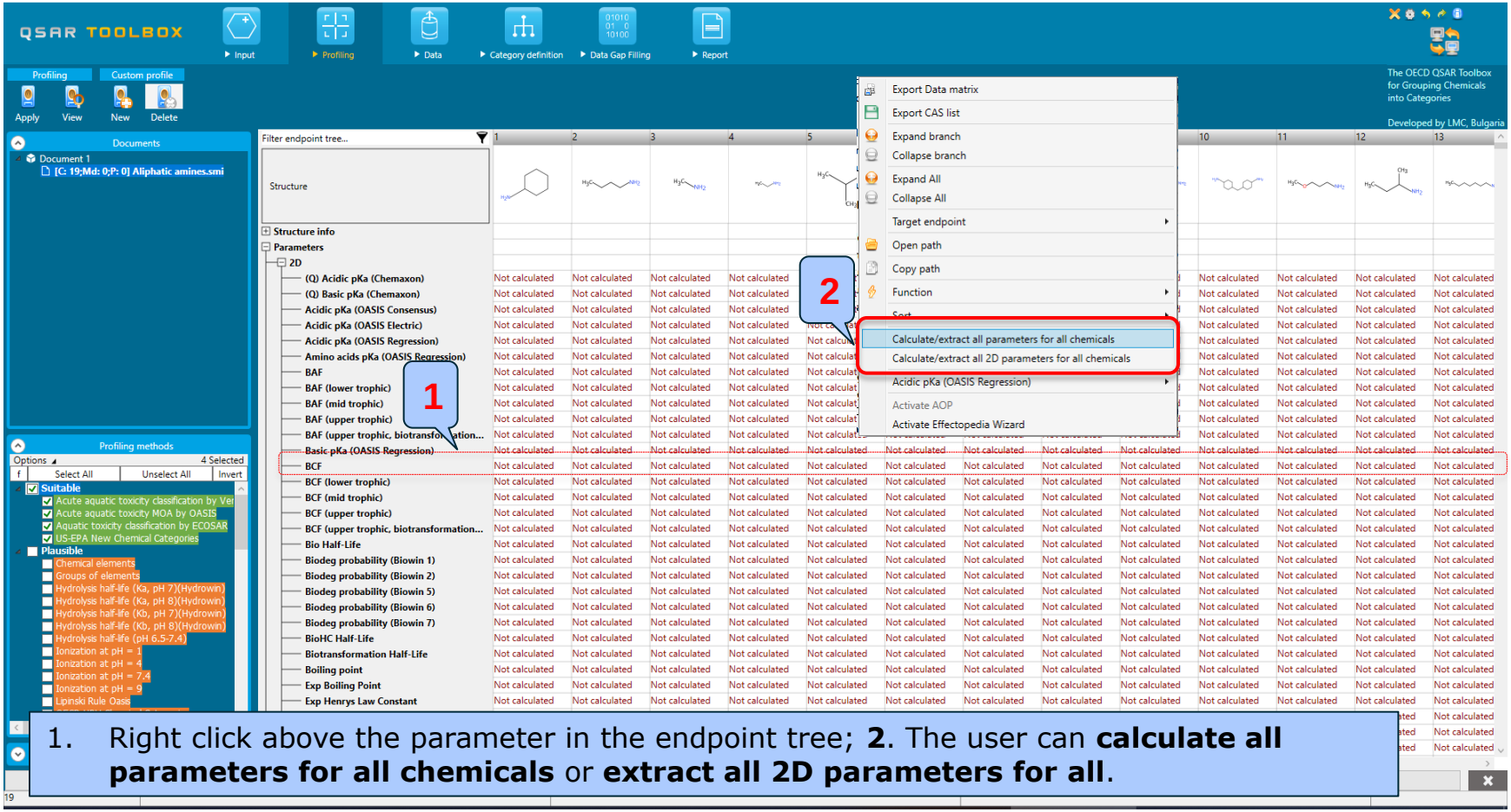


1. Expand **"Parameters"** level then open **2D** by click on the box ⊞ to.

Phase I: Evaluation of Category & Data Gap Filling

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

Extracting 2D and 3D parameters

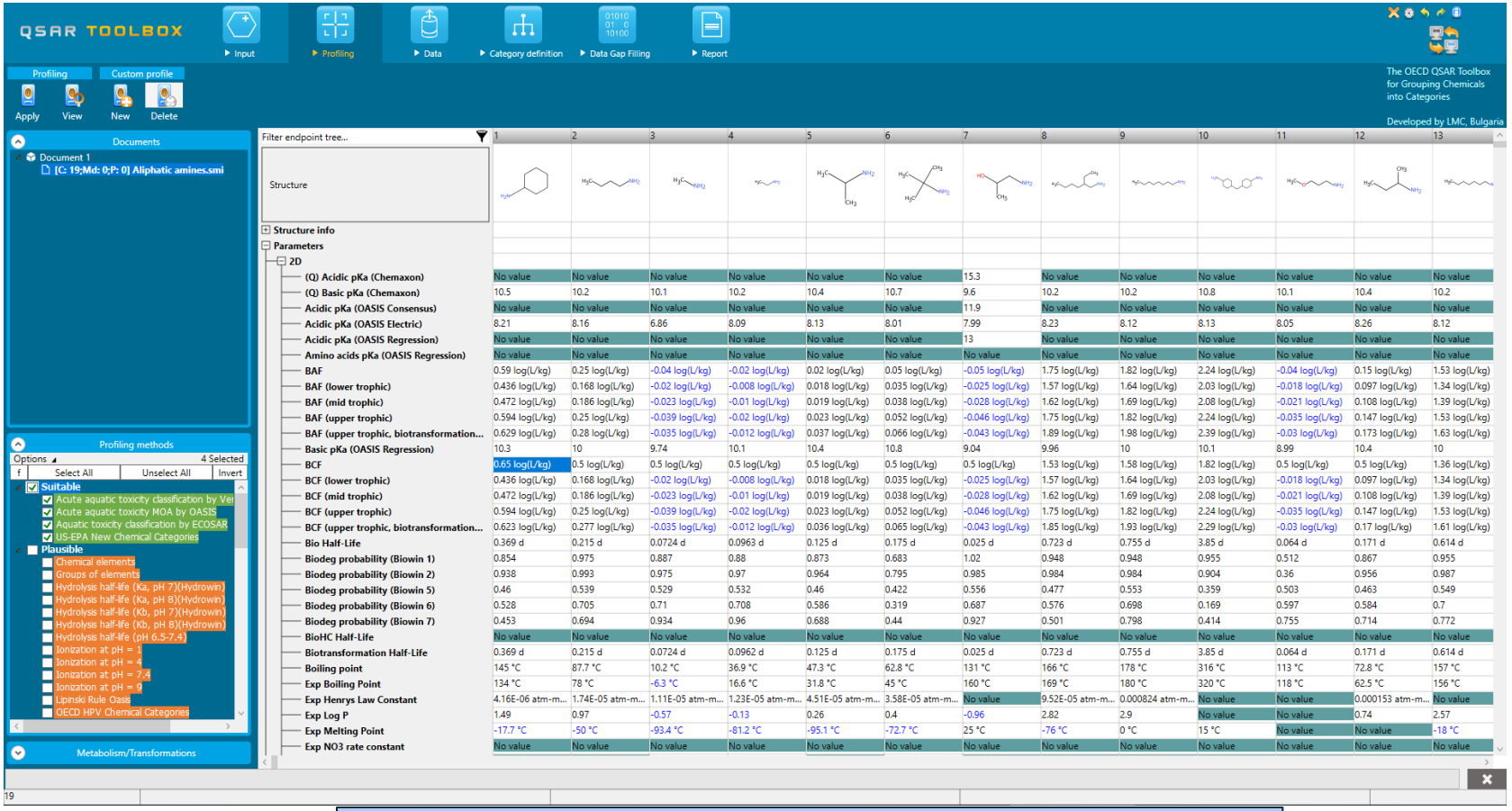


1. Right click above the parameter in the endpoint tree; **2**. The user can **calculate all parameters for all chemicals** or **extract all 2D parameters for all**.

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

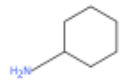
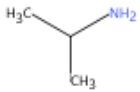
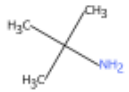
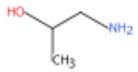
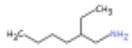


All calculated 2D parameters appears on data matrix.

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

Filter endpoint tree...	1	2	3	4	5	6	7	8
Structure								
Parameters								
2D								
log Kow	1.63	0.83	-0.64	-0.15	0.27	0.72	-1.19	2.73
Molecular Weight	99.2 Da	73.1 Da	31.1 Da	45.1 Da	59.1 Da	73.1 Da	75.1 Da	129 Da
Water Solubility	6.4E+04 mg/L	2.03E+05 mg/L	1E+06 mg/L	1E+06 mg/L	8.38E+05 mg/L	6.22E+05 mg/L	1E+06 mg/L	3.68E+03 mg/L

1

2

3

The boundaries of the three phys-chem parameters fall in the following ranges:

1. log Kow (from 0.64 to 7.71)
2. Molecular weight (from 31 to 269 Da)
3. Water solubility (from $0.48x10^{-1}$ to $1x10^{6}$ mg/l)

Outlook

- Background
- Keywords
- 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***

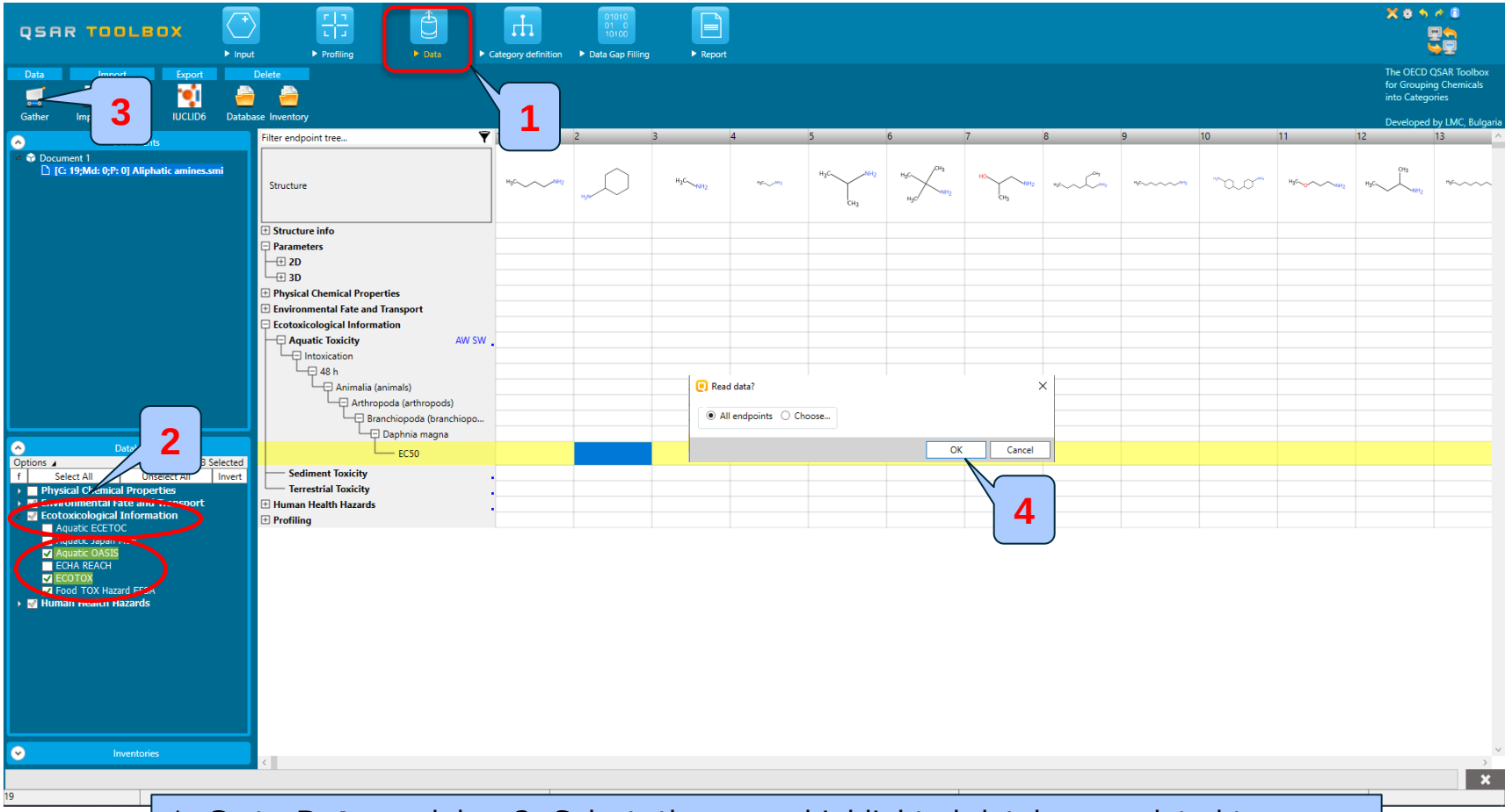
Phase I: Evaluation of Category & Data Gap Filling

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

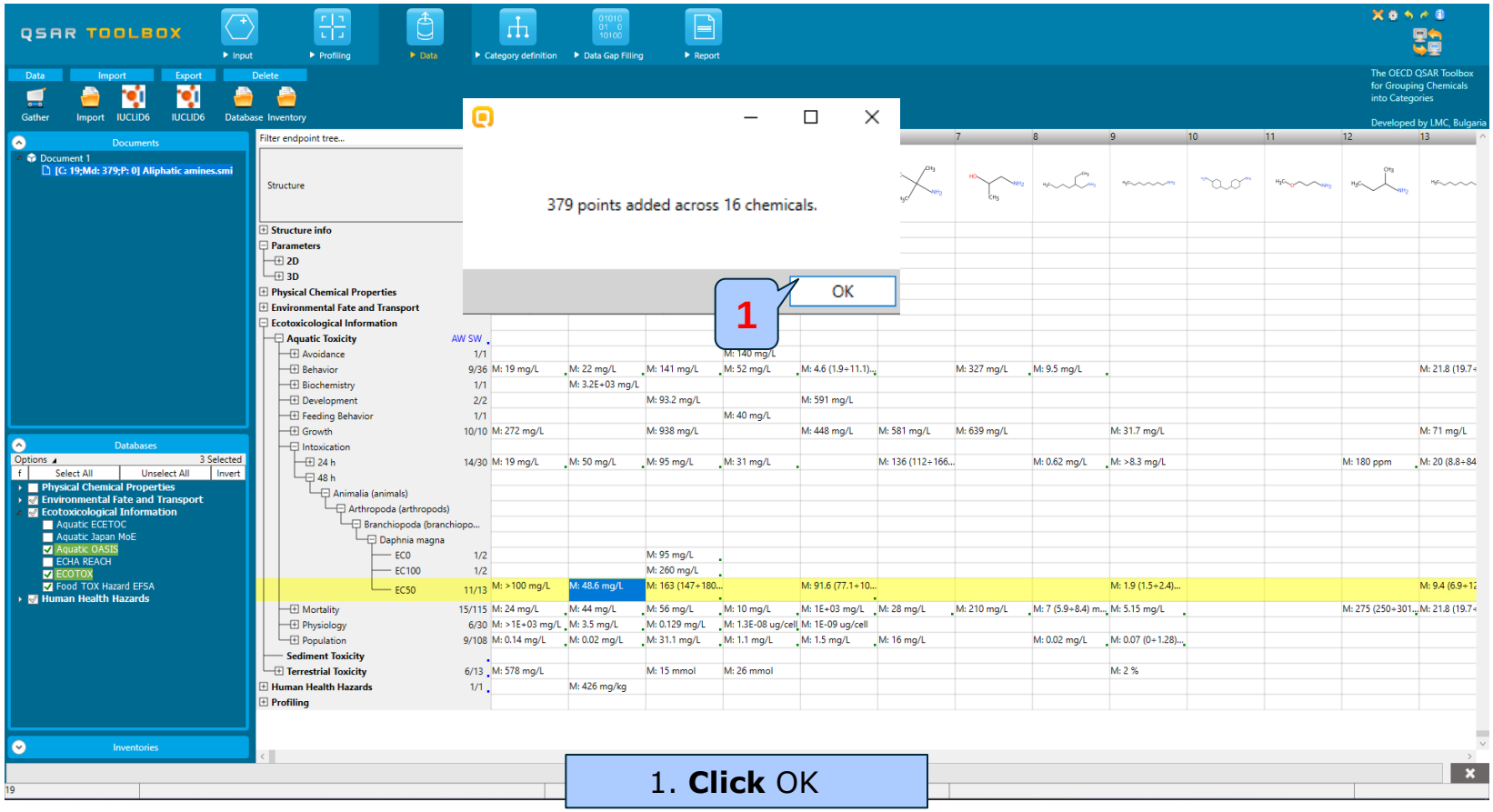


1. Go to ***Data*** module; 2. Select the green highlighted databases related to aquatic toxicity EC 50 Daphnia magna; 3. Click **Gather**; 4. Click **OK**.

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



Outlook

- Background
- Keywords
- 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

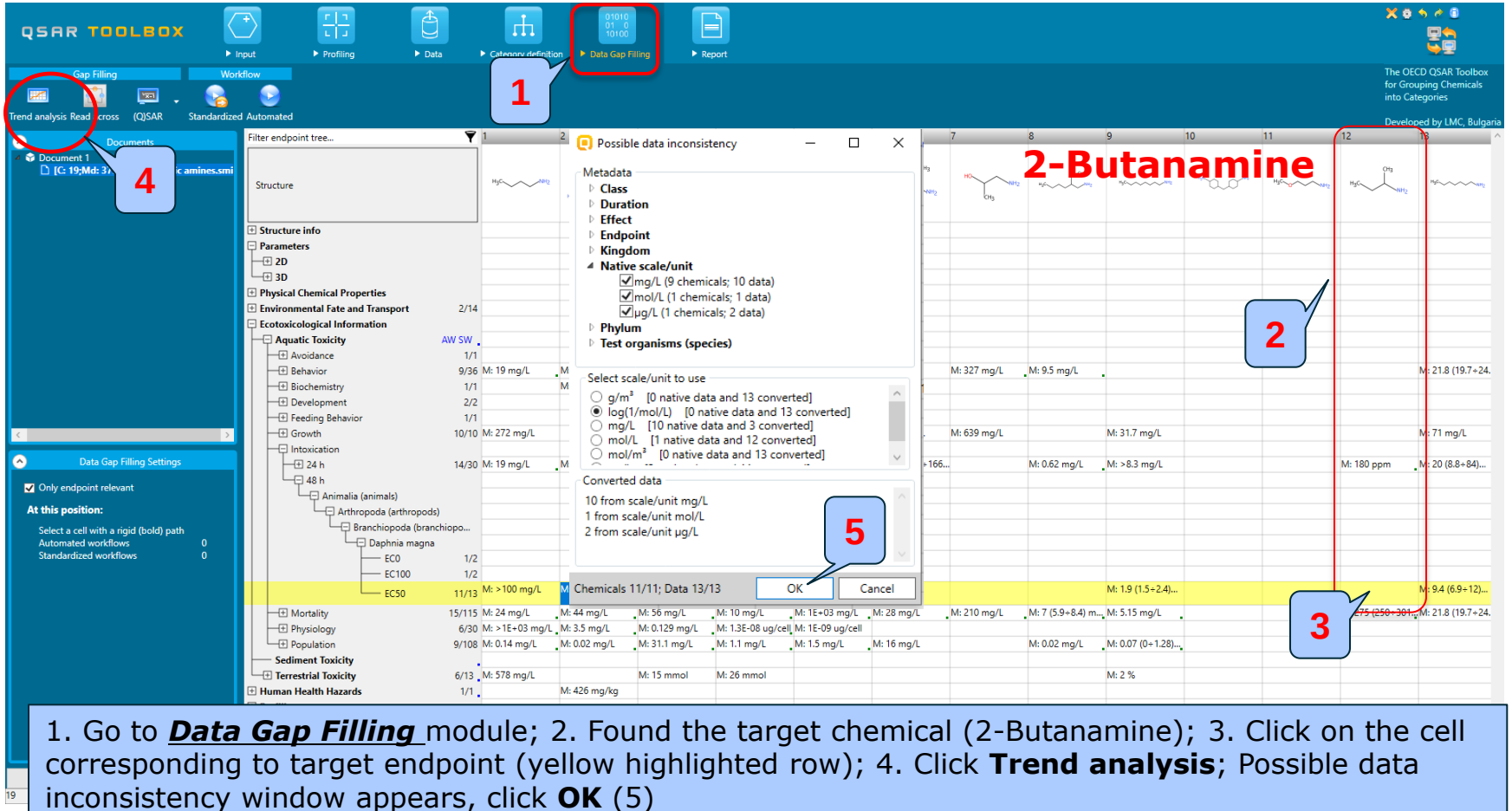
Step 4. Data gap filling for 2-Butanamine

- As mentioned before (slide 28) the category is structurally and mechanistically similar
- Hence the data gap could be filled in for analogues from the category. In this case we will fill the gap for 2-Butanamine (this will be our target chemical)
- Before proceeding with filling data gap you should navigate to the column corresponding to the target chemical (see next slide).

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine

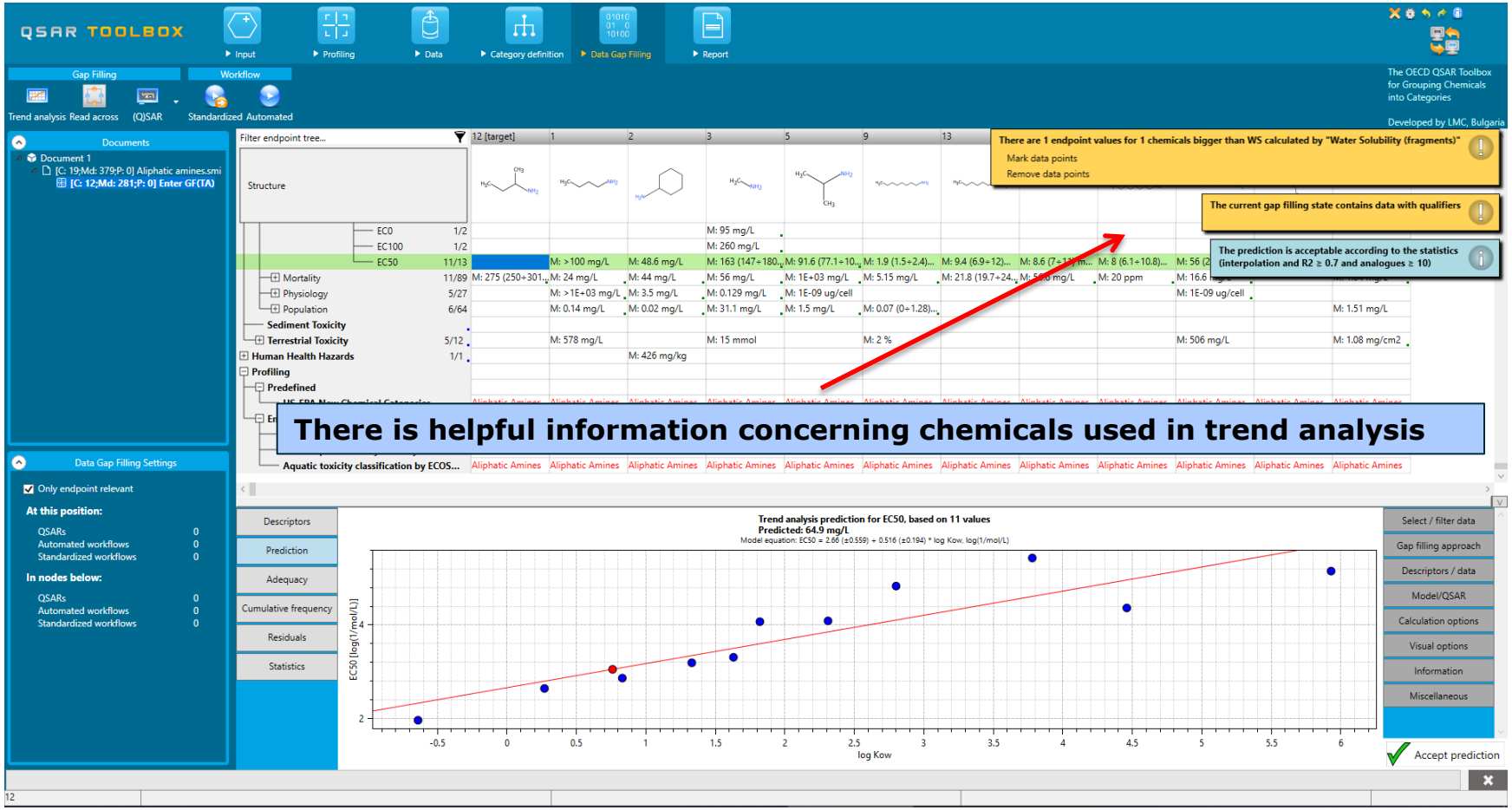
Navigate to the target chemical



1. Go to ***Data Gap Filling*** module; 2. Found the target chemical (2-Butanamine); 3. Click on the cell corresponding to target endpoint (yellow highlighted row); 4. Click **Trend analysis**; Possible data inconsistency window appears, click **OK** (5)

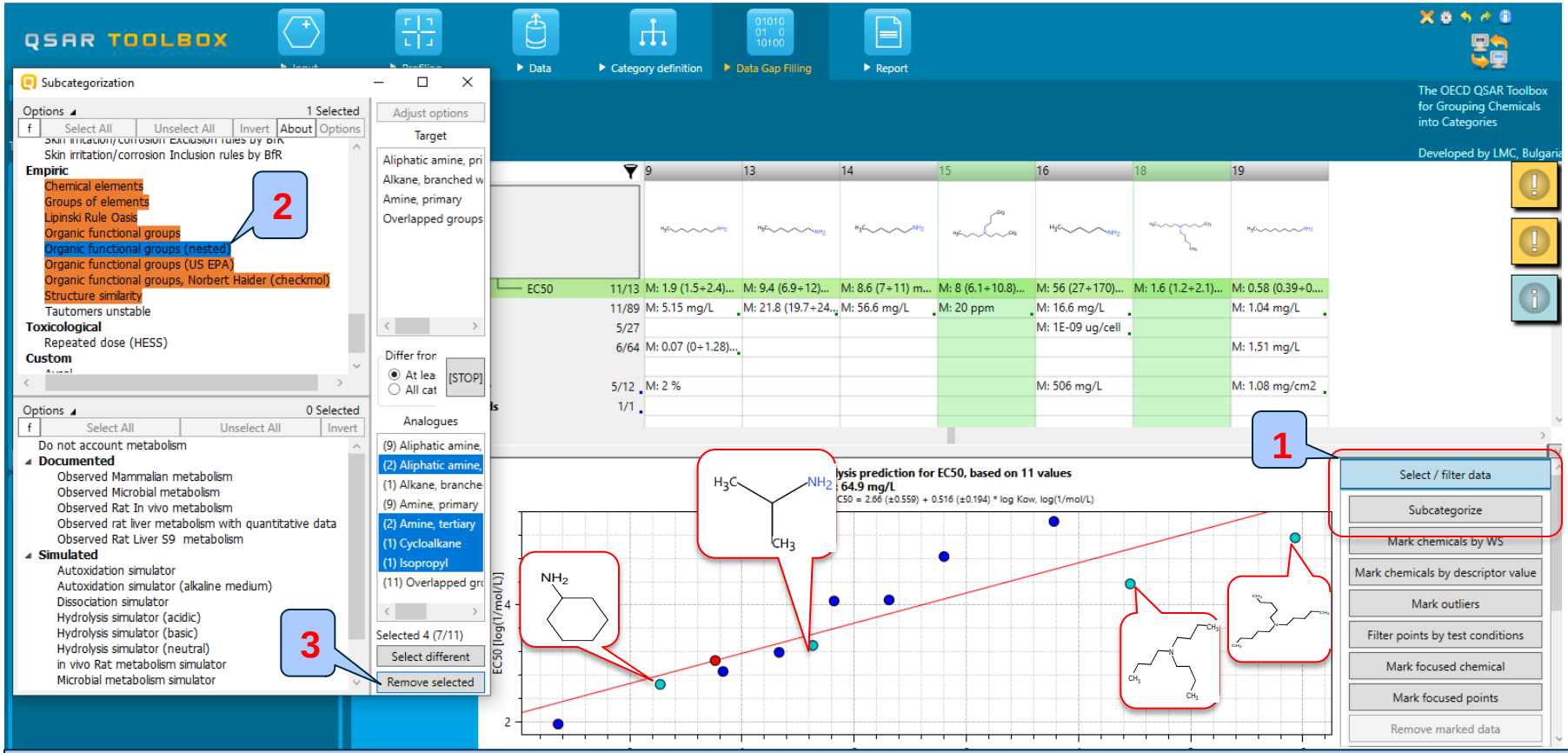
Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine



Phase I: Evaluation of Category & Data Gap Filling

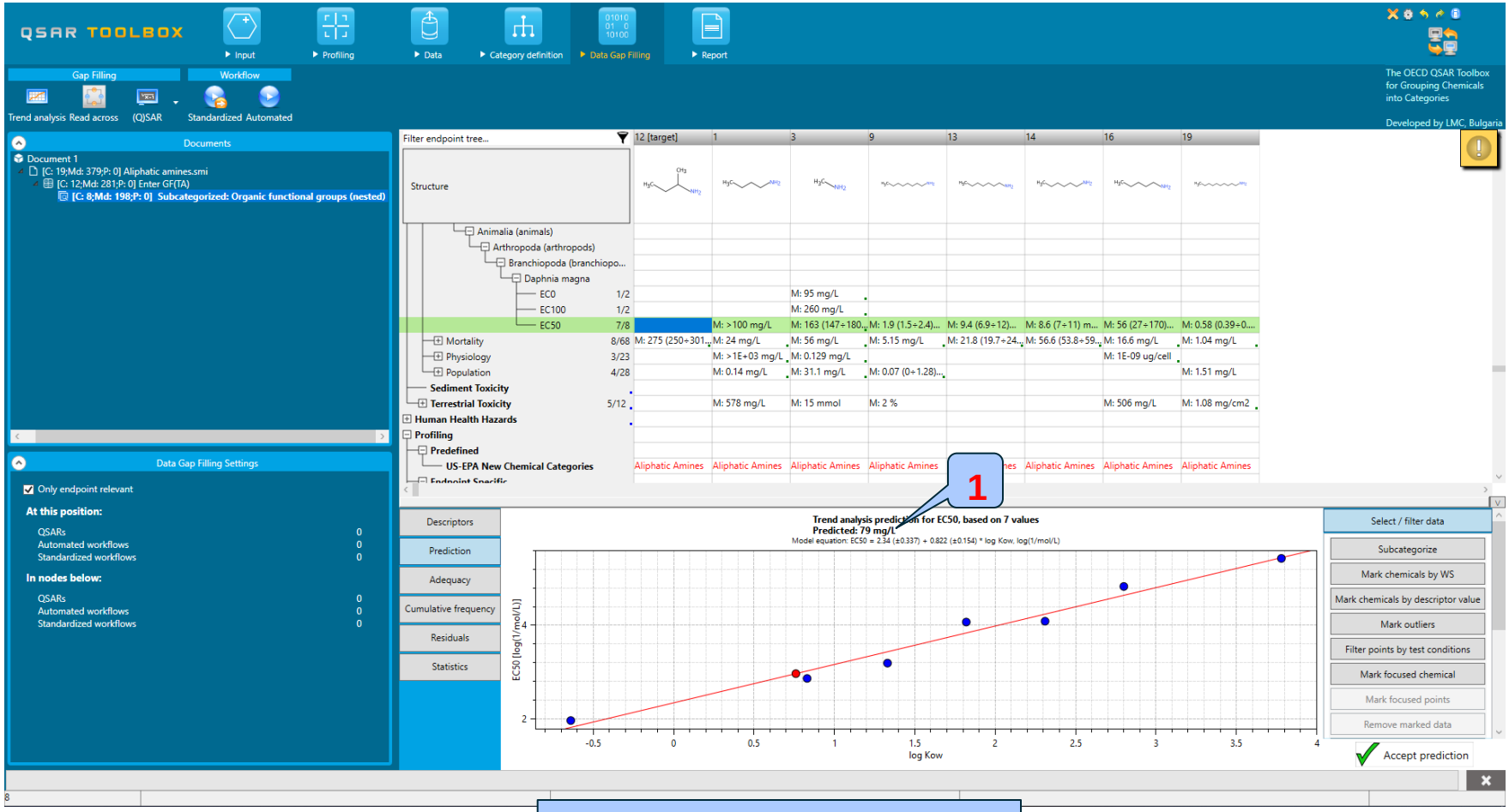
Step 4. Data Gap Filling for 2-Butanamine/*Subcategorize by OFG(nested)*



There are many profilers to be used for subcategorization. The most suitable profilers (the green ones) does not found any dissimilar chemical. In this respect the structure-based OFG (nested) profiler is used for refining the category: 1. Open **Select/filter data, Subcategorization**; 2. Click on **OFG (nested)** then 3. Click **Remove selected** button

Phase I: Evaluation of Category & Data Gap Filling

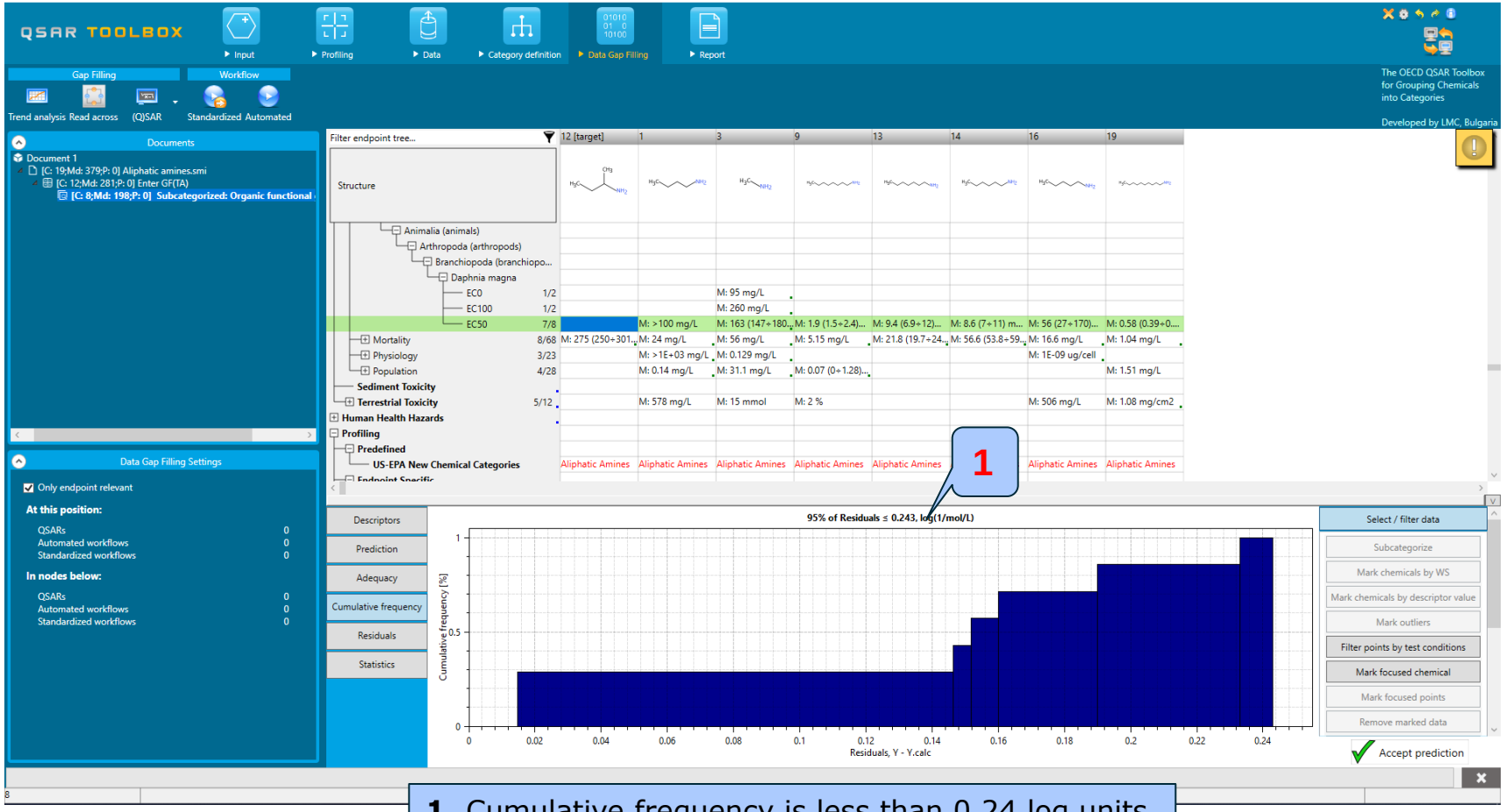
Step 4. Data Gap Filling for 2-Butanamine/*Predicted result*



1. Predicted result is 79.0 mg/l

Phase I: Evaluation of Category & Data Gap Filling

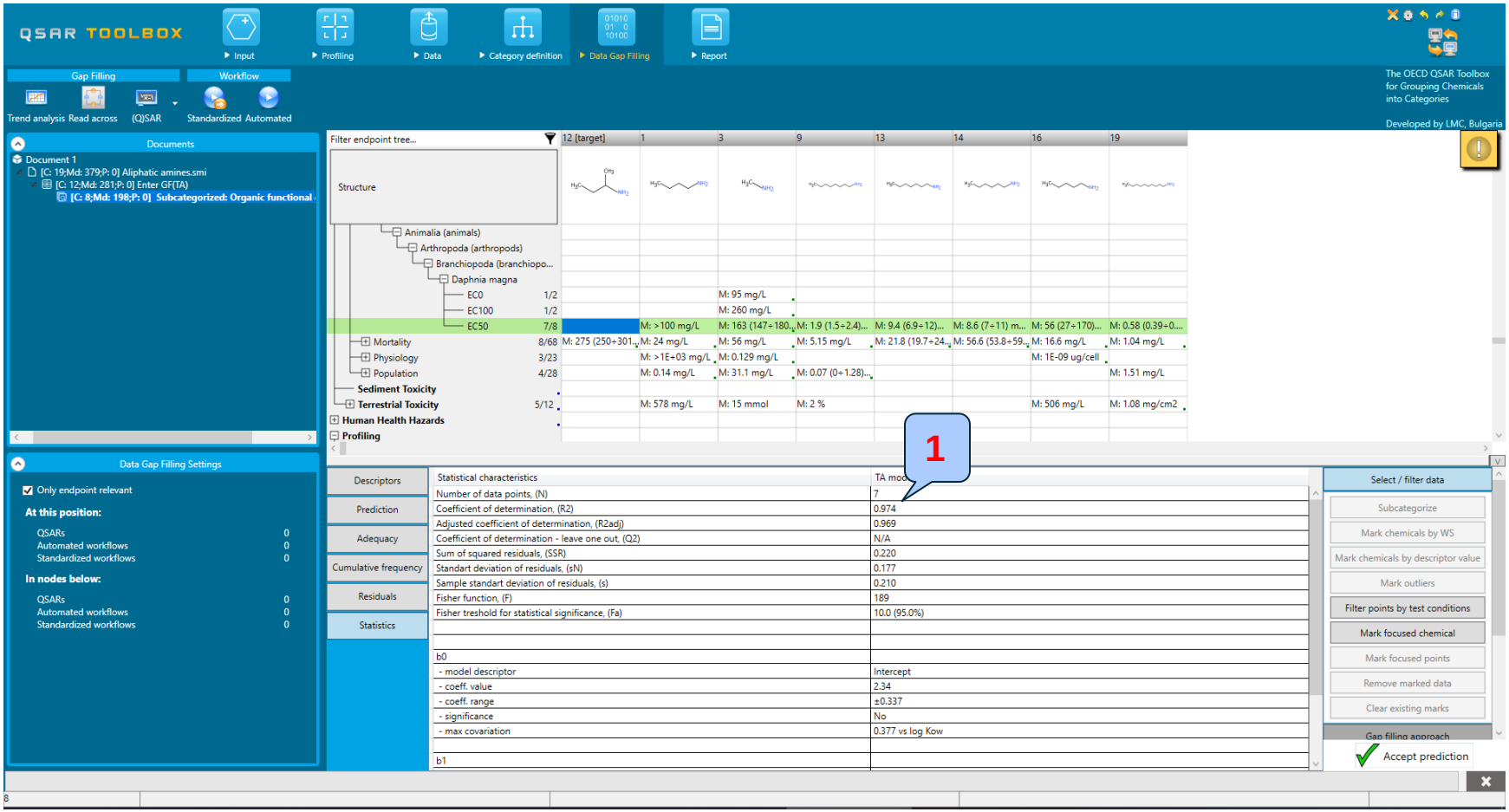
Step 4. Data Gap Filling for 2-Butanamine/*Cumulative frequency*



1. Cumulative frequency is less than 0.24 log units

Phase I: Evaluation of Category & Data Gap Filling

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



1. Coefficient of determination is higher - 0.974

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

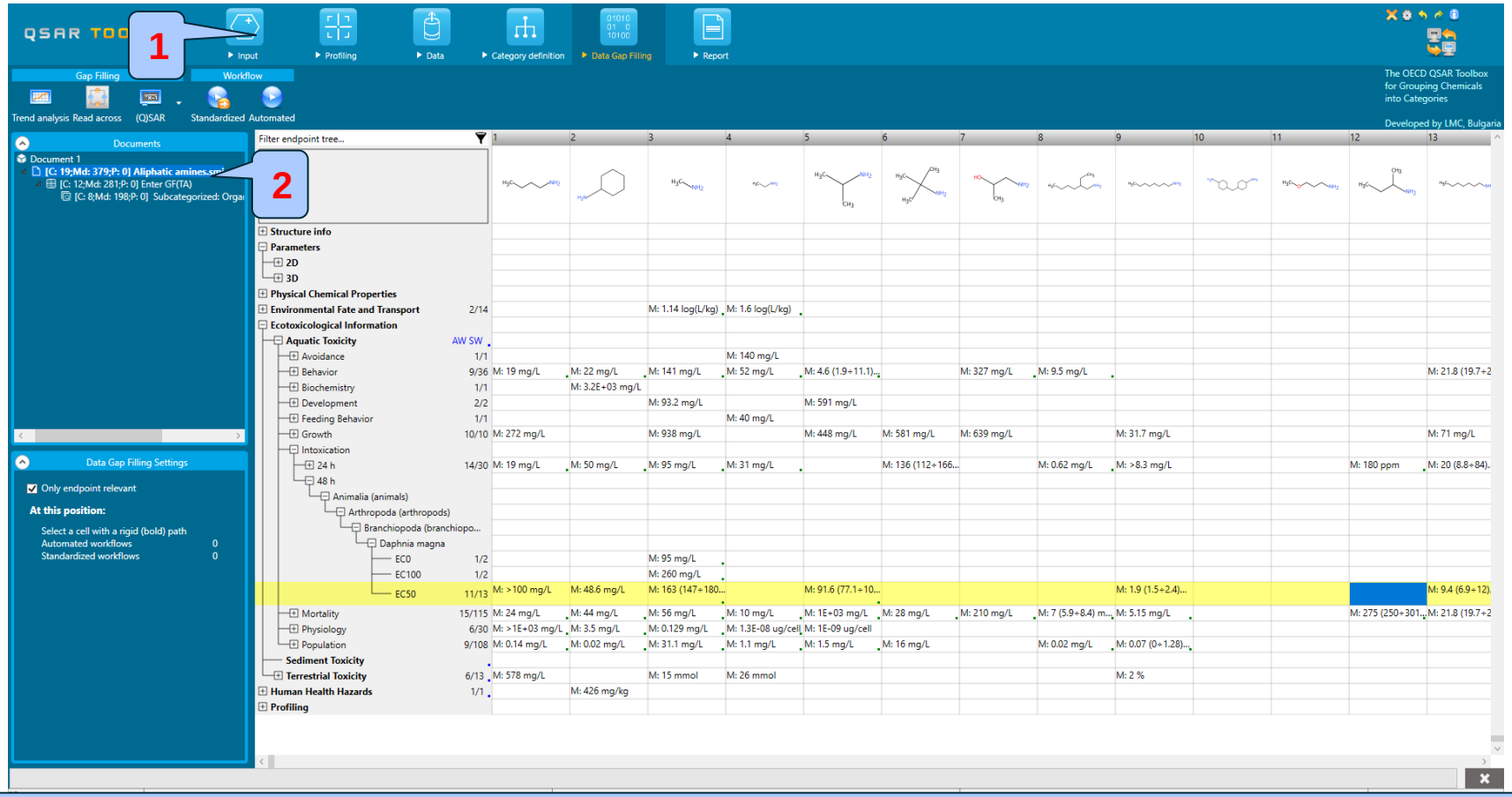
- Background
- Keywords
- Category evaluation – overview
  - Phase I. Evaluation of Category & Data Gap Filling
  - **Phase II. Extension of the Category& Data gap Filling**

Phase II: Extending the category & data gap filling

- The extension of the category is performed by using Phase II of the category evaluation process (*Extending the 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

Phase II: Extending the category & data gap filling

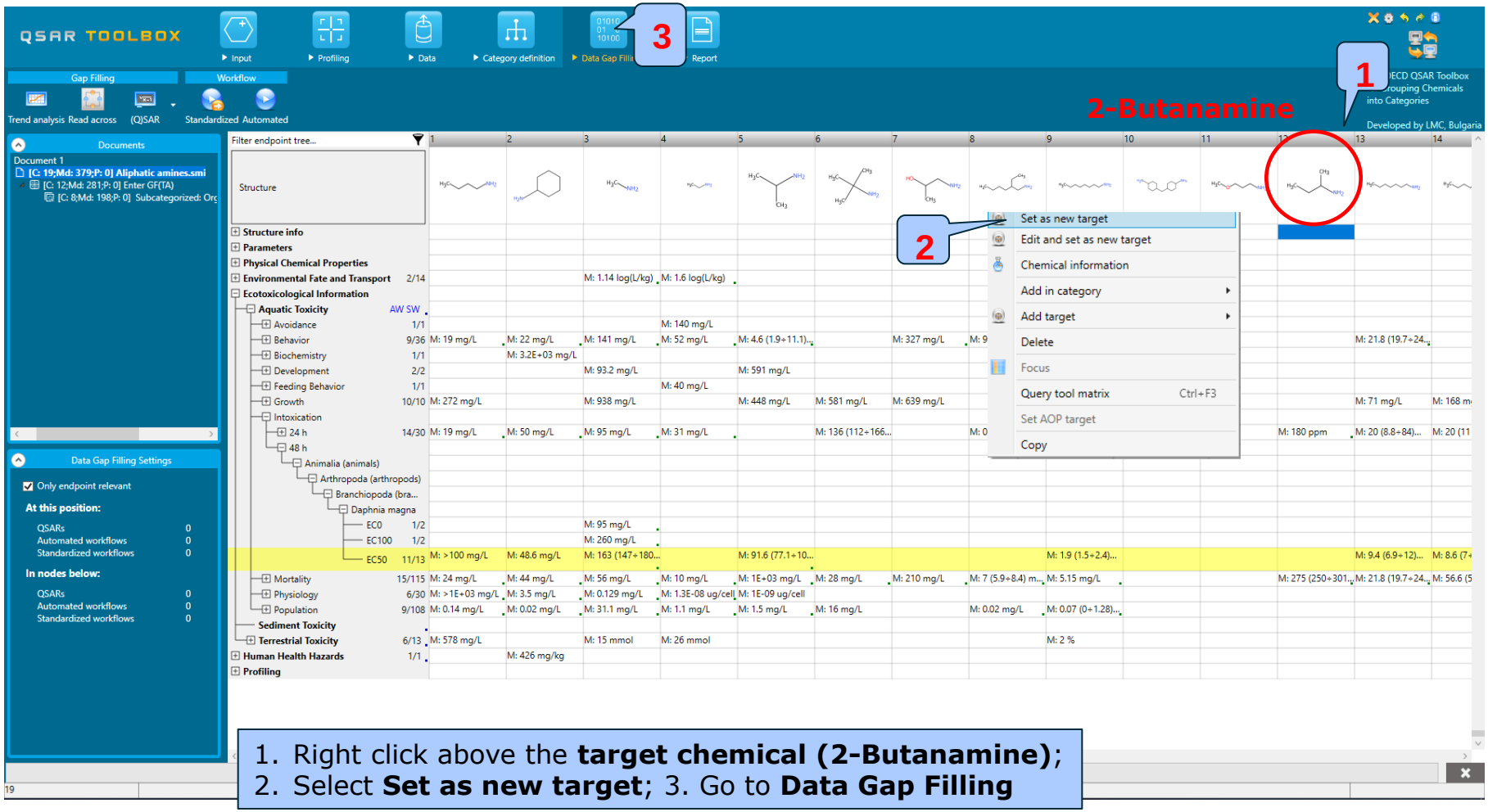
Step 1: Category definition



1. Go back to ***Input*** section; 2. Select **the row with "Aliphatic amines"** from the documented tree

Phase II: Extending the category & data gap filling

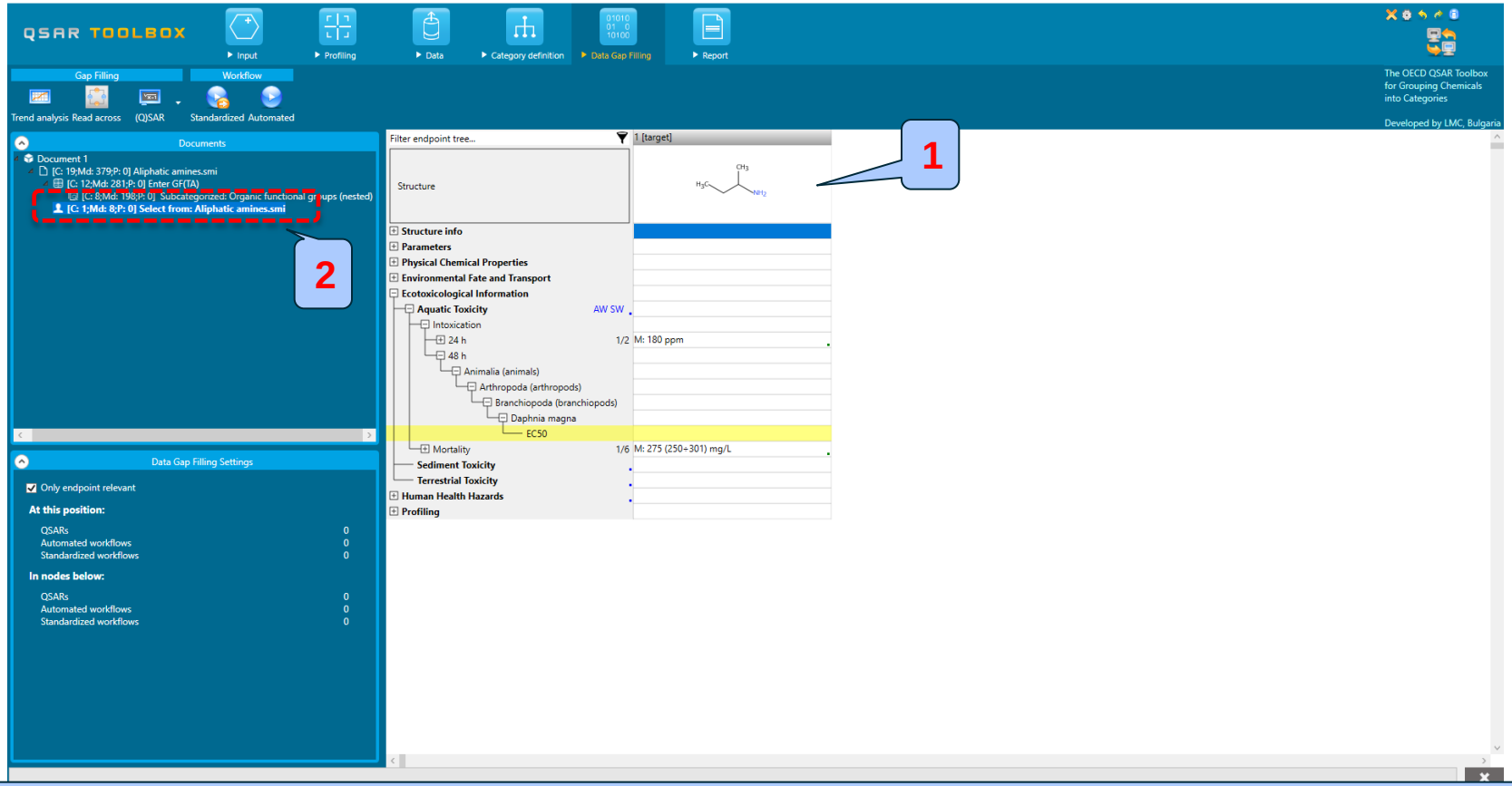
Step 1: Category definition



1. Right click above the **target chemical (2-Butanamine)**;
2. Select **Set as new target**; 3. Go to **Data Gap Filling**

Phase II: Extending the category & data gap filling

Step 1: Category definition



1. The target chemical 2-Butanamine is loaded in new data matrix; 2. Moreover a new level appears in the documented tree. The latter allows easy go back and forward to different stages of the workflow.

Phase II: Extending the category & data gap filling

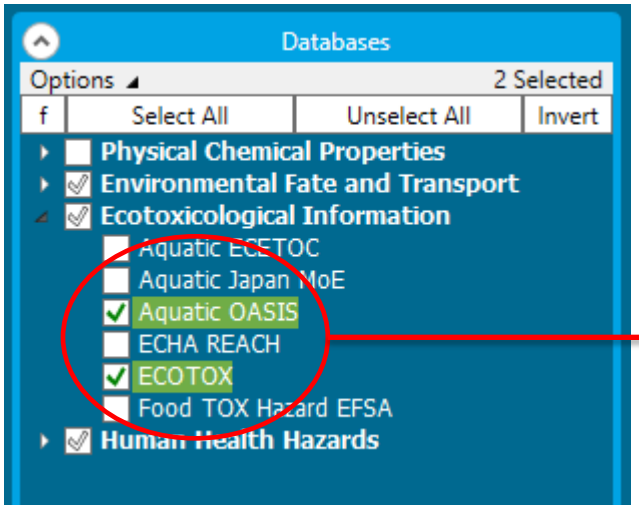
Step 1: Category definition

- The ECOSAR strict category is used to define a broader category used in further analysis. The "Strict" option means that only the defined categories should be present in the found analogues and not any others.
- The same endpoint: EC 50 48h *D.magna* will be predicted as with the predefined category

Phase II: Extending the category & data gap filling

Step 1: Category definition

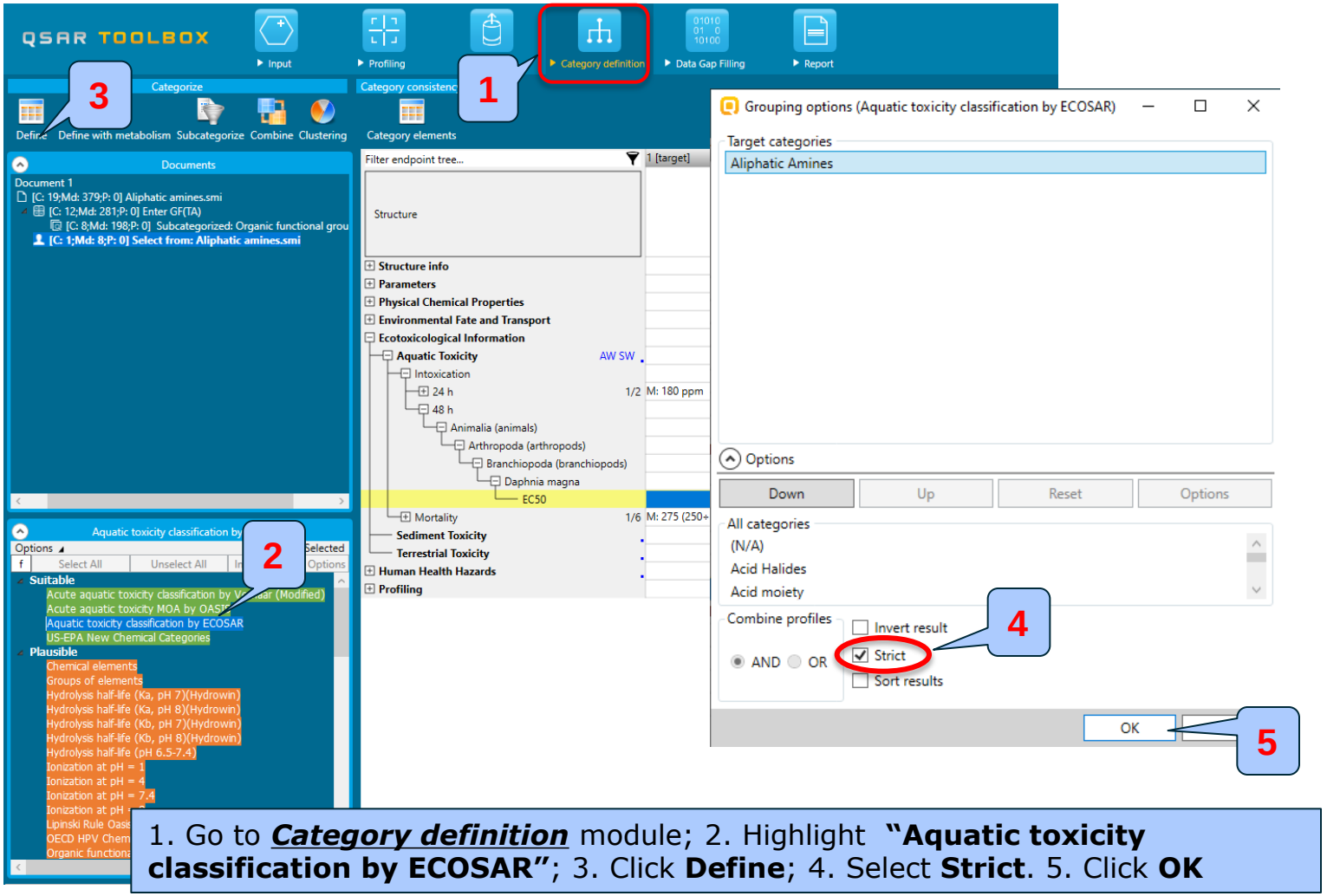
- Define ECOSAR category
- Before defining the category, the following databases related to the target endpoint are selected (databases highlighted in green are those which have data for the target endpoint):



- Aquatic OASIS
- ECOTOX

Phase II: Extending the category & data gap filling

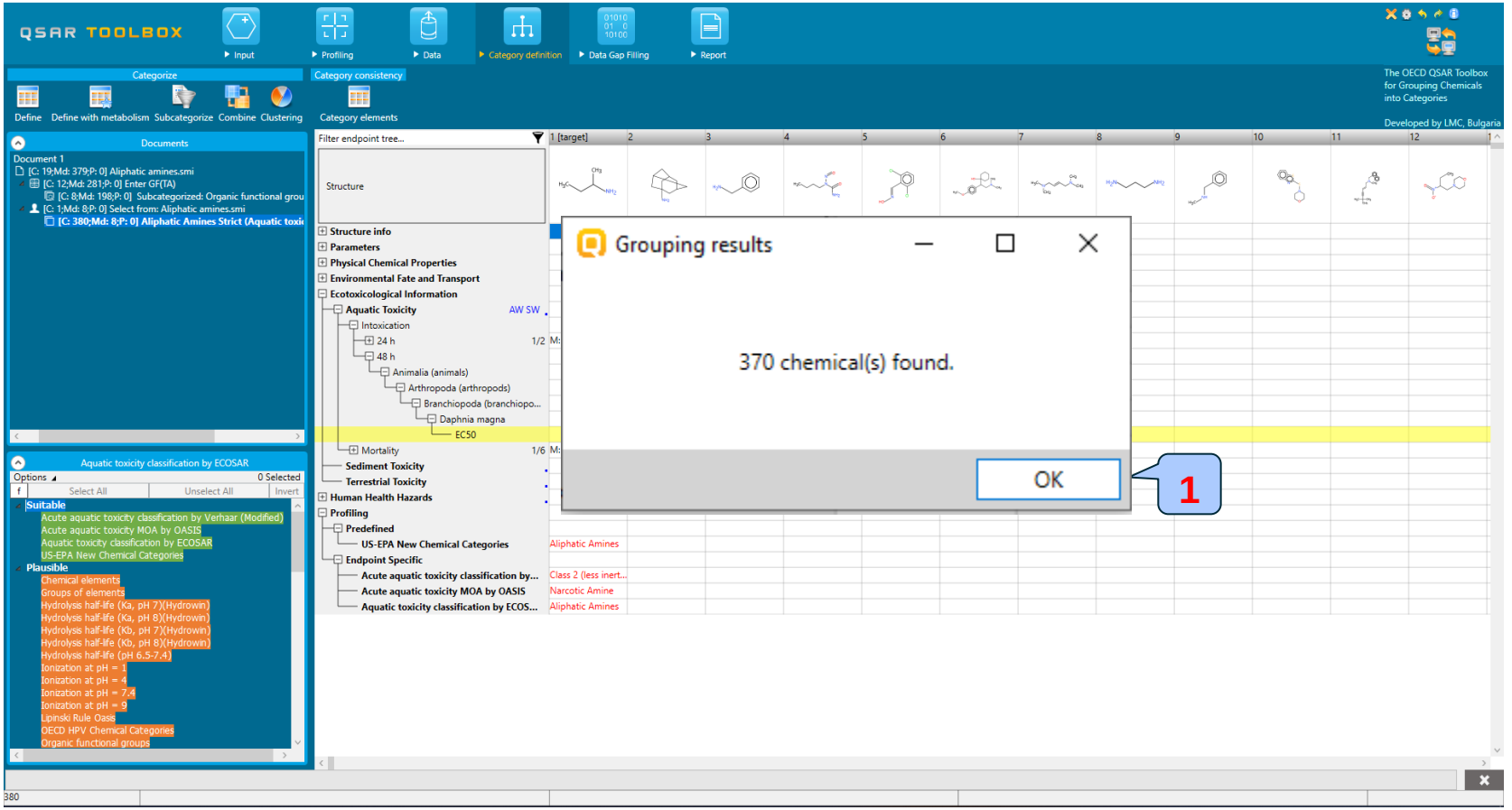
Step 1: Category definition/*Defining ECOSAR (strict)*



1. Go to ***Category definition*** module; 2. Highlight **"Aquatic toxicity classification by ECOSAR"**; 3. Click **Define**; 4. Select **Strict**. 5. Click **OK**

Phase II: Extending the category & data gap filling

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

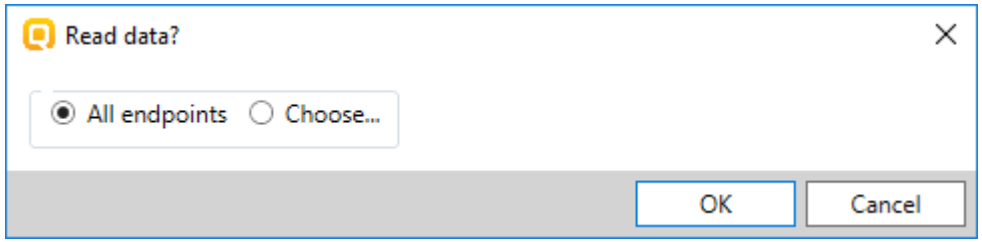


1. A message appears informing that 370 analogs are found. Click **OK**

Phase II: Extending the 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 requests the user to select the endpoint that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).

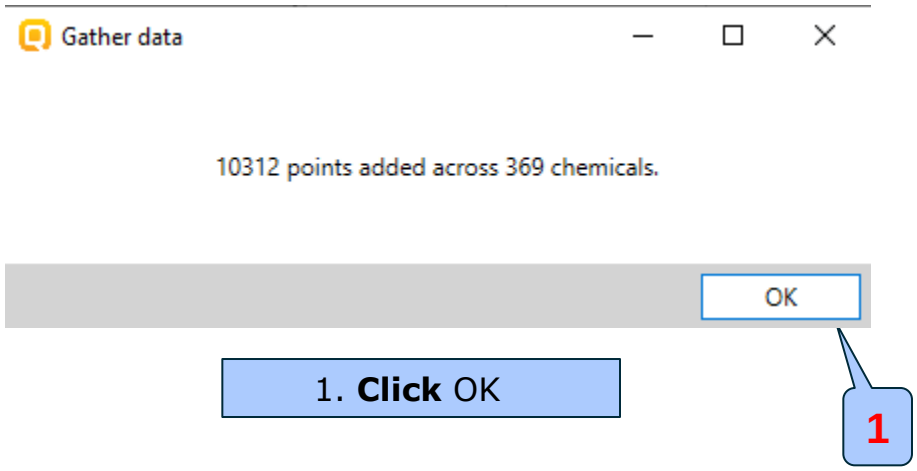


- In this example, as only databases are selected that contain information for aquatic toxicity endpoint, both options give the same results.

Phase II: Extending the category & data gap filling

Step 1: Category definition/*Reading data*

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

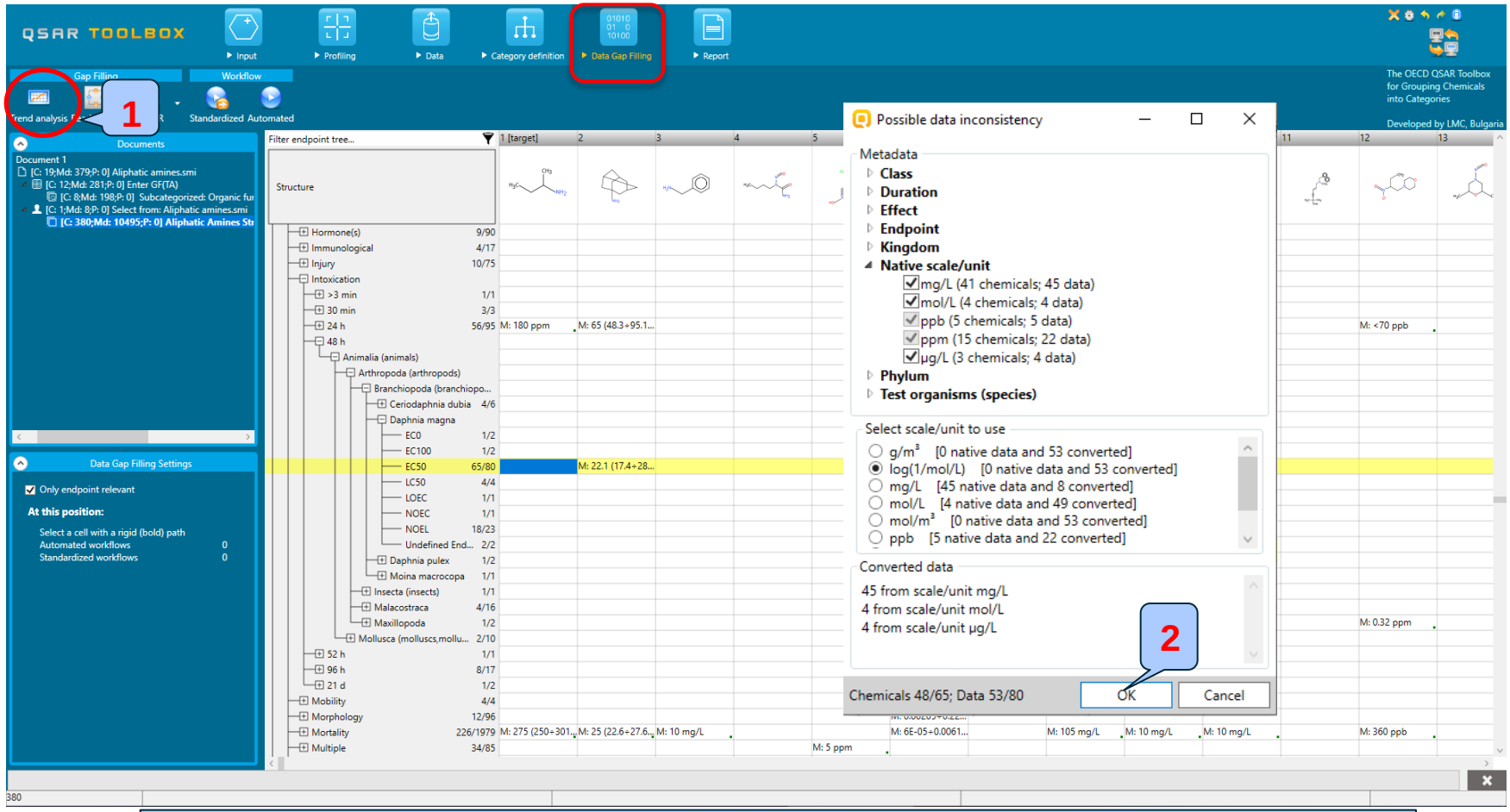


Outlook

- Background
- Keywords
- Category evaluation – overview
  - Phase I. Evaluation of Category & Data Gap Filling
  - **Phase II. Extending the Category& Data gap Filling**
    - Step 1: Category definition
    - Step 2: Navigate to the target endpoint – this is already done, so it is skipped
    - **Step 3: Data Gap Filling**

Phase II: Extending the category & data gap filling

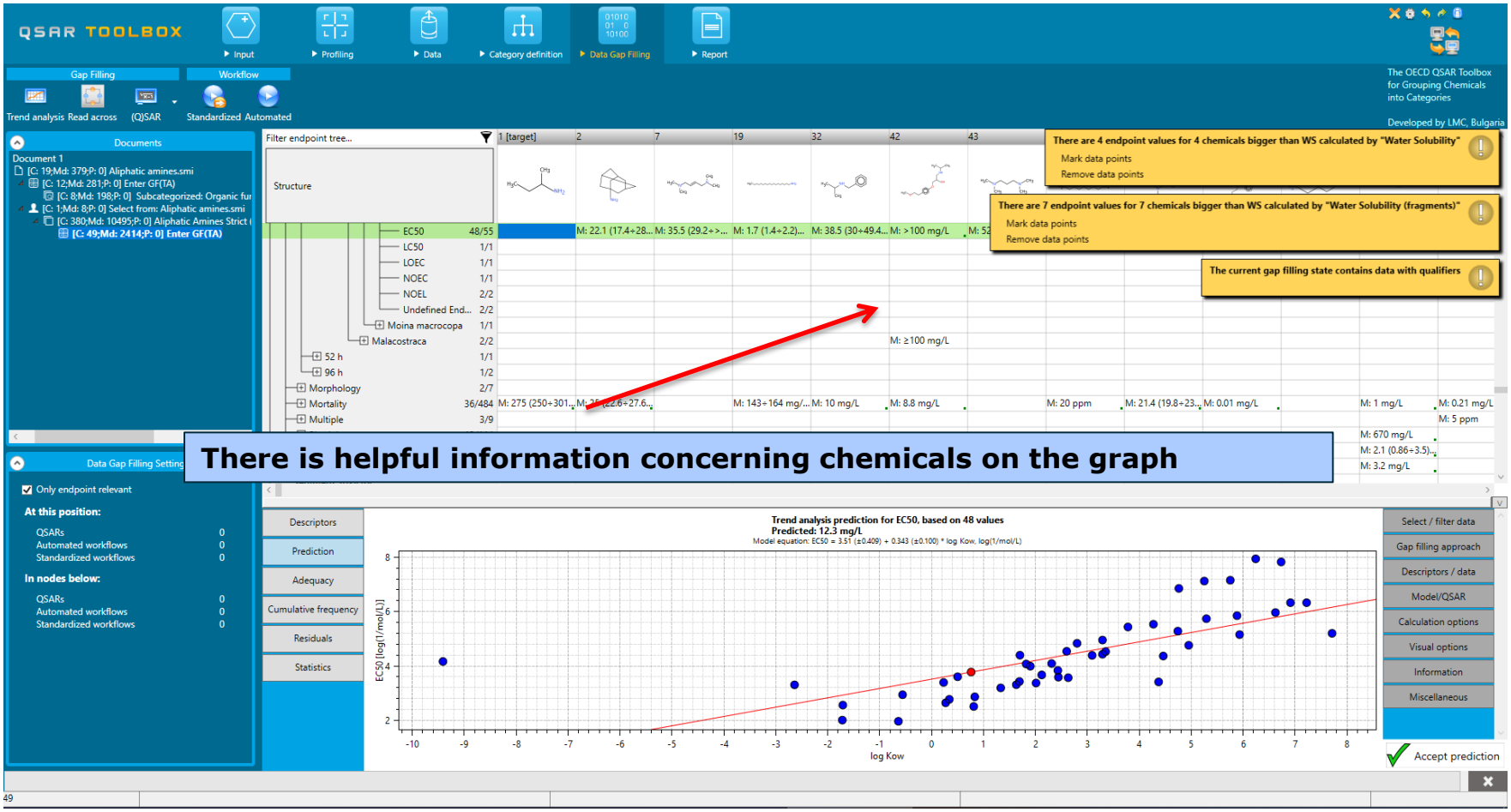
Step 3: Data Gap Filling



1. Go to ***Data Gap filling*** and select **Trend analysis**; 2. The user will be informed If there is different experimental data. Click **OK**.

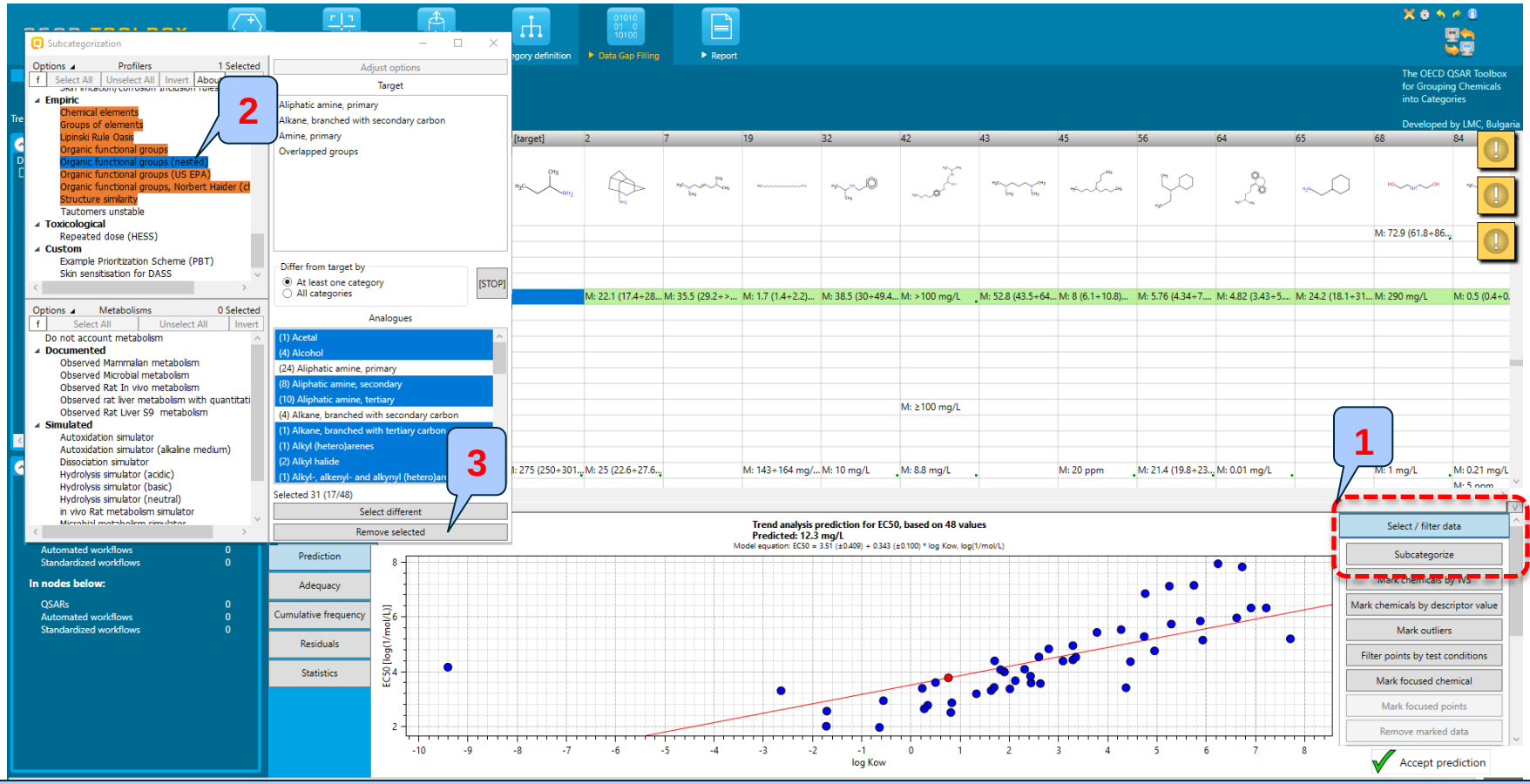
Phase II: Extending the category & data gap filling

Step 3: Data Gap Filling



Phase II: Extending the category & data gap filling

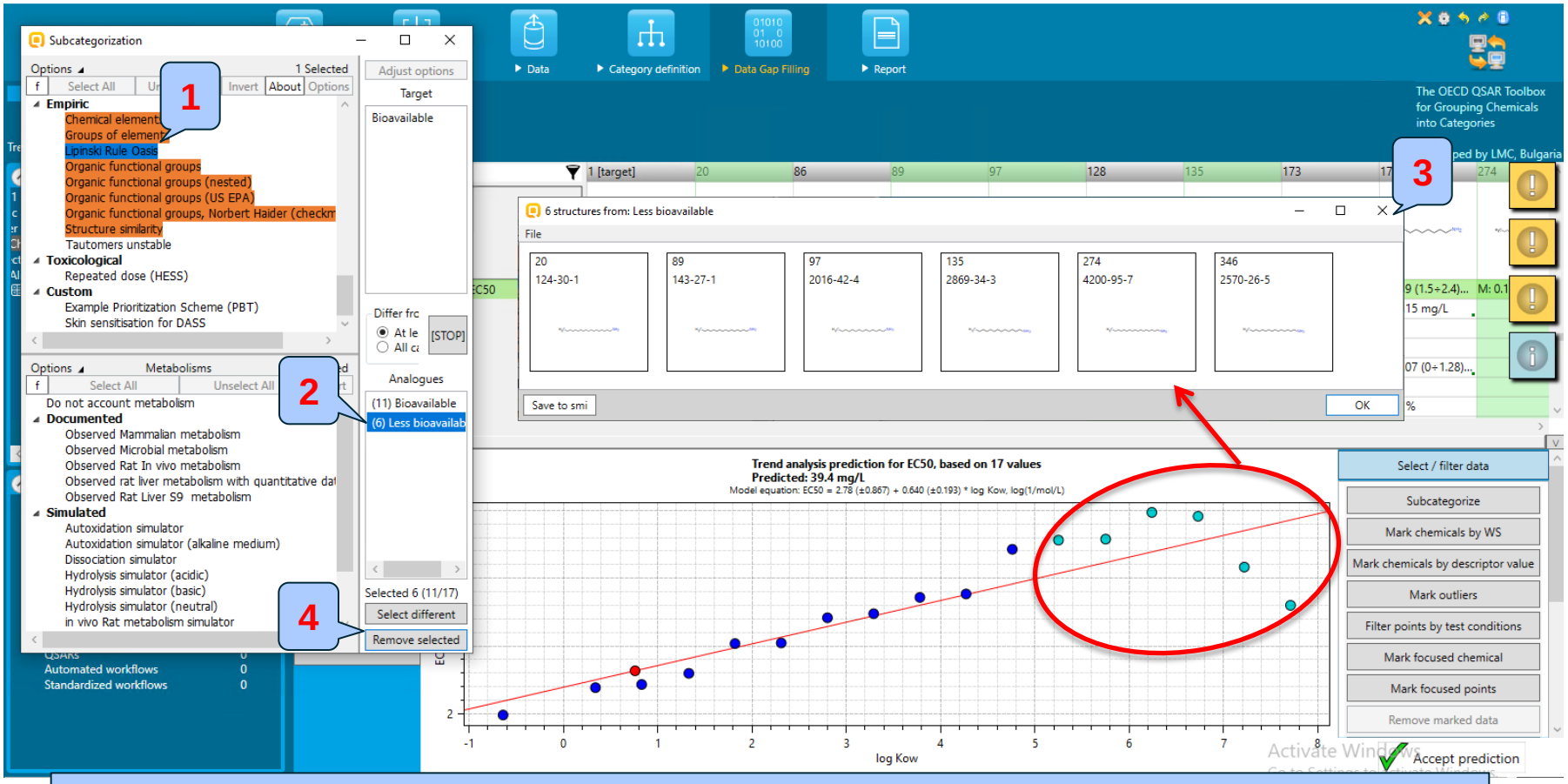
Step 3: Data Gap Filling/*Subcategorize by OFG(nested)*



In order to refine the obtained category with analogues few subcategorizations are applied: 1. Open **Select/Filter data**, then click **Subcategorize**; 2. Select **Organic functional groups(nested);** 3. Chemicals which are structurally dissimilar (highlighted in light blue on the chart) are removed.

Phase II: Extending the category & data gap filling

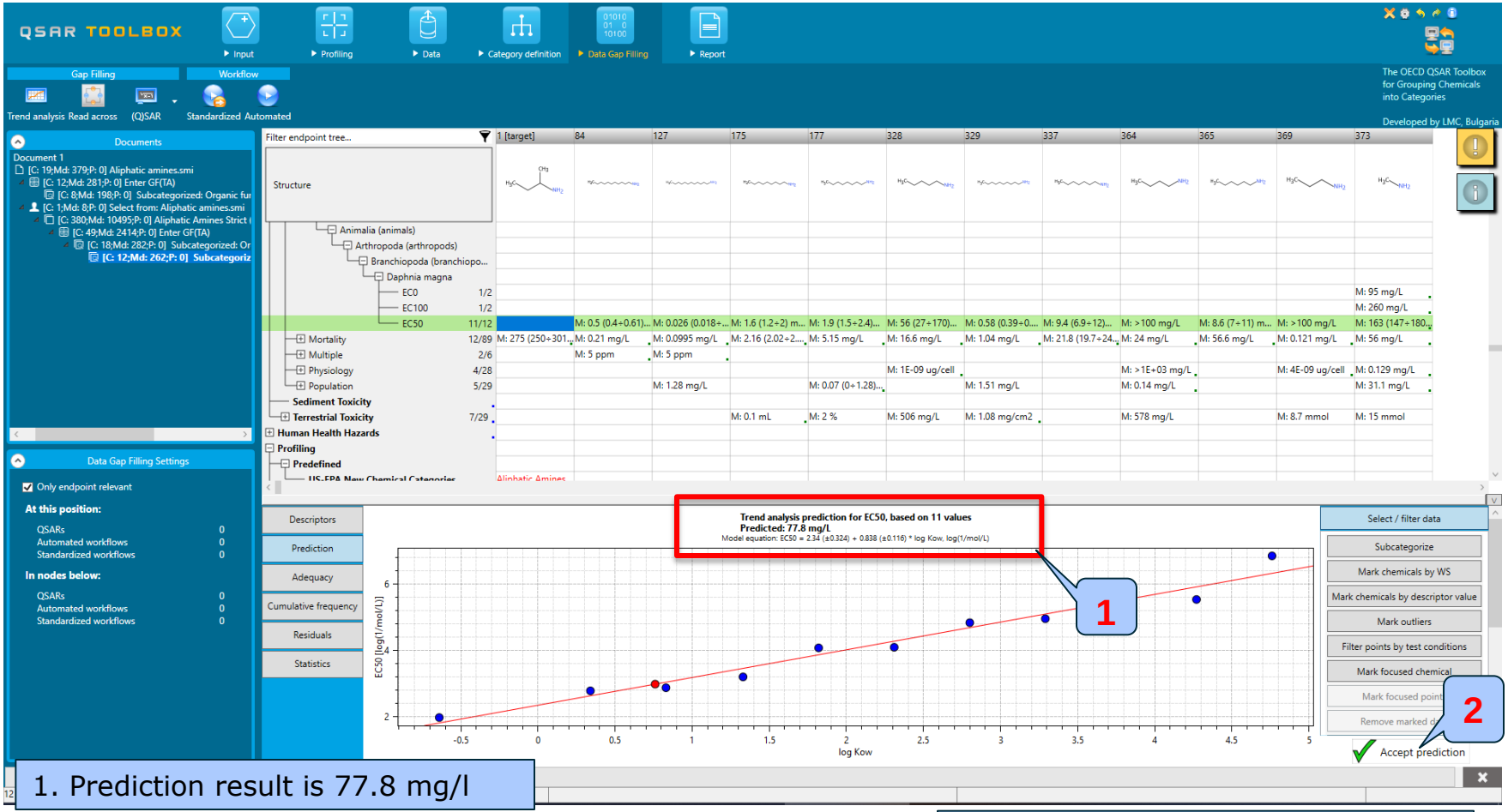
Step 3: Data Gap Filling/*Subcategorize by Lipinski rules*



Chemicals with very long chain could be removed from the category due to their less-bioavailability. In this respect: 1. Select **Lipinski rules**; 2. Double click to see "Less-bioavailable" chemicals; 3. Close the appeared window; 4. **Remove selected** 6 chemicals

Phase II: Extending the category & data gap filling

Step 3: Data Gap Filling/*Prediction result*



1. Prediction result is 77.8 mg/l

2. Click on **Accept prediction**.

Outlook

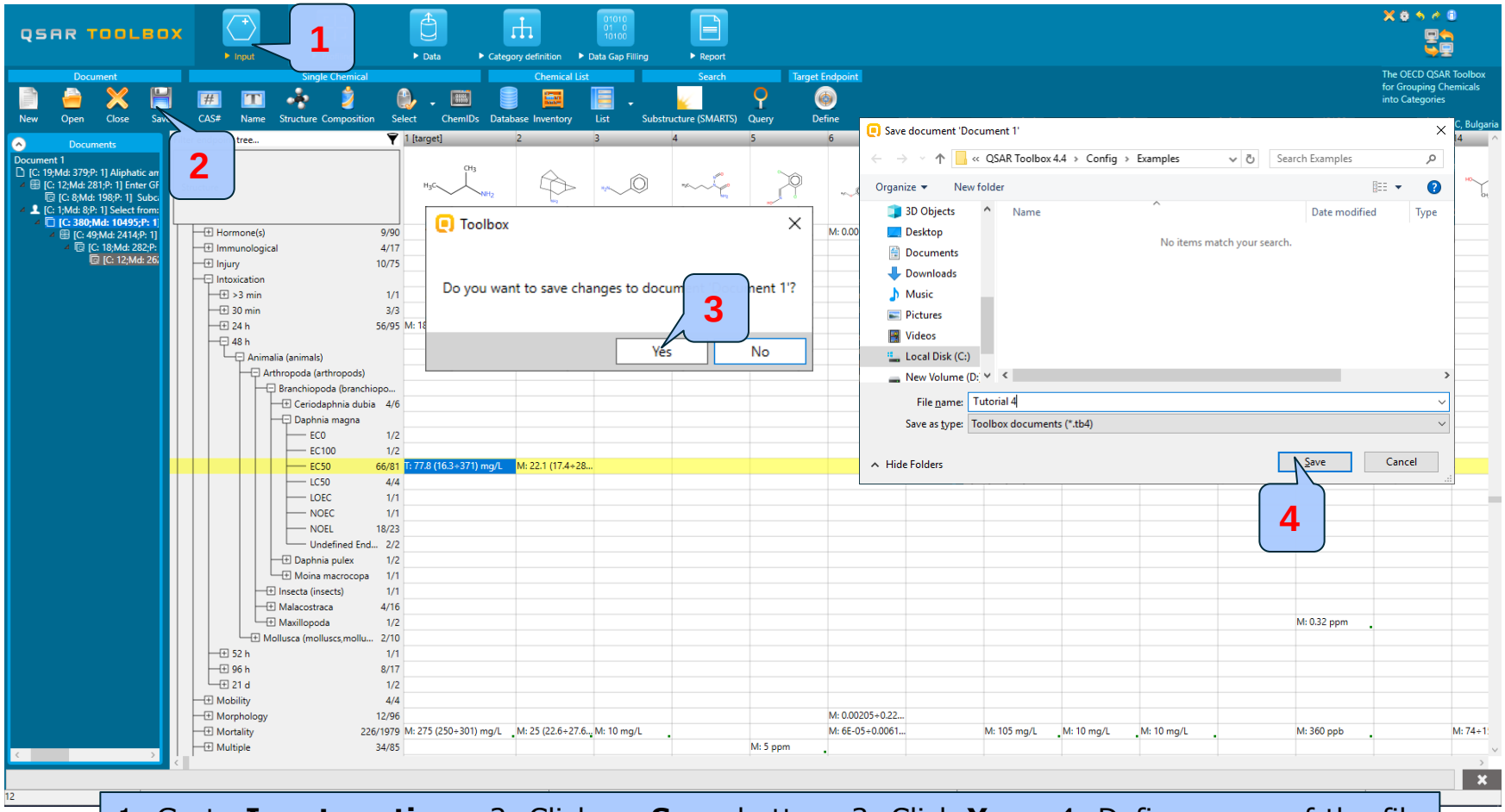
- Background
- Keywords
- Category evaluation – overview
  - Phase I. Evaluation of Category & Data Gap Filling
  - Phase II. Extension of the Category& Data gap Filling
- Save the prediction result

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).

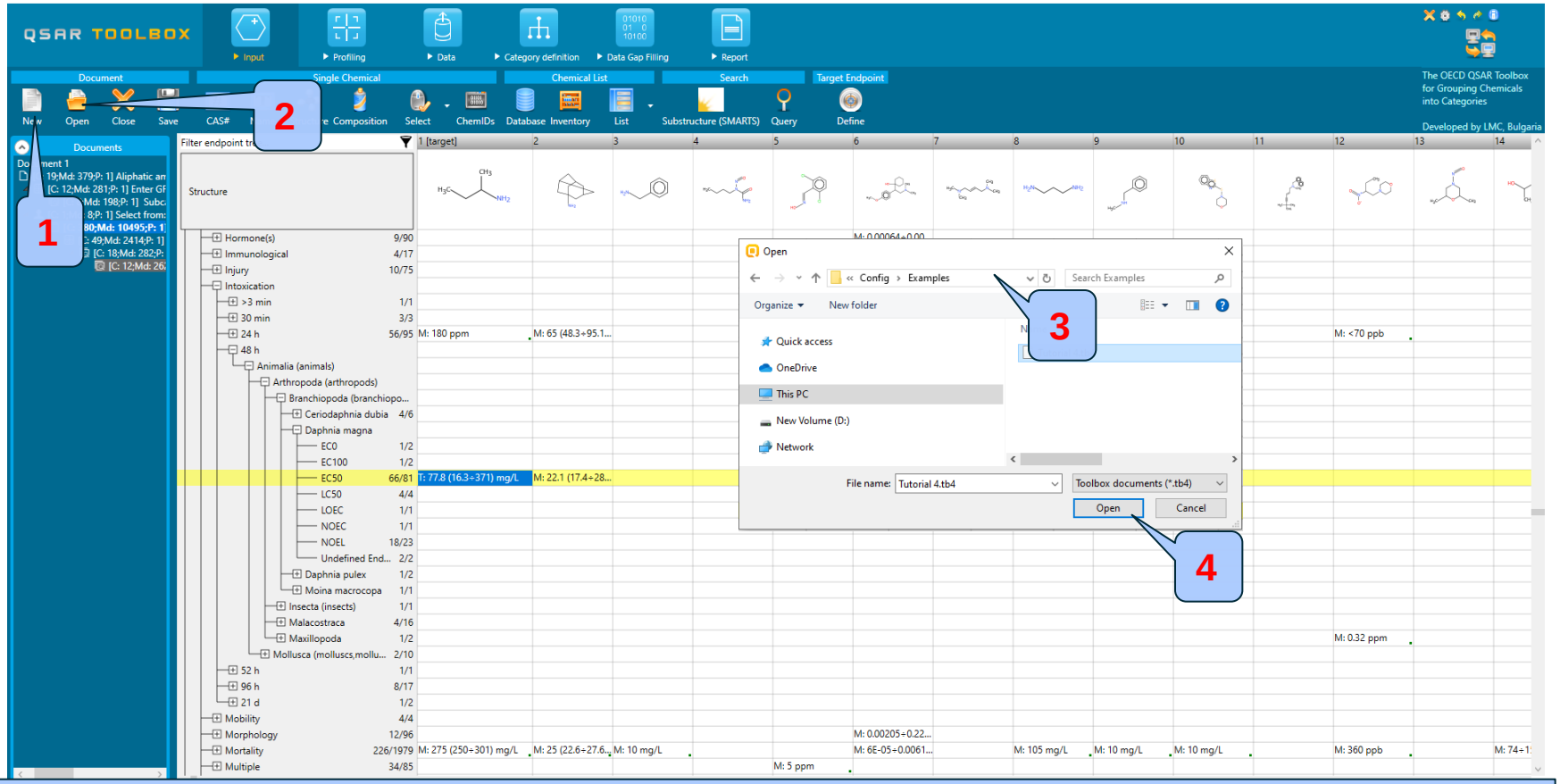
- Saving/Loading the file with TB prediction is shown on next screenshots

Saving the prediction result



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

Open saved file



1. Create **New** document; 2. Click **Open**; 3. **Find** and **select file**; 4. Click **Open.**
All the steps done during the workflow are getting executed consequentially. Once ready a message appears that the file is opened successfully, otherwise a warning message will informs you for the mistakes.