QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Step-by-step example of how to predict acute aquatic toxicity to Tetrahymena pyriformis by trend analysis using category pruning capabilities

Outlook

- **Background**
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- Workflow
- Save the prediction result

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox in a data-gap filling exercise using trend analysis based on a category formed with data pruning.
- If you are a novice user of the Toolbox you will want to review the "Getting Started" document available at [http://www.oecd.org/chemicalsafety/risk-assessment/TB3%200_GettingStarted_rev2.pdf] as well as go through tutorials 1 - 3.

Outlook

- Background
- **Objectives**
- Specific Aims
- Trend analysis
- The exercise
- Workflow
- Save the prediction result

Objectives

- **This presentation demonstrates a number of functionalities of the Toolbox :**
  - Entering a target chemical by SMILES notation and Profiling
  - Identifying analogues for a target chemical by molecular similarity
  - Retrieving experimental results available for those analogues, and for multiple endpoints
  - Filling data gaps by trend analysis

Outlook

- Background
- Objectives
- **Specific Aims**
- Trend analysis
- The exercise
- Workflow
- Save the prediction result

Specific Aims

- To reacquaint the user with the workflow and the six modules of the Toolbox.
- To reacquaint the user with the basic functionalities within each module.
- To introduce the user to new functionalities of selected modules.
- To explain to the rationale behind each step of the exercise.

Outlook

- Background
- Objectives
- Specific Aims
- **Trend analysis**
- The exercise
- Workflow
- Save the prediction result

Trend analysis

- For a given (eco)toxicological endpoint, the members of a category are often related by a trend (e.g., increasing, decreasing or constant) in potency.
- The trend could be related to molecular mass, carbon chain length, or to some other property.
- A demonstration of consistent trends in the toxicity of a group of chemicals is one of the desirable attributes of a chemical category and one of the indicators a common mechanism of action for all chemicals in the category.
- When some chemicals in a category have measured values and a consistent trend is observed, missing values can be estimated by simple scaling from the measured values to fill in the data gaps.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- **The exercise**
- Workflow
- Save the prediction result

The Exercise

- In this exercise we will predict the toxicity towards the *Tetrahymena pyriformis,* of the substance, 4-Methylhexanal [CAS# 41065-97-8], which will be the "target" chemical.
- This prediction will be accomplished by collecting experimental results for a set of chemicals considered to be in the same category as the target molecule.
- The category will initially be defined based on a protein binding mechanism.
- The initial category will be pruned via subcategorisation.
- Trend-analysis will be used for data gap filling.

Side-Bar On *Tetrahymena*

Tetrahymena pyriformis is a free-living ciliated protozoan that is found in fresh-water ecosystems throughout the world.

It is a well-studied genus with understood nutrient requirements, growth characteristics, and cell biology and genomics.

The ability to culture it axenically in simple media, coupled with its rapid doubling time of makes it a good test organism.

Strain GL-C is amicronuclated and thereby genetically extremely stable.

Side-Bar On TETRATOX

Population growth impairment testing with *Tetrahymena pyriformis* is conducted in a static assay uses population density quantified spectrophotometrically as its endpoint.

The endpoint 50% inhibitory growth concentration from this assay is one of the largest aquatic toxicity databases in the Toolbox.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
- Save the prediction result

Workflow

- **By now you should be familiar with the six modules which are used in a sequential workflow within the Toolbox. These are:**
  - Chemical Input
  - Profiling
  - Endpoints
  - Category Definition
  - Filling Data Gaps
  - Report

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
  - **Chemical Input**

Chemical Input
Overview

- This module provides the user with several ways of entering the target chemical into the Toolbox.
- This is important because all subsequent functions are based on chemical structure.
- The goal of this module is to make sure the molecular structure assigned to the target chemical is the correct one.

Chemical Input
Ways of Entering a Chemical

User Alternatives for Chemical ID:

A. Single target chemical
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases
- Chemical IDs such as EC number, Einecs number

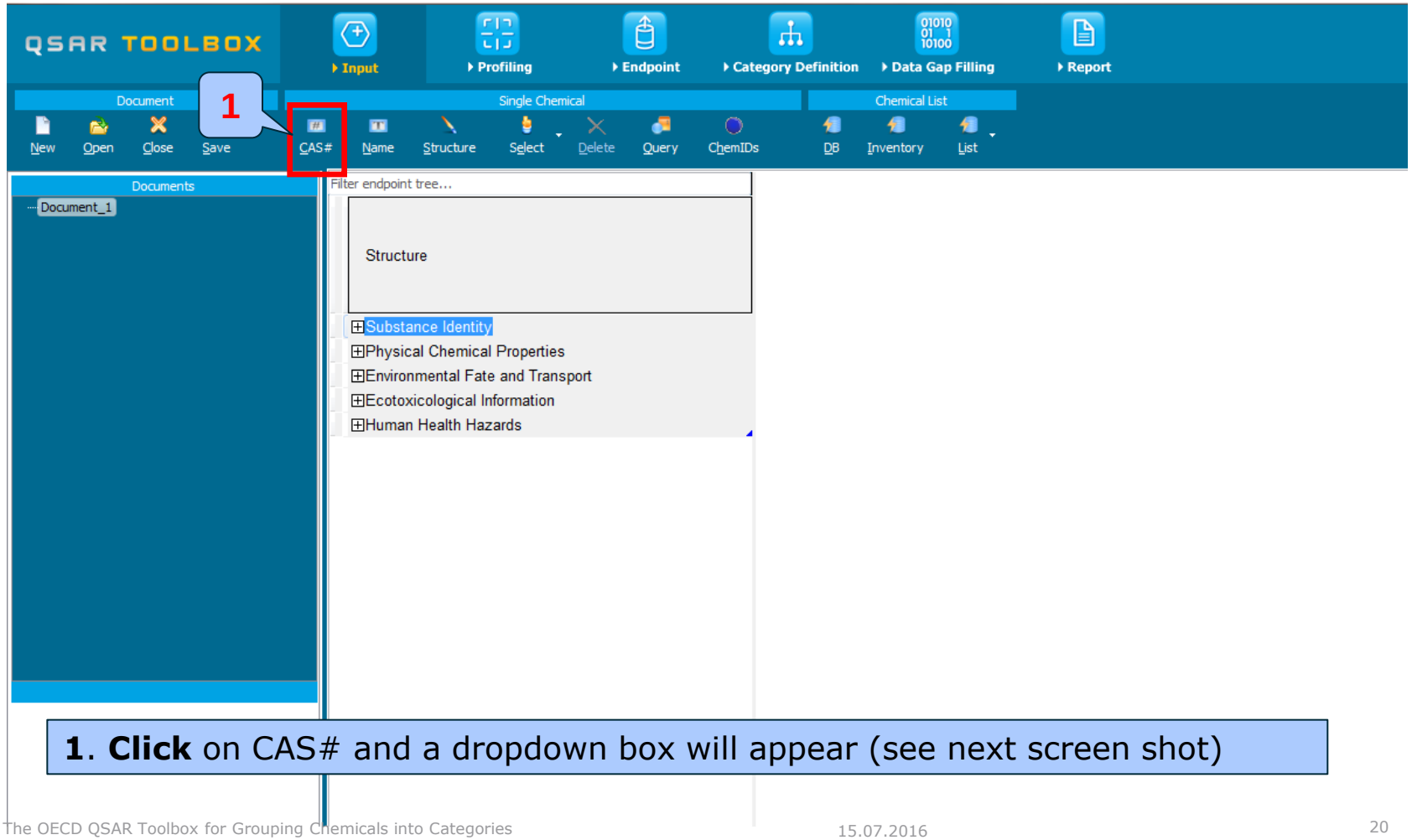
B. Group of chemicals
- User List/Inventory
- Specialized Databases

Chemical Input

- Turn on the computer and open the Toolbox. This will take some time as the databases will be populated.
- Note that the Toolbox opens to the first Module "Input" (see next screen shot).
- Since we have described this modules in the earlier tutorials, we will go through it rapidly.

Chemical Input Screen

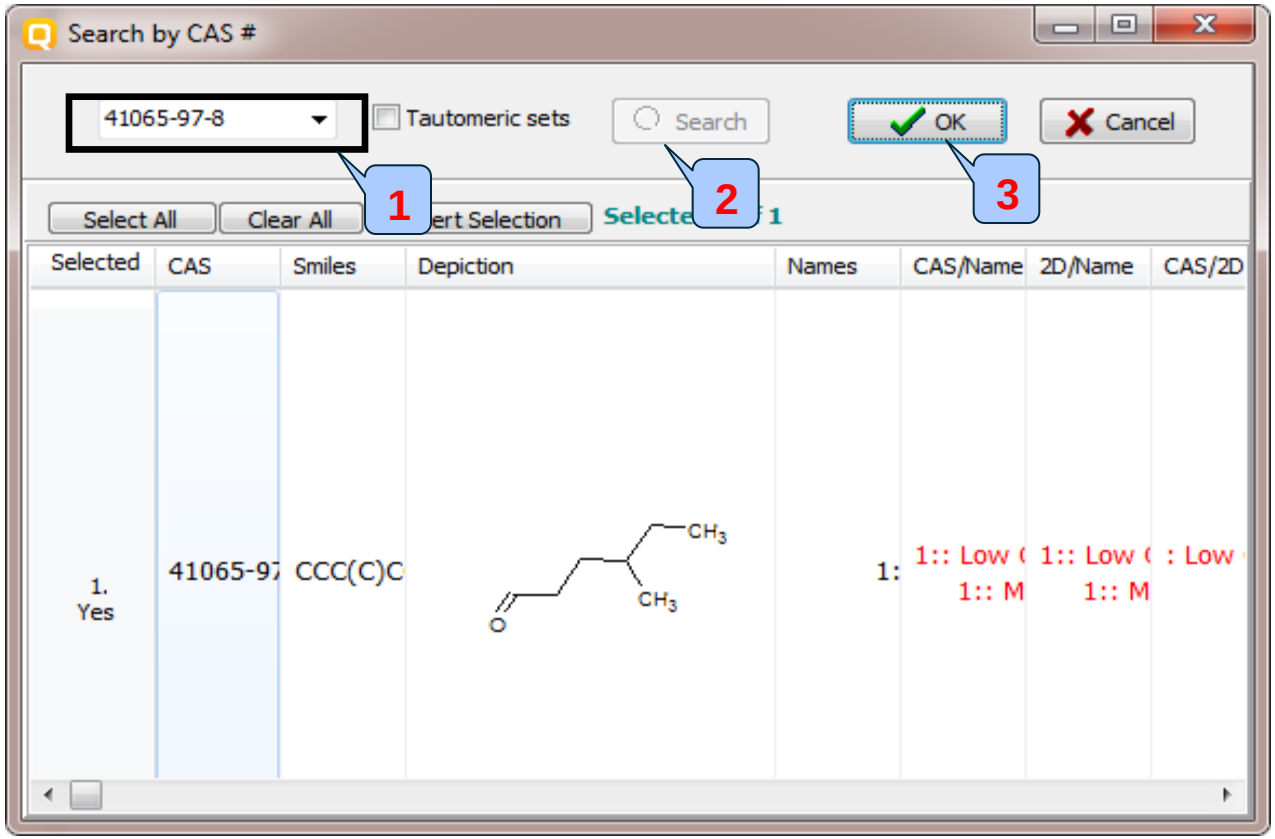
Input target chemical by CAS#



1. **Click** on CAS# and a dropdown box will appear (see next screen shot)

Chemical Input Screen

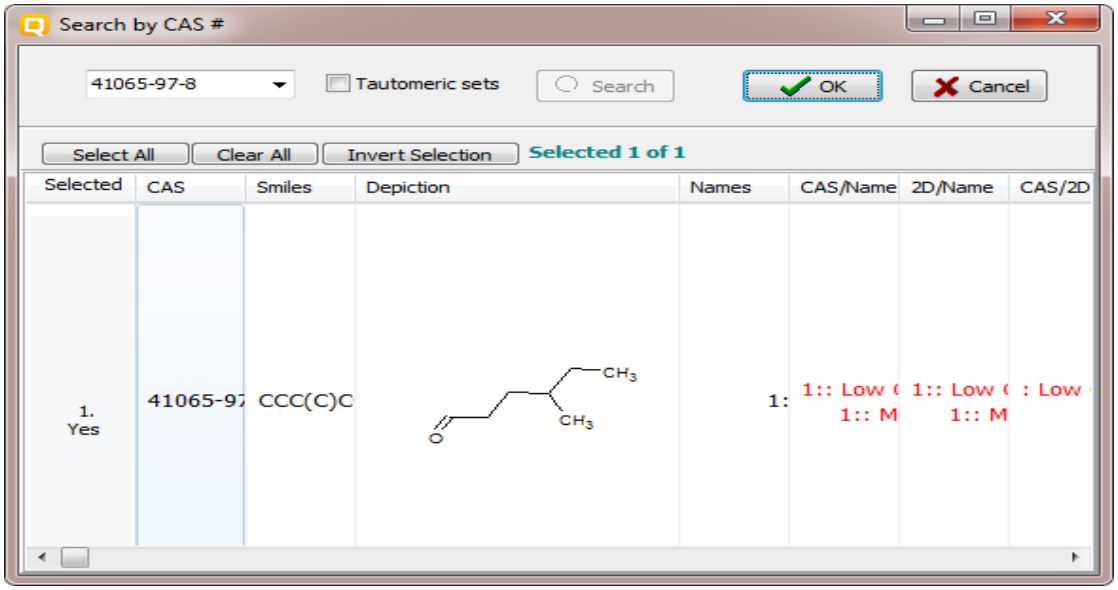
Enter CAS# 41065-97-8 of 4-Methyhexanal



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

Chemical Input
Target chemical identity

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

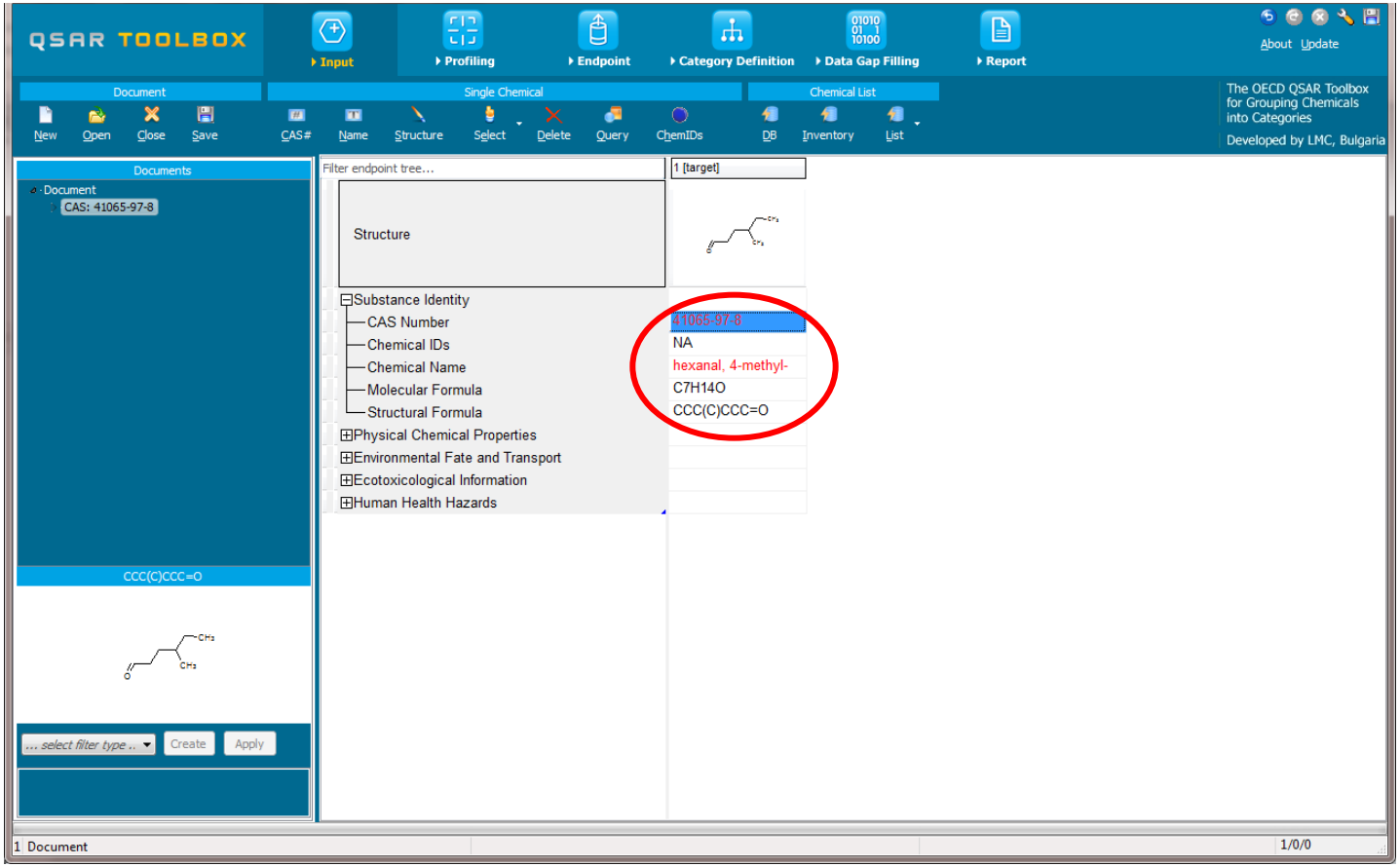


In case a structure has several CAS numbers or a structure could be related to more than one substance more than one chemical identity could be retrieved. In this case the user can decide which substance is to be retained for the subsequent workflow.

Chemical Input
Target chemical identity

- Double left click "Substance Identity"; this displays the chemical identification information (see next screen shot).
- Note that existing in the Toolbox names of target chemical are in different colours.
- The workflow on the first module is now completed, and the user can proceed to the next module (see next screen shot).

Chemical Input
Target chemical identity



Outlook

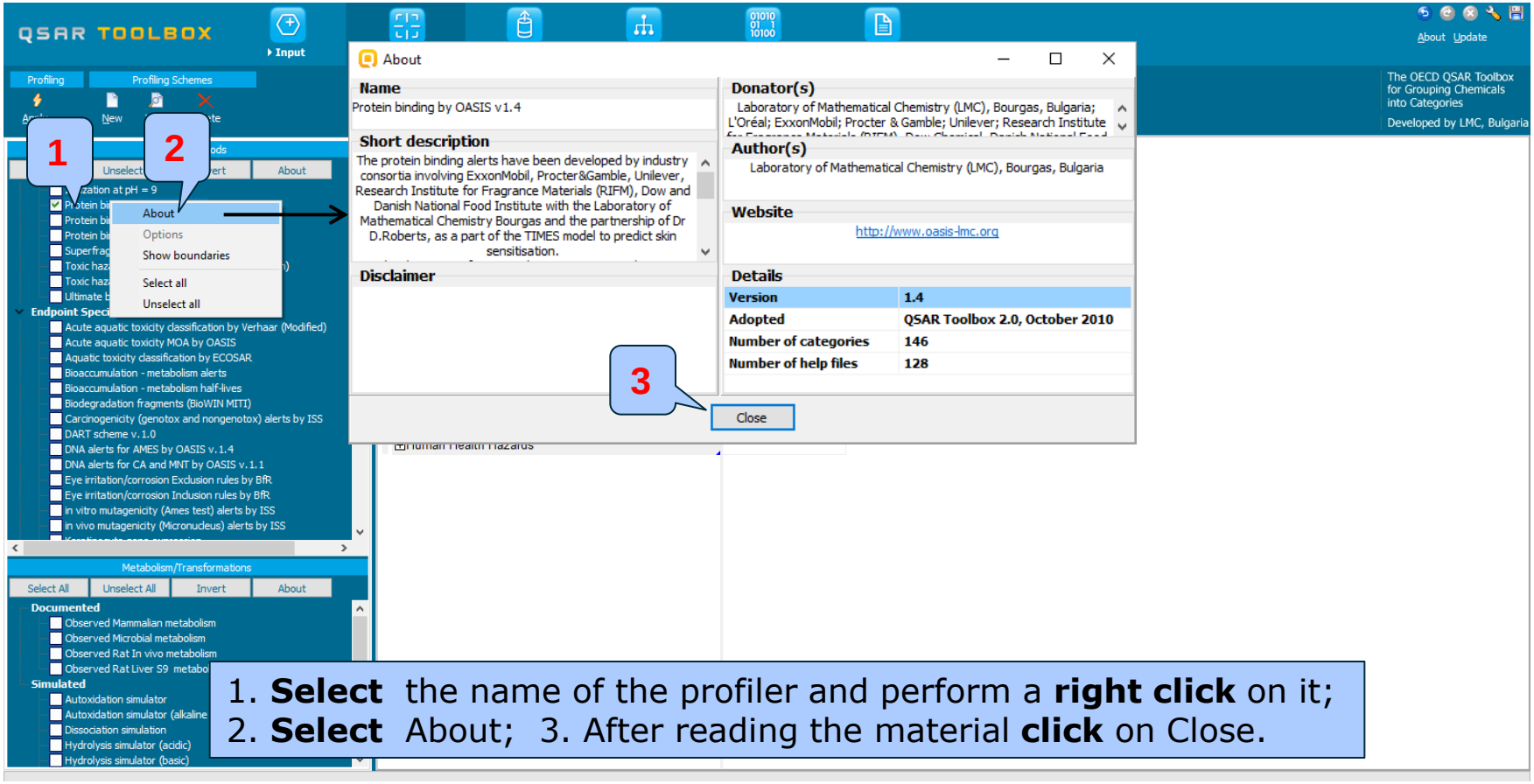
- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
  - Chemical Input
  - **Profiling**

Profiling
Overview

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

Profiling
Side-Bar to Profiling

Summary information of the different profilers is provided in the "**About**".

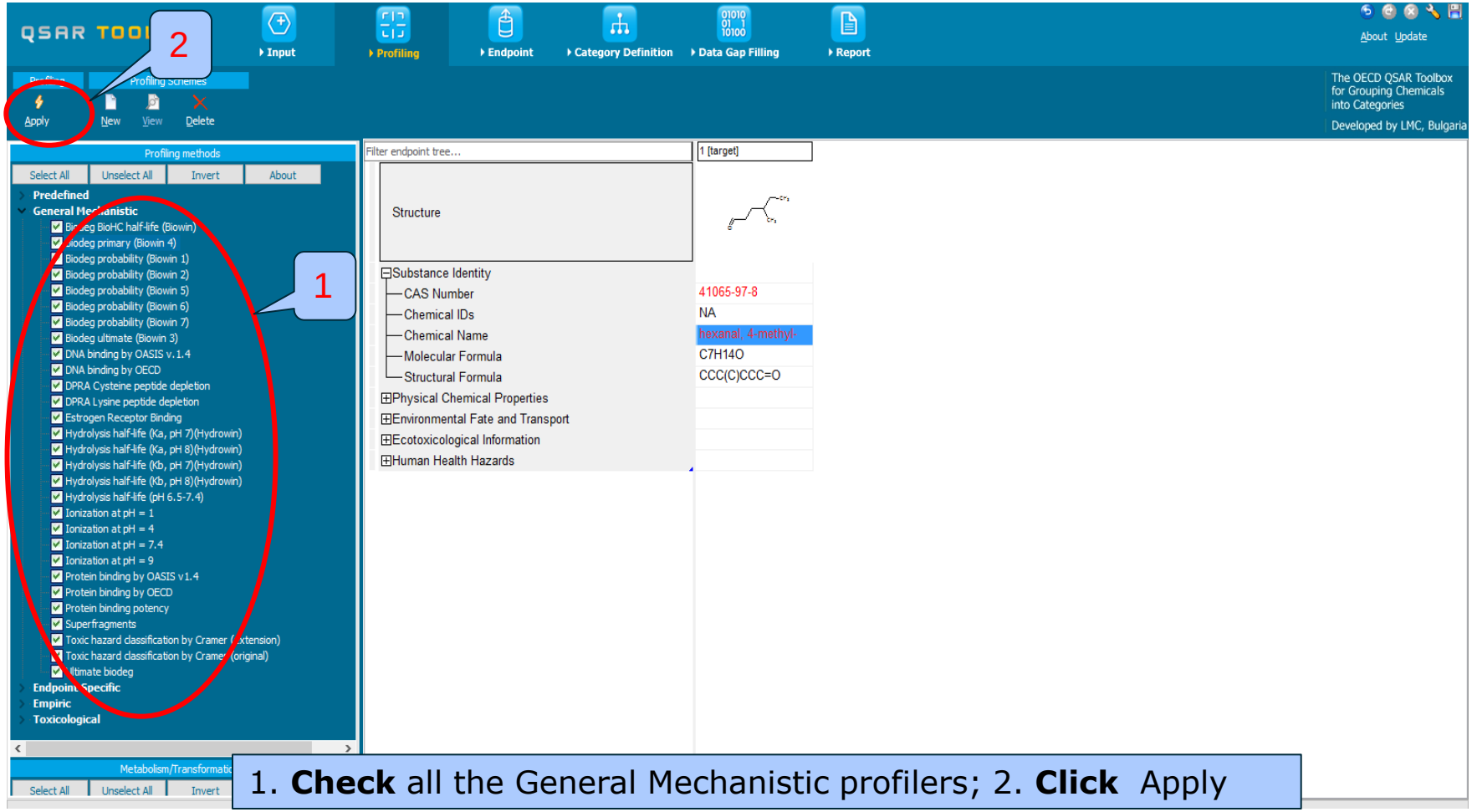


Profiling
Profiling the target chemical

- Click in the box next to the Profiling methods related to the target endpoint
- This selects (a green check mark appears) or deselects (green check mark disappears) profilers
- For this example check all the mechanistic methods.

Profiling

Profiling the target chemical



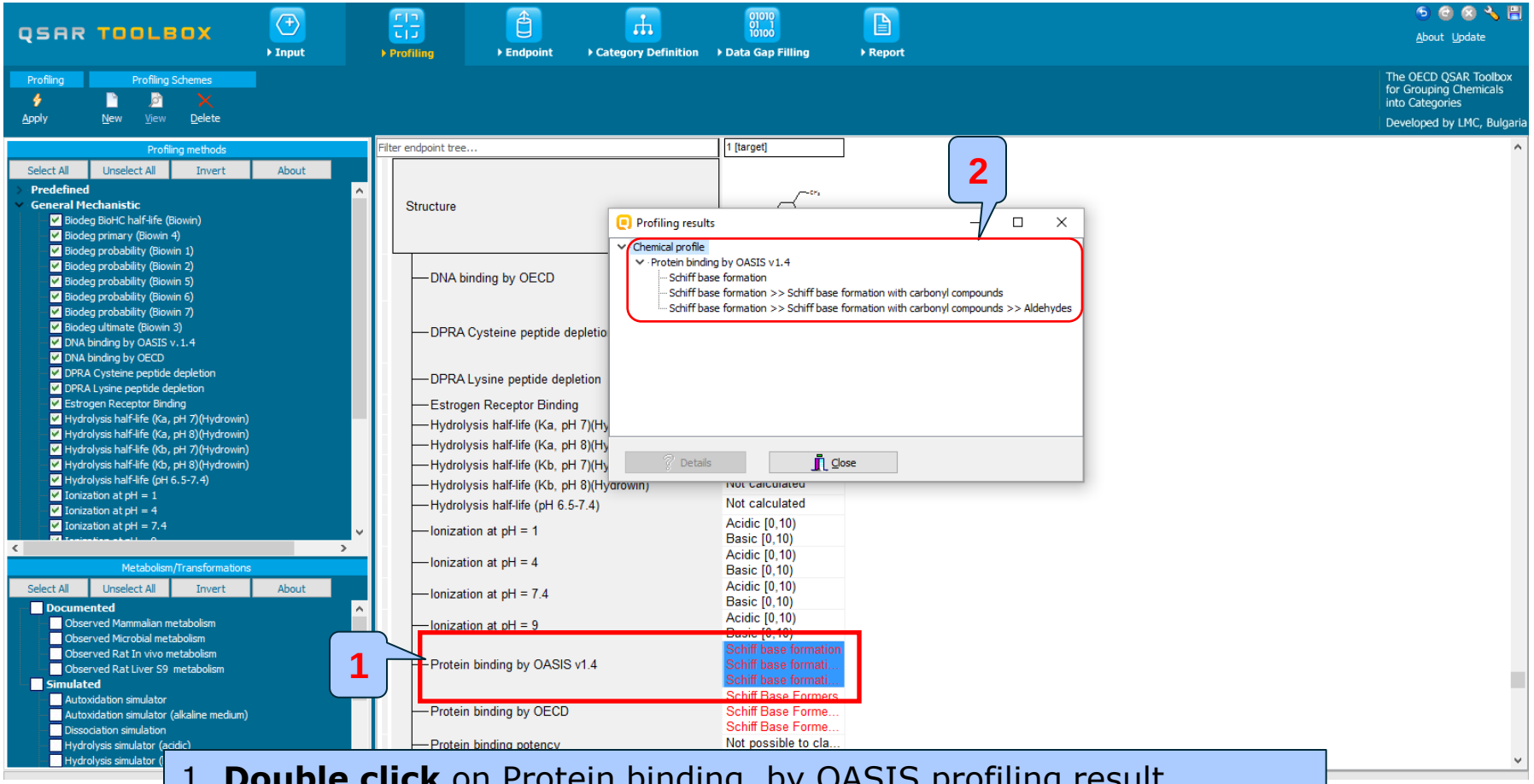
1. **Check** all the General Mechanistic profilers; 2. **Click** Apply

Profiling
Profiling the target chemical

- The actual profiling will take several seconds depending on the number and type of profilers selected.
- The results of profiling automatically appear as a dropdown box under the target chemical (see next slide)
- Please note the specific protein-binding profiler – Protein binding by OASIS v1.4 - background information can be retrieved by double click on the box with Protein binding by OASIS v1.4 result (see next slide)
- This result will be used to search for suitable analogues in the next steps of the exercise.
- Move to next module "Endpoint".

Profiling

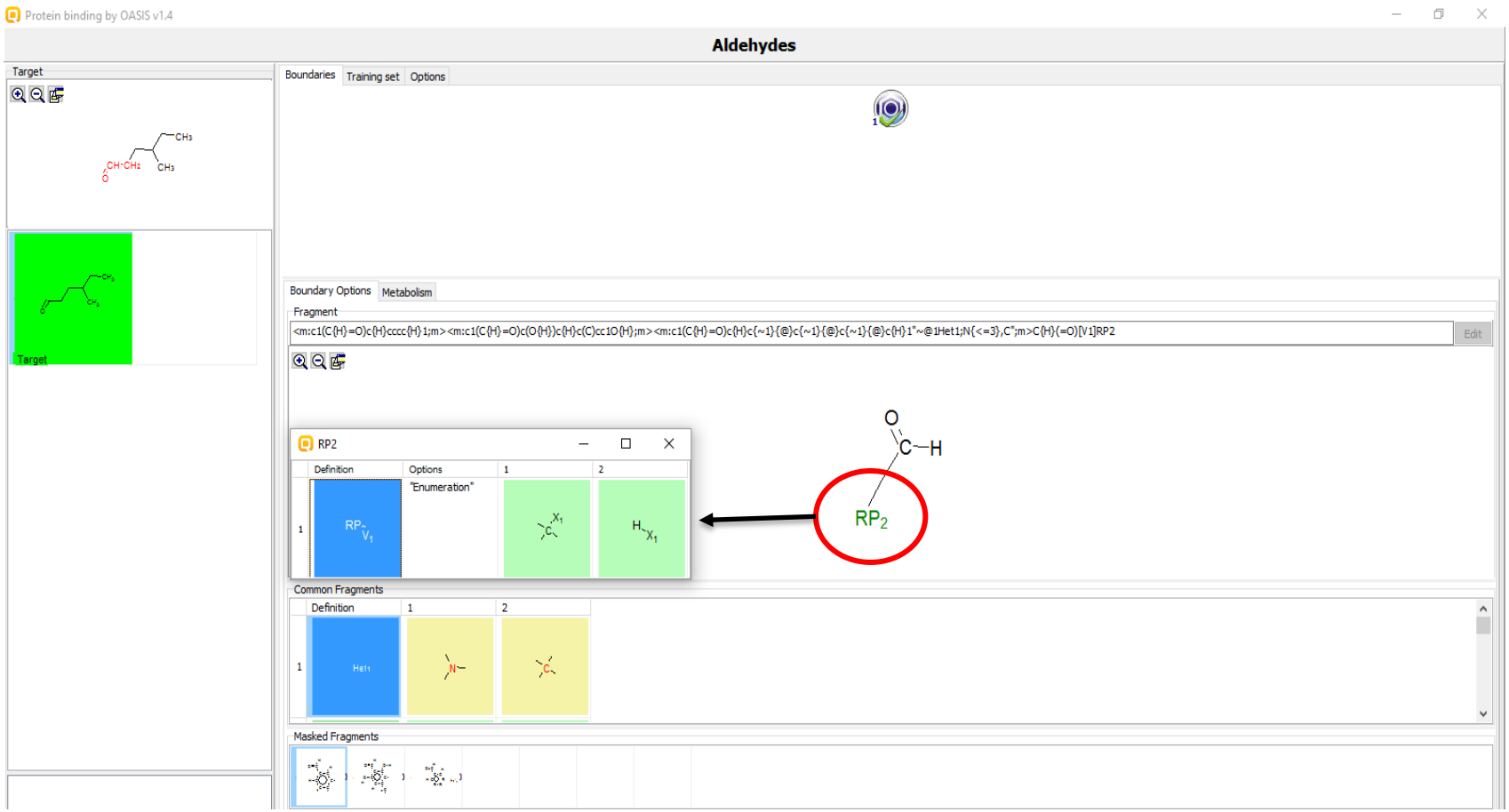
Profiling the target chemical



1. **Double click** on Protein binding by OASIS profiling result
2. **Click Details** to see the explanation of **Aldehydes** category

Profiling

Profiling results – Explain



Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - **Endpoints**

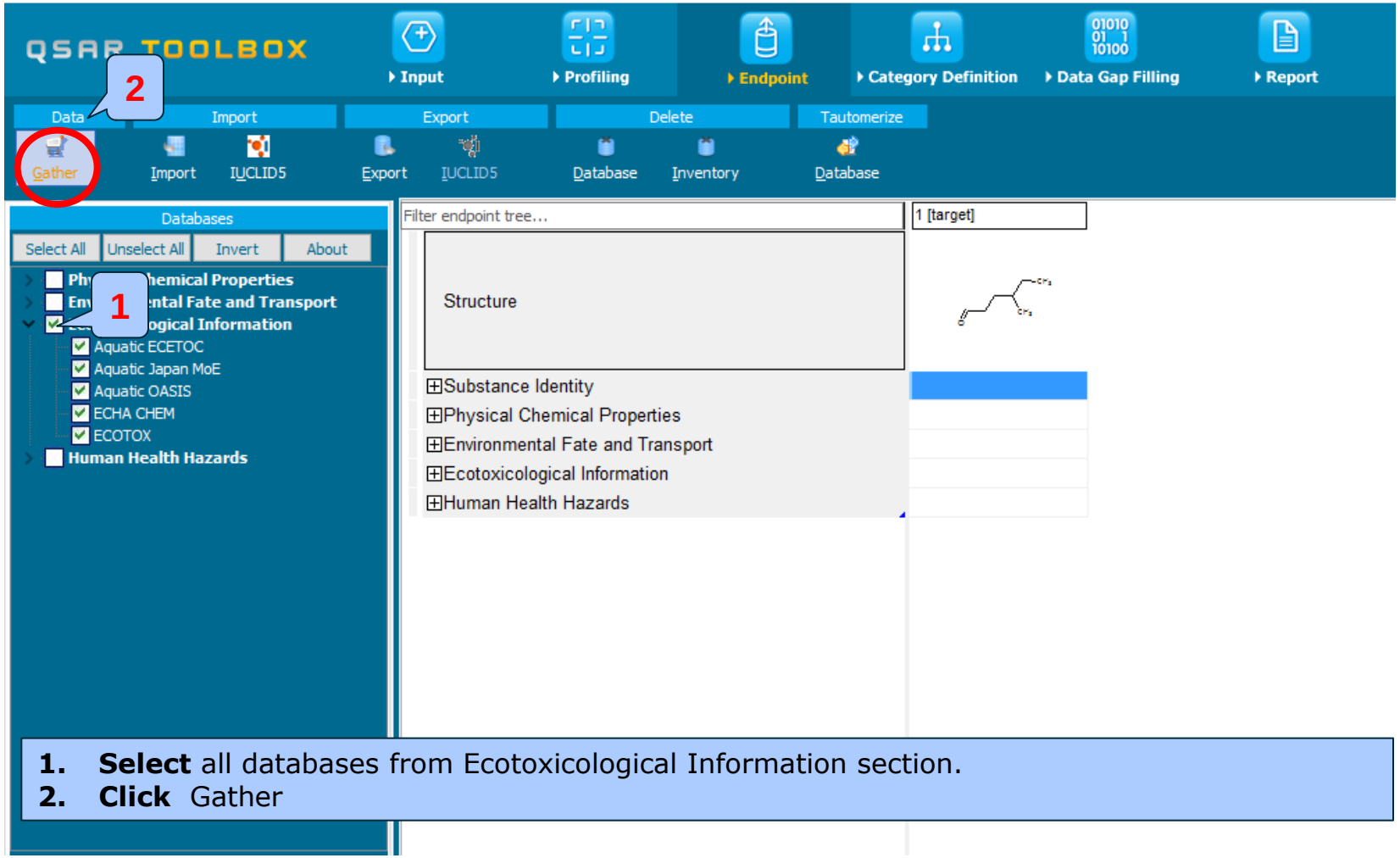
Endpoints
Overview

- As you may remember, "Endpoints" refer to the electronic process of retrieving the measured data for fate and toxicity 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).

Endpoints
Case study

- We will gather data for acute aquatic toxicity endpoints
- For this purpose, data are collected from all Ecotox
- Click on “Select All” from Ecotoxicological part
- Click on “Gather data” (see next screen shot).

Endpoints
Gather data



1. **Select** all databases from Ecotoxicological Information section.
2. **Click** Gather

Endpoints
Case study

- Aquatic Toxicity information on the target chemical is electronically collected from the selected dataset(s).
- In this example, an insert window appears stating there was “no data found” for the target chemical .
- Close the insert window by right clicking on “OK”.

Endpoints
Recap

- The workflow in this exercise is the same as you have observed with the preceding tutorials.
- You have entered the target chemical being sure of the correct structure.
- You have profiled the target chemical.
- You have found that no experimental data is currently available for this structure.
- In other words, you have identified a data gap, which you would like to fill.
- Proceed by right clicking on "Category definition" which move you to the next module.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - **Category definition**

Category Definition
Overview

- As before, this module provides the user with several means of grouping chemicals into a category that includes the target molecule.
- Remember this is the critical step in the workflow.
- Several options are available in the Toolbox to assist the user in refining the category definition. We will using several of these options in this tutorial.

Category Definition
Grouping methods

- You may want to review the information about grouping chemicals which is in (Chapter 4) on the following link. http://www.oecd.org/dataoecd/58/56/46210452.pdf
- Re-examine slide 31-32; you see that the target chemical (4-Methyl hexanal) could react with proteins via Schiff base formation and thus it has a potential for exhibiting aquatic toxicity in excess of baseline potency.
- The reaction by which a target chemical binds with proteins is relevant to deriving a group of chemicals that may mechanistically act the same way to elicit aquatic toxicity.

Category Definition
Protein binding by OASIS v1.4 grouping method

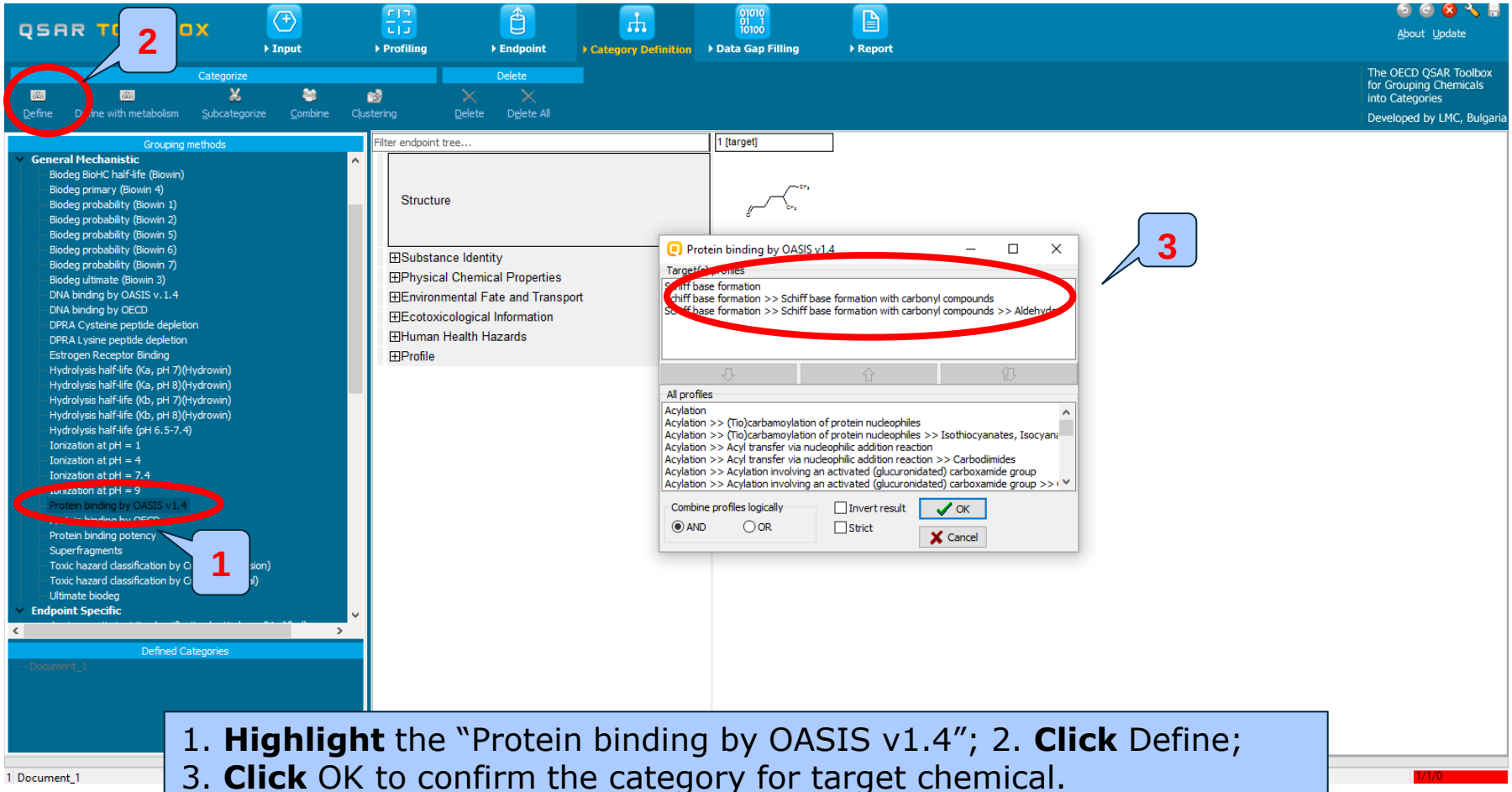
- Protein binding is one of the best grouping methods in the Toolbox. It is based on conventional organic chemical mechanisms and reactions, and as such is qualitative in character.
- This grouping method is particularly relevant for respiratory and skin sensitization and acute aquatic toxicity, but also relevant to chromosomal aberration and acute inhalation toxicity.

Category Definition
Side bar to Protein binding by OASIS v1.4 grouping method

- Protein binding by OASIS v1.4 includes 617 chemical categories.
- Each category is represented by defined 2-dimentional structural alerts that are associated with chemicals that act as electrophiles and covalently react with various moieties, in particular thiol (-SH) and amino ($-NH_2$) groups in proteins.
- Therefore, there is a sound mechanistic basis for using this grouping method.

Category Definition

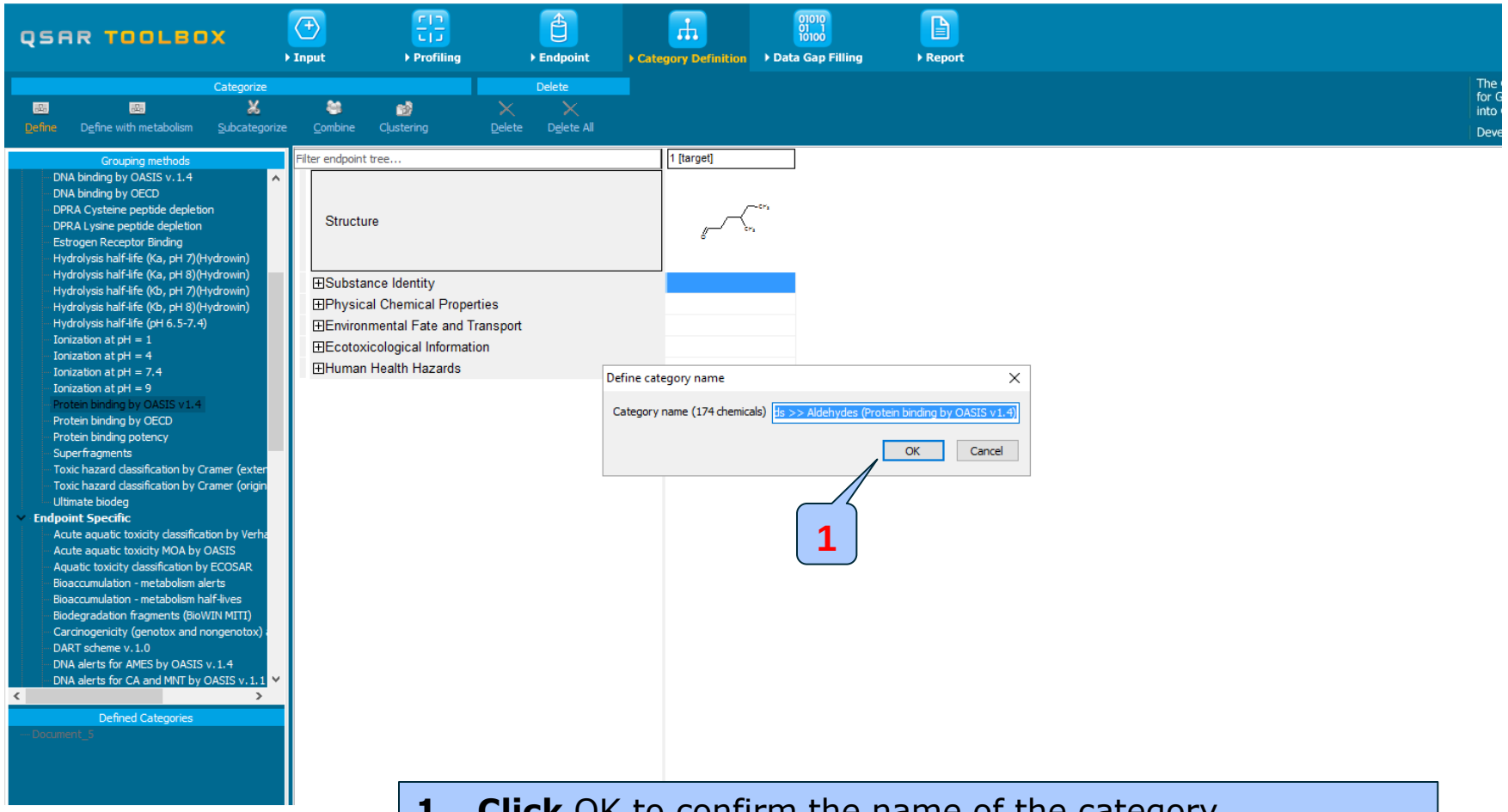
Defining Protein binding by OASIS v1.4 category



1. **Highlight** the "Protein binding by OASIS v1.4"; 2. **Click** Define;
3. **Click** OK to confirm the category for target chemical.

Category Definition

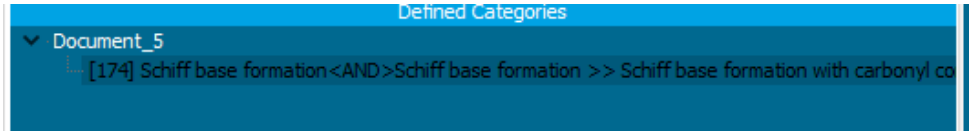
Defining Protein binding by OASIS v1.4 category



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

Category Definition
Analogues

- The data is automatically collected.
- Based on the defined category (Schiff base formation<AND>Schiff base formation >> Schiff base formation with carbonyl compounds<AND>Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes) 174 analogues* (plus the target chemical) have been identified.
- These 174 compounds along with the target chemical form a category, which can be used for data filling (see next screen shot).



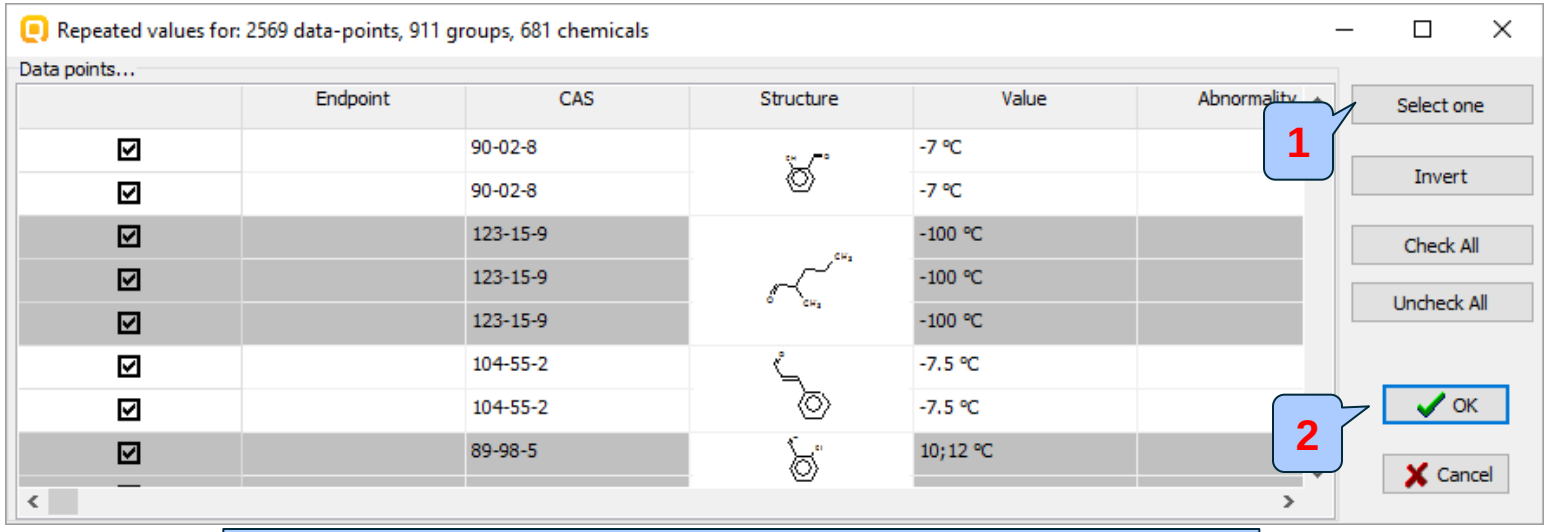
- The name of the category appears in the "Defined Categories" window, indicating the number of substances belonging to the category.

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).

Category Definition
Read data for Analogues

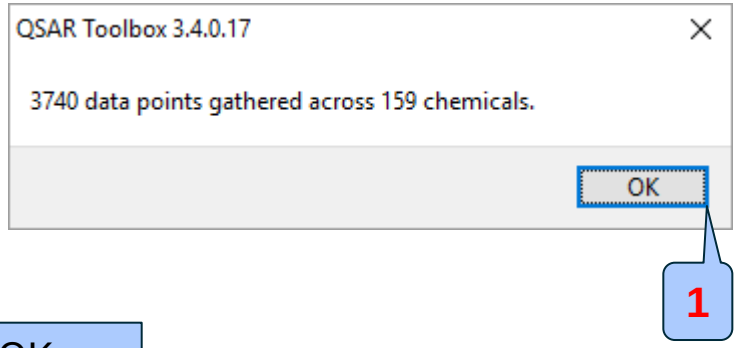
Due to the overlap between the Toolbox databases same data for intersecting chemicals could be found simultaneously in more than one databases. The data redundancy is identified and the user has the opportunity to select either a single data value or all data values.



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

Category Definition
Read data for Analogues

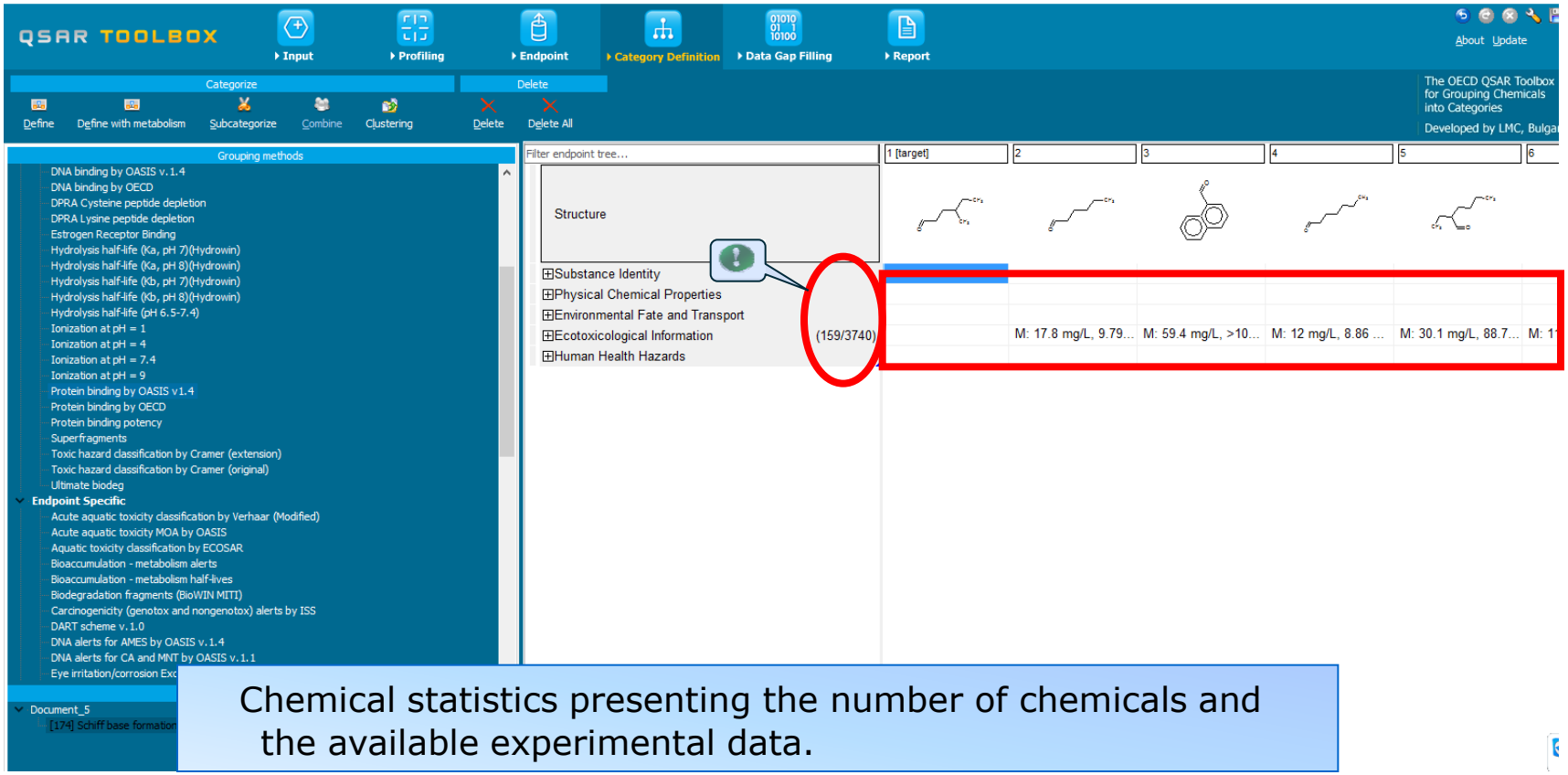
The system automatically gives indication for the number of gather experimental data points



1. **Click** OK

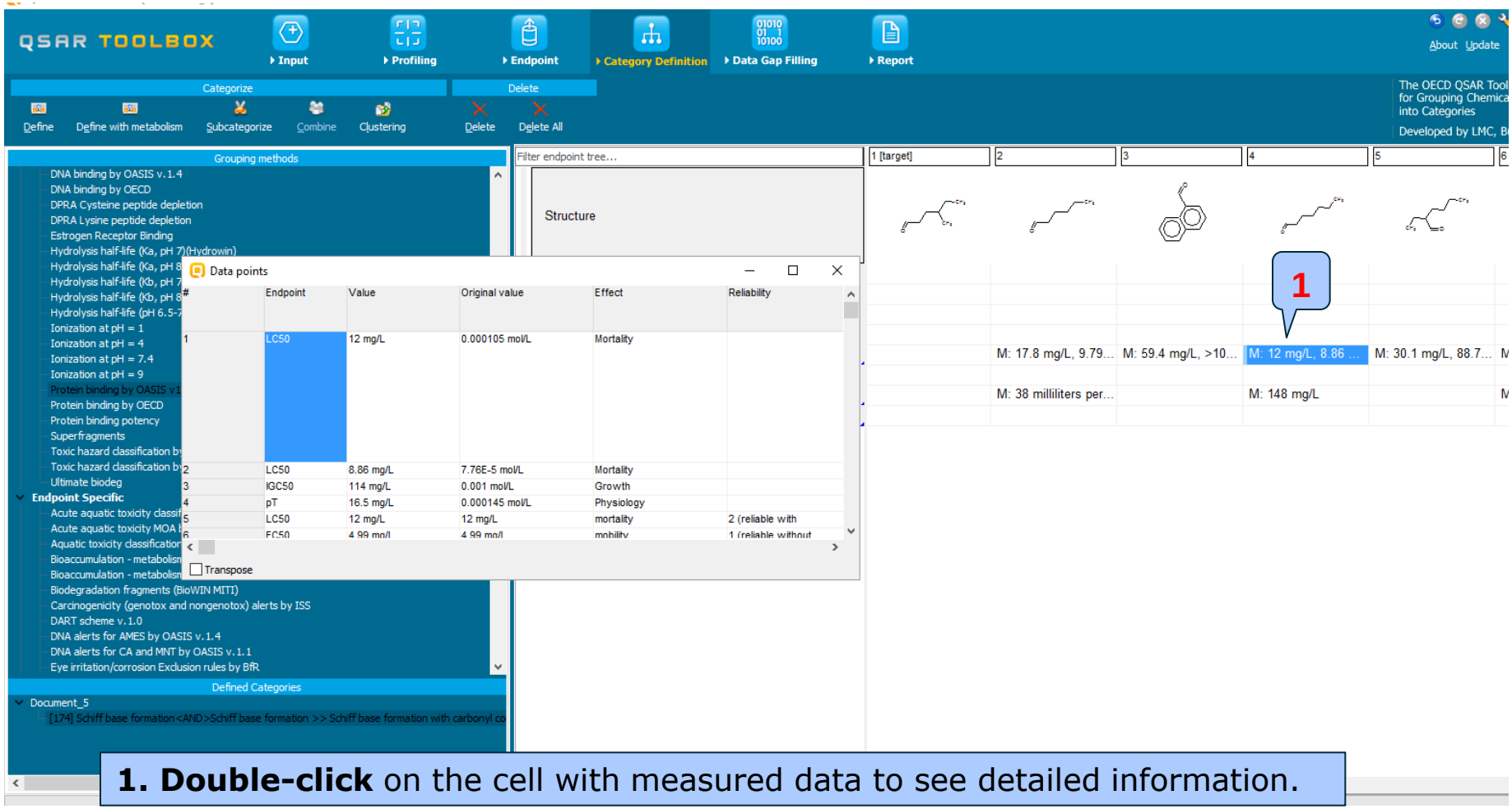
Category Definition
Summary information for Analogues

- The experimental results for the analogues are inserted into the data matrix.



Category Definition

Side bar of experimental data



1. Double-click on the cell with measured data to see detailed information.

Category Definition
Recap

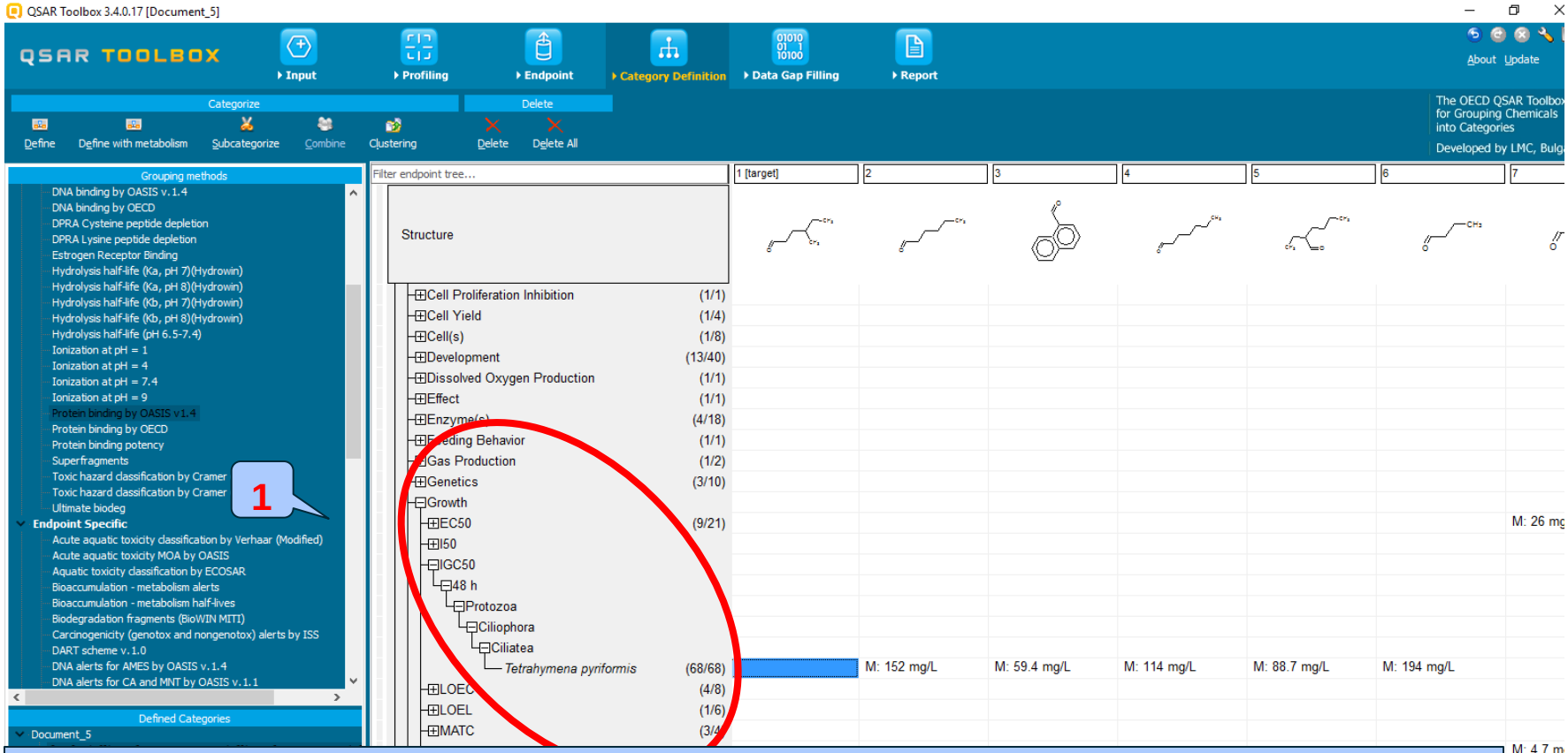
- You have identified a mechanistic category by Protein binding by OASIS v1.4 for the target chemical (4-Methy hexanal).
- The available experimental results for these 174 analogues are collected from the previously selected databases.
- The user can then proceed to the module "Data gap filling", but before that the user should navigate through the endpoint tree and find the gap that will be filled in.

Category Definition
Navigation throw the endpoint tree

- You can proceed to the module "Data gap filling", but before that we will navigate through the endpoint tree and find the gap that will be filled.
- As before we navigate through the data tree by closing or opening the nodes of the tree.
- In this example, the IGC50-48h Growth for *Tetrahymena pyriformis* is the target endpoint (see next screen shot).

Category Definition

Navigation throw the endpoint tree



1. **Double click** on the specific nodes to reach the leaf of the target endpoint – Ecotoxicological information; Aquatic toxicity; Growth; IGC50; 48h; Protozoa; Ciliophora; Ciliatea; Tetrahymena pyriformis.

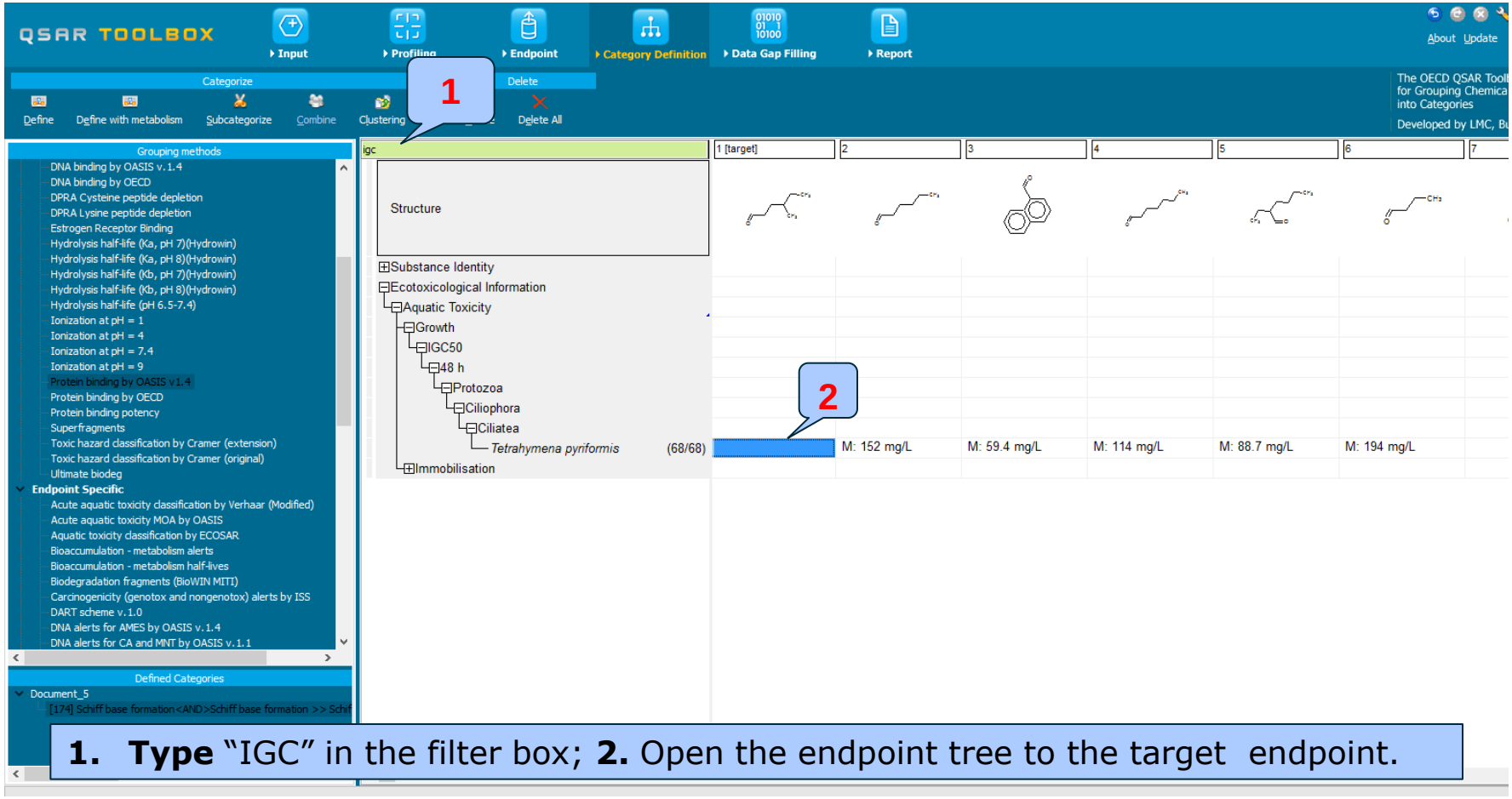
Category Definition

Navigation throw the endpoint tree with the filter box

- Remember you can also navigate through the endpoint tree by typing the endpoint of choice in the filter box and then double clicking through the endpoint tree to IGC50-48h Growth for *Tetrahymena pyriformis* (see next screen shot).

Category Definition

Navigation through the endpoint tree - use filter



1. Type "IGC" in the filter box; **2.** Open the endpoint tree to the target endpoint.

Outlook

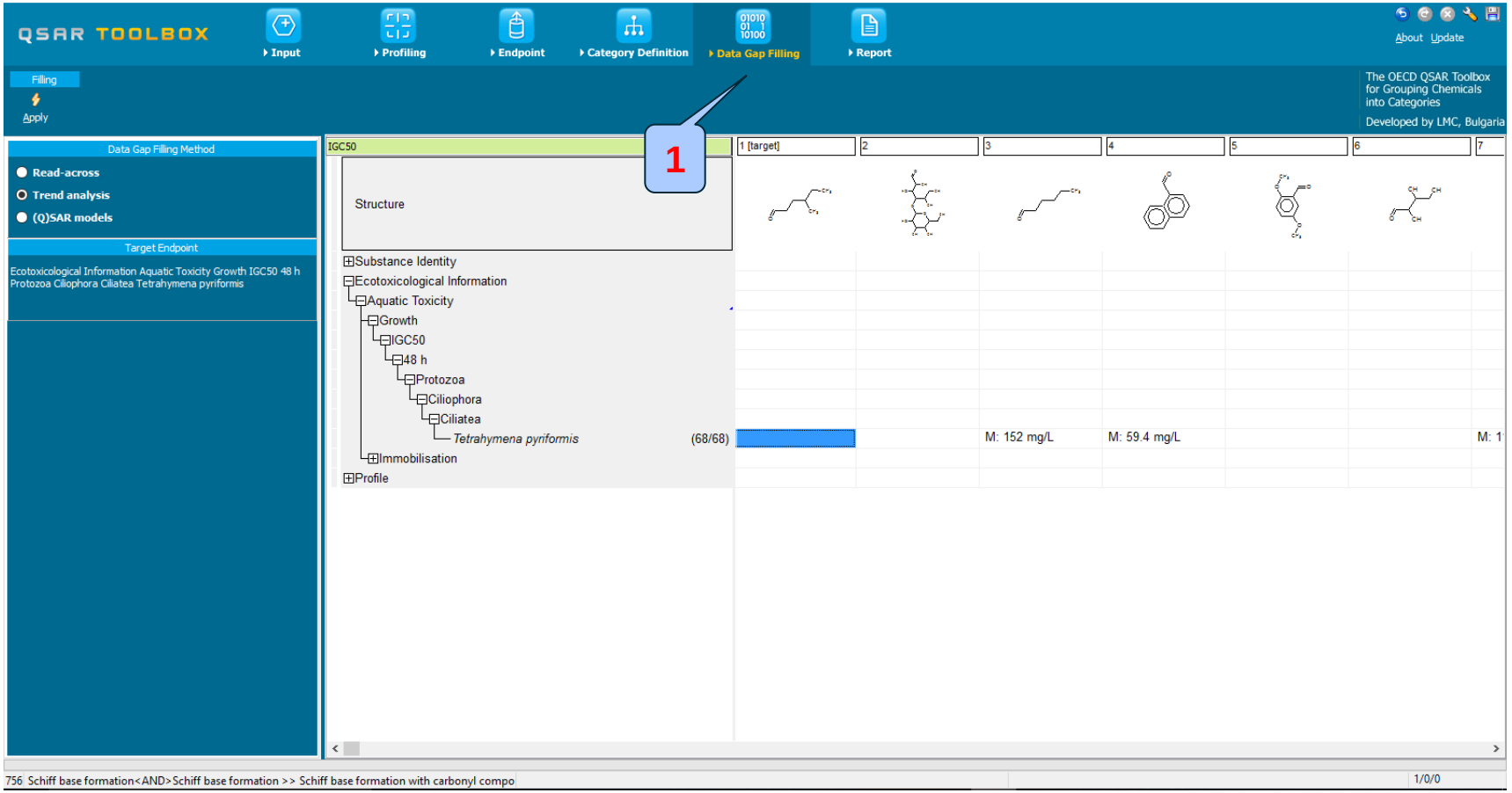
- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - Category definition
  - **Data Gap Filling**

Data Gap Filling
Overview

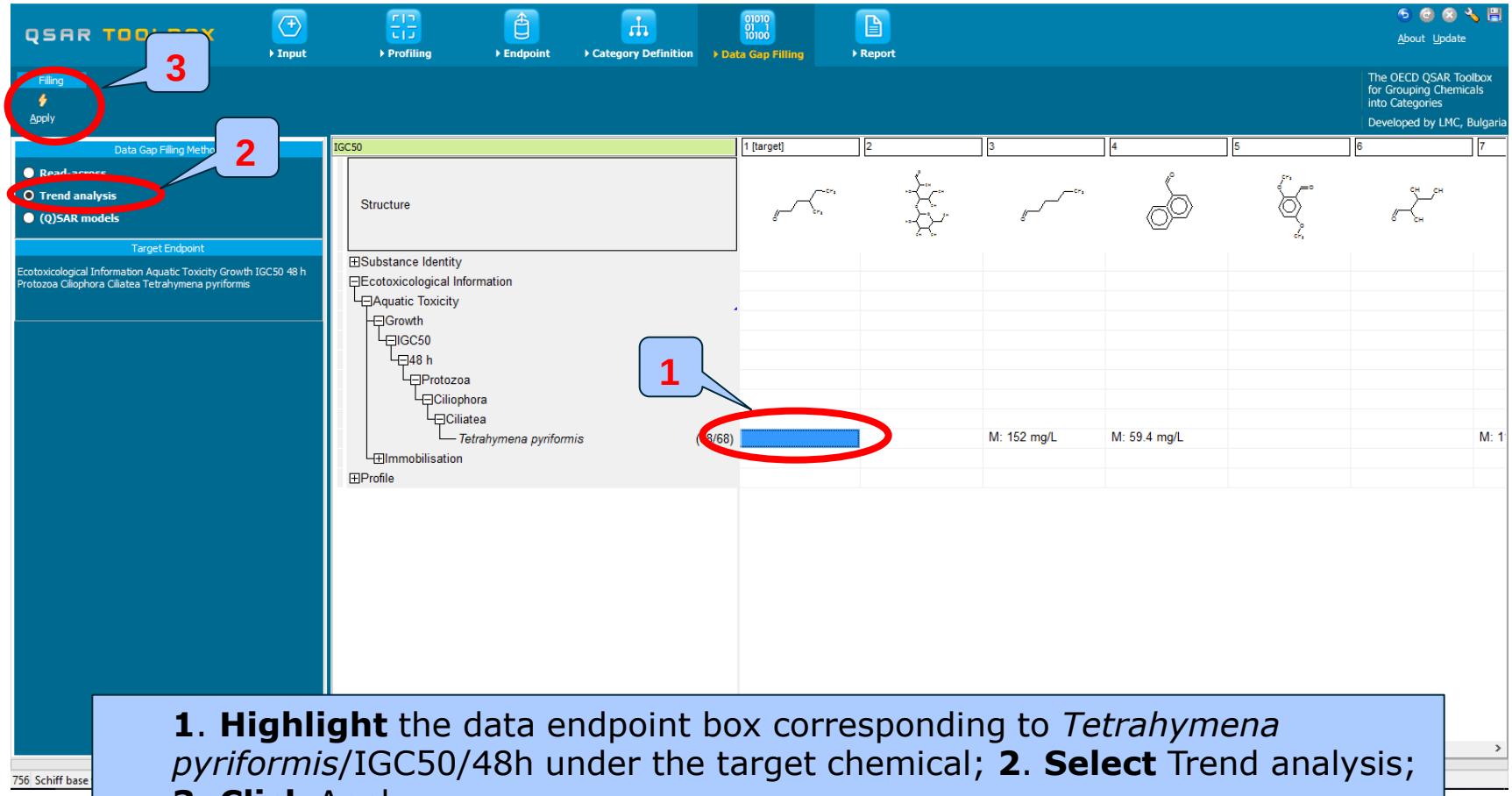
- **"Data Gap Filling" module give access to three different data gap filling tools:**
  - Read-across
  - Trend analysis
  - (Q)SAR models
- **Depending on the situation, the most relevant data gap mechanism should be chosen, taking into account the following considerations:**
  - Read-across is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitization or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore read-across is recommended for "quantitative endpoints" (e.g., 96h-LC50 for fish) if only a low number of analogues with experimental results are identified.
  - Trend analysis is the appropriate data-gap filling method for "quantitative endpoints" (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.
  - (Q)SAR models can be used to fill a data gap if no adequate analogues are found for a target chemical.
- **In our case trend analysis is applied**

Data Gap Filling
Data Gap window

1. Click on Data gap filling.



Data Gap Filling
Apply Trend analysis

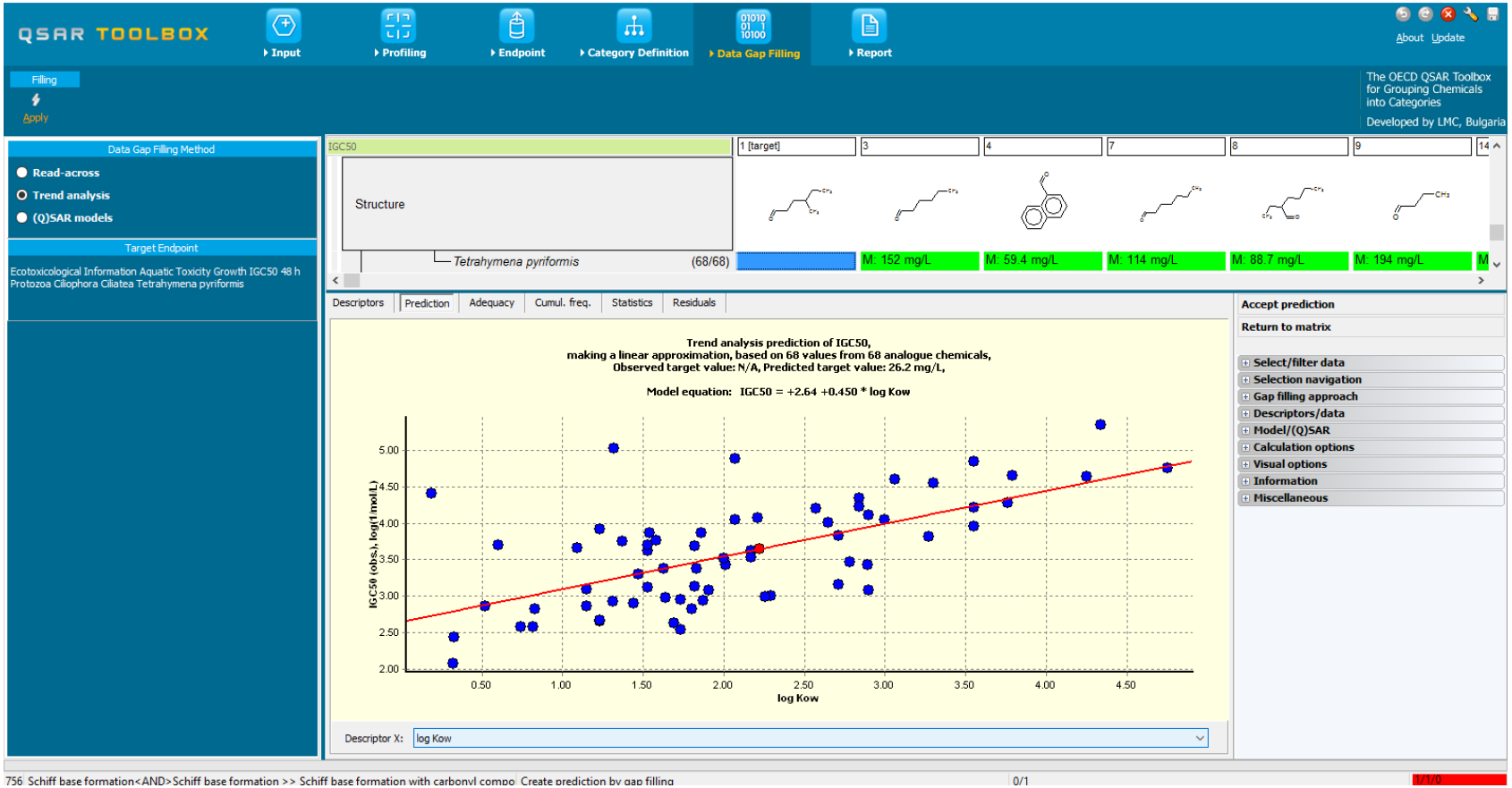


1. **Highlight** the data endpoint box corresponding to *Tetrahymena pyriformis*/IGC50/48h under the target chemical; **2**. **Select** Trend analysis; **3**. **Click** Apply.

Data Gap Filling
Interpreting Trend analysis

- The resulting plot outlines the experimental IGC50 results of all analogues (Y axis) according to a descriptor (X axis) with Log Kow being the default descriptor (see next screen shot).
- The **RED** dot represents the estimated result for the target chemical.
- The **BLUE** dots represent the experimental results available for the other analogues.
- The **GREEN** dots (which we will see later) represent analogues belonging to different subcategories.
- Before accepting the estimated result for the target chemical, the trend analysis should be further refined by subcategorisation.

Data Gap Filling
Results of Trend analysis



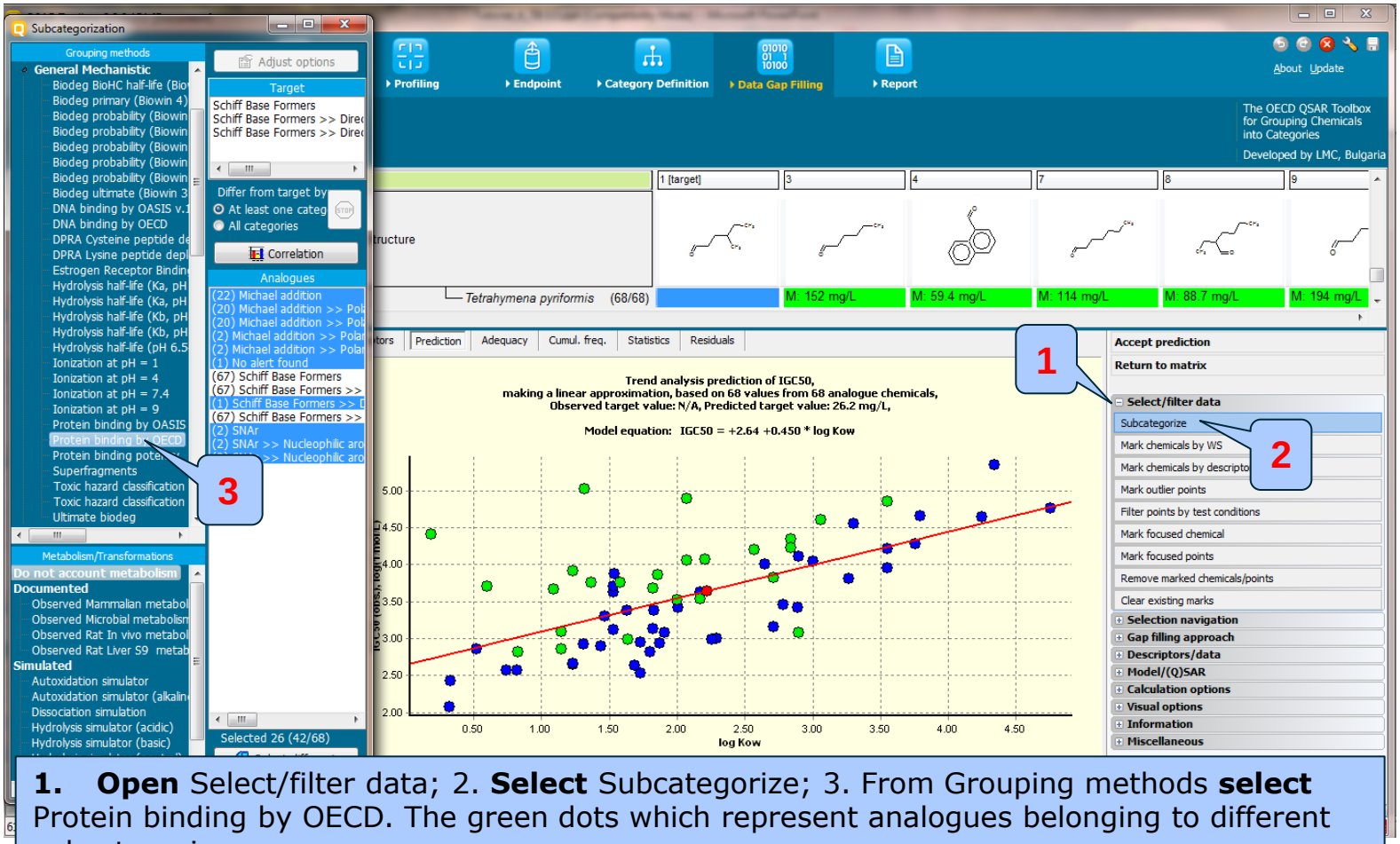
Data Gap Filling
Side Bar of Subcategorization

- Subcategorisation refers to the process of applying additional profilers to a previously defined category.
- The goal here is to identify chemicals which have differing profiling results, eliminating these chemicals from the category, and eventually getting a more homogenous chemical category.
- In this example, two subcategorisations are applied to prune the analogues.

Data Gap Filling
Subcategorization

- Due to polyfunctionality of molecules, there are analogues which may undergo protein binding reactions different from those of the target compound.
- In addition, there are analogues which contain organic functional group that differ from those of the target compound.
- These analogues can be identified and later removed via subcategorisation.

Data Gap Filling
Subcategorization 1 (Protein binding by OECD)



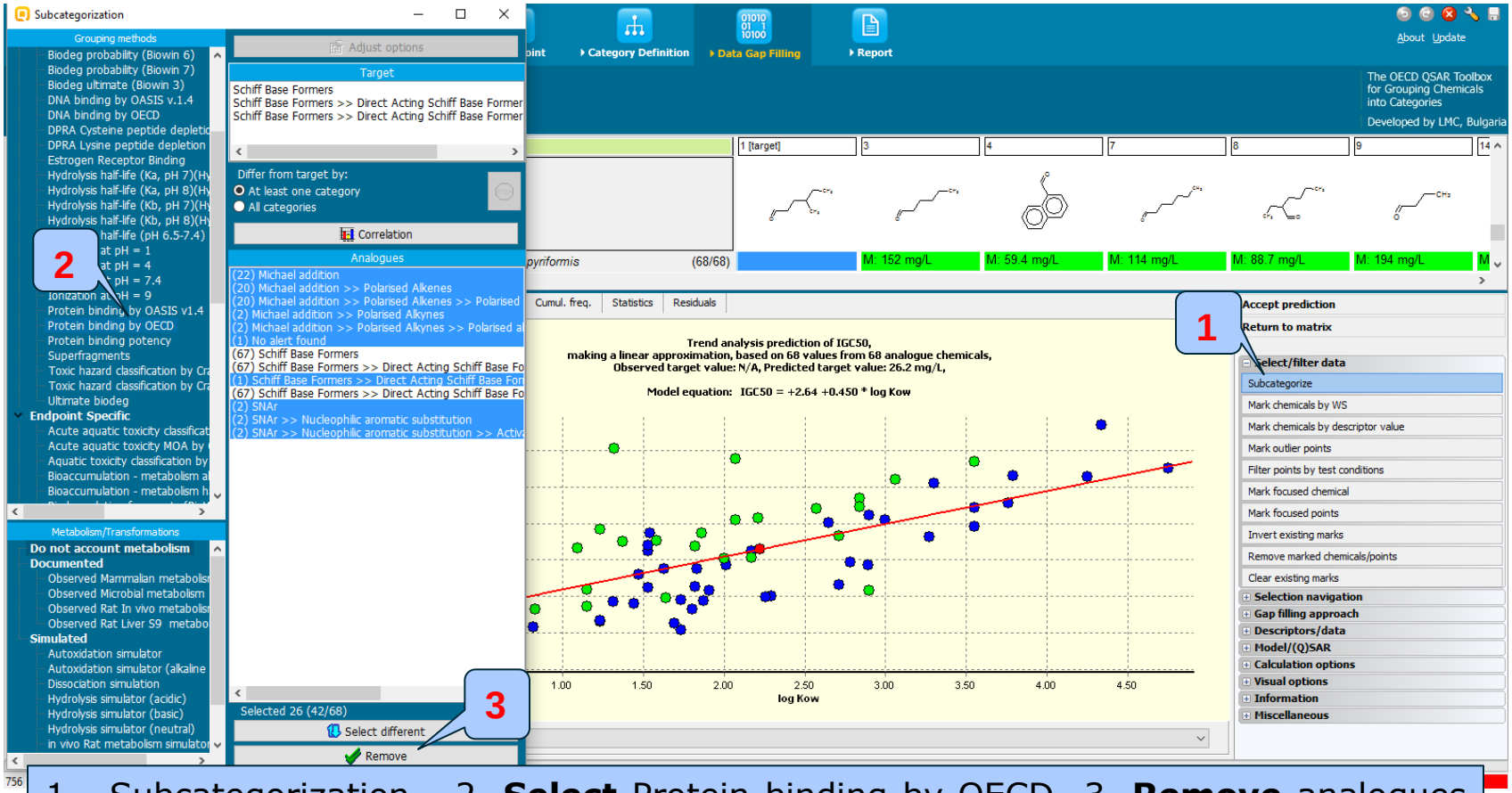
1. Open Select/filter data; 2. **Select** Subcategorize; 3. From Grouping methods **select** Protein binding by OECD. The green dots which represent analogues belonging to different subcategories.

Data Gap Filling
Information on an Analogue

- Additional information may be observed for any analogue in the trend analysis.
- Analogues which are different from the target chemical may be removed from the category.
- In this example, we will examine the upper most data point (see the next series of screen shot).

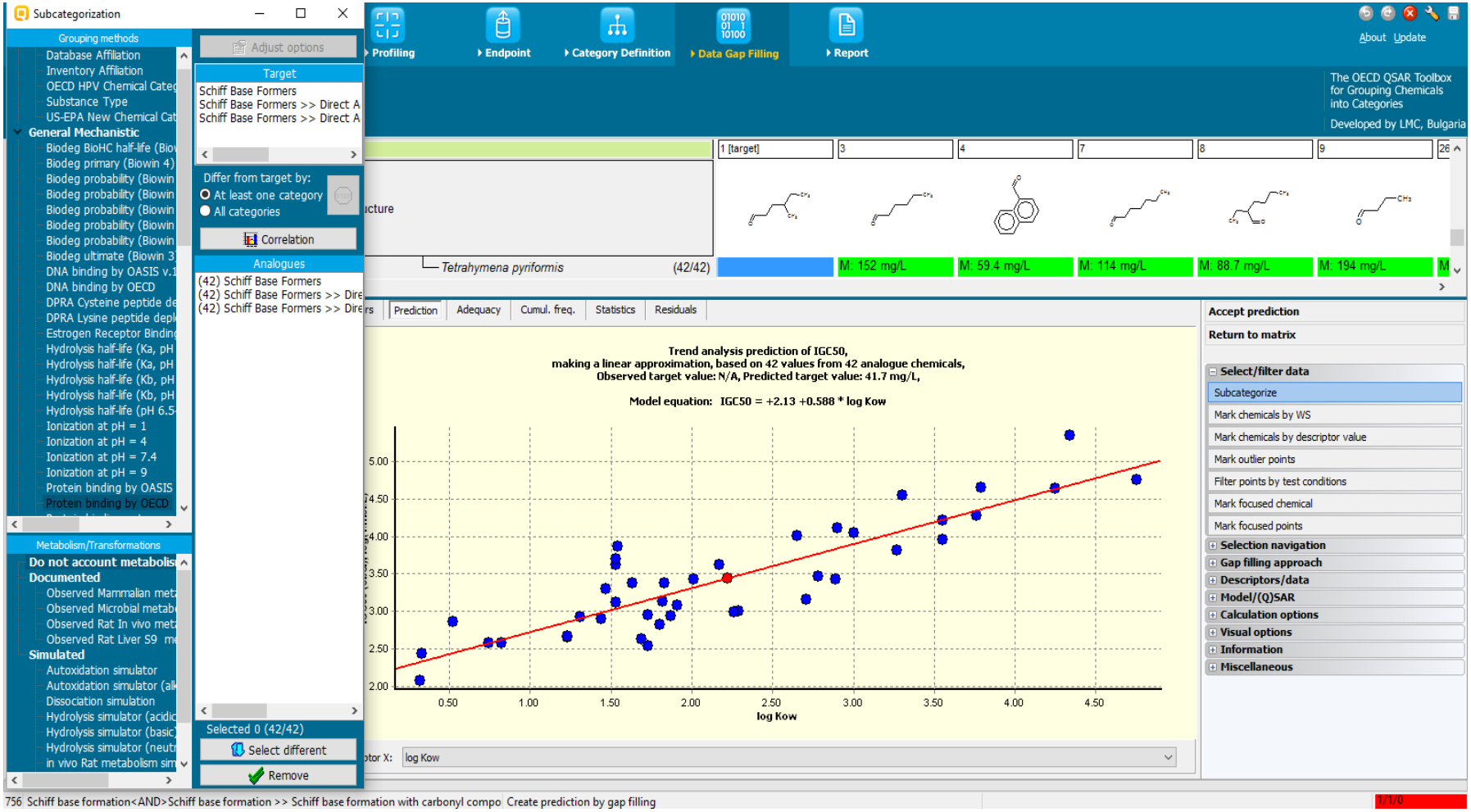
Data Gap Filling

Subcategorization 1 (Protein binding by OECD)



1. Subcategorization 2. **Select** Protein binding by OECD 3. **Remove** analogues which have different protein binding mechanism than the target chemical.

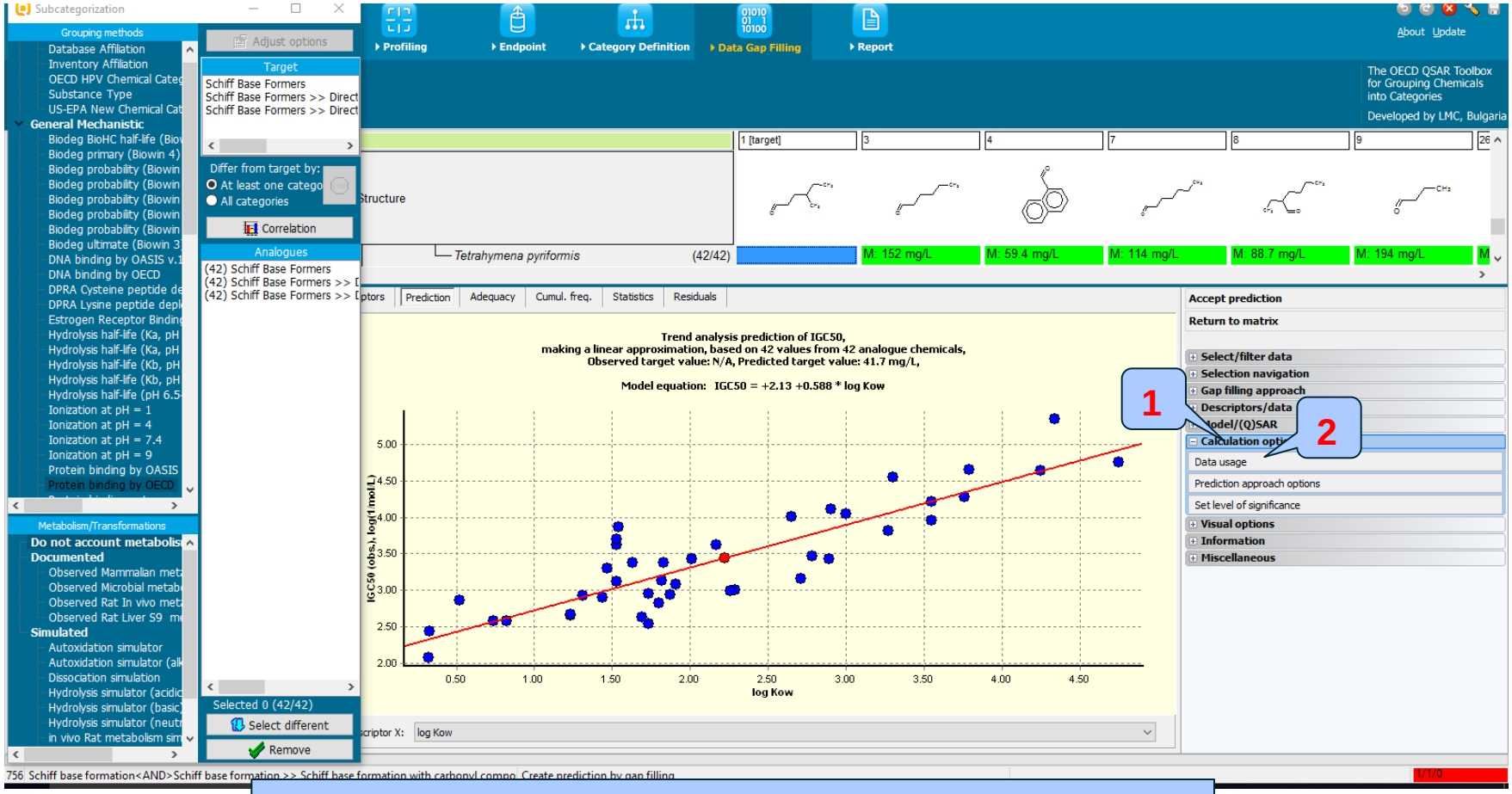
Data Gap Filling
Subcategorization 1 (Protein binding by OECD)



Data Gap Filling
Selecting the Data Type

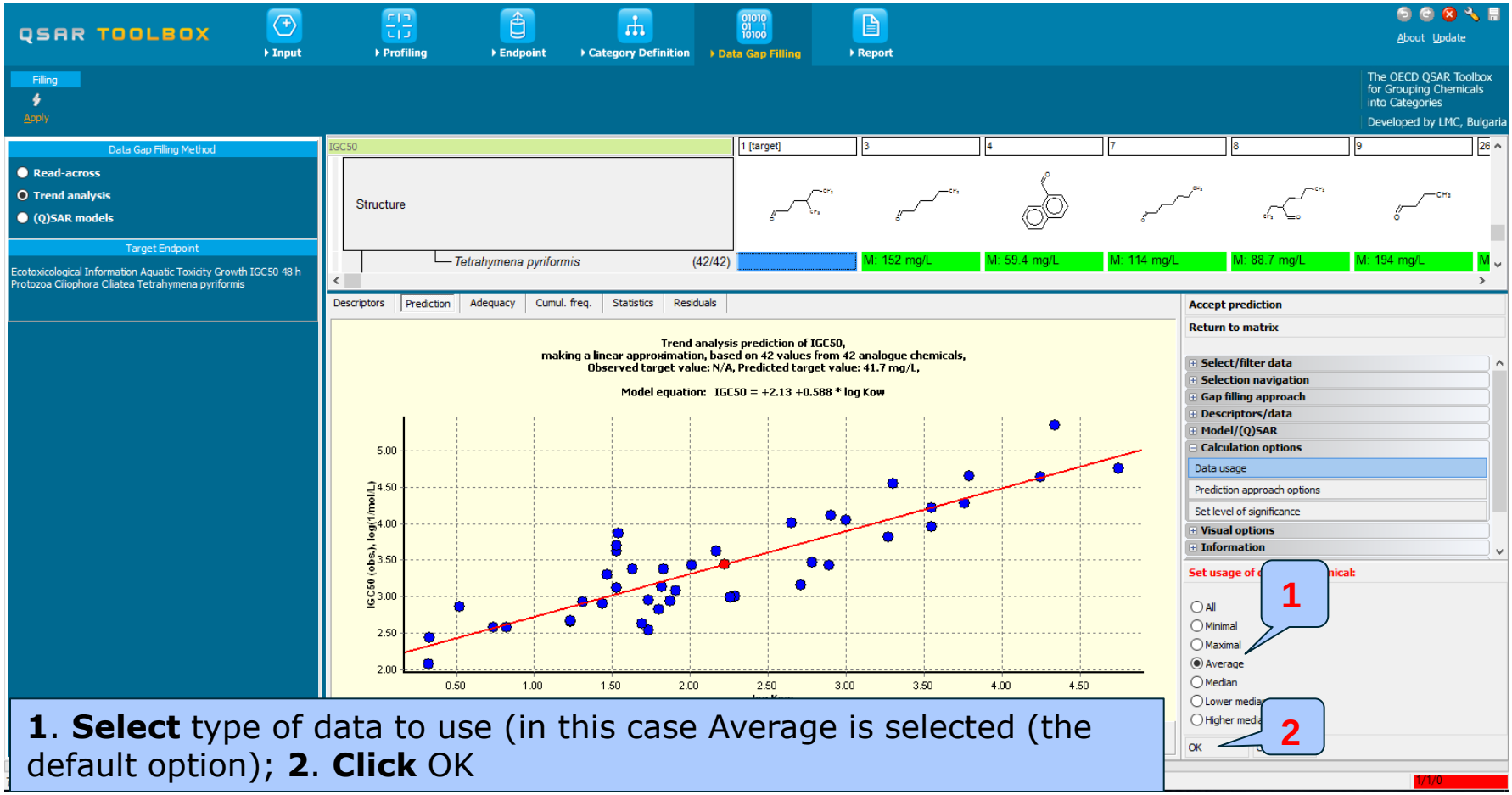
- The Toolbox allows the user to decide which type of data to be used in cases where more than one result are available for any analogue.
- This can be all values, average values, minimum or maximum results (see next screen shot).
- It should be noted that averaging results is only useful for quantitative endpoints, which is the case in this example.

Data Gap Filling
Selecting the Data Type



1. **Open** Calculation options; 2. **Select** Data usage

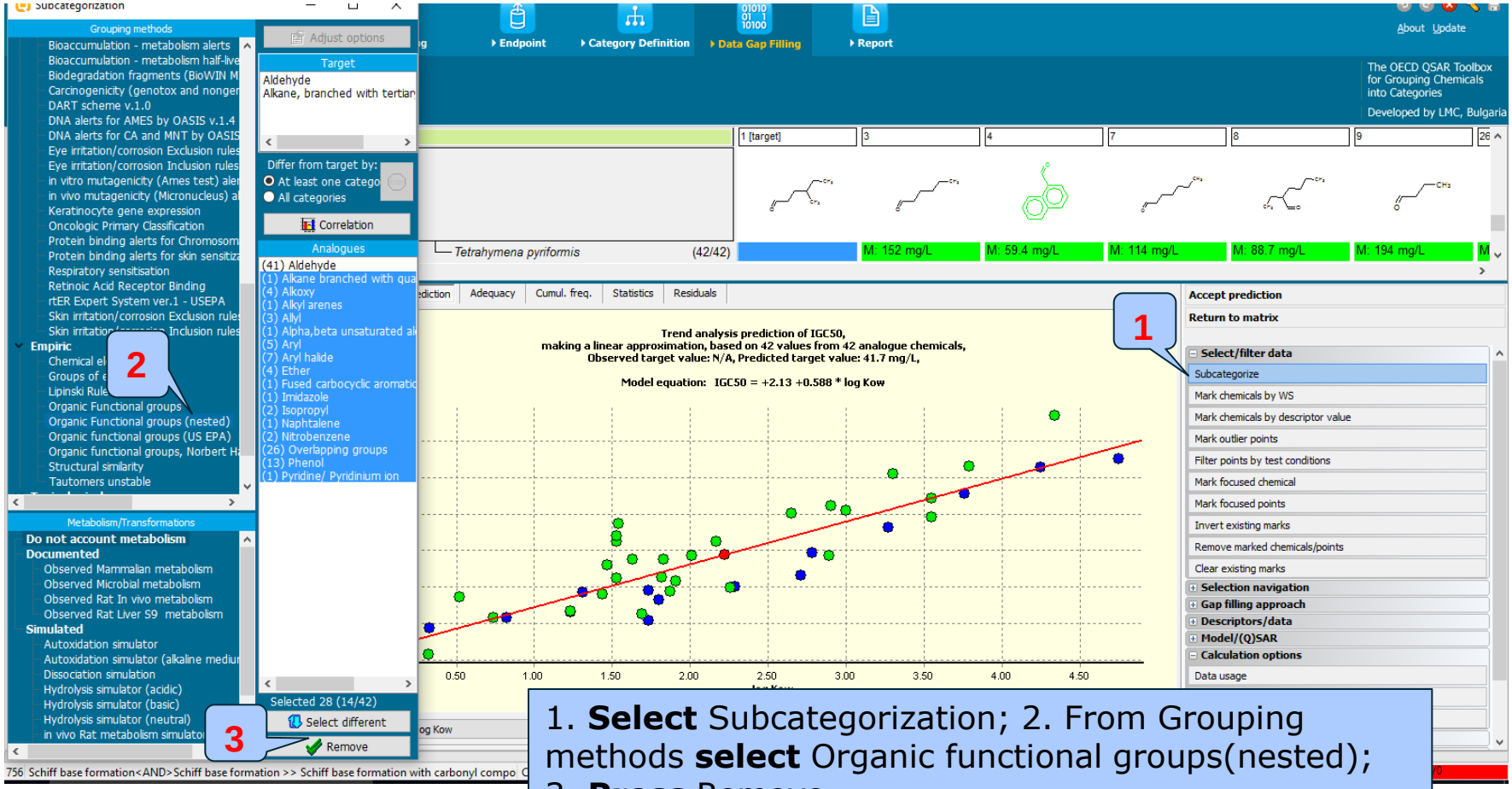
Data Gap Filling
Selecting the Data Points



1. **Select** type of data to use (in this case Average is selected (the default option); **2**. **Click** OK

Data Gap Filling

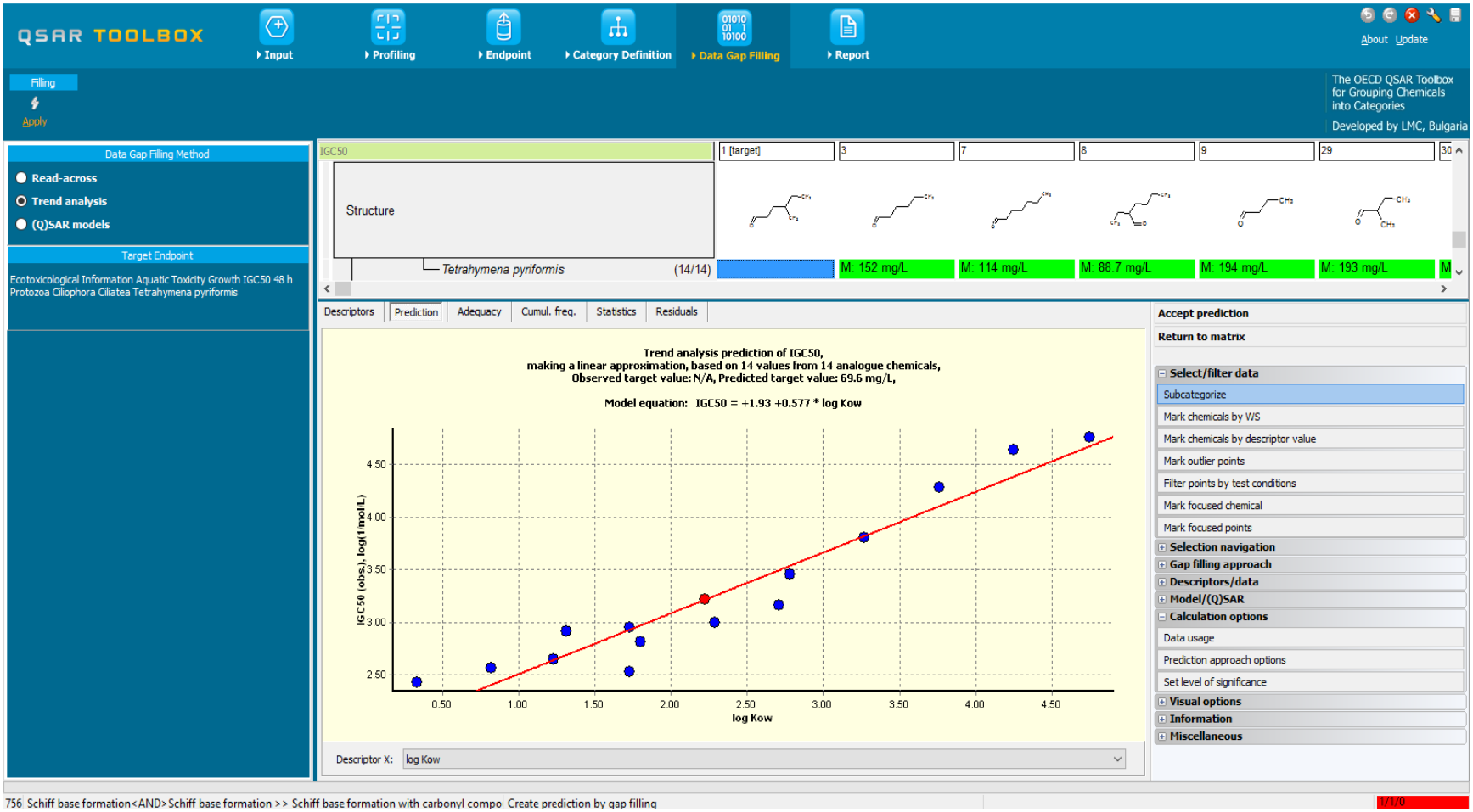
Subcategorization 2 (Organic functional groups (nested))



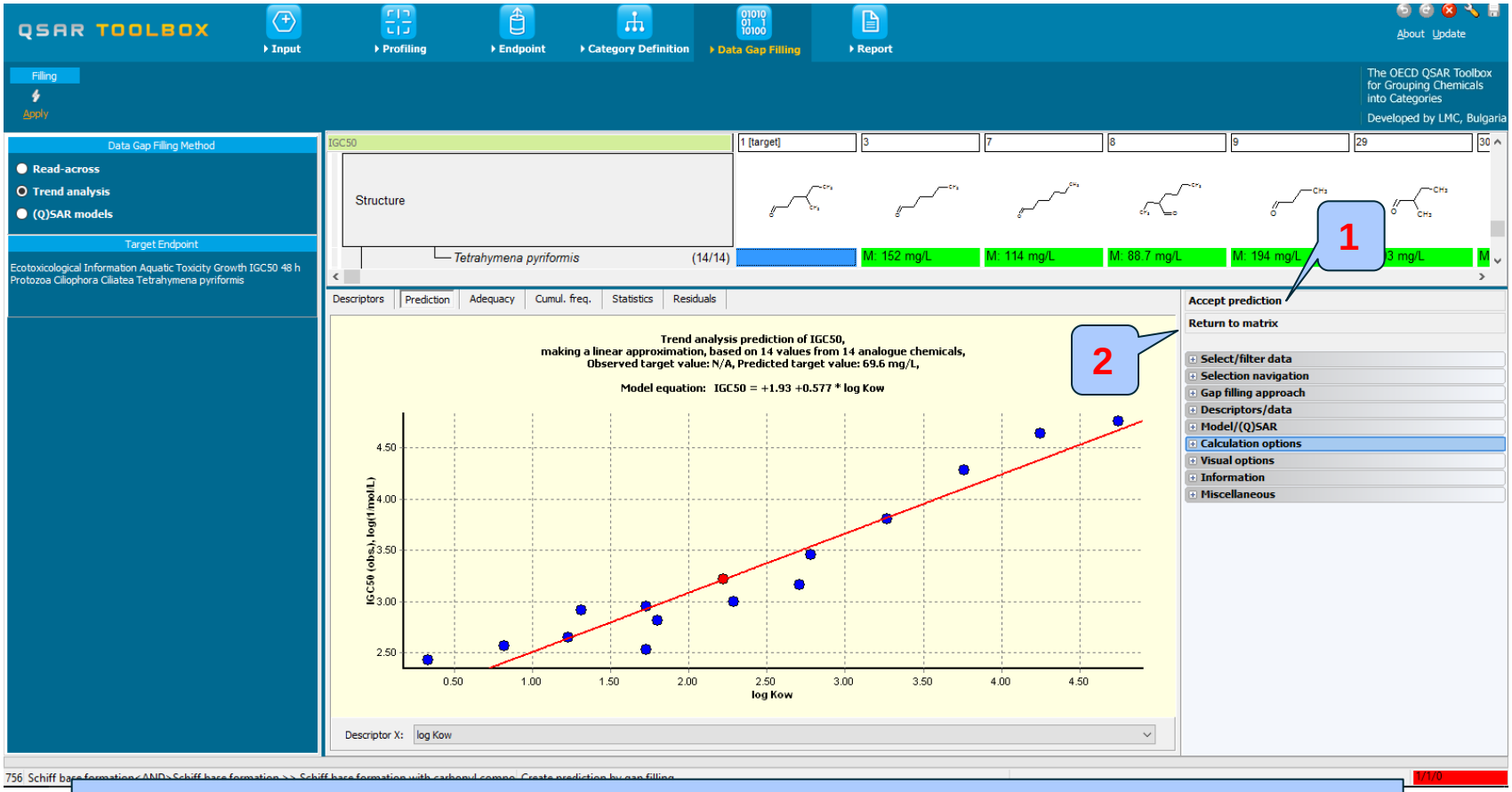
1. **Select** Subcategorization; 2. From Grouping methods **select** Organic functional groups(nested); 3. **Press** Remove.

Data Gap Filling

Subcategorization 2 (Organic functional groups (nested))



Data Gap Filling
Accepting the prediction result



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

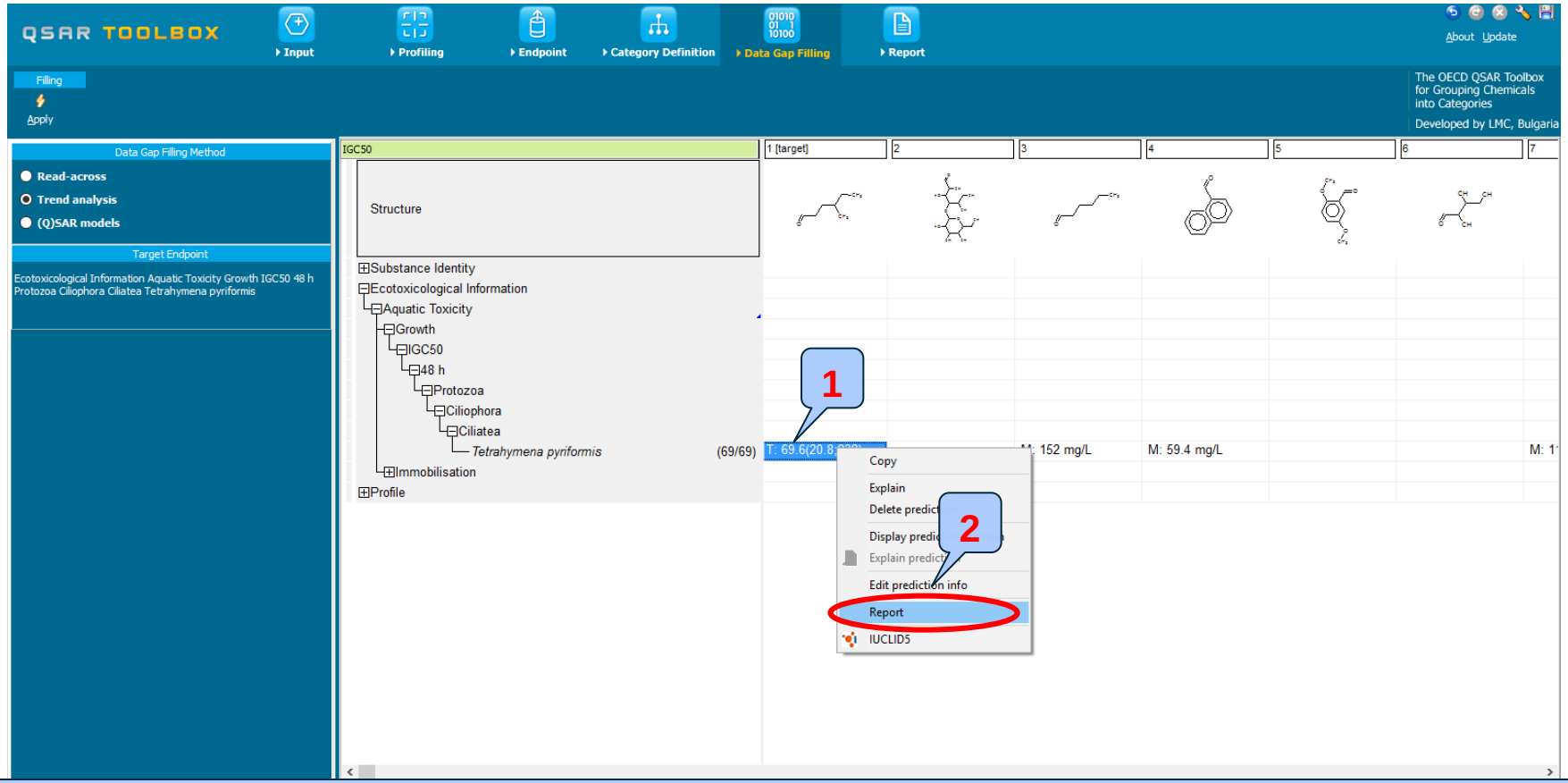
Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - Category definition
  - Data Gap Filling
  - **Reporting**

Report
Overview

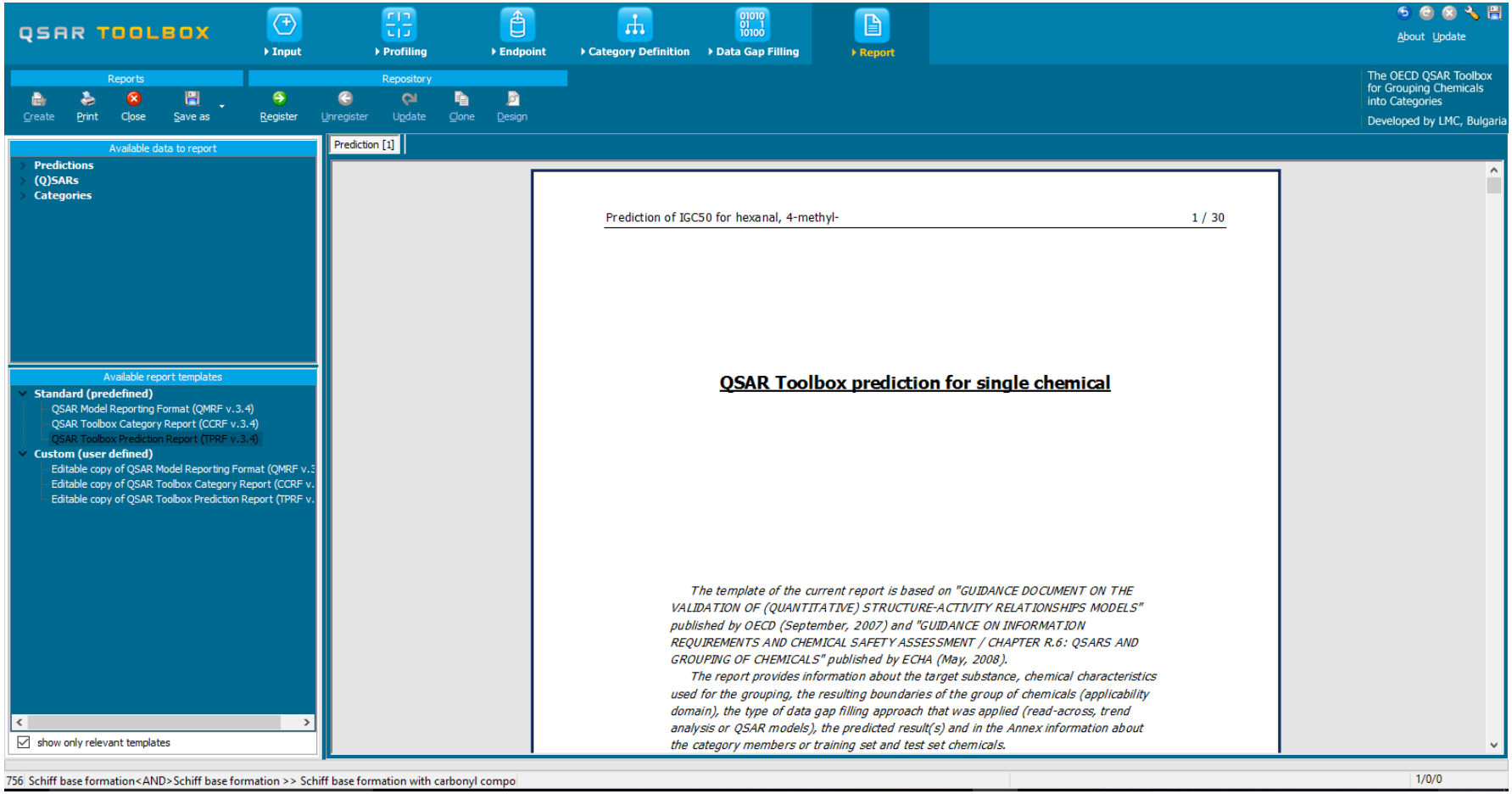
- Remember, the report module generates reports on any of predictions made with the Toolbox.
- The report module contains predefined report templates as well as a template editor with which users can define their own “user defined” templates.
- The report can then be printed or saved in different formats.

Report
Generate Report



1. **Right click** on the cell which contains the predicted value;2. **Select** Report from the menu.

Report
Generate Report



Outlook

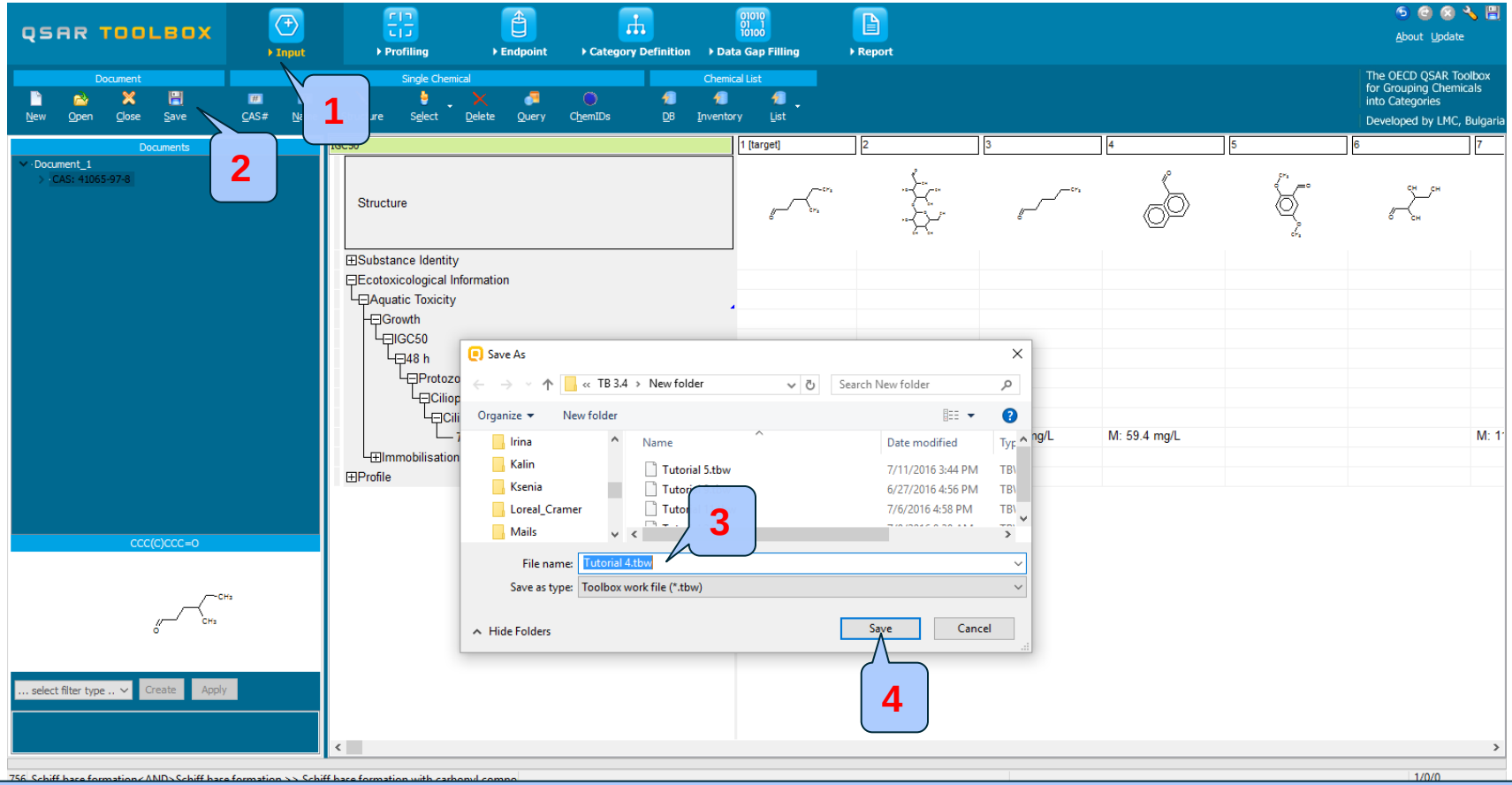
- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- Workflow
- **Save the prediction result**

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).

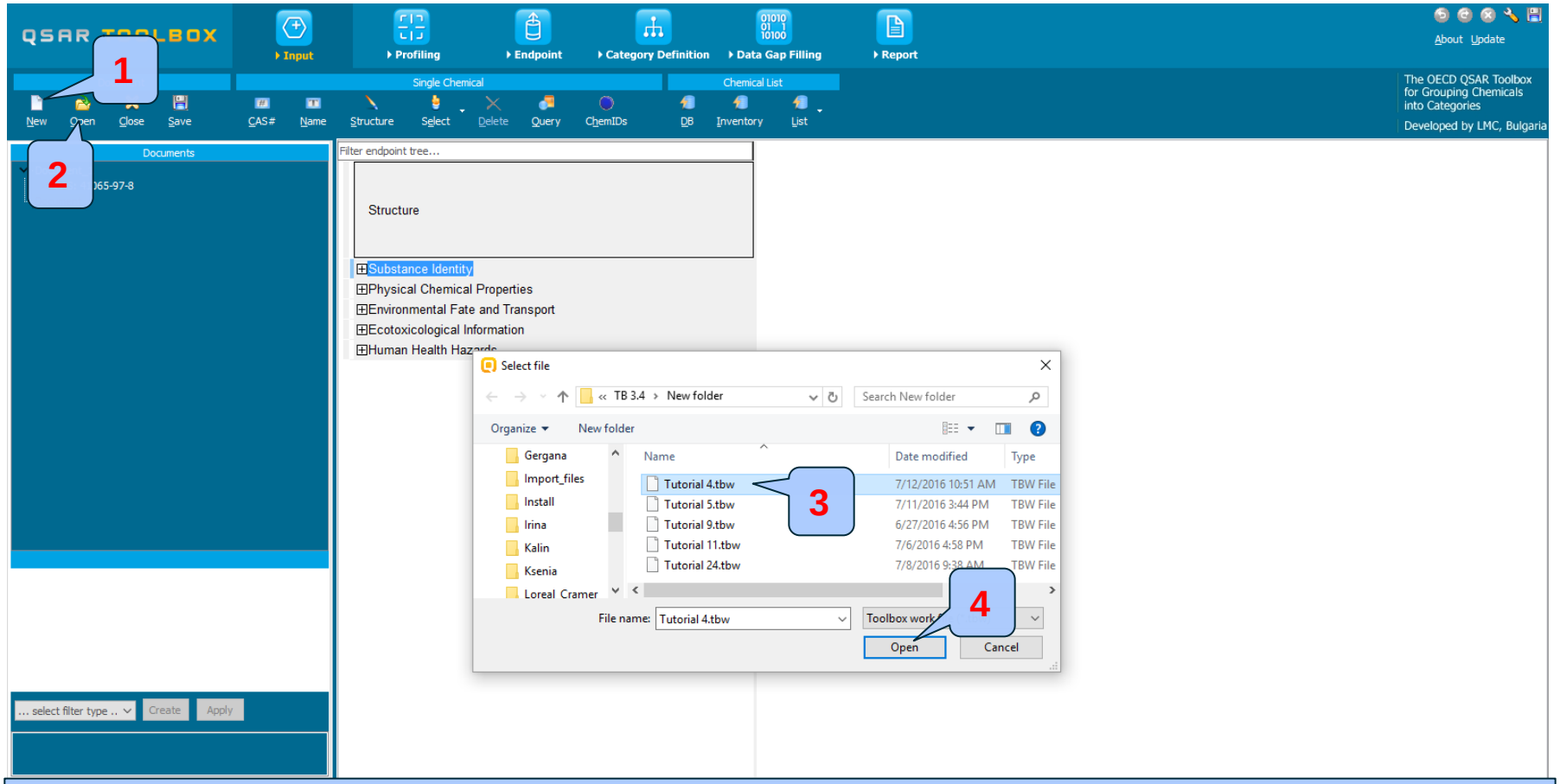
- Saving/Loading the file with TB prediction is shown on next screenshots

Saving the prediction result



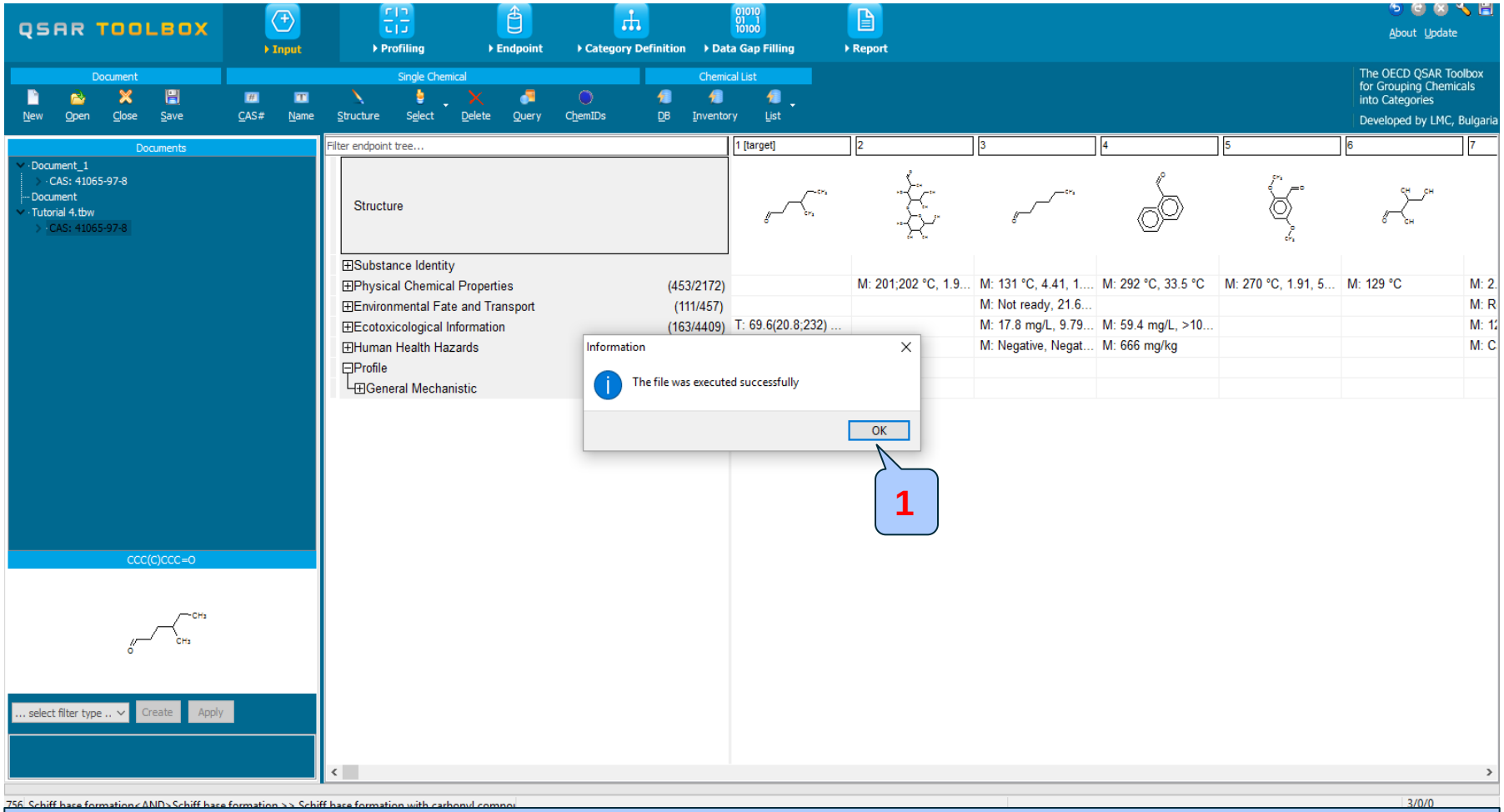
1. **Go** to Input section 2.**Click** on Save button 3. **Define** name of the file; 4. **Click** Save button

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

Congratulation

- By now you should feel comfortable with the six basic modules of the toolbox and how they form the workflow of the Toolbox.
- In this tutorial you have been introduced to several additional function in the Toolbox.
- You have used different profilers in subcategorizing the initial category of the target chemical.
- Remember proficiency in using the Toolbox comes with practice.