QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD (Q)SAR Toolbox v.4.4.1

Manipulation of data matrix and manual transferring of data to the target outside of the data gap filling module

Outlook

- **Background**
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example

Background

- This is a step-by-step presentation designed to introduce the user to the newly created functionalities for manipulation of the data matrix.

- A simple example illustrating how to manually transfer a read-across prediction from source (analogue) to target chemical.

Outlook

- Background
- **Keywords**
- Objectives
- Specific aim
- Manipulation of data matrix
- Example

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
- **Objectives**
- Specific aim
- Manipulation of data matrix
- Example

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Rearrangement of columns of the data matrix;
- Filtering of parameters/experimental data or profiling results of the analogues within the category;
- Hide/show gap-filling chart while in Data gap filling module;
- Transferring of data to the target chemical outside Data gap filling module.

Outlook

- Background
- Keywords
- Objectives
- **Specific aim**
- Manipulation of data matrix
- Example

Aim

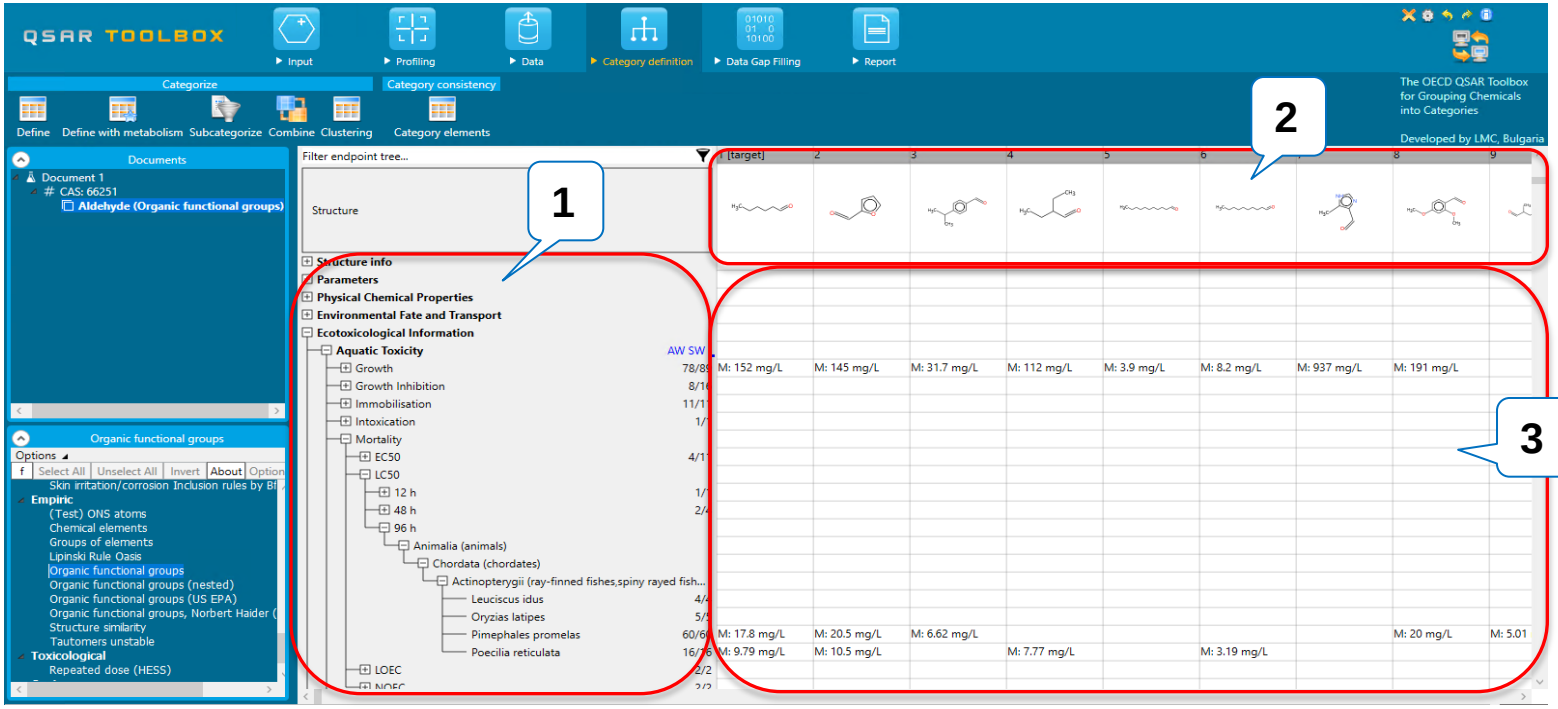
- To introduce and make the user familiar with:
  - Manipulation of the data matrix;
  - Filtering the data matrix with respect to parameters, experimental data that appear for the selected analogues;
  - Hide/Show the gap filling chart while the user is in the Gap filling module for better screening and analysis of data on the data matrix;
  - Transfer of data to target chemical outside Data gap filling.

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- **Manipulation of data matrix**
- Example

Data matrix
Overview

- The data matrix window has three main parts:
  - Area with the Endpoint tree (1)
  - Area with the selected chemicals (2) and
  - Area with data (experimental, predicted) (3)



Manipulation of data matrix
Implementation in Toolbox

The functionality allows the user manually to manipulate the matrix via:

1) Filtering the parameters and/or experimental data and/or profiling results which appear on the data matrix for the selected analogues;

2) Reordering the columns with analogues in the category in order to more effectively analyze the data between the target and analogues;

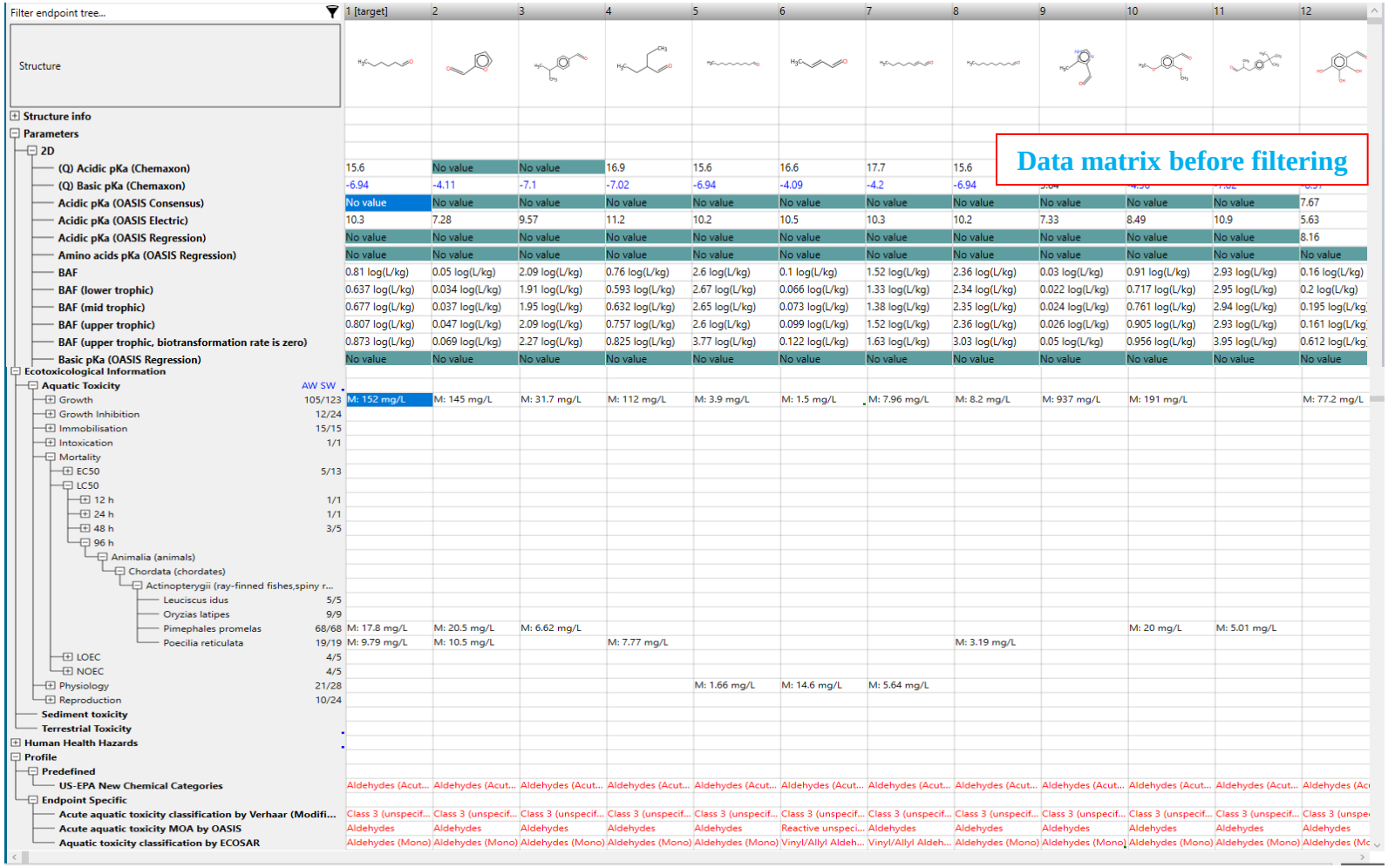
3) Transferring of data (experimental/predicted) from analogues to the target chemical outside the gap filling module;

4) Hide/Show the data matrix once the user is in the stage of Data gap filling module;

Illustration of the functionalities are shown on the next few slides.

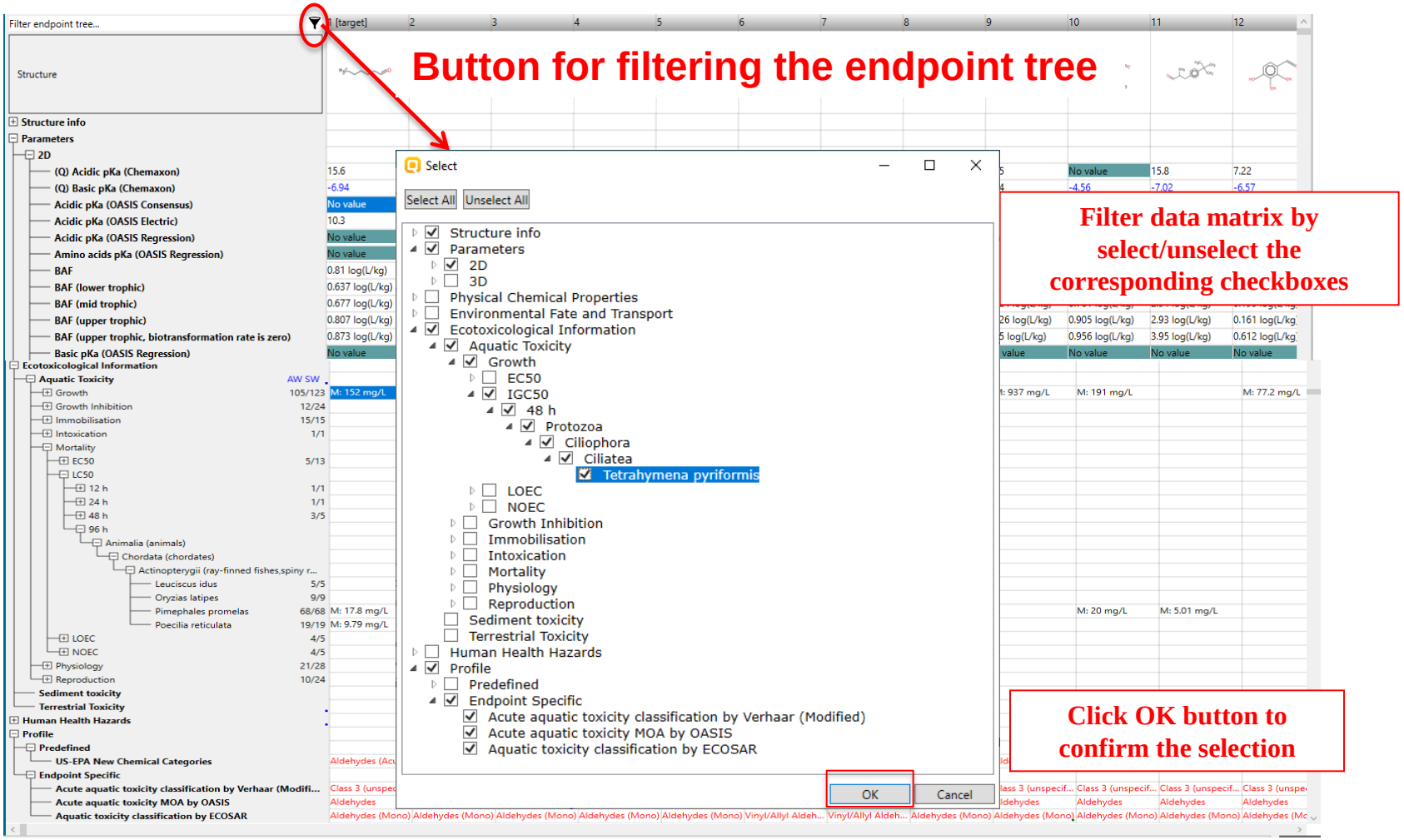
Manipulation of data matrix
Implementation in Toolbox

1) *Filtering the data matrix*



Manipulation of data matrix
Implementation in Toolbox

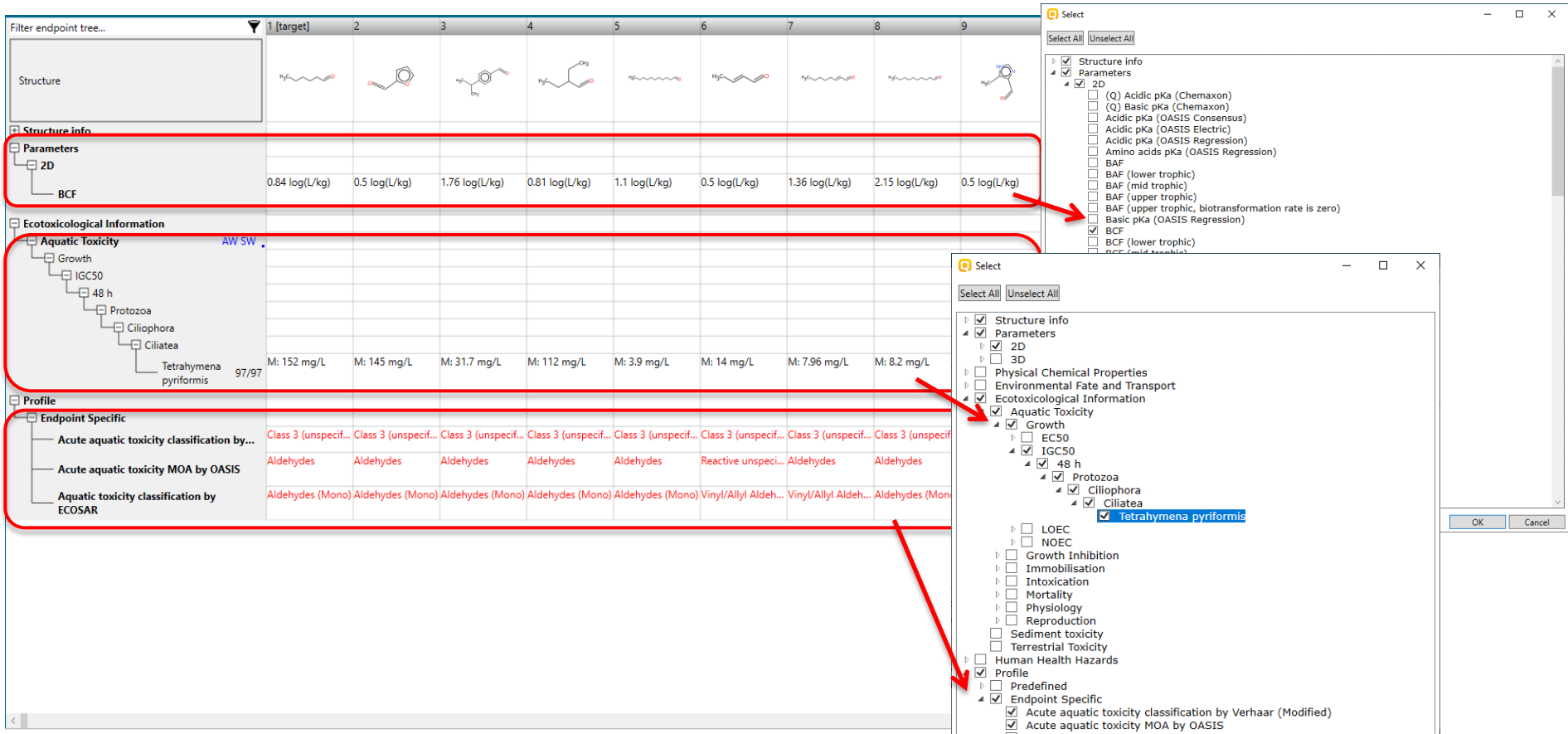
1) *Filtering the data matrix*



Manipulation of data matrix
Implementation in Toolbox

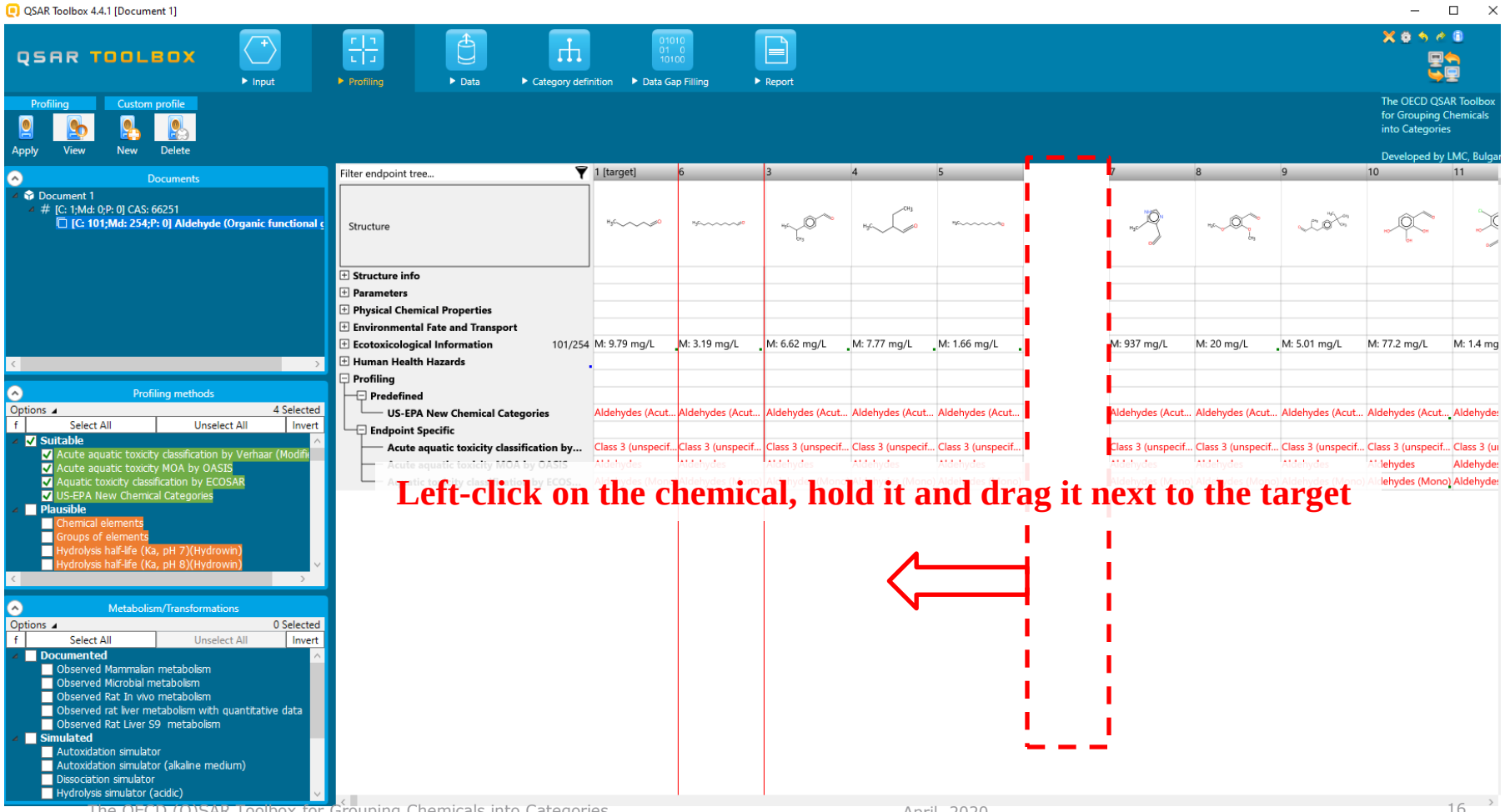
1) *Filtering the data matrix*

Data matrix after filtering includes the selected items (parameters/data/profilers) only



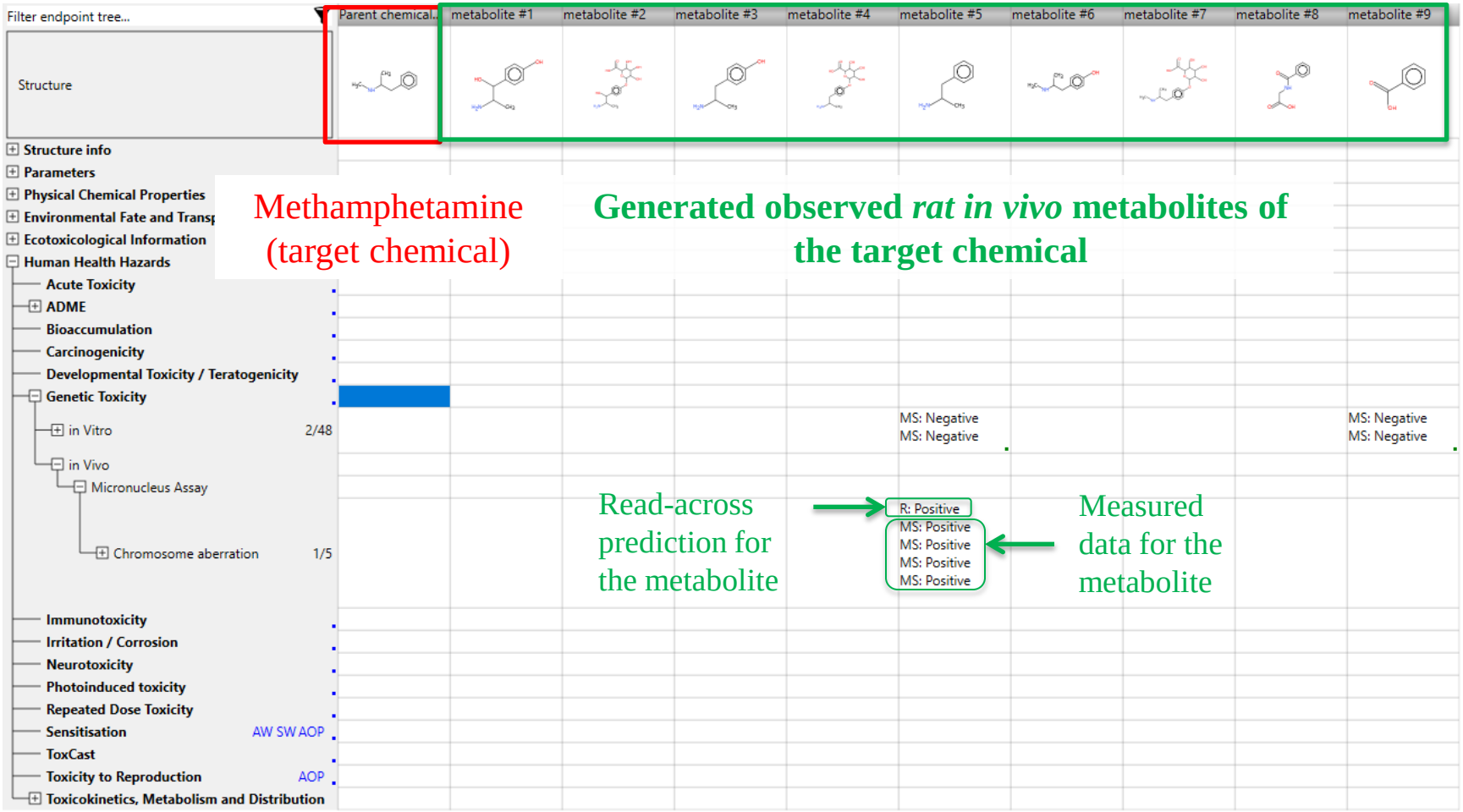
Manipulation of data matrix
Implementation in Toolbox

2) Reordering of the analogues on data matrix



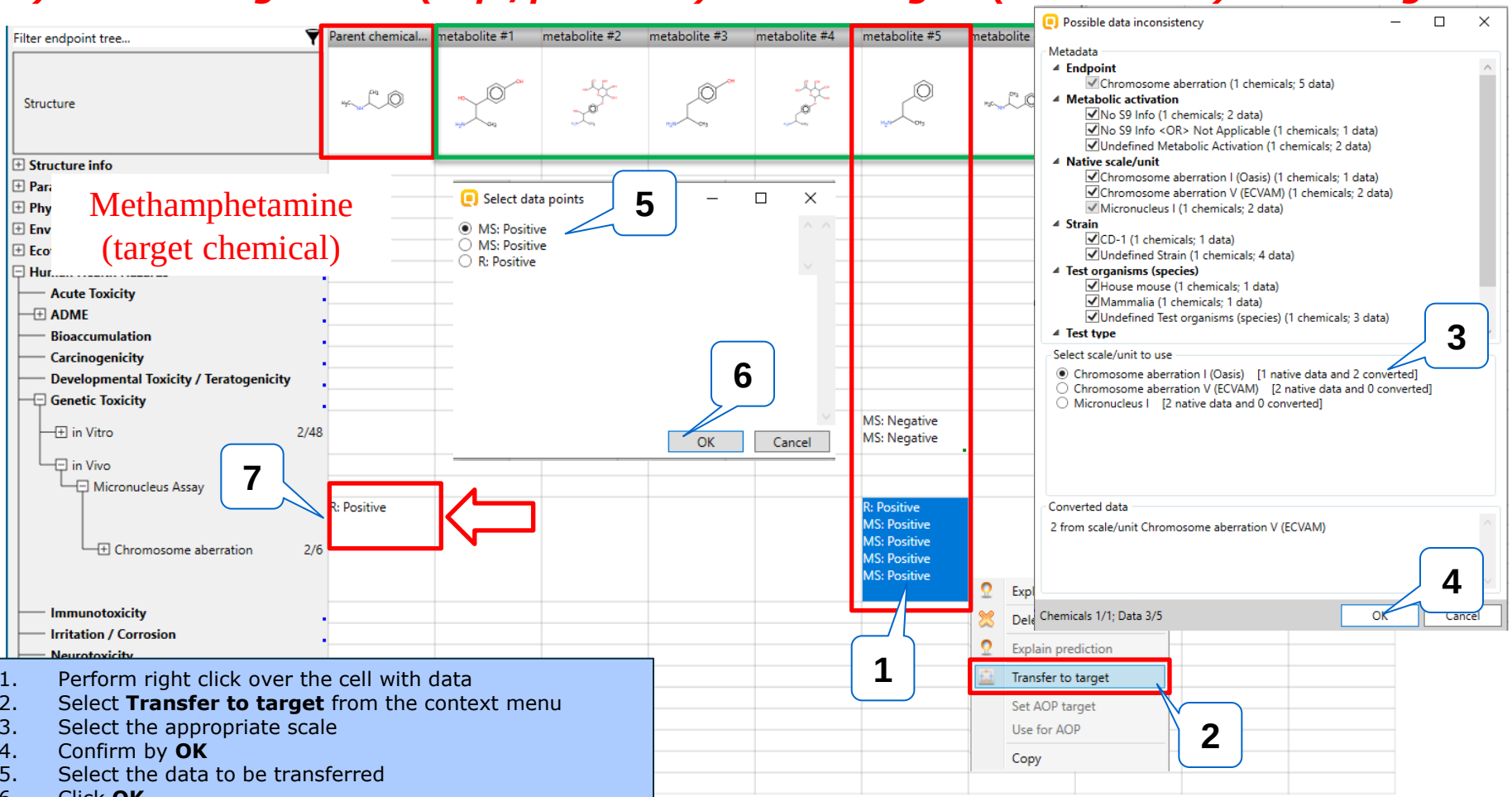
Manipulation of data matrix
Implementation in Toolbox

3) Transferring of data (exp./predicted) from analogues(metabolites) to the target



Manipulation of data matrix
Implementation in Toolbox

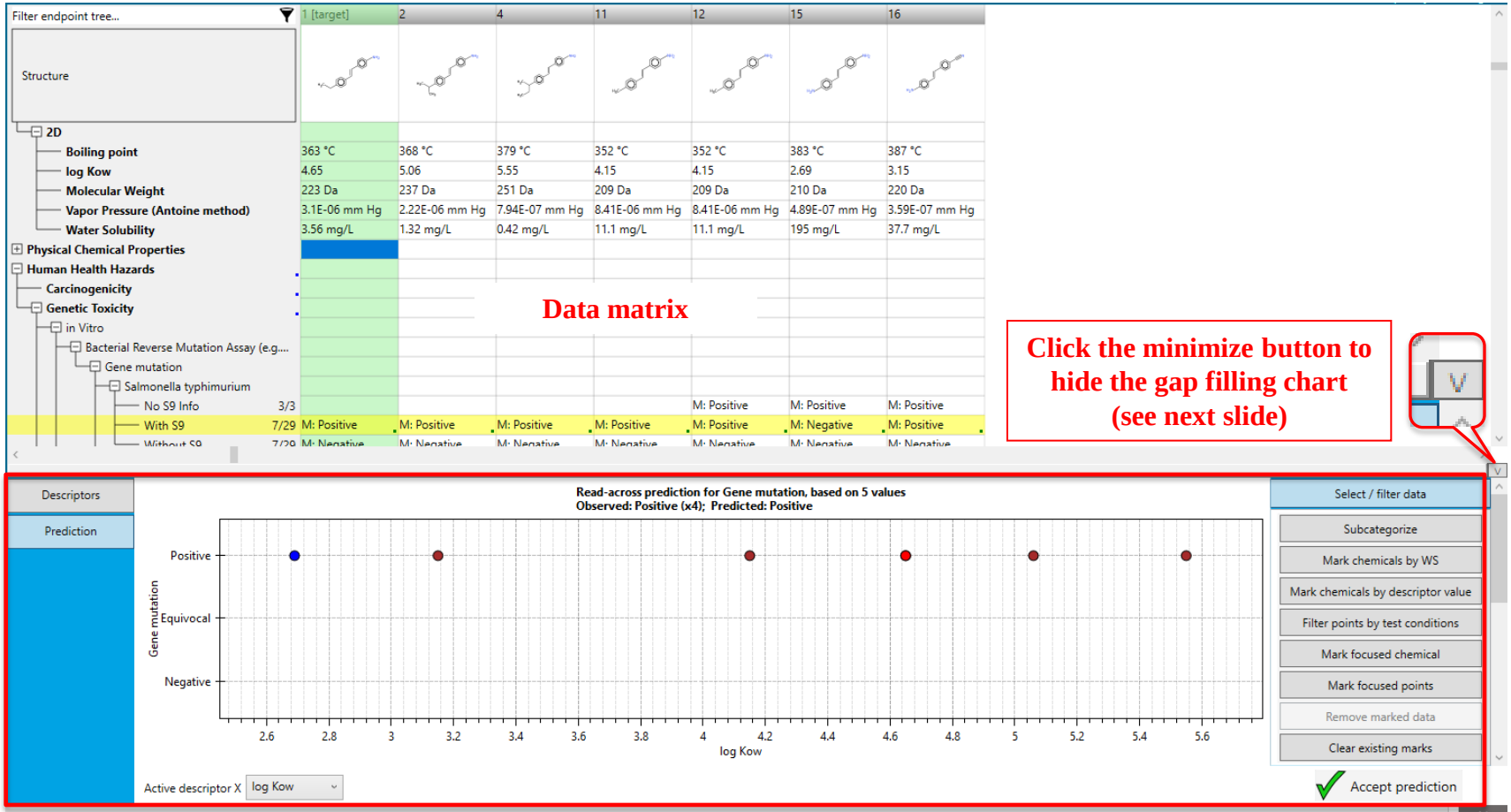
3) Transferring of data (exp./predicted) from analogues(metabolites) to the target



1. Perform right click over the cell with data
2. Select **Transfer to target** from the context menu
3. Select the appropriate scale
4. Confirm by **OK**
5. Select the data to be transferred
6. Click **OK**
7. Selected data is transferred to the target chemical

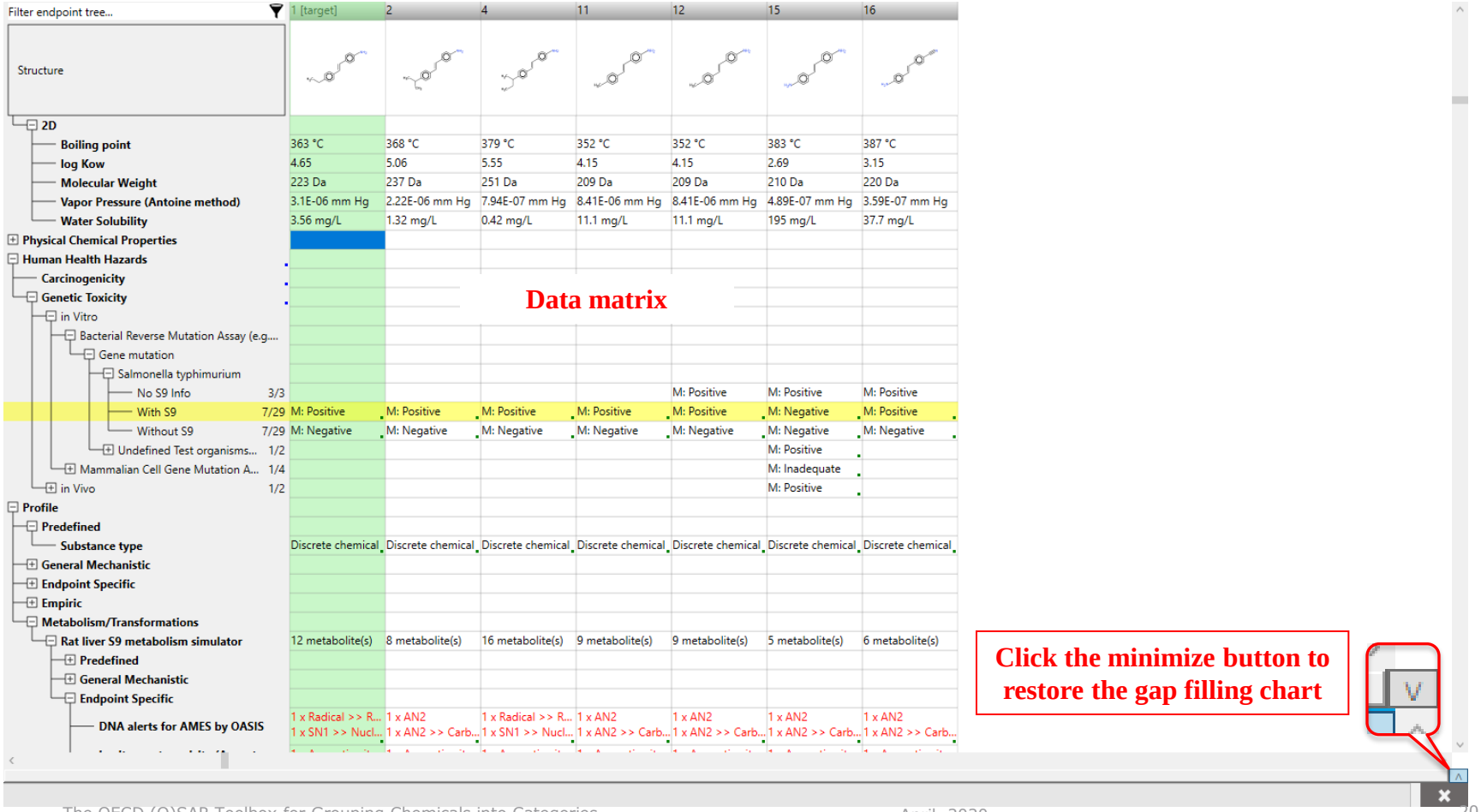
Manipulation of data matrix
Implementation in Toolbox

4) Hide/Show the chart while in data gap filling module



Manipulation of data matrix
Implementation in Toolbox

4) Hide/Show the chart while in data gap filling module



Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- **Example**

Example

- In this example we will predict *AMES mutagenicity* of Naphthyl N-methylcarbamate [CAS# 27636-33-5], which will be the "target" chemical

- Collect data and profiling results for the target according to the suitable profilers and databases

- Generate *in vitro* rat liver metabolites of the target chemical

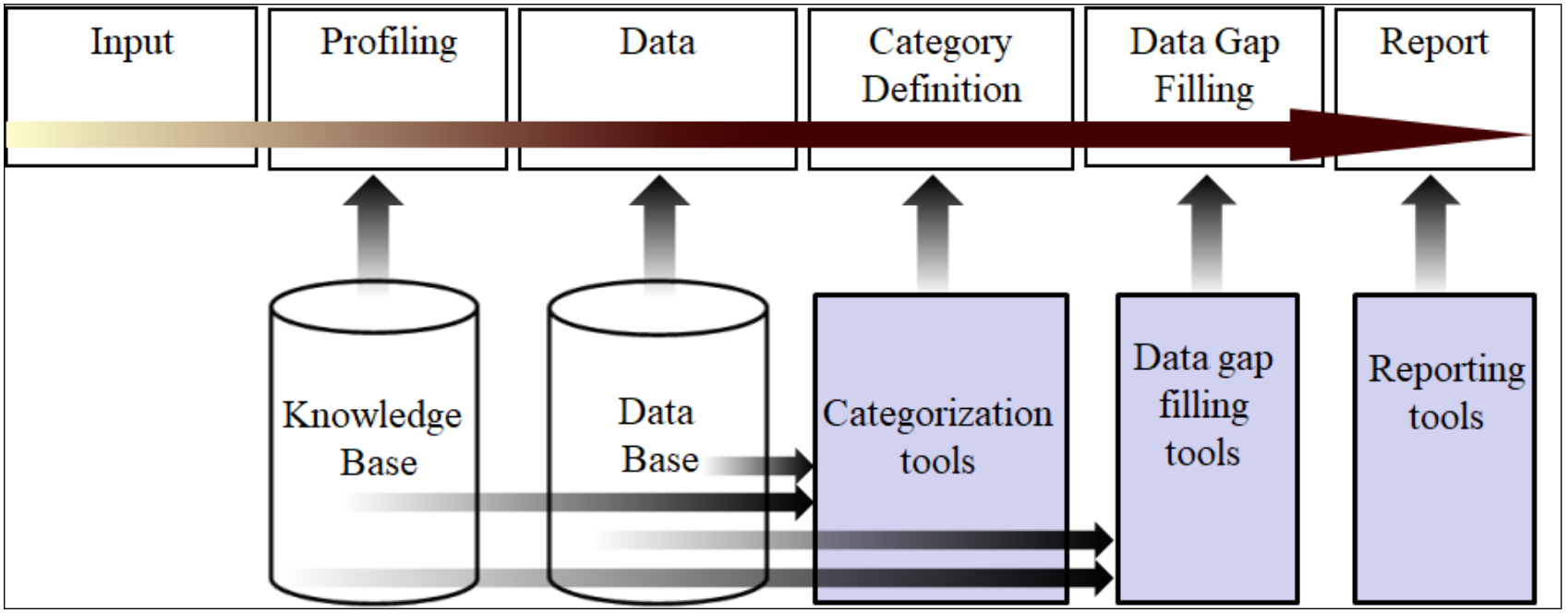
- Make read-across prediction for the target by transferring observed data of the preliminary generated metabolites (by *in vitro* rat liver metabolic simulator) to the target outside data gap filling module

Workflow

- **The Toolbox workflow include six modules :**
  - Input
  - Profiling
  - Data
  - Category Definition
  - Data Gap Filling
  - Report
- **In this example we will use only the first three modules in order to fulfil the aims of the example.**

Workflow

Scheme illustrating the Toolbox workflow



Outlook

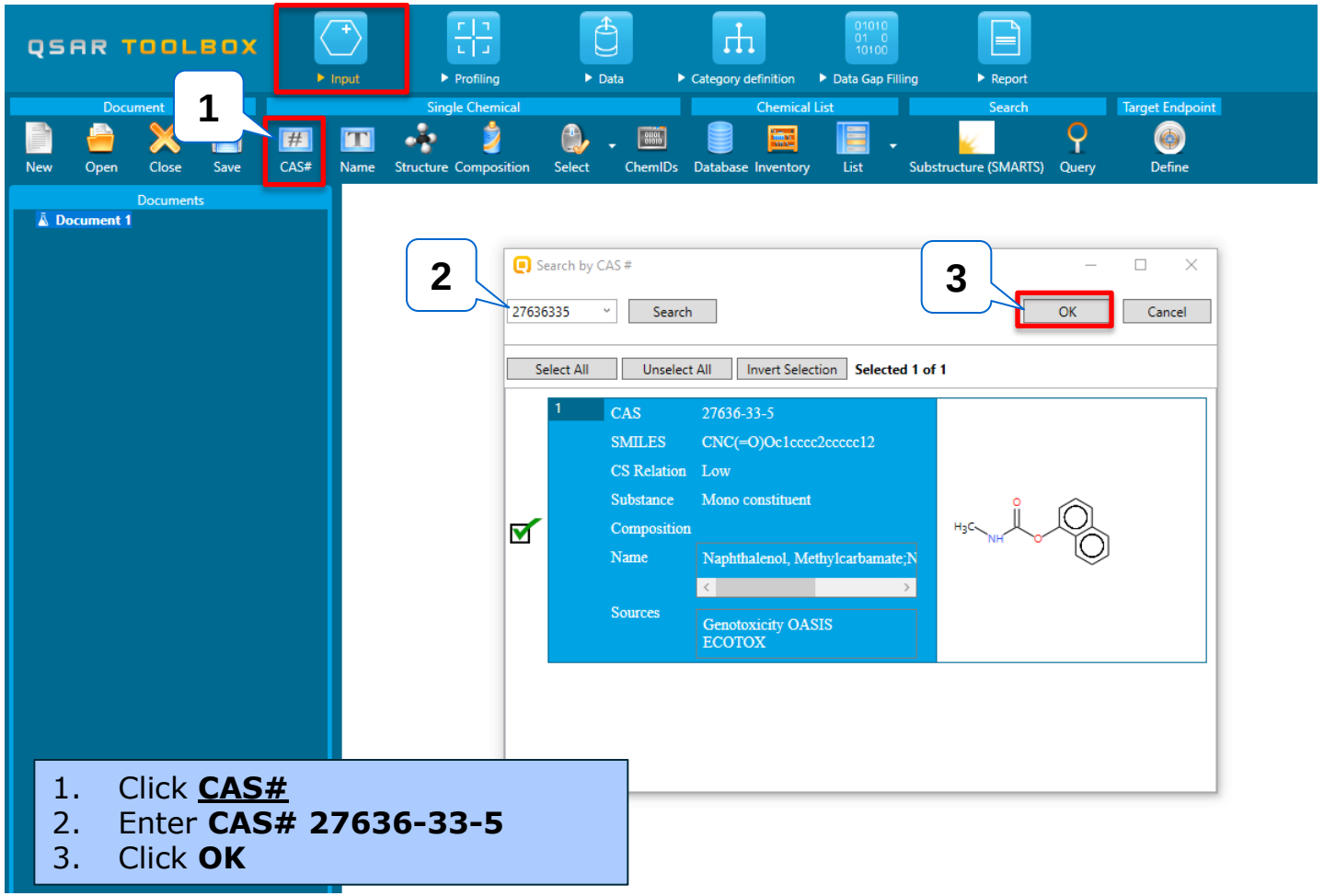
- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
  - **Input**

Input
Overview

- This module provides the user with several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.

Input

Entering a target chemical by CAS#



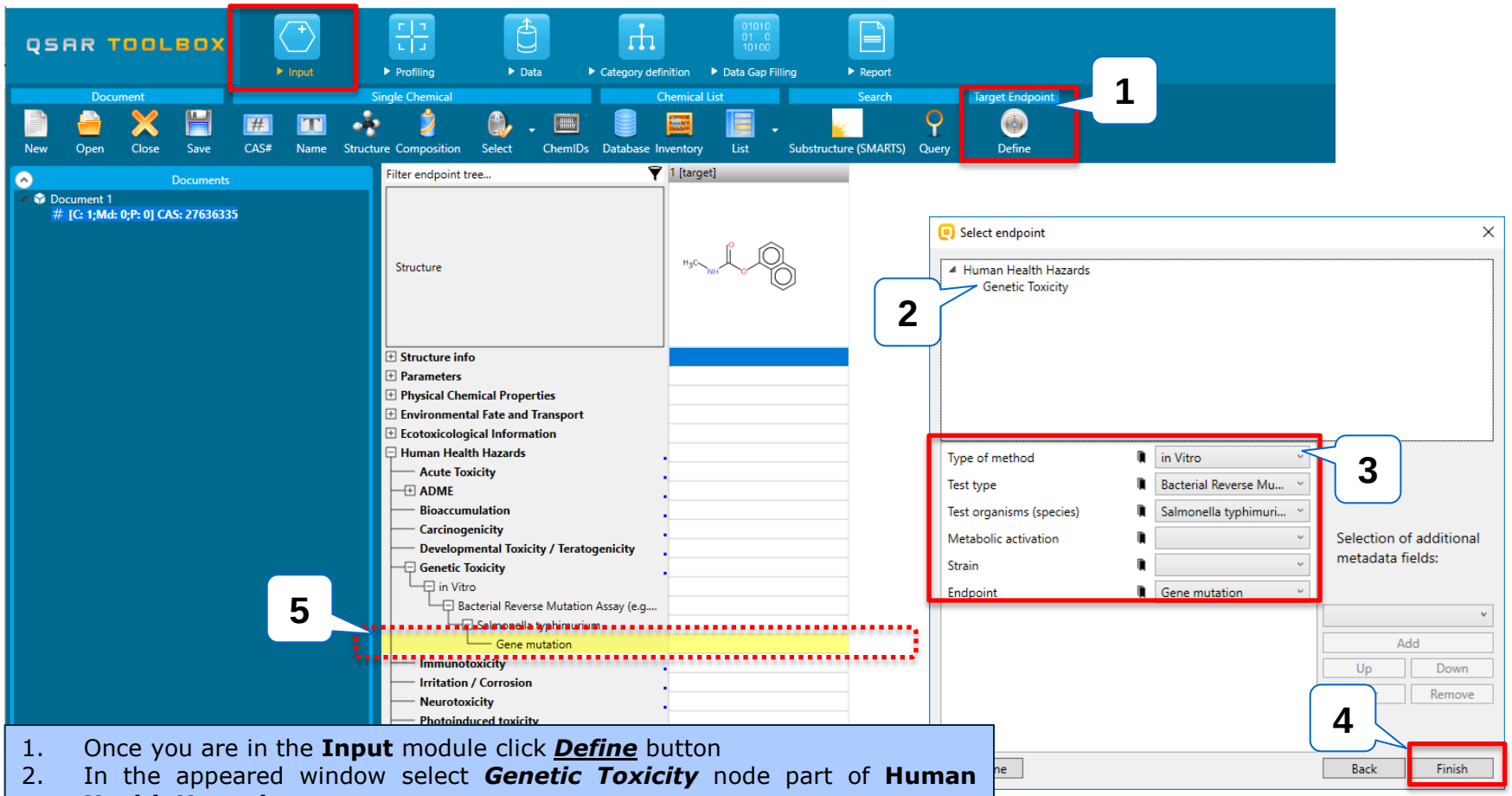
Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
  - Input
    - **Define target endpoint**

Define target endpoint
Overview

- This functionality allows entering the endpoint of interest e.g., EC3, LC50, gene mutation etc.
- The relevant profiles and databases are getting color highlighted once the targeted endpoint is preliminary defined by this functionality
- There are different ways for defining the target endpoint (via the button from the Input module or by right click from the endpoint tree). For more details press F1 button in order to see the online help.

Define target endpoint



1. Once you are in the **Input** module click ***Define*** button
2. In the appeared window select ***Genetic Toxicity*** node part of **Human Health Hazard**
3. Select the specific data from the pop-up menu as shown
4. Click **Finish**
5. The row related to the defined target endpoint is getting yellow highlighted

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- The exercise
  - Input
  - **Profiling**

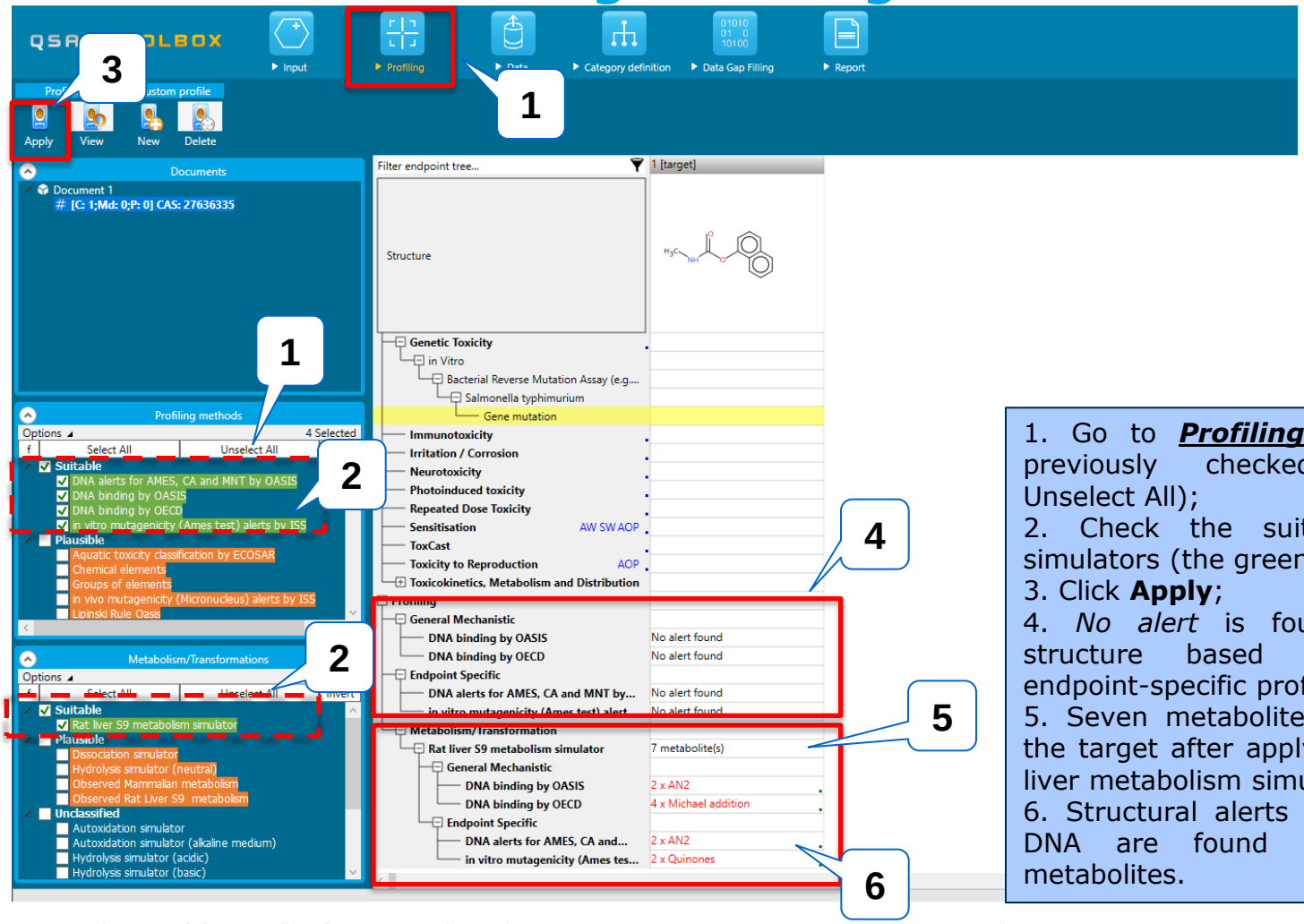
Profiling
Overview

- "Profiling" refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- "Profiling" module contains all the knowledge in the system coded in profiling schemes (profilers);
- "Profilers" are a collection of empirical and mechanism knowledge (expertly derived) which could be used to analyse the structural properties of chemicals;
- The "profilers" identify the affiliation of the target chemical(s) to preliminary defined categories (functional groups/alerts);
- The "Profiling" module contains also observed and simulated metabolisms/transformations, which could be used in combination with the profiling schemes;
- The outcome of the profiling determines the most appropriate way to search for analogues, but they are also useful for preliminary screening or prioritization of substances;
- The "profilers" are not (Q)SARs, i.e. they are not prediction models themselves;
- Based on the defined target endpoint the most relevant profilers are getting colour highlighted: greens are the most suitable, orange are plausible*
- For the purpose of our example only suitable profilers in combination with suitable metabolism simulator are used (see next slide).

*For more details regarding relavancy of profilers see tutorial: *Example for predicting skin sensitization taking into account alert performance*

Profiling

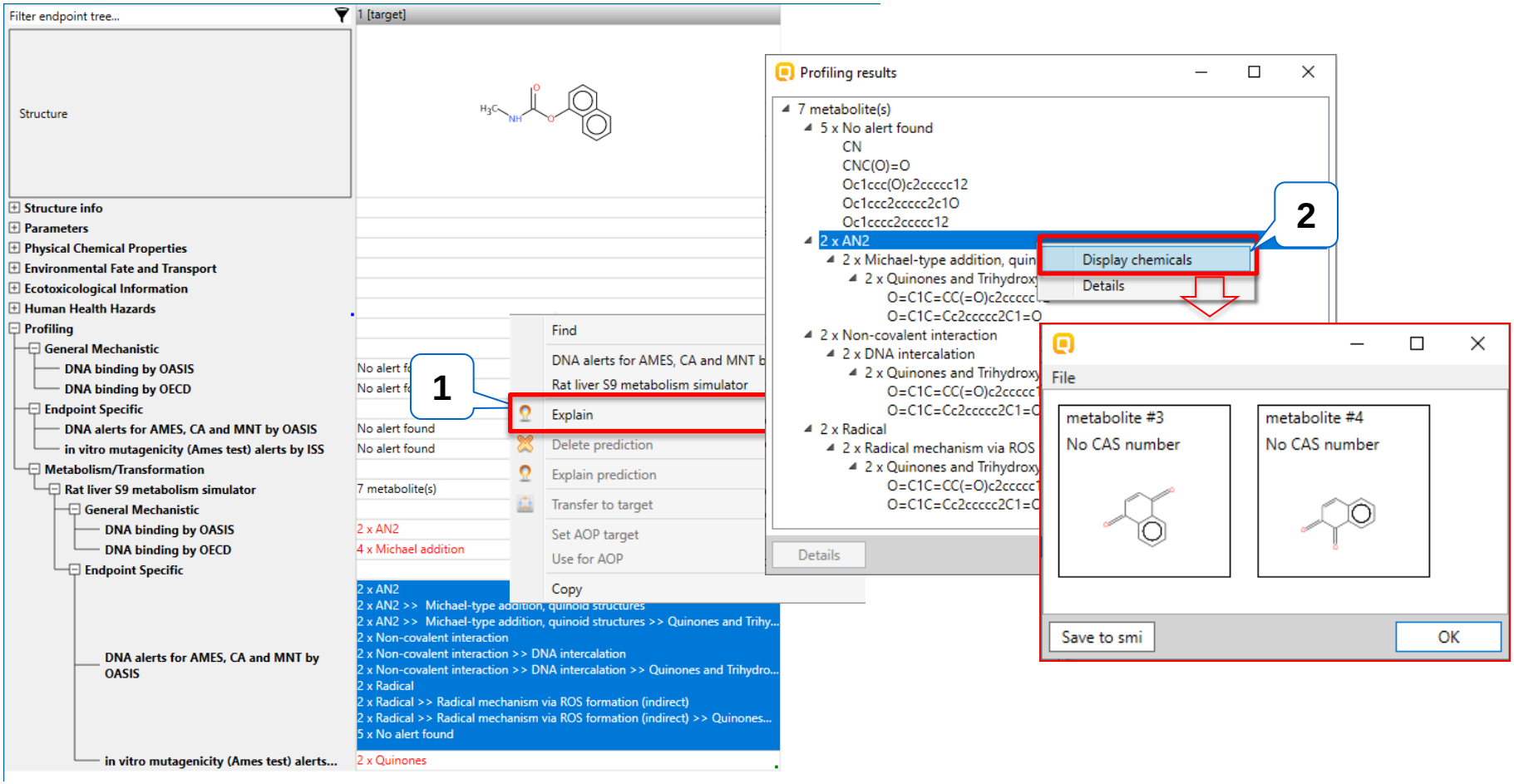
Profiling the target chemical



1. Go to ***Profiling***; and unselect all previously checked profiler (click Unselect All);
2. Check the suitable profiles and simulators (the green ones);
3. Click **Apply**;
4. *No alert* is found in the target structure based on general and endpoint-specific profilers;
5. Seven metabolites are produced for the target after applying the *in vitro* rat liver metabolism simulator;
6. Structural alerts for interaction with DNA are found in the generated metabolites.

Profiling

Explain of profiling results



1. Right-click over the results and select ***Explain*** for more details.
2. Right-click over the domain information (e.g. AN2) and select ***Display chemicals*** to see the metabolites belonging to this domain (in this case this is 'Quinones and Trihydroxybenzenes' alert part of nucleophilic addition type mechanism (AN2)).

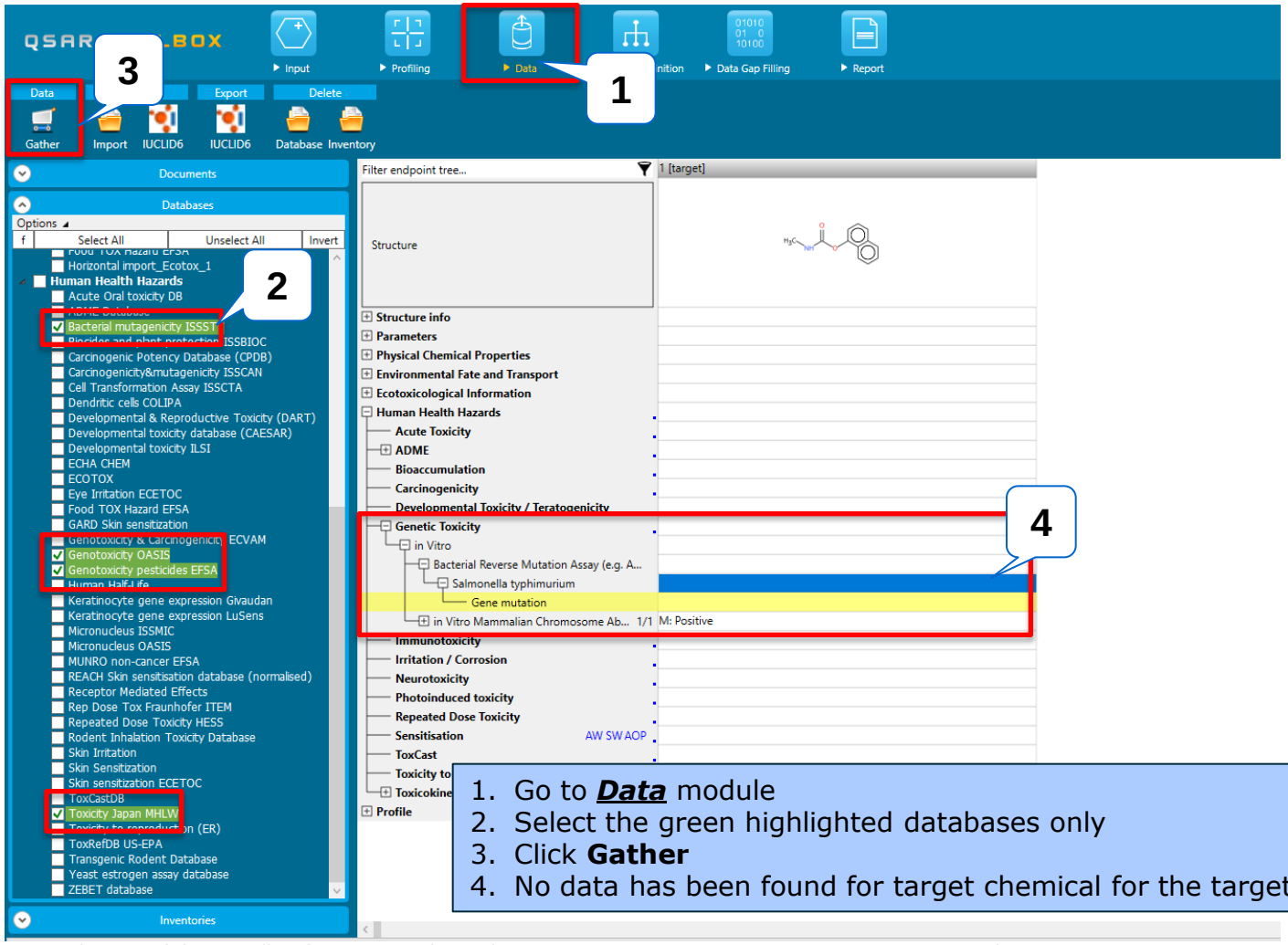
Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
  - Input
  - Profiling
  - **Data**

Data
Overview

- The "Data" module refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox databases.
- Data gathering can be executed in a global fashion (i.e., collecting all data for all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).
- Once the endpoint is selected, the databases, which contain such type of data are highlighted in green (see next slide).

Data
Collect data



1. Go to ***Data*** module
2. Select the green highlighted databases only
3. Click **Gather**
4. No data has been found for target chemical for the target endpoint

Recap

- The first module (Input), which introduces the target chemical, ensure correctness of the structure.

- The second module (Profiling) shows that there is no DNA binding alert for target chemical itself, but structural alerts responsible for DNA interaction have been found in the generated rat liver metabolites. The latter determines the forthcoming actions of the workflow.

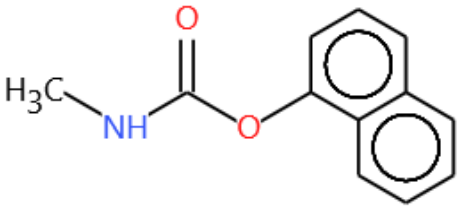
- In the third module (Data), you have found that the target chemical has no data associated with the target endpoint

- Due to the fact that AMES test accounts S9 metabolism the study continues with investigating profiling results and data of generated rat liver metabolites of the target chemical (see next slides).

Recap

Parent

Naphthyl N-methylcarbamate
CAS# 27636-33-5



DNA binding alerts

No alert found

Profiling Result

No alert found (no AMES mutagenicity effect is expected)

Rat liver metabolism – 2 of the metabolites have quinone type DNA binding alert

Investigating DNA prof + data for the metabolites

Outlook

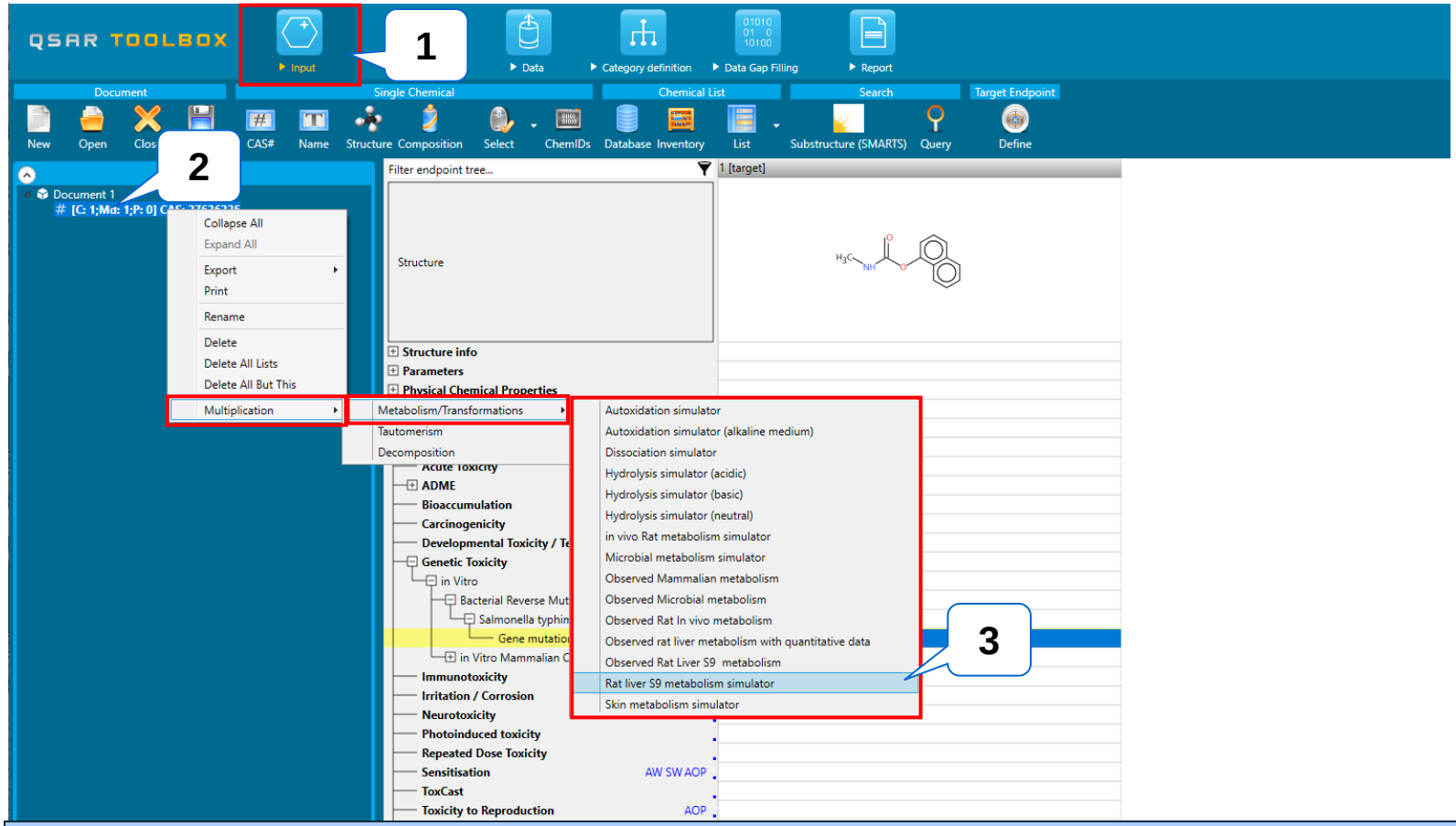
- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
  - Input
  - Profiling
  - Data
  - **Simulation of rat liver S9 metabolism**

Handling of rat liver S9 metabolism of target chemical

- The next actions in the workflow is metabolizing the target chemical by Rat liver S9 metabolism simulator
- The simulation of the rat liver metabolism of the target chemical is accomplished in section **Input**
- The generated metabolites appear in tree like form (see next slide)

Handling of rat liver S9 metabolism of target chemical

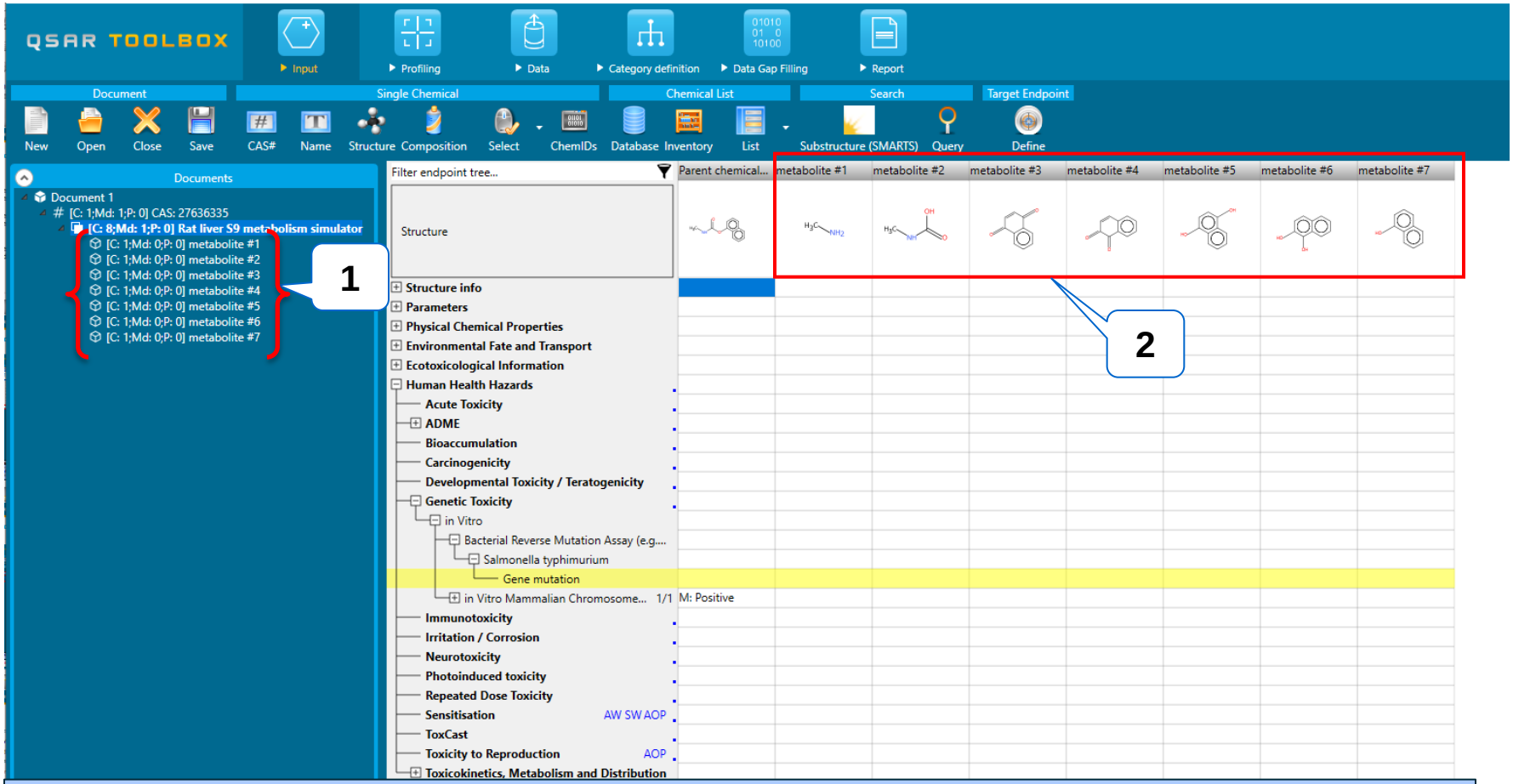
Multiplication of the target chemical



1. Go to ***Input*** module
2. Click on the level with **CAS #** of the target chemical and right-click on it, then
3. Select **Multiplication / Metabolism/Transformations / Rat liver S9 metabolism simulator**
4. Generated metabolites appear in tree like form and also are aligned next to the target (shown on next slide)

Handling of rat liver S9 metabolism of target chemical

Multiplication of the target chemical



1. Generated metabolites appear in tree like form in the documented tree
2. Also metabolites are aligned next to the target

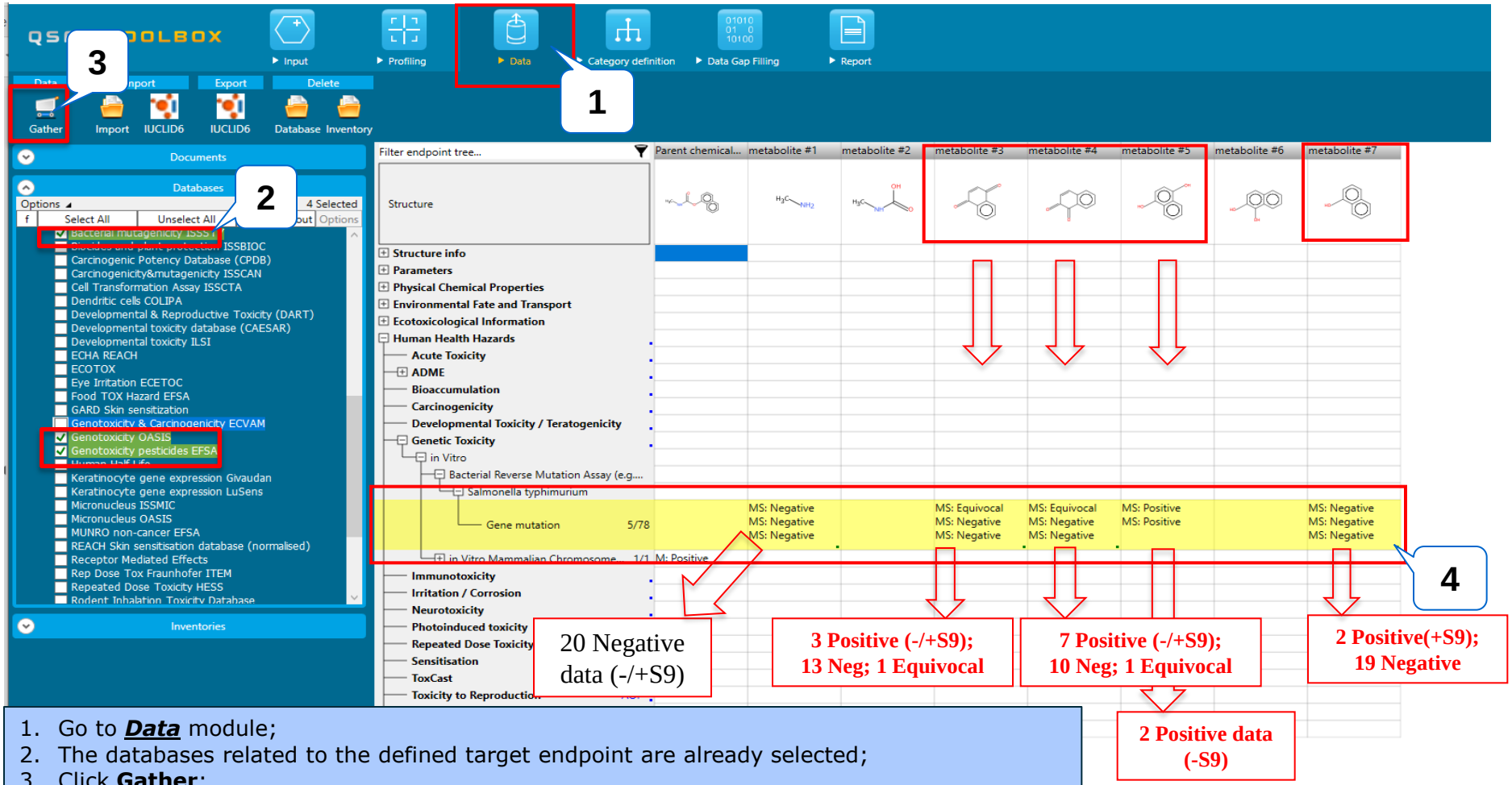
Next actions are focused on investigating the profilers of the generated metabolites and collecting data for them

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
  - Input
  - Profiling
  - Data
  - Generation of rat liver S9 metabolism
  - **Profiling and collecting data for metabolites**

Handling of rat liver S9 metabolism of target chemical

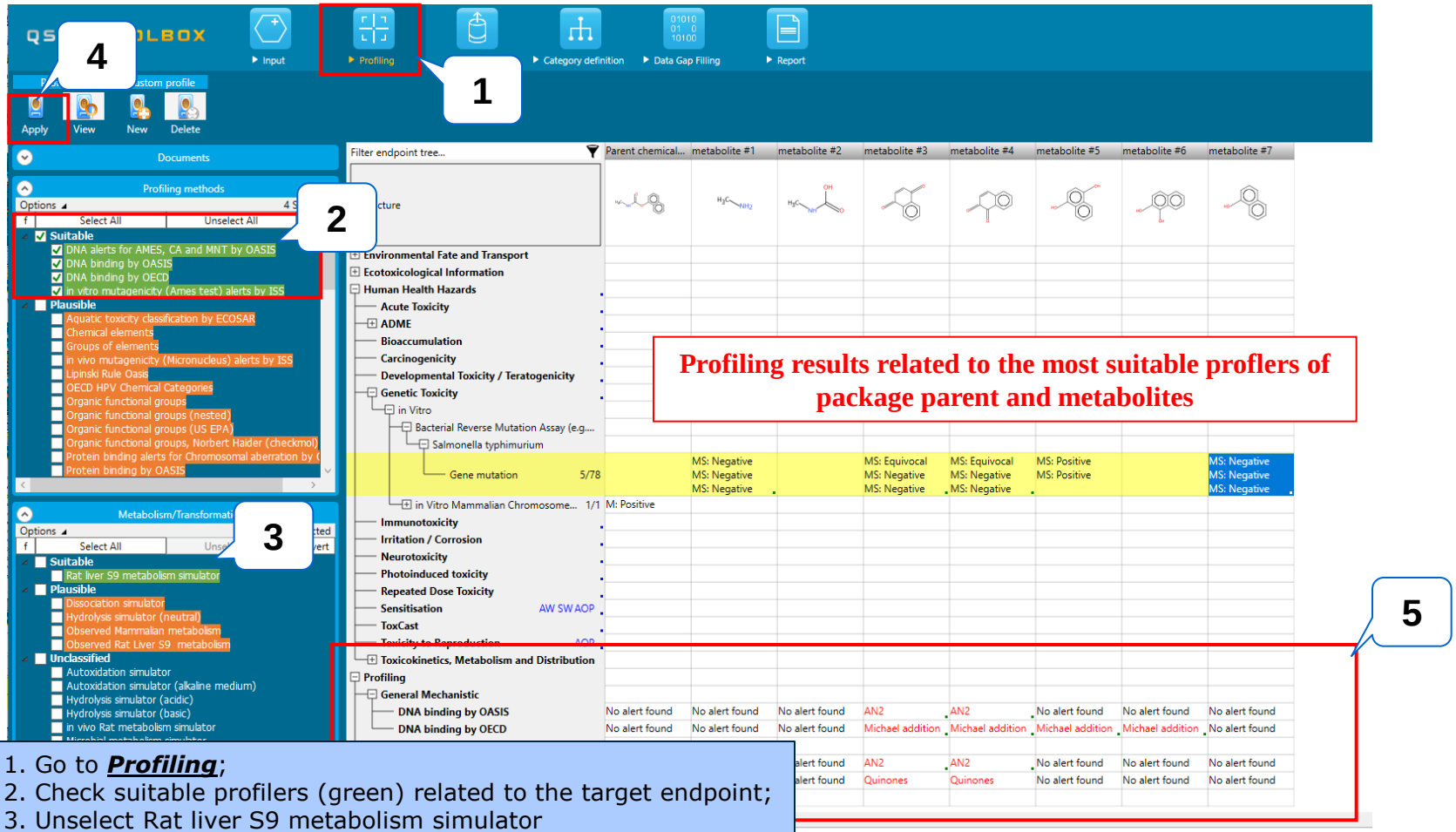
Collect data for metabolites



1. Go to ***Data*** module;
2. The databases related to the defined target endpoint are already selected;
3. Click **Gather**;
4. The data for the parent and metabolites appears in data matrix. Only one of the generated analogues is pure negative chemical according to the collected experimental data

Handling of rat liver S9 metabolism of target chemical

Profiling the package of metabolites

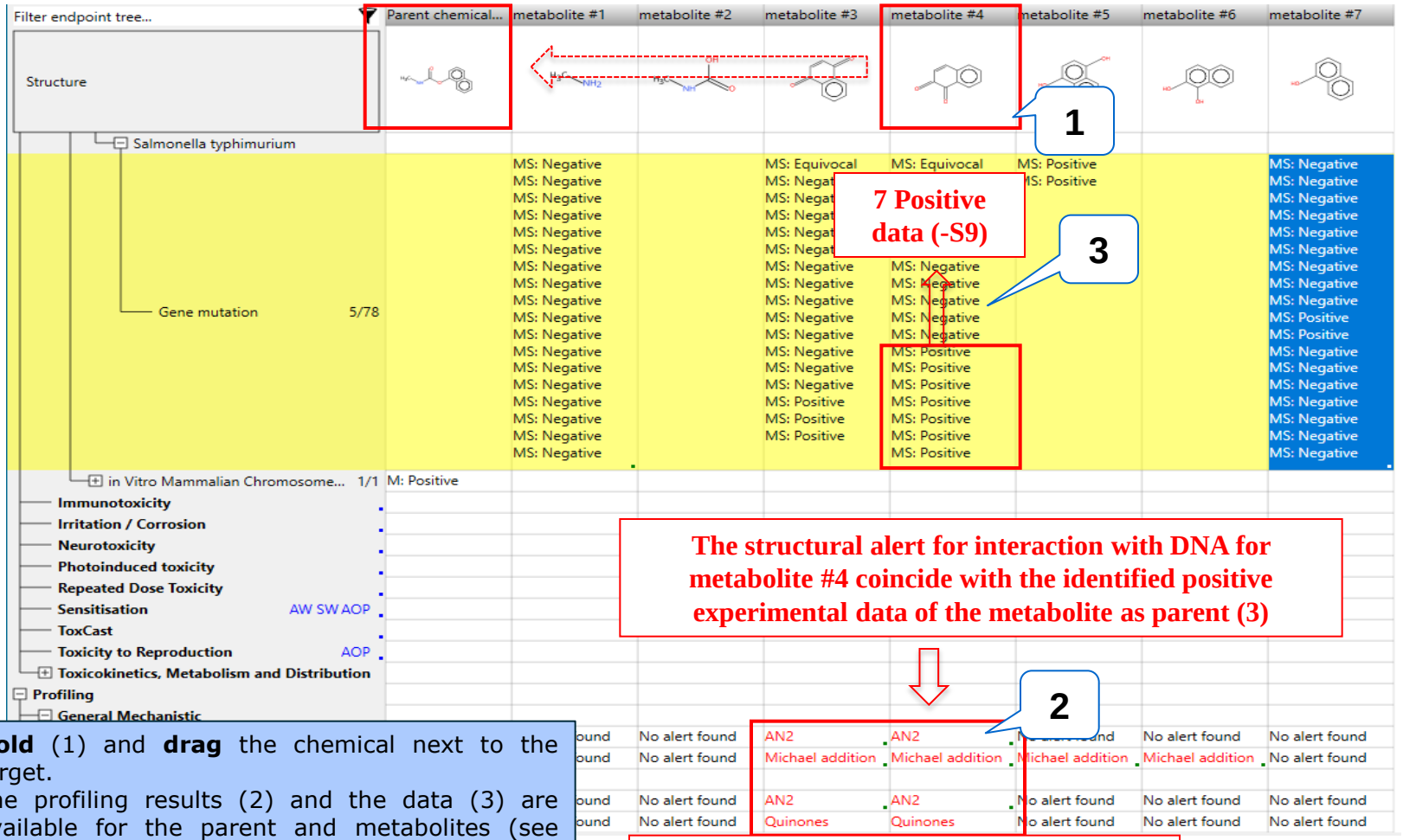


1. Go to ***Profiling***;
2. Check suitable profilers (green) related to the target endpoint;
3. Unselect Rat liver S9 metabolism simulator
4. Click **Apply**;
5. The profiling results appears in the data matrix.

Outlook

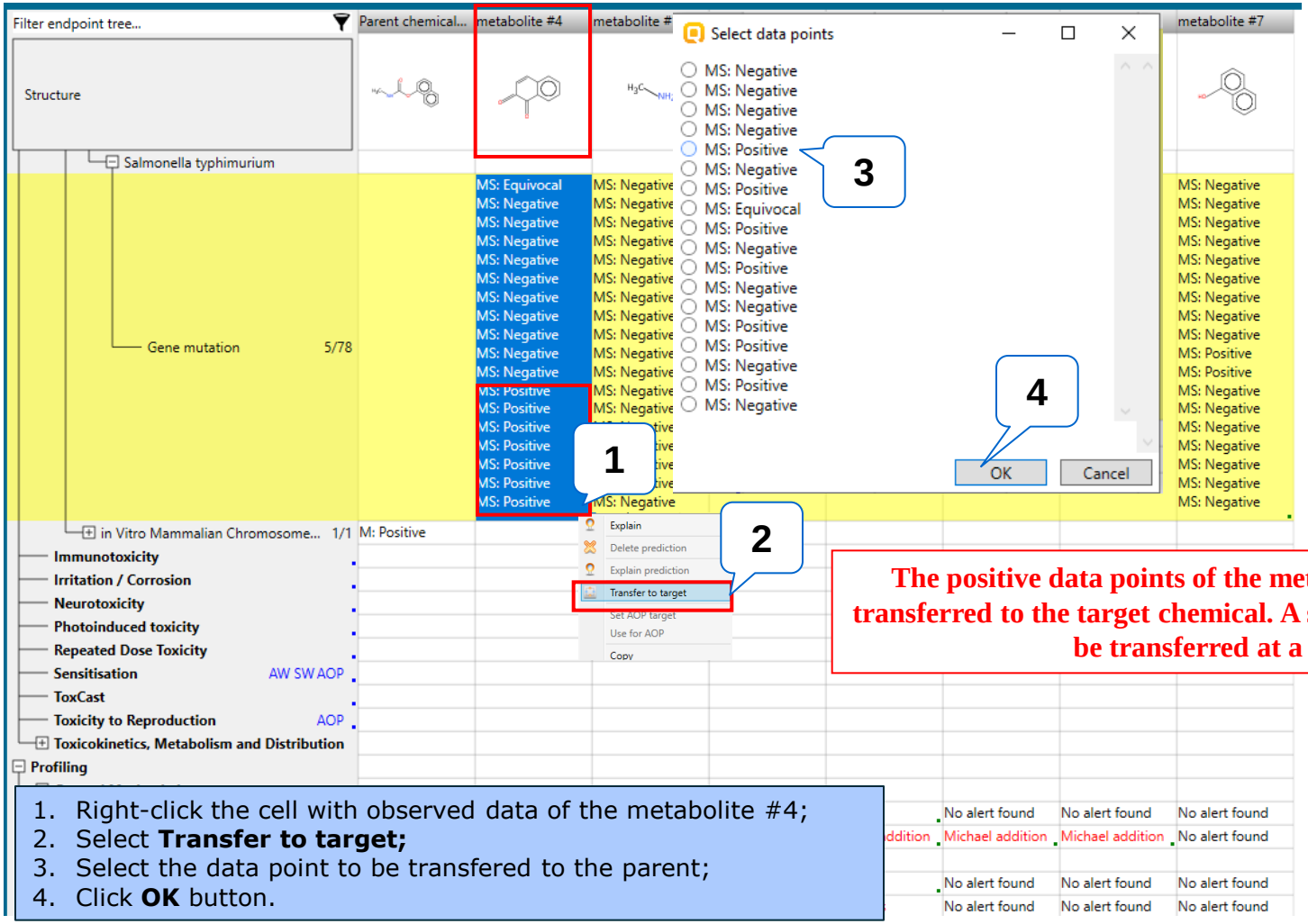
- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
  - Input
  - Profiling
  - Data
  - Generation of rat liver S9 metabolism
  - Profiling and collecting data for metabolites
  - **Transferring data to the target outside data gap filling module**

Transferring experimental data of metabolite to the target outside data gap filling



Hold (1) and **drag** the chemical next to the target.
The profiling results (2) and the data (3) are available for the parent and metabolites (see next slide).

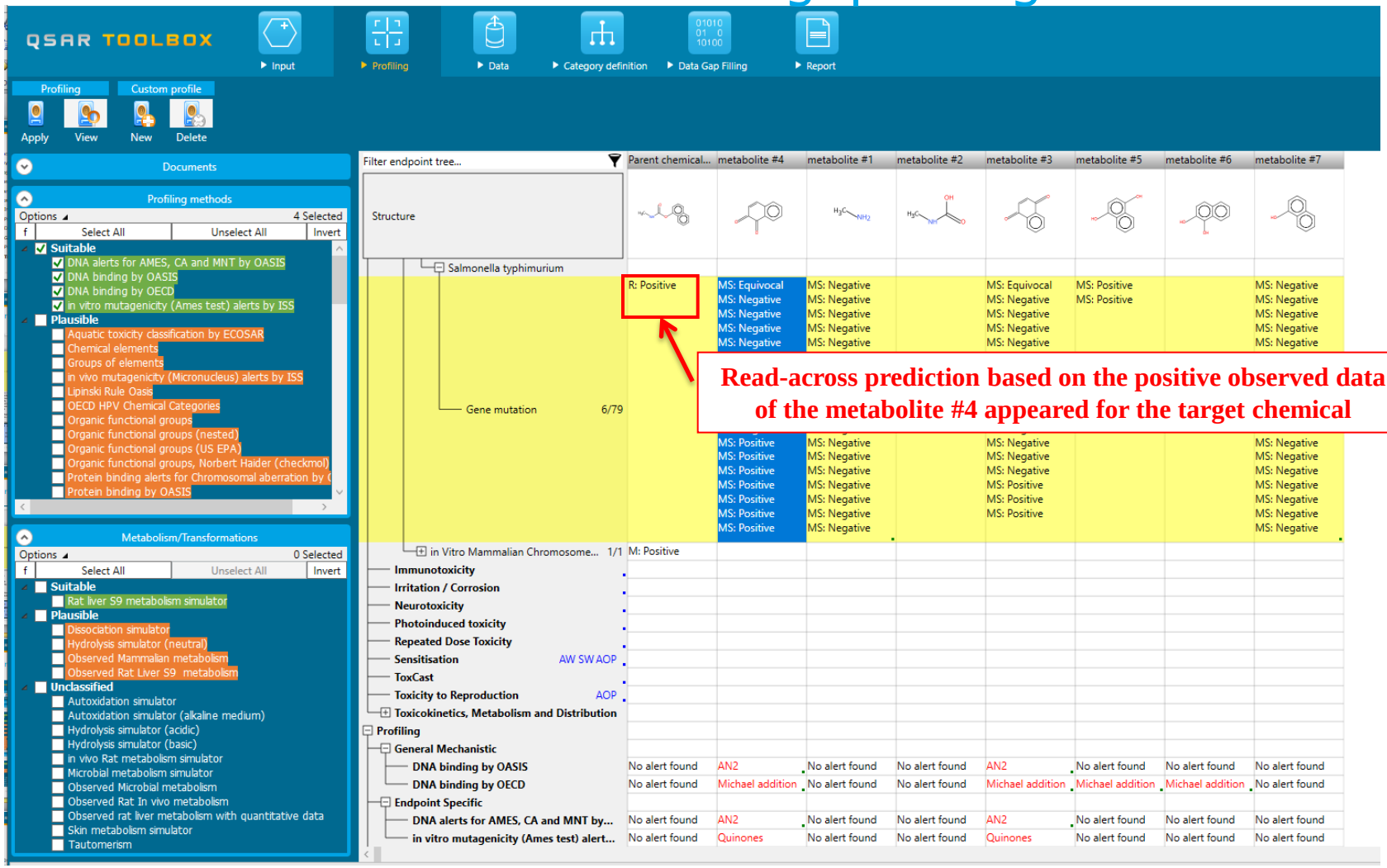
Transferring experimental data of metabolite to the target outside data gap filling



The positive data points of the metabolite #4 could be transferred to the target chemical. A single data point could be transferred at a time

1. Right-click the cell with observed data of the metabolite #4;
2. Select **Transfer to target;**
3. Select the data point to be transfered to the parent;
4. Click **OK** button.

Transferring experimental data of metabolite to the target outside data gap filling



Congratulations!

- You have now been introduced to the Data matrix manipulation options;
- You have now been introduced to the transfer of a read-across prediction to the target chemical outside of the gap filling module.
- Note, proficiency comes with practice!