QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.4.1

Step-by-step example for predicting Ames mutagenicity by making use of read-across

Outlook

- **Background**
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- Workflow of the exercise
- Save the prediction

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 read-across based on molecular similarity with data pruning.

- If you are a novice user of the Toolbox you may wish to review the "Getting Started" document [click here] as well as go through tutorials 1 and 3.

Outlook

- Background
- **Keywords**
- Objectives
- Specific Aims
- Read-across
- The exercise
- Workflow of the exercise
- Save the prediction

Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options

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 Aims
- Read-across
- The exercise
- Workflow of the exercise
- Save the prediction

Objectives

- **This presentation demonstrates a number of functionalities of the Toolbox:**
  - Entering a target chemical by SMILES notation and Profiling;
  - Identifying analogues to a target chemical by molecular similarity;
  - Retrieving experimental results available for the identified analogues;
  - Filling data gaps by read-across.

Outlook

- Background
- Keywords
- Objectives
- **Specific Aims**
- Read-across
- The exercise
- Workflow of the exercise
- Save the prediction

Specific Aims

- To review the workflow of the Toolbox.
- To reacquaint the user with 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 the rationale behind each step of the exercise.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- **Read-across**
- The exercise
- Workflow of the exercise
- Save the prediction

Read-across & the Analogue Approach

- Read-across is a method that can be used to fill a data gap of a chemical using an analogue approach.
- In the analogue approach, experimental endpoint information for a single or small number of tested chemicals is used to predict the same endpoint for an untested chemical that is considered to be "similar" (i.e., within the same category).

Analogous Chemicals

- Previously you have learned that analogous sets of chemicals are often selected based on the hypothesis that the toxicological effects of each member of the set will show a common behaviour.
- For this reason, the mechanistic profilers have been shown to be of great value in using for category definition.
- However, there are cases where the mechanistic profilers and grouping methods are inadequate and one is forced to rely on molecular similarity to form a category.
- The Toolbox allows one to develop a category by using either mechanistic knowledge (e.g. for DNA binding) or structural similarity.
- Since there is no preferred way of identifying structural similarity, the user is guided to use DNA binding as a first option.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- **The exercise**
- Workflow of the exercise
- Save the prediction

Exercise

- In this exercise we will predict the Ames mutagenicity potential for an untested compound - *n-hexanal* [SMILES: CCCCCC=O], which is the "target" chemical.
- This prediction will be accomplished by collecting a small set of test data for chemicals considered to be in the same category as the target molecule.
- The category will be defined by empirical similarity, with respect to the "Organic functional groups" profiler.
- The prediction itself will be made by "read-across" analysis.

On Mutagenesis

- Mutations within a gene are generally base-substitutions or small deletions/insertions (i.e., frame shifts).
- Such alteration are generally called point mutations.
- The Ames scheme based on strains of *Salmonella typhimurium* provides the corresponding experimental data.

On Mutagenesis

- The Ames mutagenicity assay (see OECD guideline 471) is designed to assess the ability of a chemical to cause point mutations in the DNA of the bacterium *Salmonella typhimurium*.
- The Ames test includes a number of strains (TA1537, TA1535, TA100, TA98 and TA97) that have been engineered to detect differing classes of mutagenic chemicals.
- The basic test only detects direct acting mutagens (i.e., those chemicals able to interact with DNA without the need for metabolic activation).

on Metabolic Activation

- The inclusion of a S9 mix of rodent liver enzymes is designed to assess those chemicals requiring metabolic activation in order to be mutagenic.
- Typically, chemicals are assayed both - without S9 and with S9 and the results are reported in a binary fashion.
- A positive result in any of the bacterial strains with or without S9 indicates mutagenic potential.
- In the current example both cases with and without S9 will be exemplified.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
- Save the prediction

Workflow

- **The Toolbox has six modules which are used in a sequential workflow:**
  - Input
  - Profiling
  - Data
  - Category Definition
  - Data Gap Filling
  - Report

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
  - **Input**

Input
Overview

- This module provides the user with several means of entering the chemical of interest (i.e. 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.

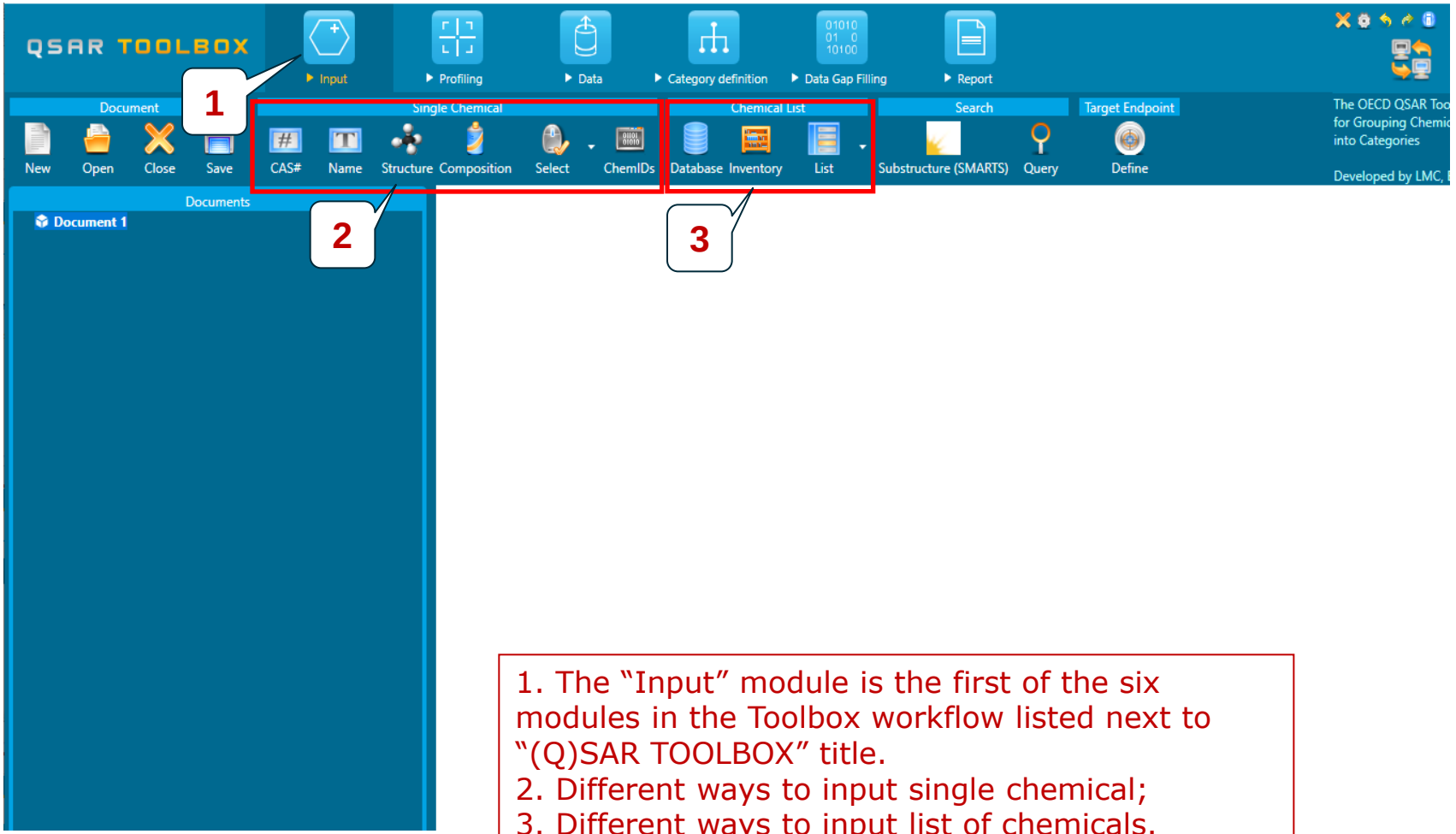
Input
Input chemical(s)

Alternative ways to input chemical(s):

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

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

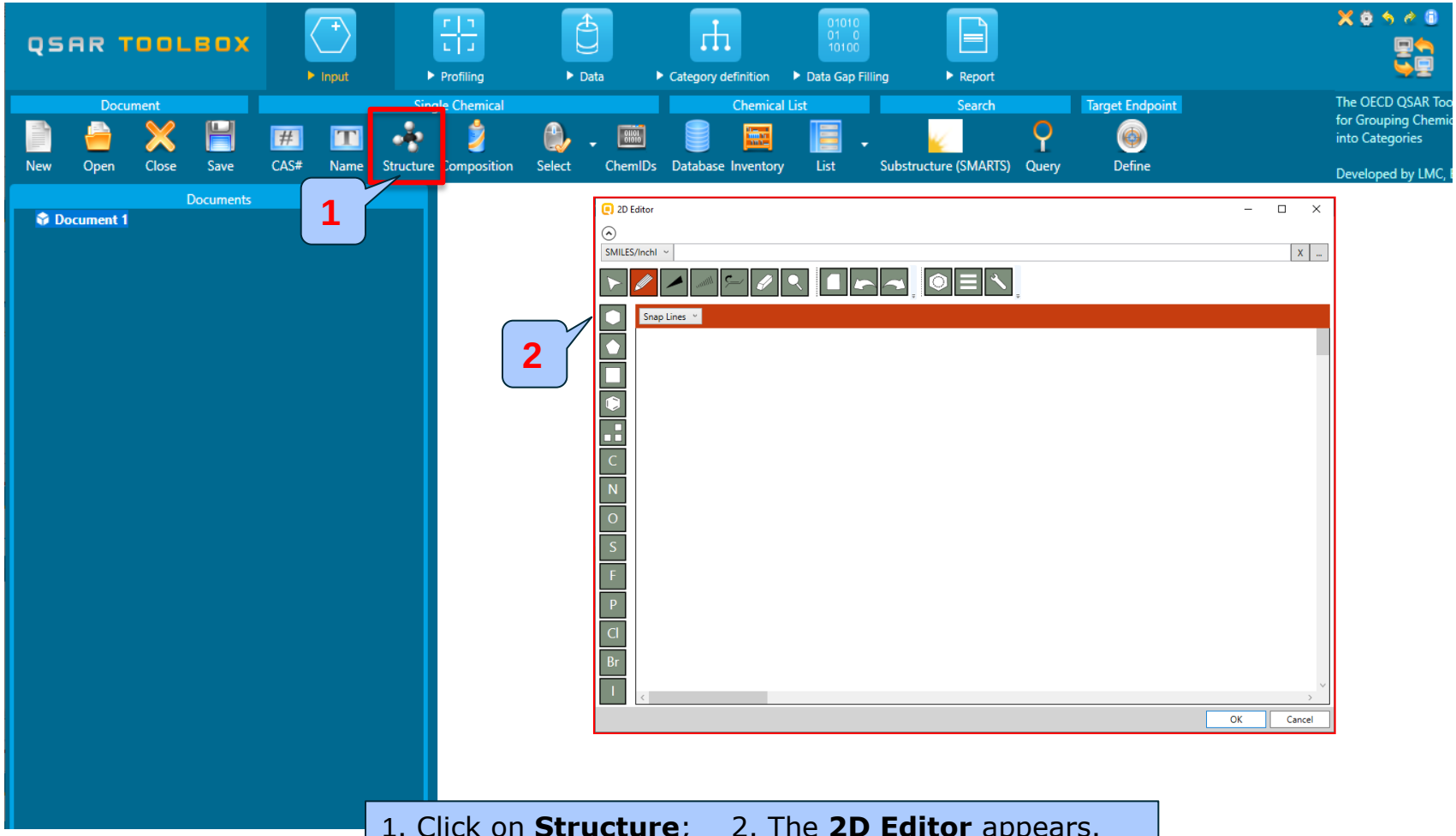
Input
Input Screen



1. The "Input" module is the first of the six modules in the Toolbox workflow listed next to "(Q)SAR TOOLBOX" title.
2. Different ways to input single chemical;
3. Different ways to input list of chemicals.

Input

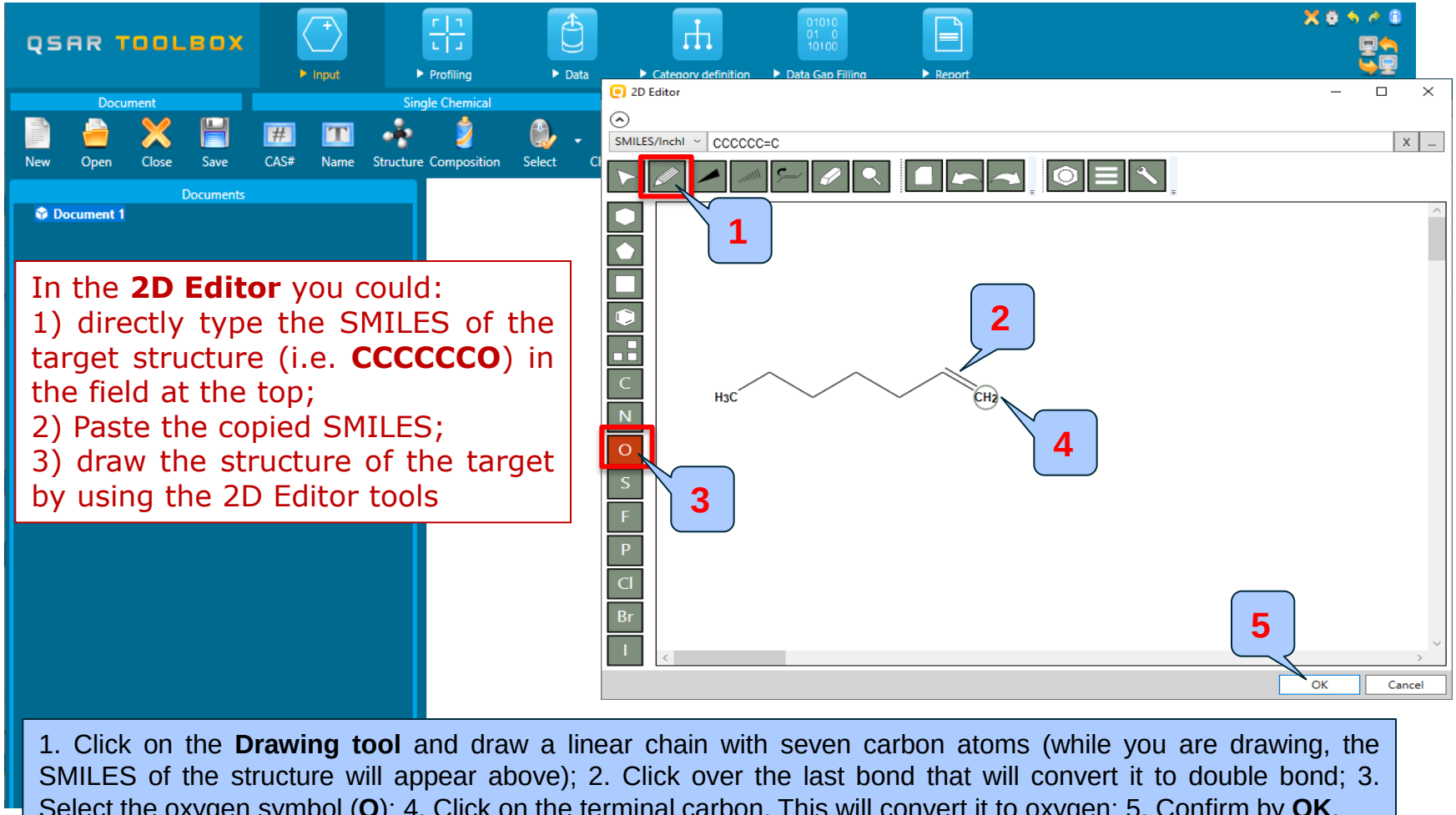
Input target chemical by drawing



1. Click on **Structure**; 2. The **2D Editor** appears.

Input

Input target chemical by drawing



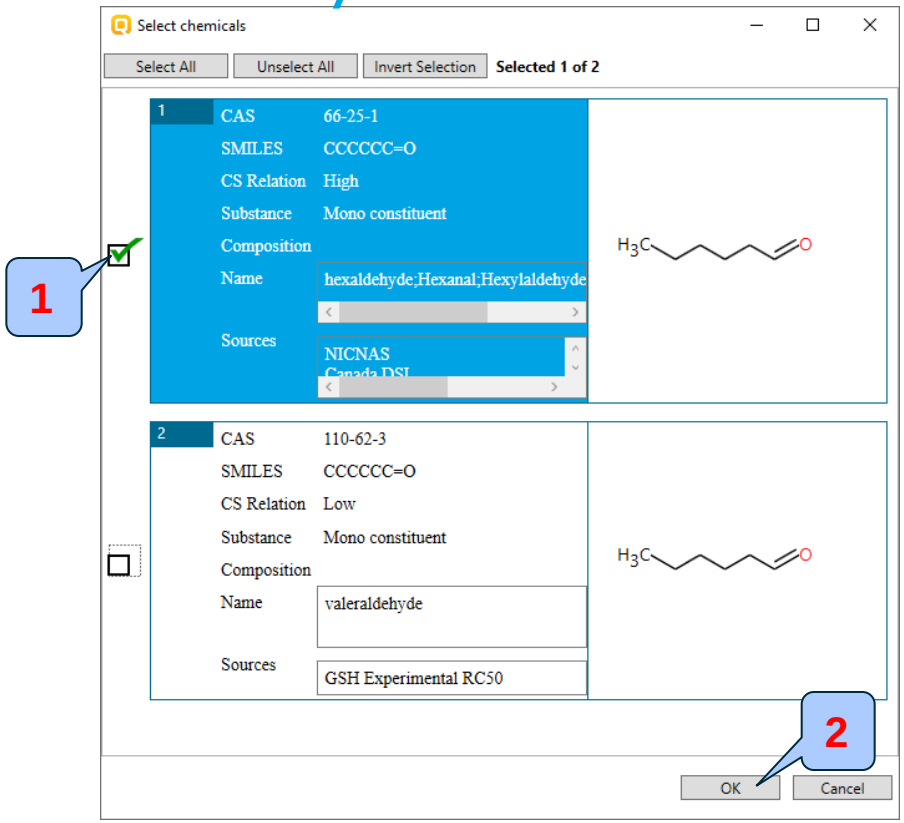
In the **2D Editor** you could:
1) directly type the SMILES of the target structure (i.e. **CCCCCCO**) in the field at the top;
2) Paste the copied SMILES;
3) draw the structure of the target by using the 2D Editor tools

1. Click on the **Drawing tool** and draw a linear chain with seven carbon atoms (while you are drawing, the SMILES of the structure will appear above); 2. Click over the last bond that will convert it to double bond; 3. Select the oxygen symbol (**O**); 4. Click on the terminal carbon. This will convert it to oxygen; 5. Confirm by **OK**.

Input
Input target chemical by SMILES

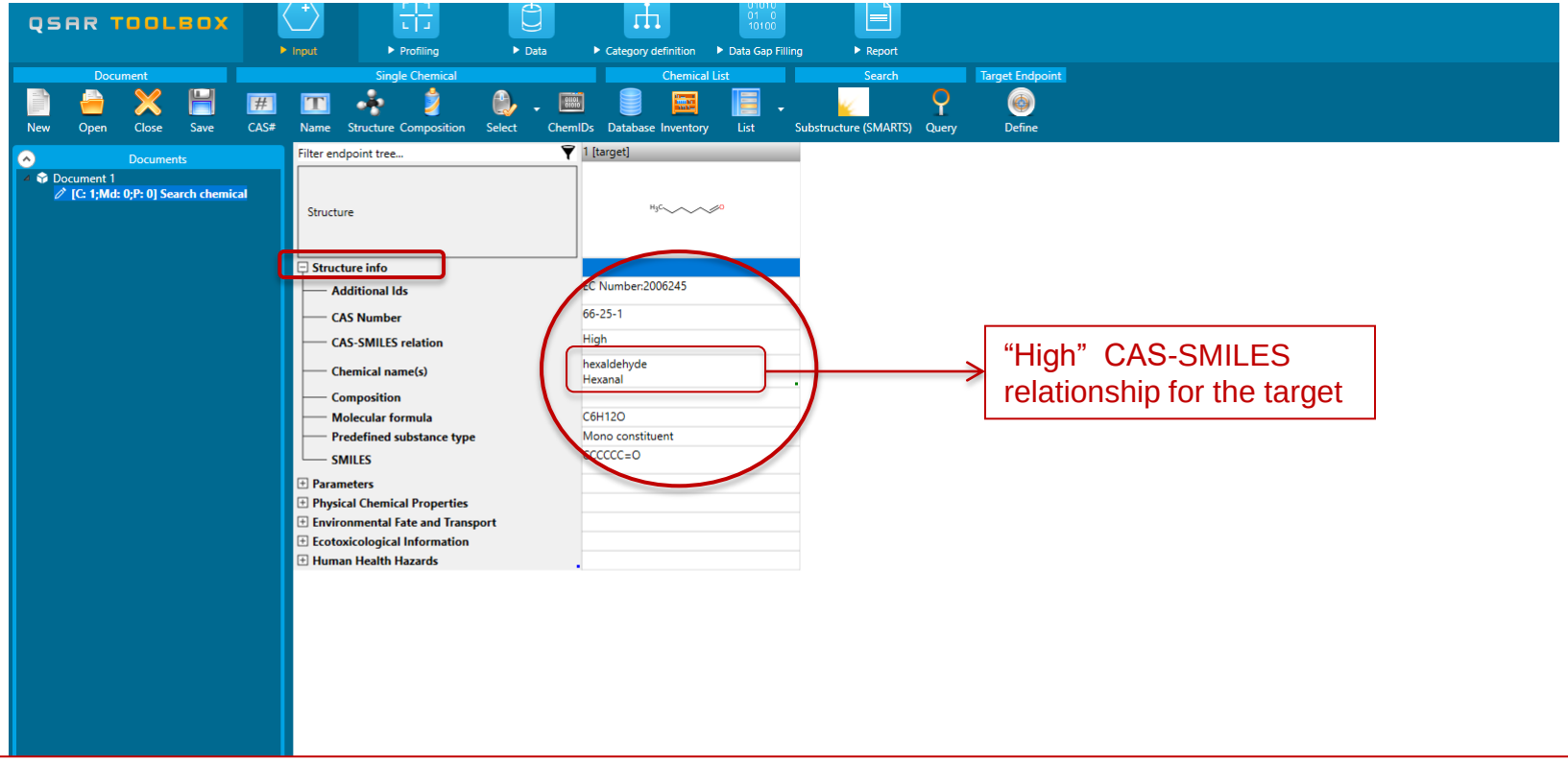
The Toolbox now searches the Toolbox databases and inventories for the presence of a chemical with structure related to the current SMILES notation. It is depicted as a 2D image.

Two chemicals are found. By default they are unselected. Select the chemical with high CAS-SMILES relationship (CS Relation).



1. **Select** the first chemical by clicking on the box in front it; 2. **Click** OK.

Input
Target chemical identity



- Expand "Structure info" field to display chemical identification information.
- It is important to remember that the workflow is based on the structure coded in SMILES.
- See the next slide for more details on the CAS-SMILES relationship.

Chemical Input
Target chemical identity

The code indicates the reliability of the chemical identifier:

- **High:** This reliability corresponds to high reliability of CAS-SMILES relation. This label is assigned if the chemical belongs to at least one high quality data source (database or inventory)

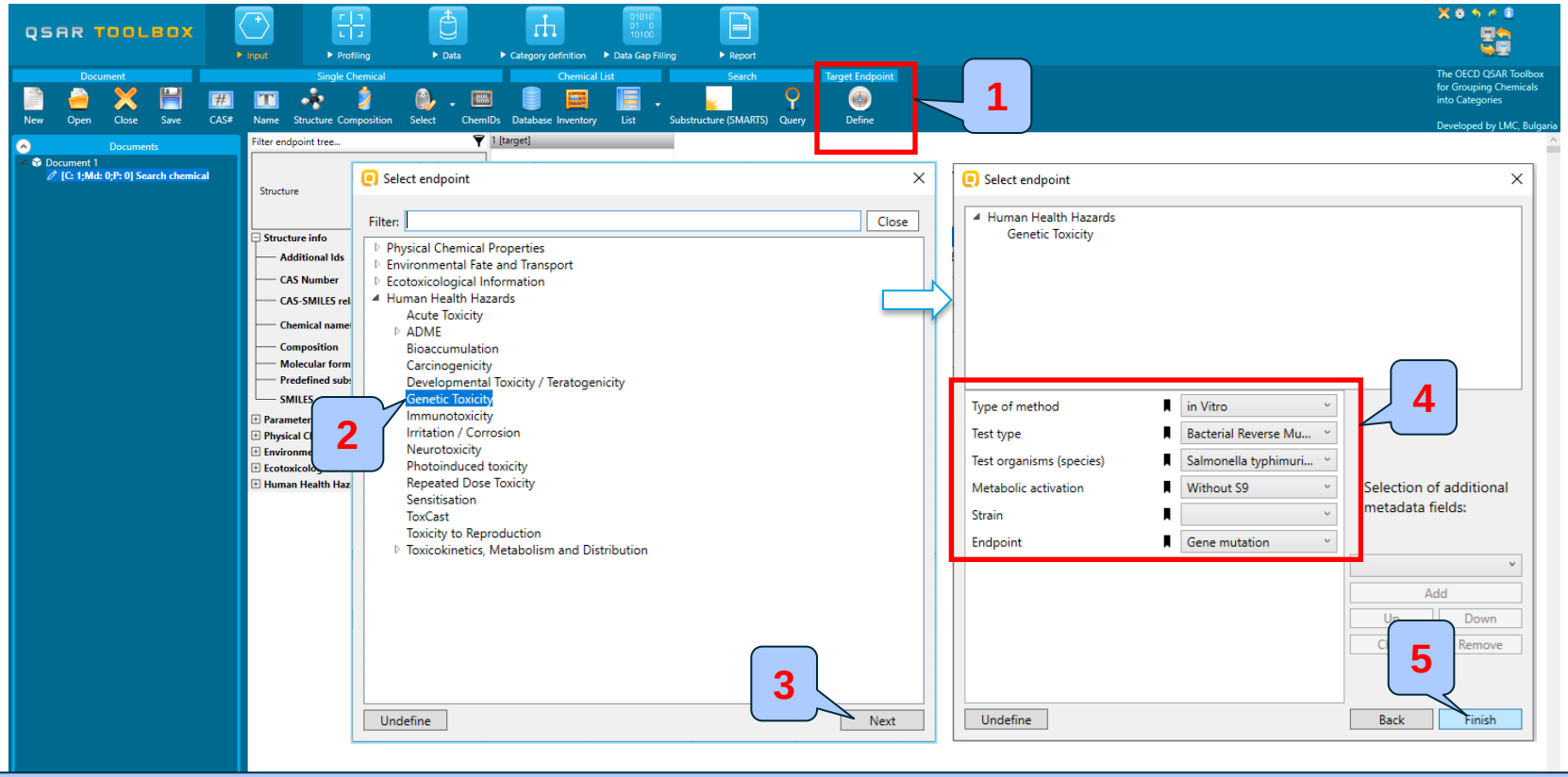
- **Moderate:** This reliability corresponds to moderate reliability of CAS-SMILES relation. The moderate label is assigned if the chemical belongs to three or more sources with unknown quality (marked with "Distribute to QA").

- **Low:** This reliability corresponds to poor reliability of CAS-SMILES relation. This label is assigned if the chemical belongs to less than three, but at least one source with unknown quality ("Distribute to QA").

Input
Define target endpoint - overview

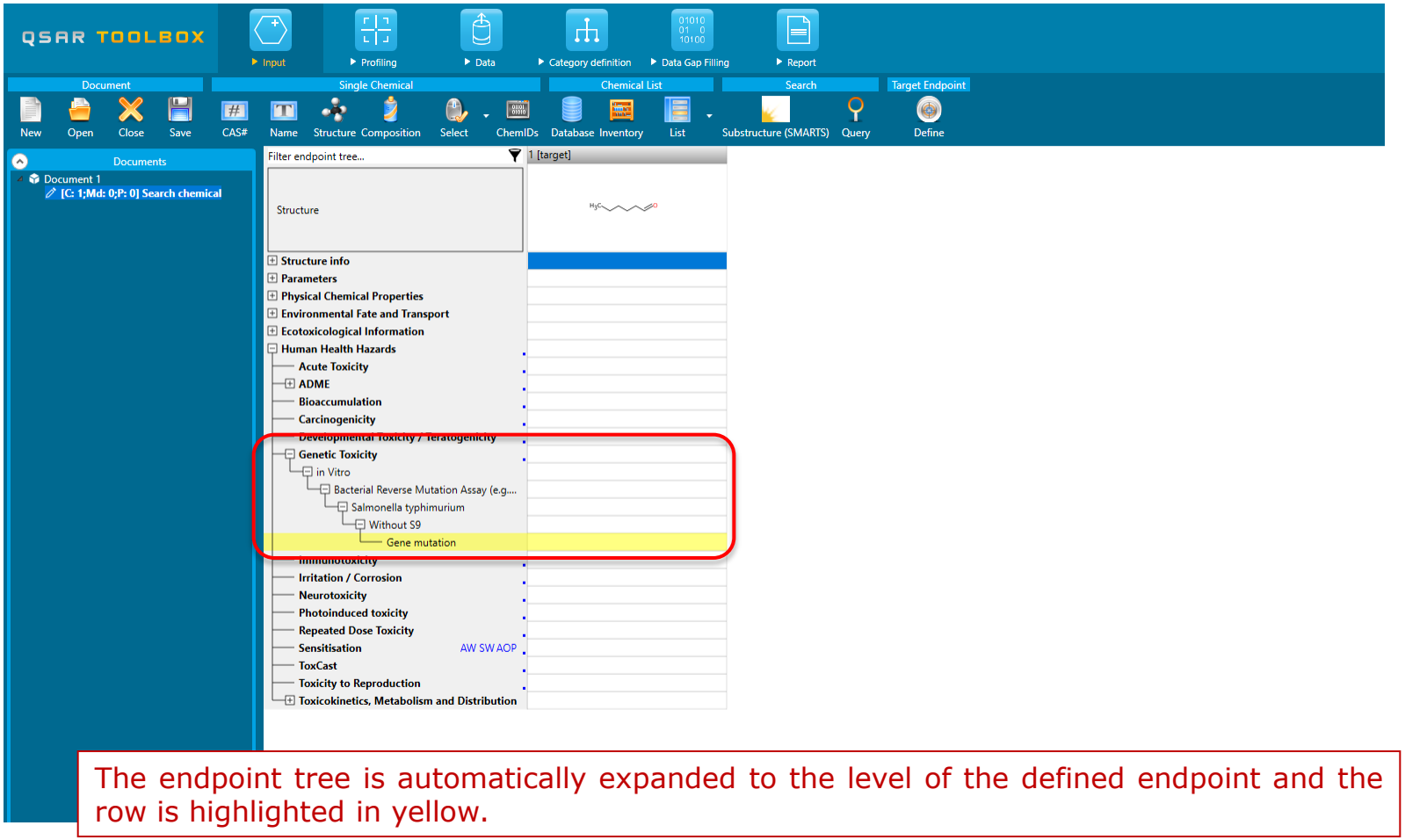
- The Define target endpoint functionality allows for entering the endpoint of interest e.g., EC3, LC50, gene mutation etc.
- The relevant profiles and databases become highlighted in colour once the targeted endpoint is preliminary defined by this functionality;
- Calculation of alert performance (AP) is only possible if the target endpoint is preliminary defined;
- 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.

Input
Define target endpoint



1. Click on the **Define** button; 2. Select "***Genetic Toxicity***" from *Human health hazard* level; 3. Click on **Next**; 4. Specify the endpoint using the drop-down menus as follows: For *Endpoint* select **Gene mutation**, for *Type of method* – **in Vitro**; for *Test type* - **Bacterial Reverse Mutation Assay (e.g. Ames Test)**, for *Test organism (species)* - **Salmonella typhimurium**, for *Metabolic activation* – **Without S9**; 5. Finally click on **Finish**.

Input
Define target endpoint



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

Input
Input results

1) In module *Input*, you have entered the target chemical. The target has high CAS-SMILES relationship.
2) The target endpoint (gene mutation) is defined using "Define target endpoint" functionality.
3) Based on the defined target endpoint the relevant profiles and databases become highlighted (see next slides).

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
  - Input
  - **Profiling**

Profiling
Overview

- "Profilers" are a collection of empirical and mechanism knowledge 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).
- "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.
- 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.
- The "Profiling" module contains also observed and simulated metabolisms/transformations, which could be used in combination with the profiling schemes.

Profiling
Overview

- To help the user to choose the most appropriate profiling methods, the profilers are highlighted in different colors:
  - **in green** – the most suitable for the target endpoint profilers. These are the profilers developed using data/knowledge associated with the mechanisms conditioning the target endpoint (e.g. Protein binding alerts for skin sensitization by OASIS);
  - **in orange -** are indicated the plausible profilers. These are the profilers for which data/knowledge used for building them is known to be somehow related to the target endpoint, i.e. these which are not directly related to the target endpoint, but still could be used (e.g. Organic functional groups),
  - unclassified – these are profilers for which there is no evidence for the relation data/knowledge used for building them and the target endpoint.
- The profilers identifying structural groups are marked as plausible, while the profilers based on mechanistic knowledge are highlighted in green (see next few slides).

Profiling

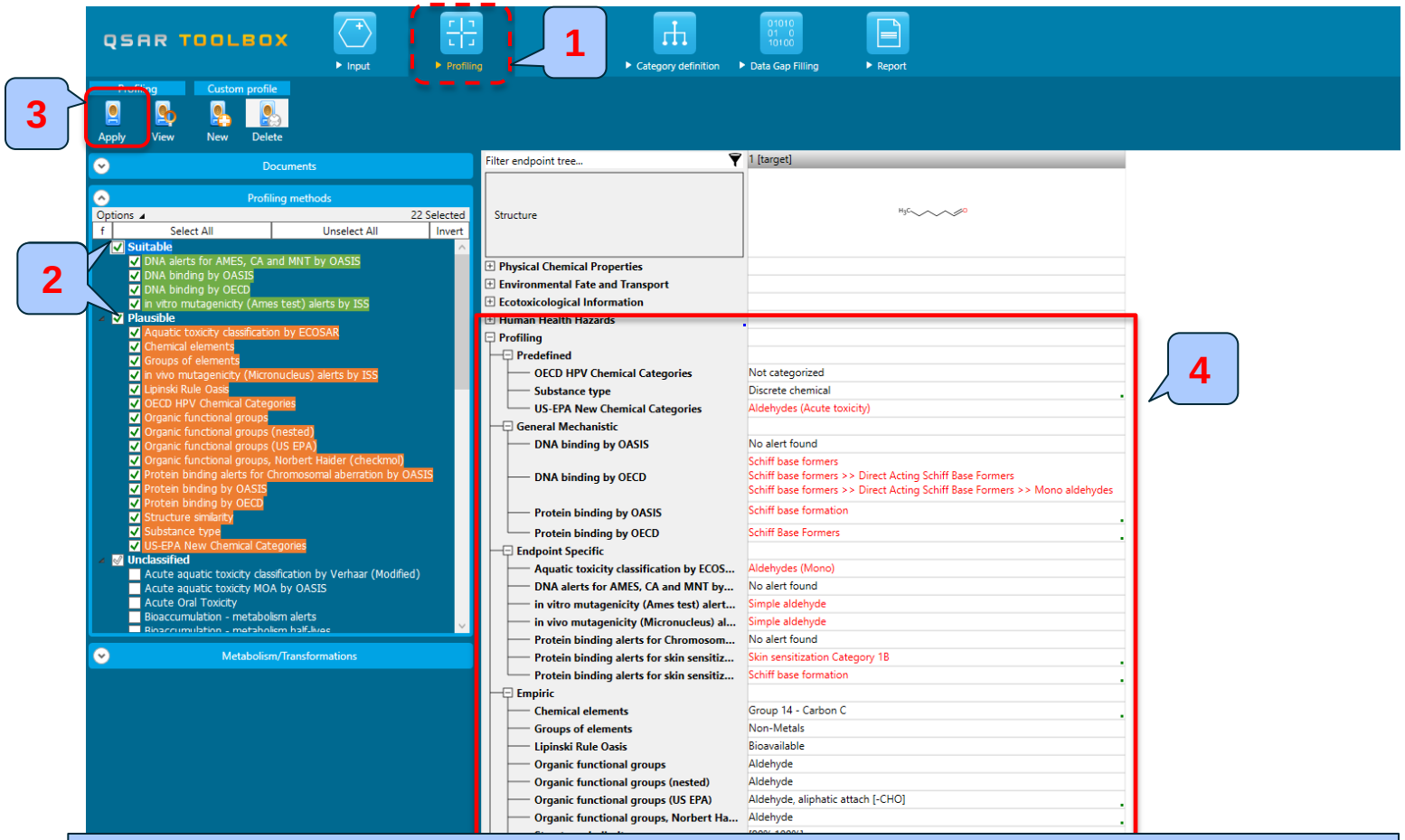
Profiling the target chemical - background

- The following profiling methods are relevant to the defined target endpoint (Ames mutagenicity):
  - Suitable profilers
    - DNA alerts for AMES, CA and MNT by OASIS
    - DNA binding by OASIS
    - DNA binding by OECD
    - in vitro mutagenicity (Ames test) alerts by ISS
  - Plausible profilers
    - Aquatic toxicity classification by ECOSAR*
    - Organic function groups
    - ……

*Aquatic toxicity classification by ECOSAR is a rule-based profiler, which very well describe the functional groups present in the target molecules (It is a method for identifying chemical classes). Hence this method is one of the most robust of the mechanistic grouping method and it is often the method of choice for different hazards endpoints (e.g. acute aquatic toxicity and skin sensitization).

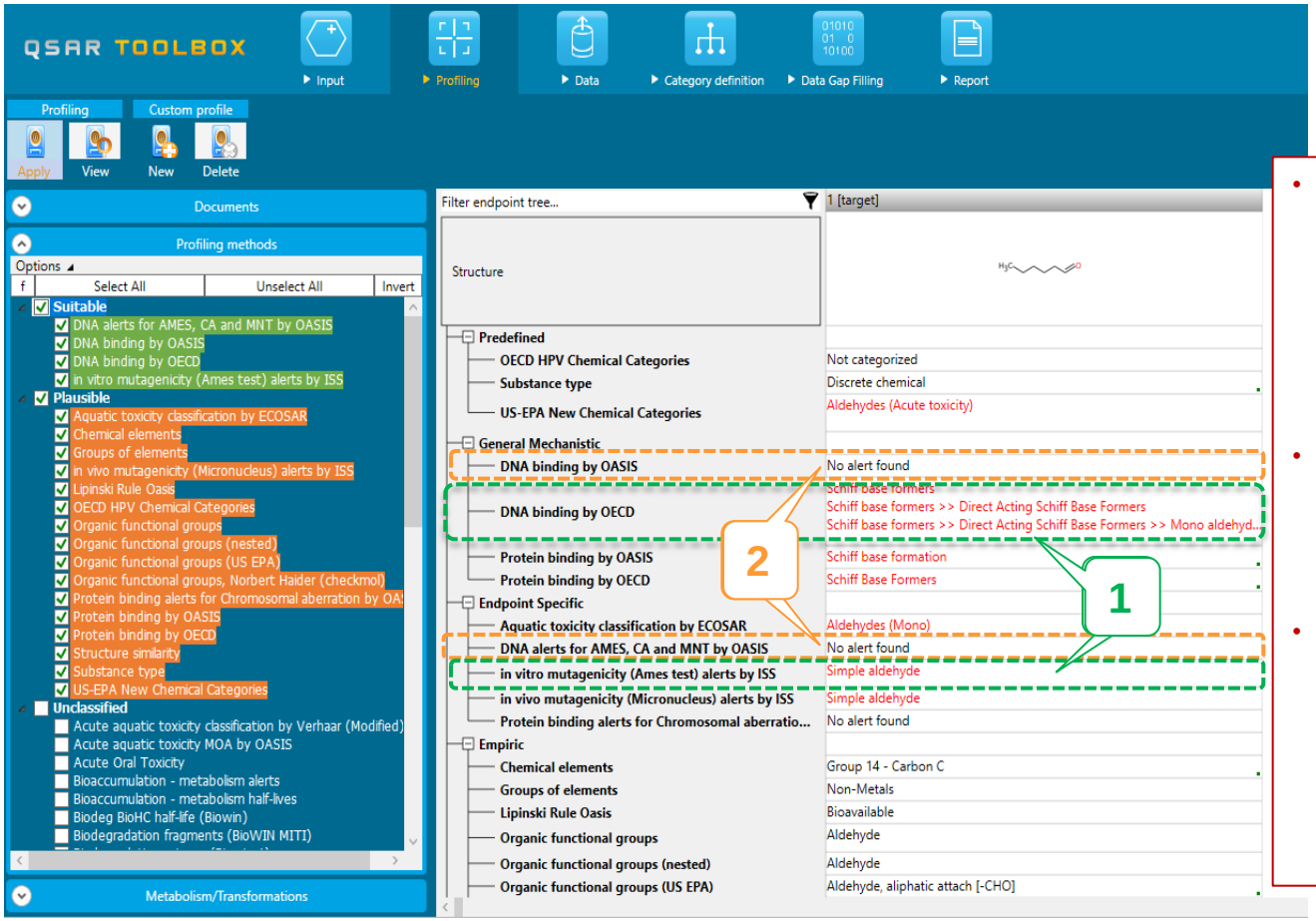
Profiling

Profiling the target chemical - background



1. Go *Profiling* module 2. Check all highlighted (in green and in orange) profilers; 3. Click **Apply;** 4. The results appear on data matrix

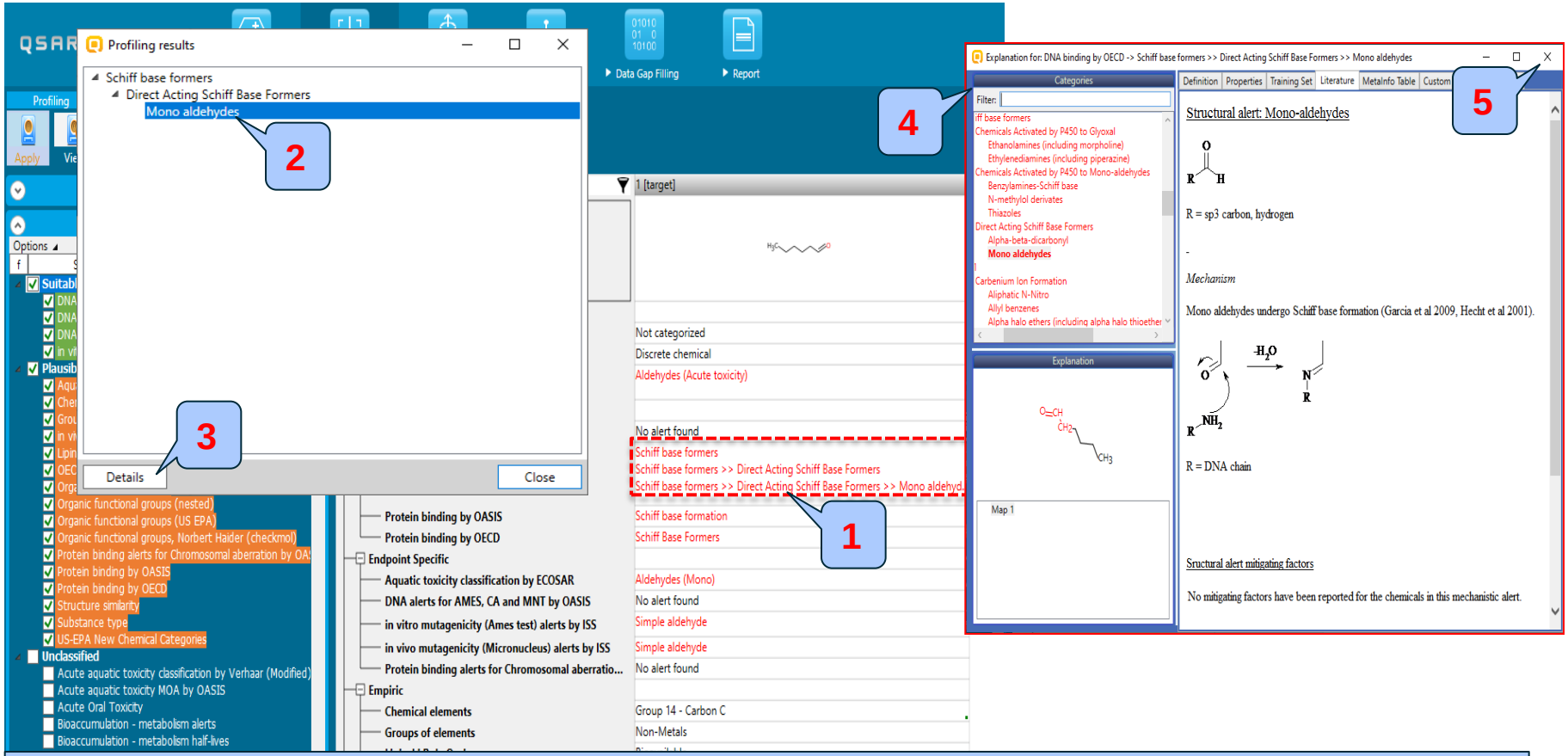
Profiling
Profiles of n-hexanal



- In this case an alert associated with DNA binding (i.e. Mono aldehydes acting as Schiff base formers) is identified by the general mechanistic *DNA binding by OECD* profiler as well as by the endpoint specific *in vitro mutagenicity (Ames test) alerts by ISS* (1)
- However, no alerts are found by the general mechanistic *DNA binding by OASIS* profile as well as by the endpoint specific *DNA alerts for AMES, CA and MNT by OASIS* (2)
- Due to the contradictory results, in the current example, the *Organic functional groups* profiler will be used for categorization purposes. This structure-based profiler will allow as to form broader primary group of analogues that could be further reduced.

Profiling

Mechanistic justification of profiling results



1) Double click on the cell and to see why the target is profiled as "Mono aldehydes" by *DNA binding by OECD* profiler; 2) Select the category **"Mono aldehydes"**; 3) Click **Details**; 4) A new window with literature info appears. 5) Explore the mechanism and close the window.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - **Data**

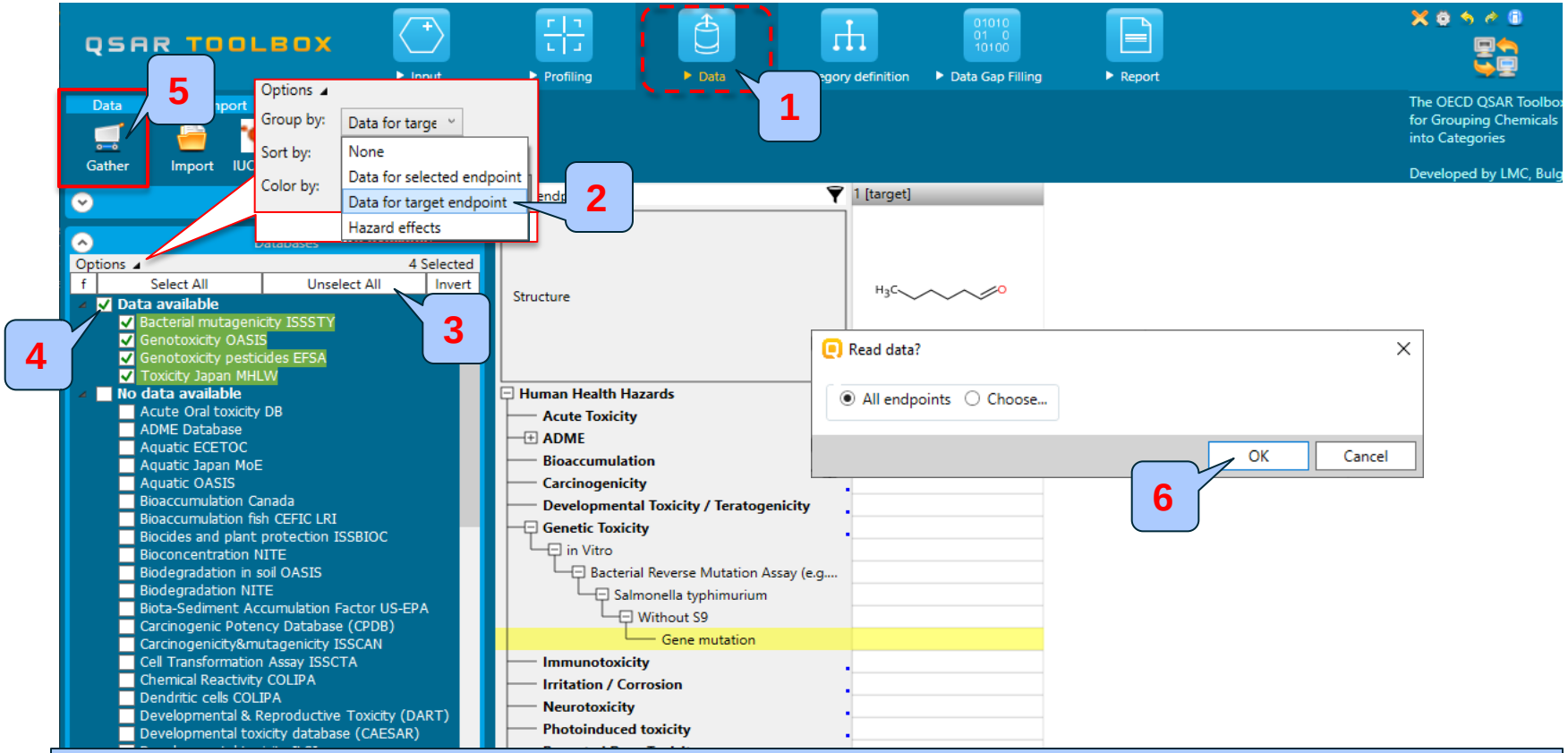
Data
Overview

- "Gather data" refers to the electronic process of retrieving the fate and toxicity data that are stored in the Toolbox database.

- Note, data can be gathered in a global fashion (i.e., collecting all data of all endpoints) or on more narrowly defined settings (e.g., collecting data for a single or limited number of endpoints)

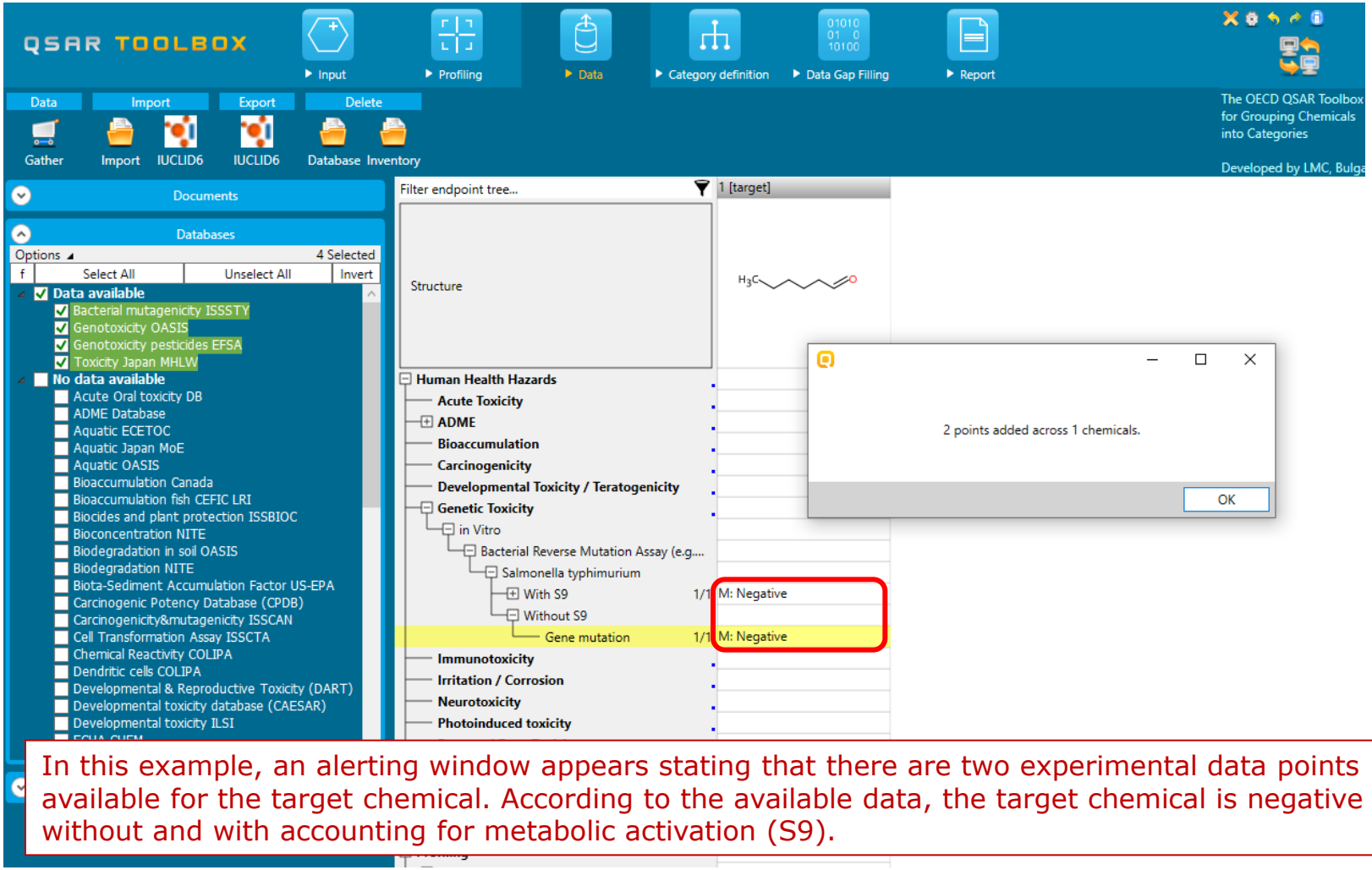
- Database "relevancy" is determined based on the defined target endpoint (see next slide).

Data
Gather data



1. Go to ***Data*** module. The databases containing data for the defined target endpoint will be highlighted in green. 2. Group the databases according to the target endpoint (all green databases will appear at the top); 3. Unselect all previously selected databases; 4. Check the box in front of "Data available"; 5. Click **Gather**; A window with "Read data?" appears.; 6. Click **OK** to collect all data for the target chemical available in the selected databases.

Data
Process of collecting data



In this example, an alerting window appears stating that there are two experimental data points available for the target chemical. According to the available data, the target chemical is negative without and with accounting for metabolic activation (S9).

Data
Recap

- All databases containing data for the endpoint (Ames mutagenicity) have been highlighted and grouped according to the target endpoint.
- Two negative experimental data results for n-hexanal (the target chemical) have been found in the selected databases.
- We will try to reproduce the experimental data by searching of analogues and making use of a read-across approach.
- Click on “Category definition” to move to the next module.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - Data
  - **Category definition**

Category Definition
Overview

- As stated in the previous tutorial, this module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- This is the critical step in the workflow of the Toolbox.
- Several options are available in the Toolbox to assist the user in defining the category definition.
- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of "similarity" so that within a category data gaps can be filled by read-across.

Category Definition
Grouping methods

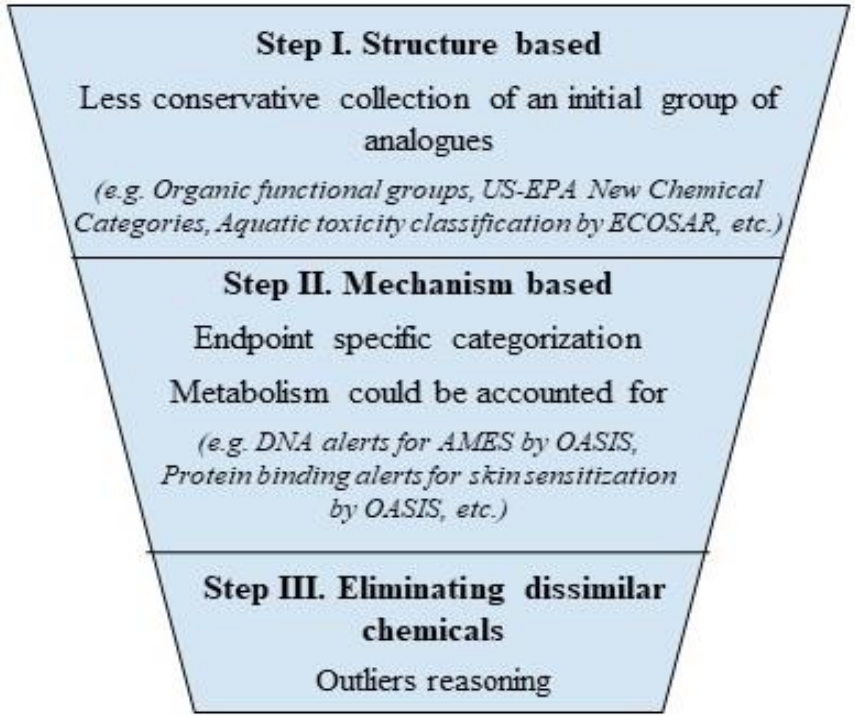
- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of "similarity" so that within a category data gaps can be filled by read-across.

- For this example, we will start from a broad group based on Organic functional group and after that

- Will refine the category by a specific DNA binding mechanism identified for the target chemical and find analogues which can bind by the same mechanism and for which experimental results are available.

Category Definition

Suitable Categorization/Assessment Phases



Broad grouping
Endpoint Non-specific

Subcategorization
Endpoint Specific

Subcategorization
Endpoint Non-specific

Note: As long as an acceptable level of structural and mechanistic similarity is achieved, it is not mandatory to follow all the stages described in the order given above; they can be executed differently or even skipped.
For example, in case the target chemical or any of its metabolites interact with biomacromolecules via a clearly defined mechanism relevant for the endpoint to predict, Stage I could be skipped and Stage II to be used for the primary categorization step.

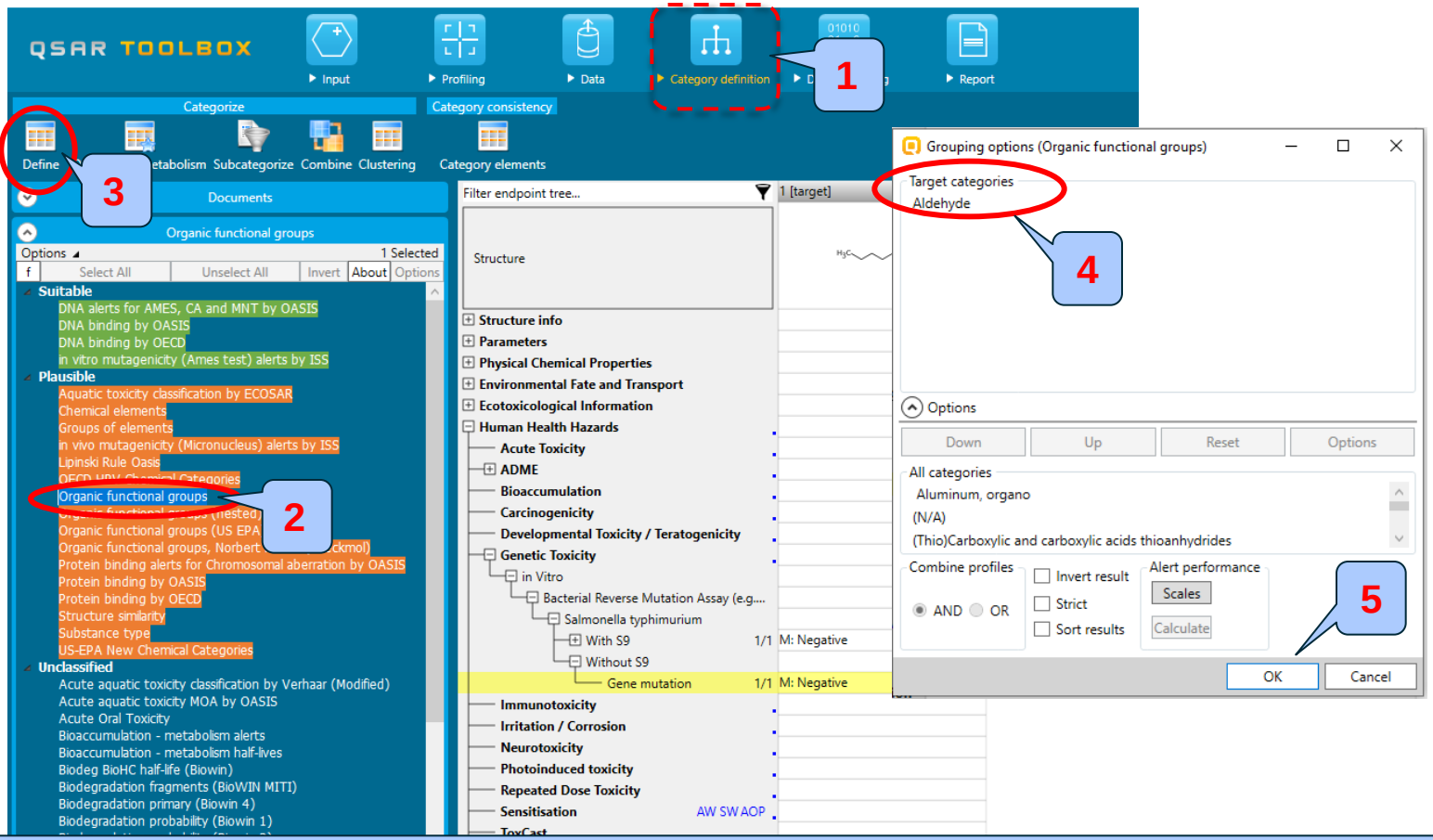
Category Definition
Suitable Categorization/Assessment Phases

- In the current example we will define a broad group of analogues using the *Organic functional groups* (OFG) profiler (step I).

- After that we will refine the category by the specific DNA binding mechanism identified in the target chemical and will keep the analogues which can bind by the same mechanism and for which experimental results are available.

Category Definition

Defining Organic functional group

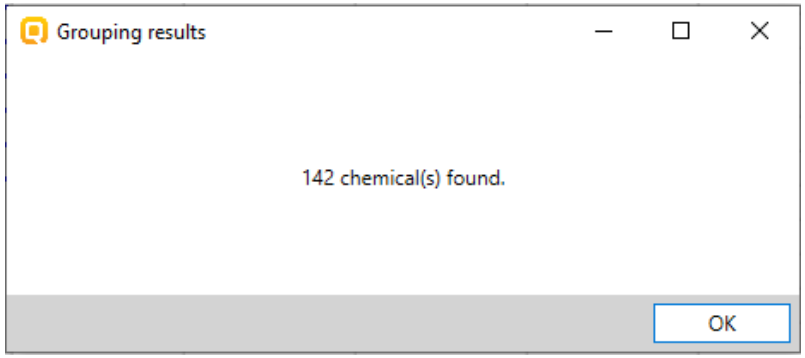


1. Go to the ***Category definition*** module; 2. Select **Organic functional groups** profiler; 3. Click **Define**; 4. The target category is *Aldehydes*; 5. Click **OK**

Category Definition
Analogues

- The Toolbox now searches for all chemicals corresponding to category "Aldehydes" by the same profiler (i.e. *Organic functional groups*) listed in the databases selected under "Data" (*see slide 42*).

- A message with numbers of found analogues (including the target chemical) appears.



Category Definition
Read data for Analogues

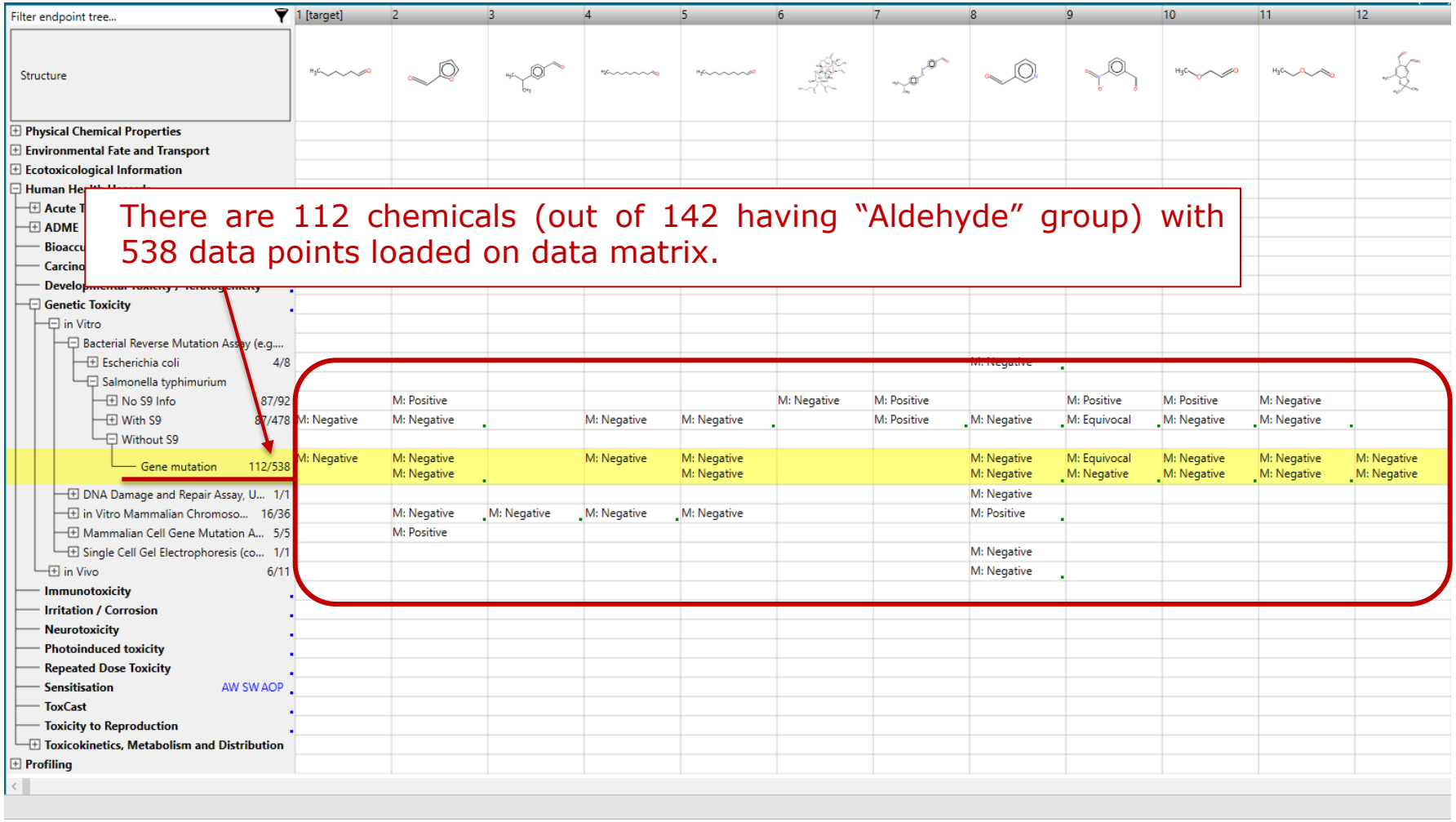
- The Toolbox automatically requests the user to select the endpoint for which the data should be retrieved.

- The user can either select the specific endpoint or by default choose to gather data for all endpoints (see below)

- In this example, because only databases containing information for genetic toxicity endpoint are selected, both options will give the same results.

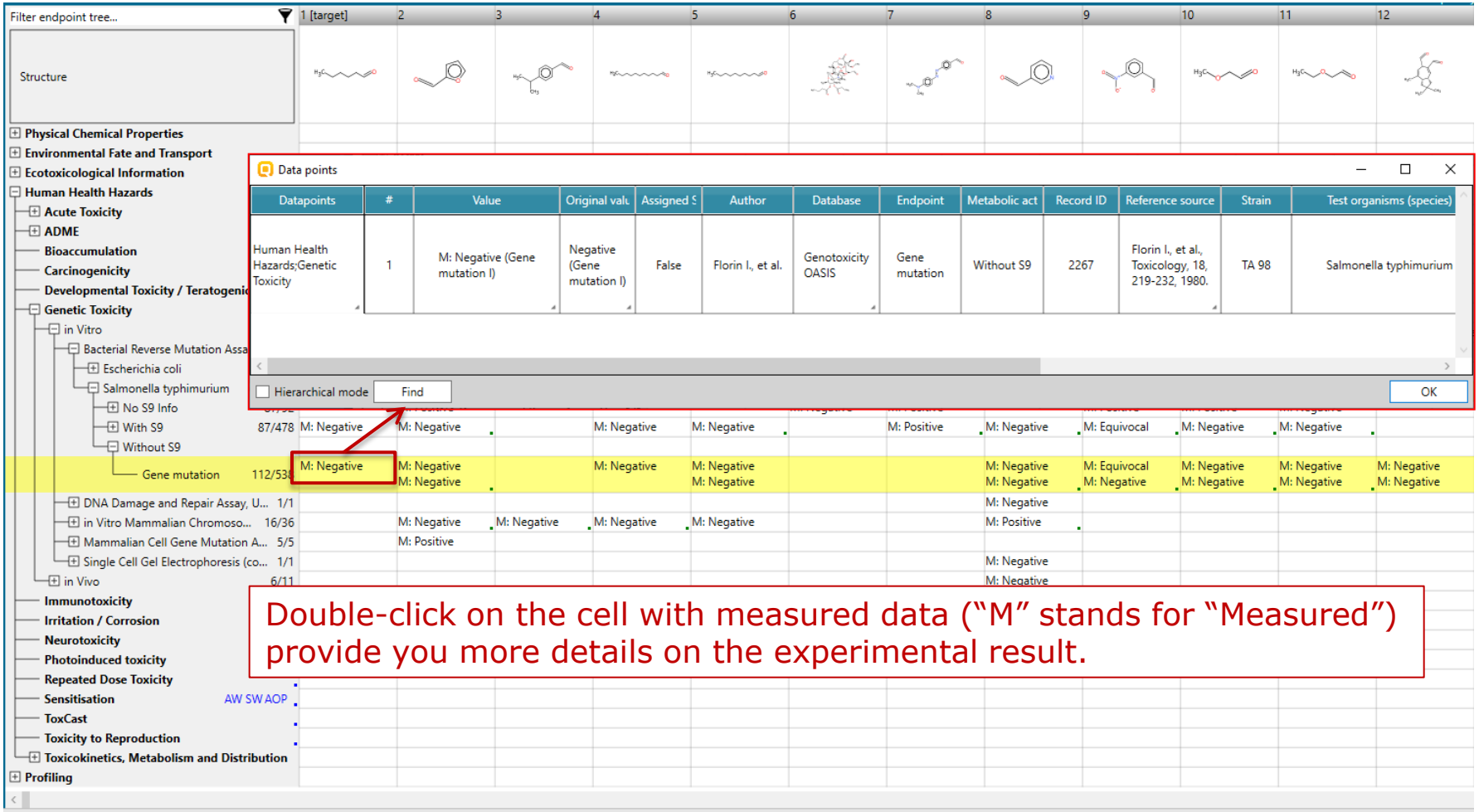
Category Definition

Summary information for Analogues



Category Definition

Side-Bar of experimental data



Double-click on the cell with measured data ("M" stands for "Measured") provide you more details on the experimental result.

Category Definition
Recap

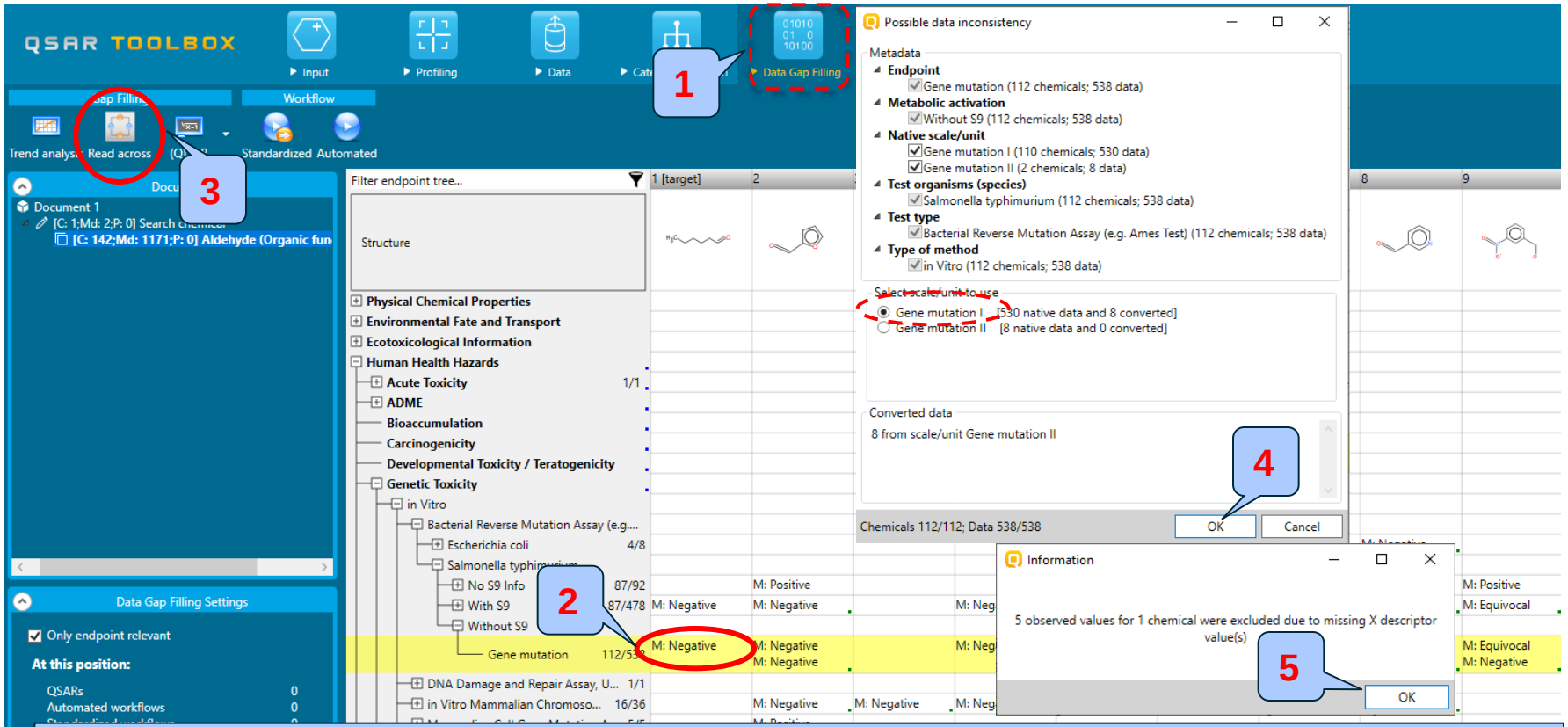
- You have defined a category consisting of 142 analogues ("Aldehydes" by OFG classification) with the target chemical (n-hexanal).
- The available experimental data for these 142 similar chemicals are collected from the previously selected databases under Data section.
- 538 data points for the target endpoint are available for 112 out of 142 chemicals.
- You are now ready to fill in the data gap and trying to reproduce the experimental data of the target.
- In this example we will apply the read-across method due to the qualitative mutagenicity data (the trend analysis method is applicable only to quantitative estimates, e.g. LC50).
- Two predictions will be done – without and with accounting for metabolic activation (i.e. S9).

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - Data
  - Category definition
  - **Data Gap Filling**
    - **Ames without S9**

Data Gap Filling

Continue with read-across for Ames without S9

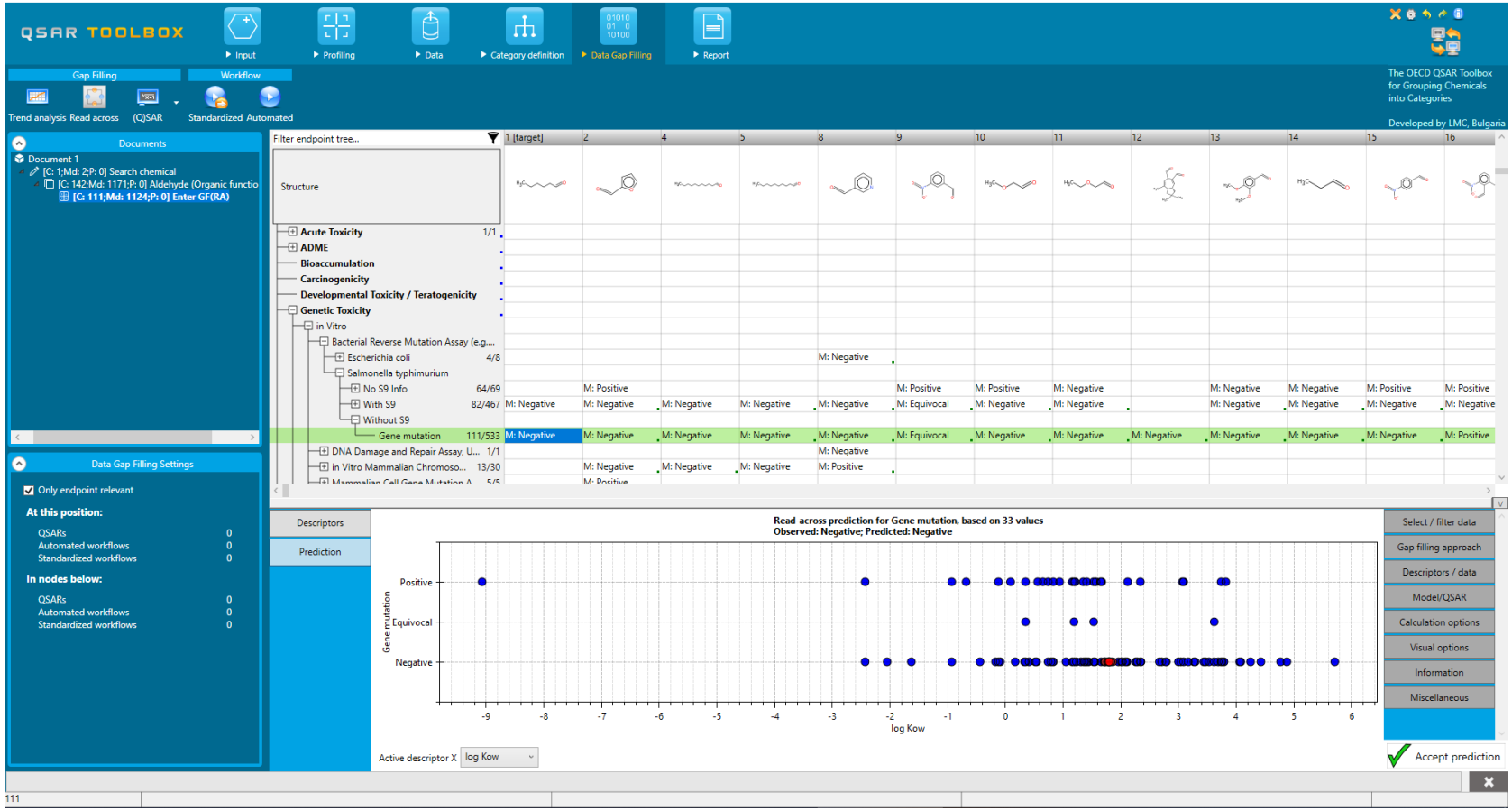


1. Go to the ***Data Gap Filling*** module; 2. Click on the cell from data matrix corresponding to the target chemical and the target endpoint; 3. Click **Read across** button; 4. Keep the default scale selection "***Gene mutation I***" – this is the most general scale converting the data only to positive/negative. Click **OK**; 5. An alerting message informing the user that 1 chemical with 5 observed value is excluded due to missing X values for it. By default the X descriptor is logKow. The excluded substances could be mixtures, UVCB, etc, for which the logKow could not be calculated. Click **OK**

Data Gap Filling (Ames without S9)
Interpretation of the Read across

- The resulting plot outlines the experimental Ames results of all analogues (Y axis) according to a descriptor (X axis). Note, Log Kow is on the X-axis; while this descriptor is not significant to Ames data, it is the default descriptor for data gap filing (see next screen shot).
- The **RED** dot represents the predicted value for target chemical (see next screen shot).
- The **BROWN** dots represent the observed value for the target neighbours (analogues) used for read-across.
- The **BLUE** dots represent the experimental results available for the analogues but not used for read-across (see next screen shot).
- Please note **LIGHT BLUE** dots (which you will see shortly) represent analogues belonging to different subcategories.

Data Gap Filling
Results for Ames without S9

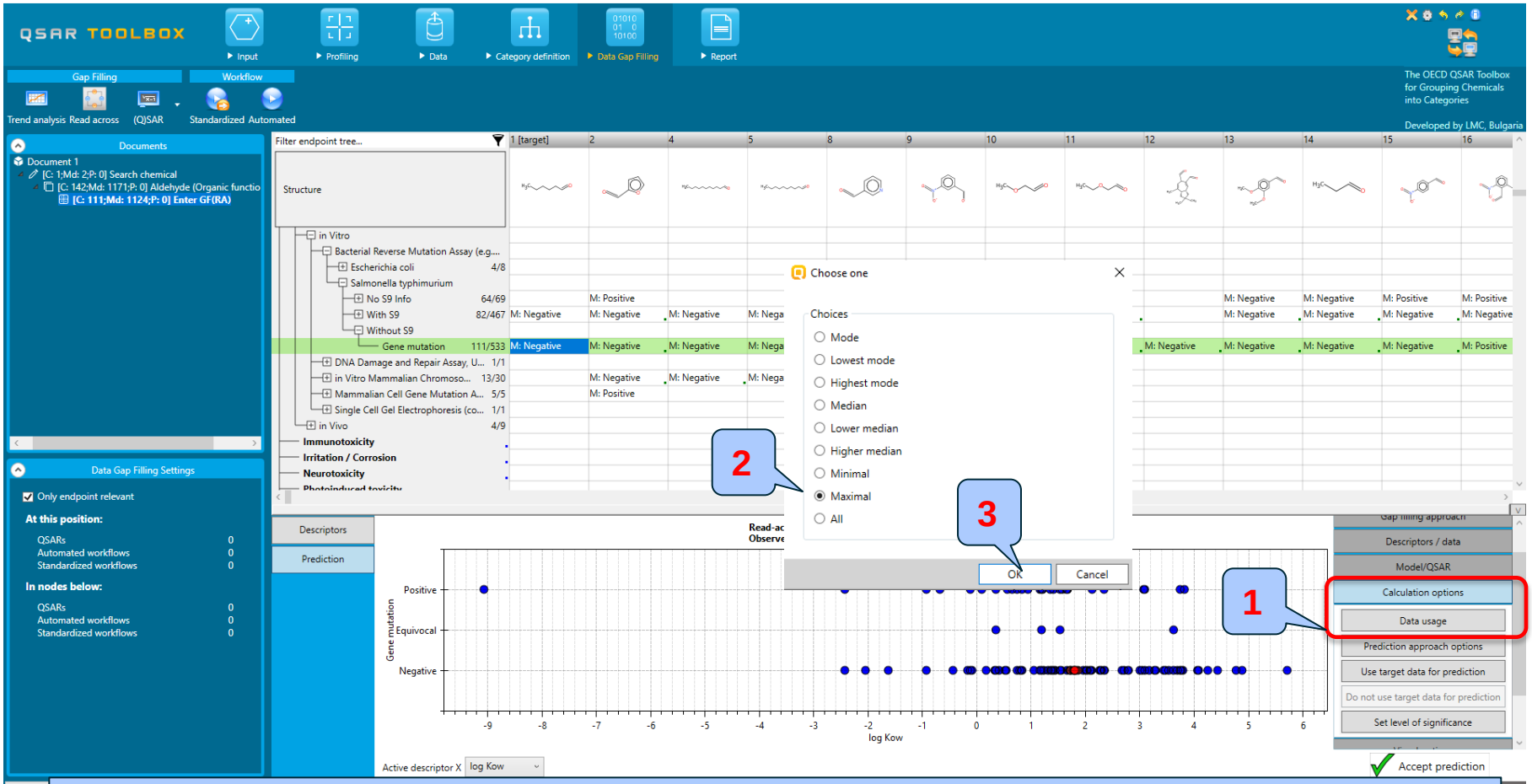


Data Gap Filling (Ames without S9)
Interpretation of the Read across

- As seen there are mutagenic and not mutagenic analogues of the target chemical
- Therefore, the fist step will be to apply the worst-case scenario, i.e. if a chemical has positive and negative data simultaneously, then the positive one to be taken into account.
- Before data gap filling it is also recommended to check the similarity of the analogues used for the prediction. This is performed in order to assure the category consists of analogues that are both mechanistically and structurally similar. The latter is done by a so called “subcategorization”.

Data Gap Filling (Ames without S9)

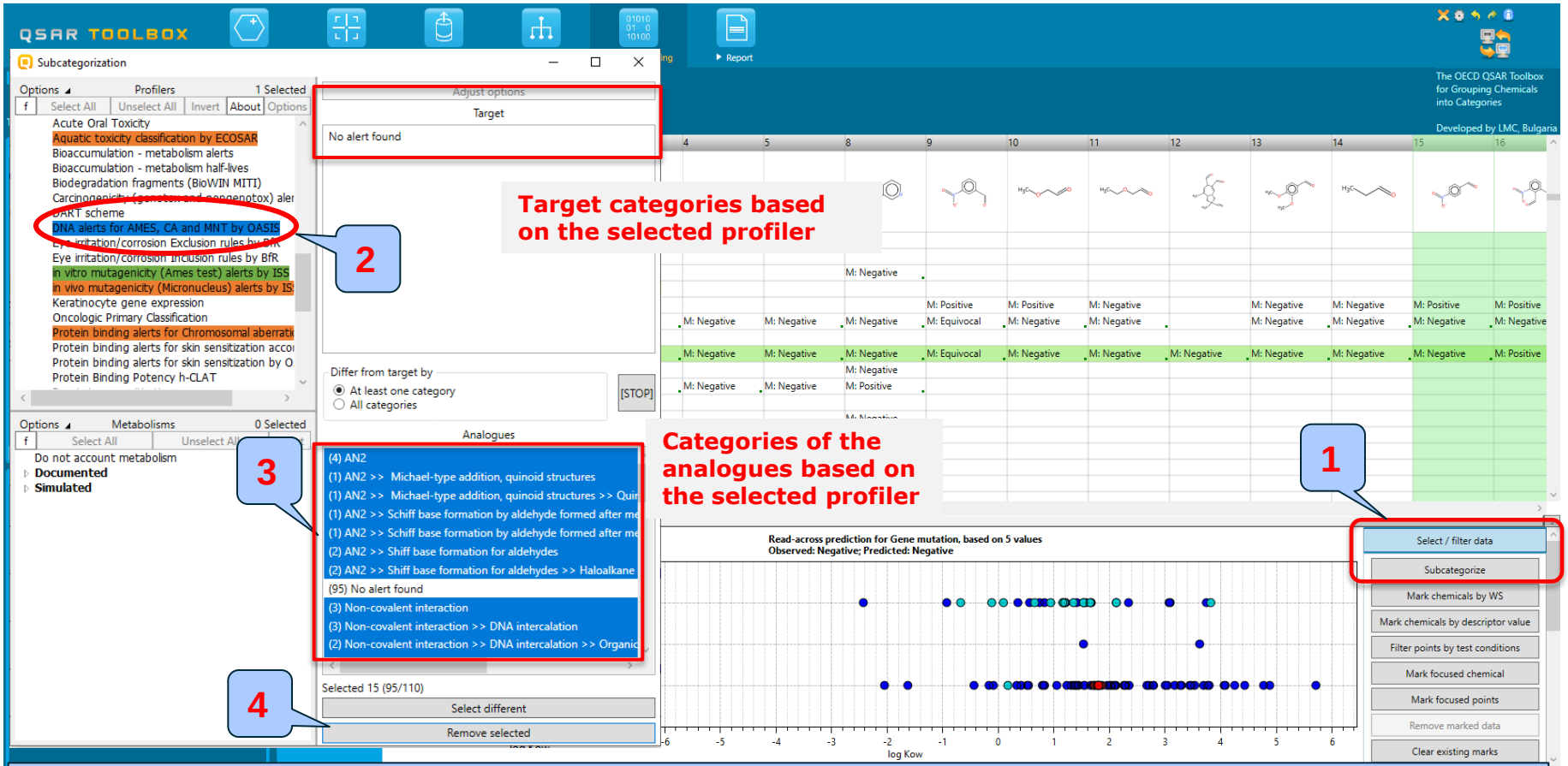
Take the worst-case scenario



1. Go to **Calculation options >> Data usage**; 2. Select **Maximal** value to be taken into account (i.e. in case of positive/negative data, the positive is the maximal data); 3. Click **OK**.

Data Gap Filling (Ames without S9)

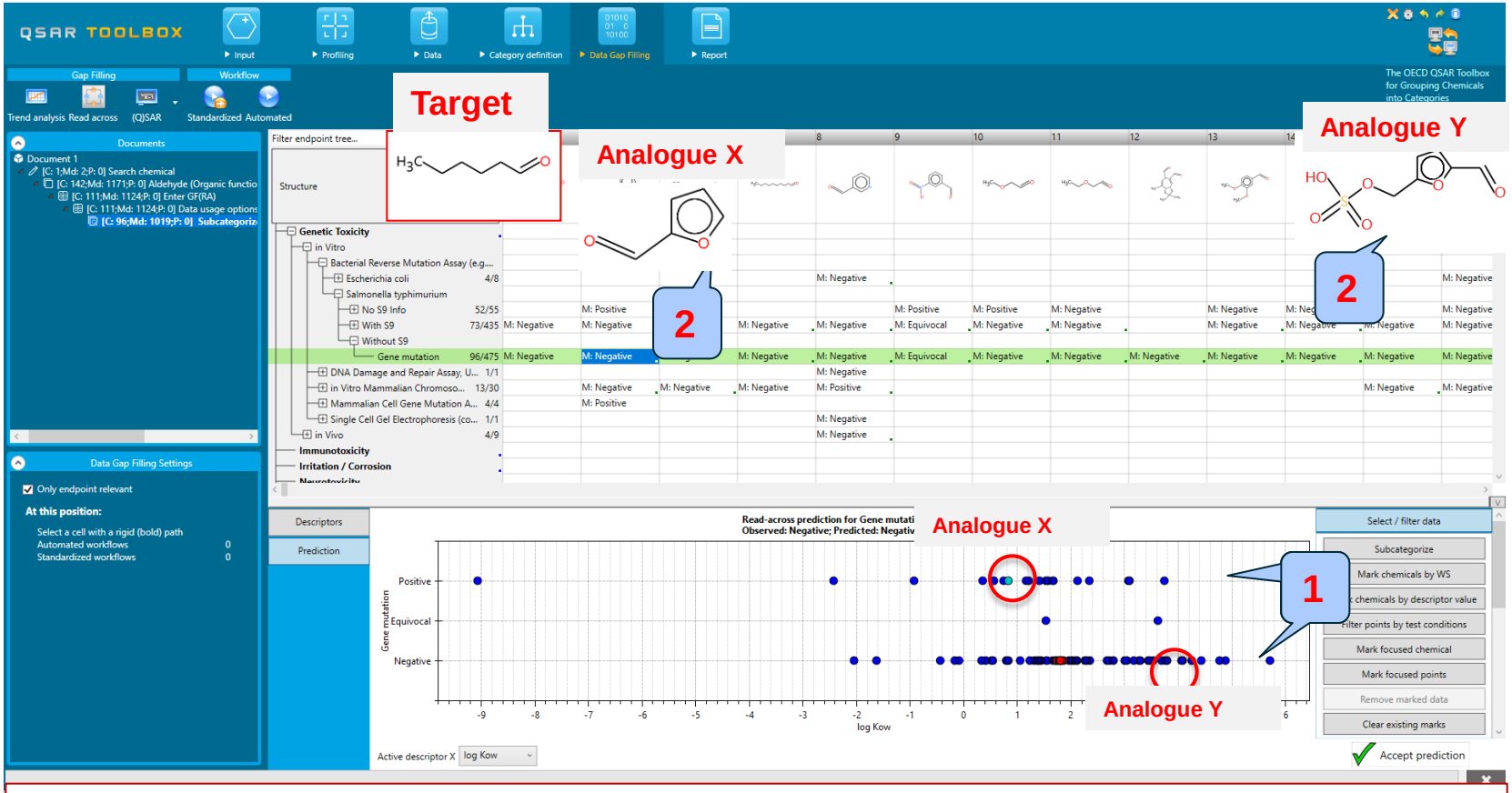
Subcategorization by DNA alerts for AMES, CA and MNT by OASIS (endpoint-specific)



1. Go to **Select / filter data** >> **Subcategorize** 2. Select ***DNA alerts for AMES, CA and MNT by OASIS*** profiler; 3. The chemicals different to the target according to this profiler will be highlighted in light blue; 4. **Click** Remove selected.

Data Gap Filling (Ames without S9)

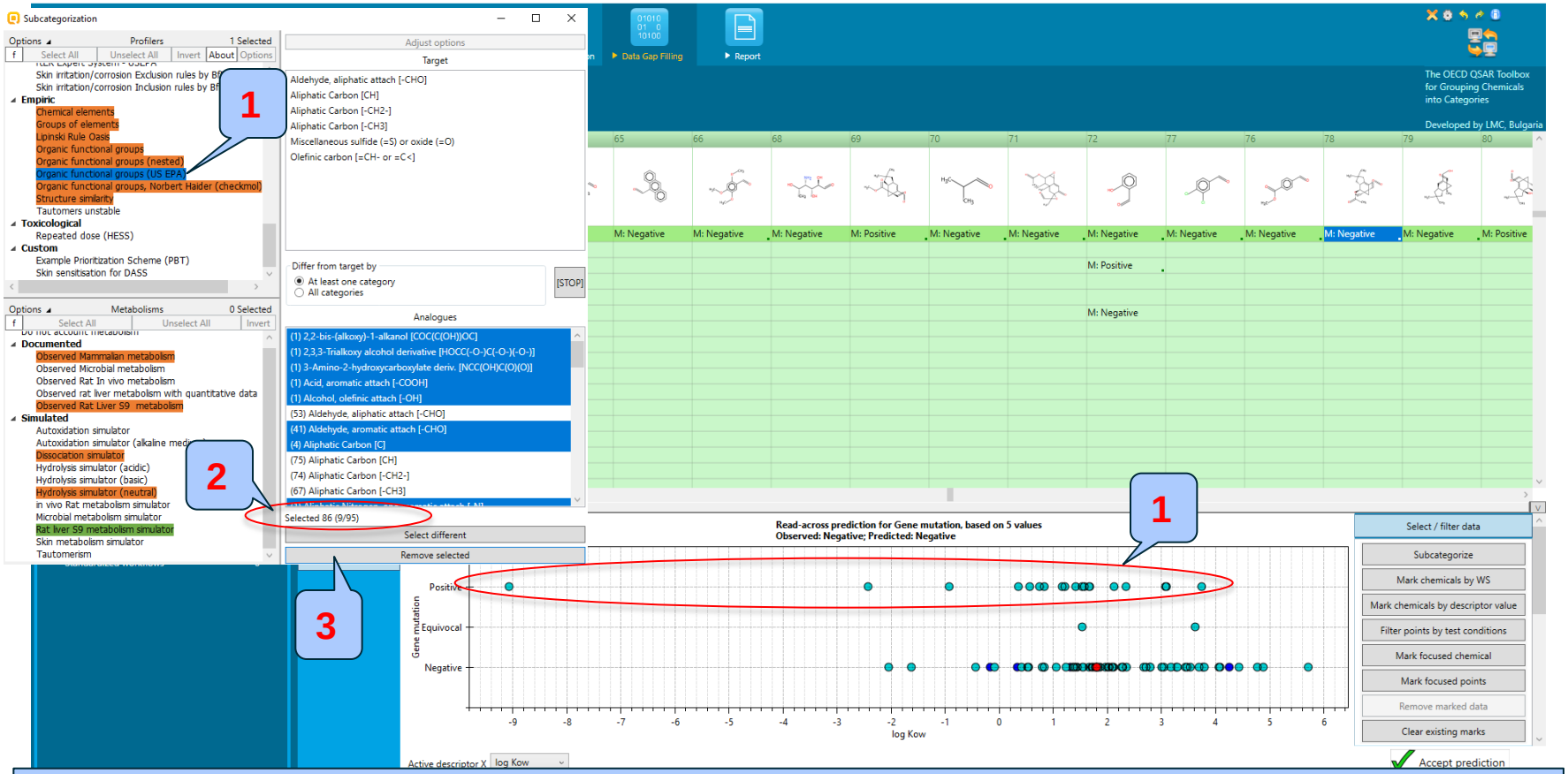
Analysis of the remaining analogues



As seen there are still positive and negative analogues (1). Also there are many of the analogues which are quite dissimilar to our target chemical (2). In order to eliminate dissimilar analogue (structurally dissimilar), we will use one of the organic functional groups profiler for subcategorization (see the next slide).

Data Gap Filling (Ames without S9)

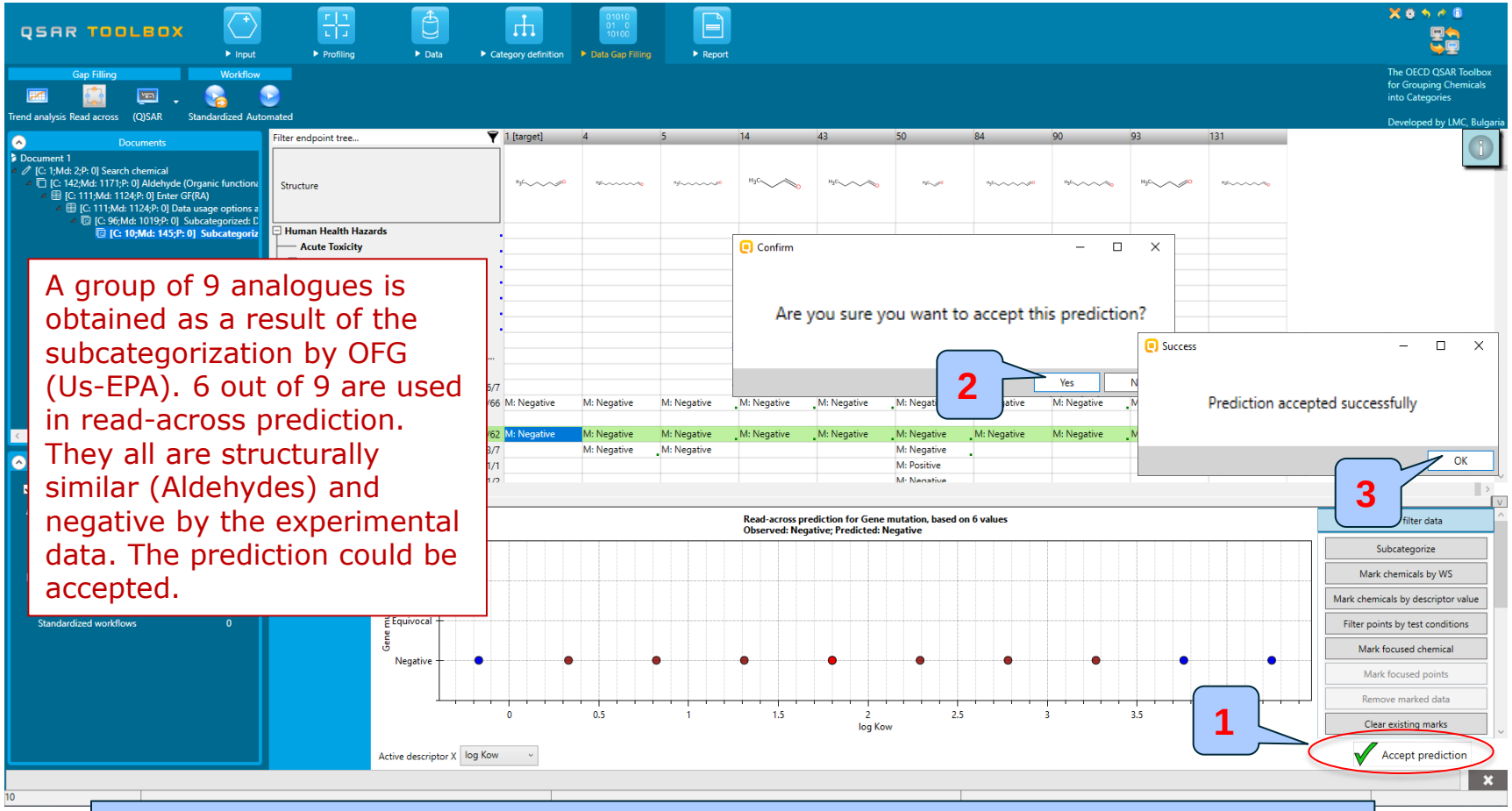
Subcategorization by Organic functional groups (US-EPA) (empiric)



1. Explore the different types of organic functional groups profilers. We can see that ***Organic functional groups (US-EPA)*** profiler removes all positive analogues, i.e. it leads to consistent results; 2. The statistic shows that 86 chemicals are identified as different to the target and nine chemicals out of 95 analogues will remain after applying of subcategorization with this profiler. 3. Click ***Remove selected***.

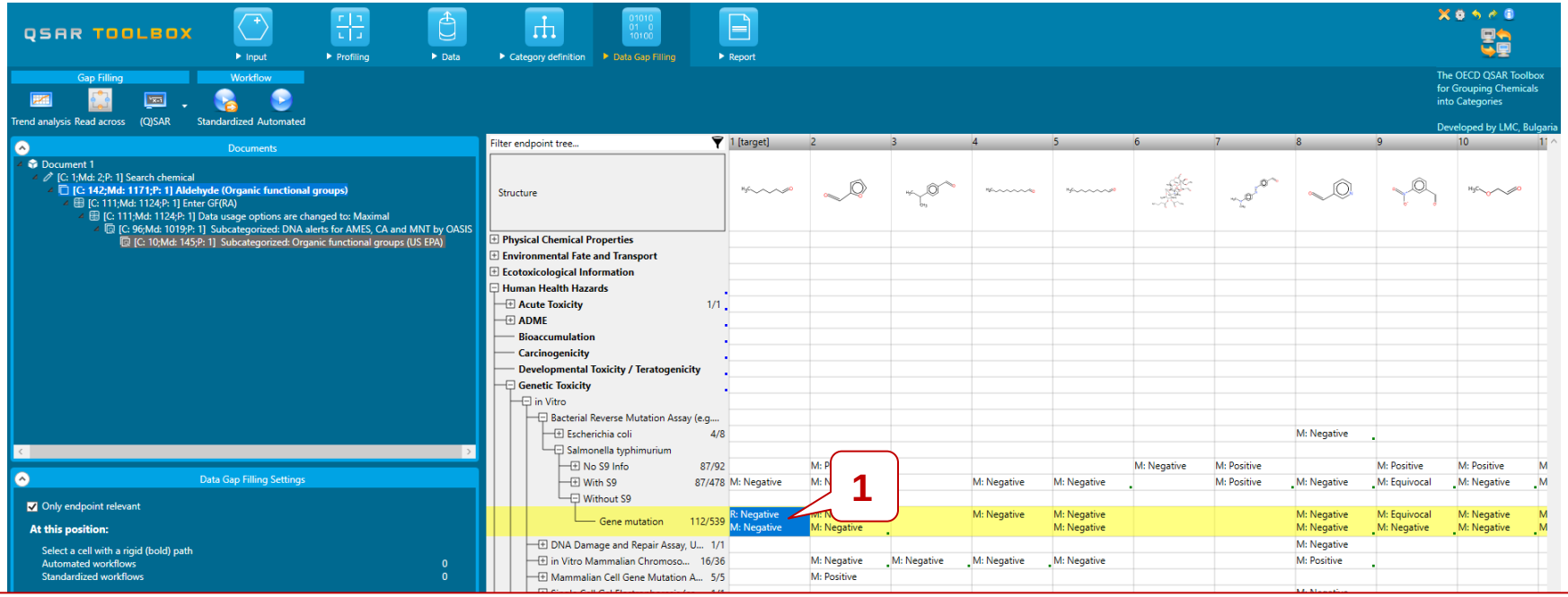
Data Gap Filling (Ames without S9)

Read across result



1. Click on **"Accept prediction"**; 2. Click **Yes** to confirm he acceptance; 3. Click **OK**.

Data Gap Filling
Results



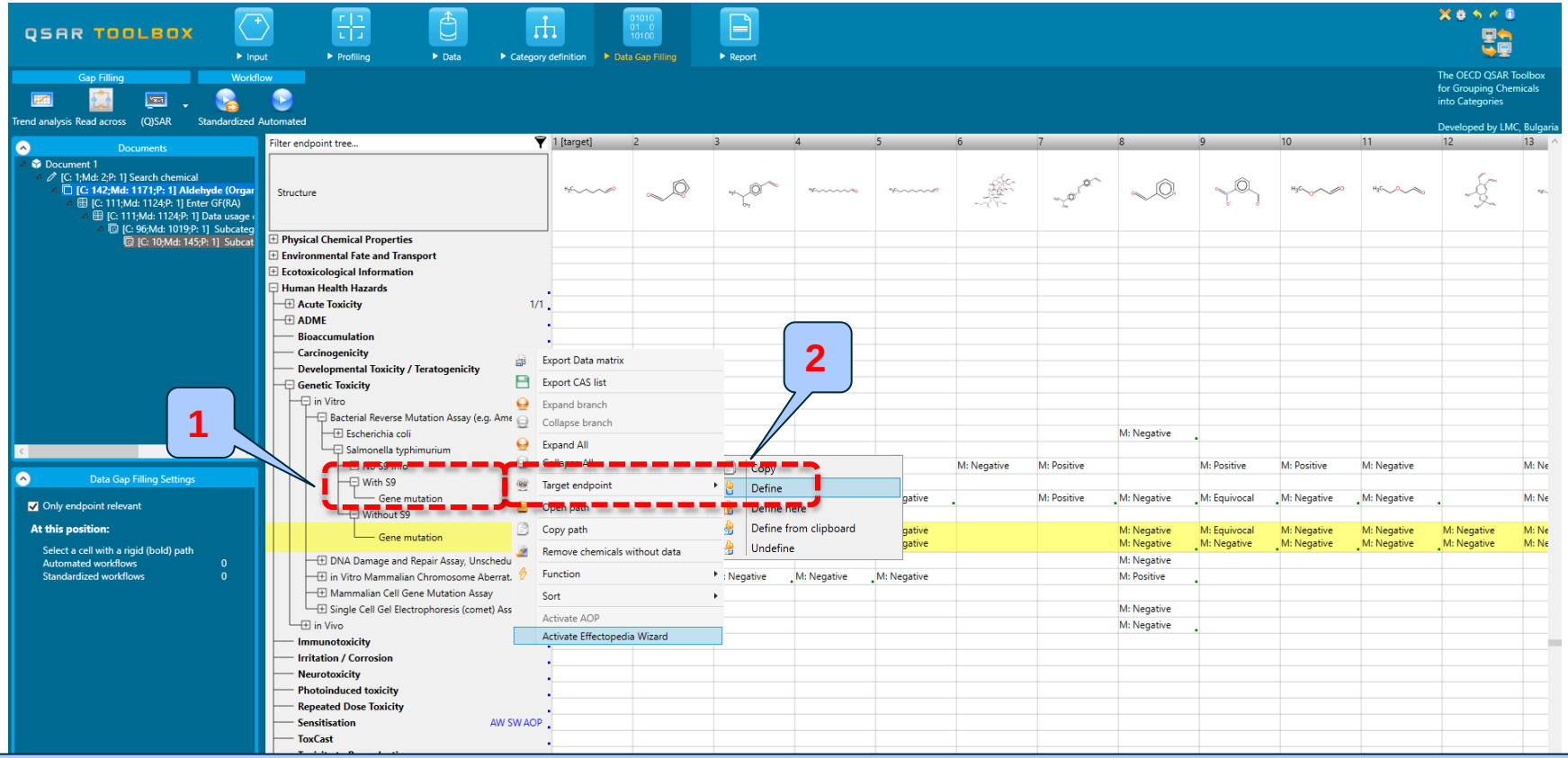
- By accepting the prediction read-across for the current endpoint (Ames mutagenicity without S9 activation) is finished and the data gap is filled ("R" stands for read-across) (1). The target is predicted as *Ames mutagenicity negative* based on the read-across analysis.
- The user can proceed with the workflow for the second endpoint, which in this case will be **"Ames with S9"**
- Because the analogues are selected by the structure-based profiler – *Organic functional groups*, we can use the same analogues for the prediction.
- In this case we need just to change the target endpoint.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - Data
  - Category definition
  - **Data Gap Filling**
    - Ames without S9
    - **Ames with S9**

Data Gap Filling (Ames with S9)

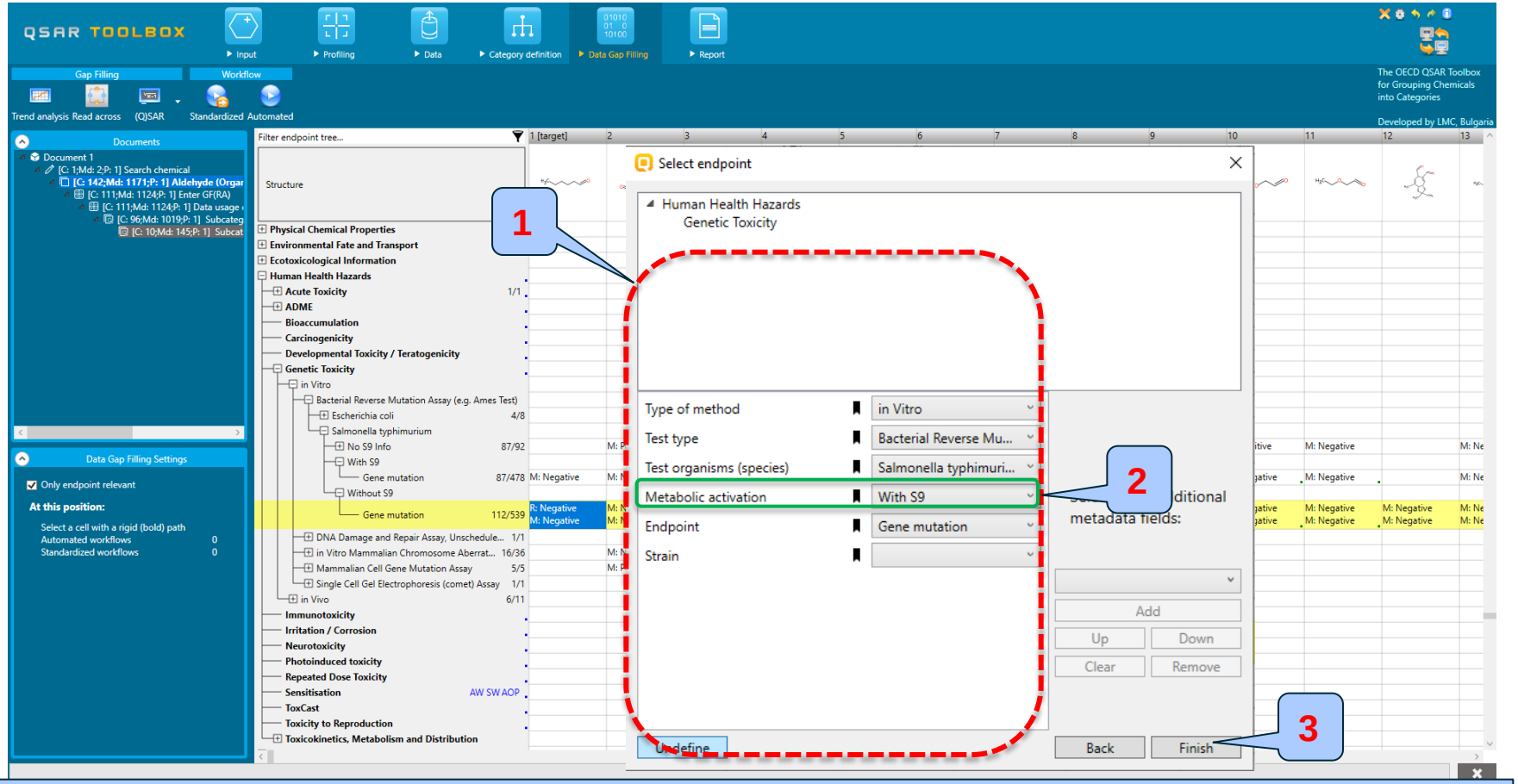
Define the target endpoint



1. Apply right click to the row corresponding to *Human Health Hazards >> Genetic Toxicity >> in Vitro >> Bacterial Reverse Mutation Assay (e.g. Ames Test) >> Salmonella typhimurium >>* ***With S9*** *>> Gene mutation*; 2. Select ***Target endpoint >> Define***. The window with the defined target endpoint will appear (see next slide)

Data Gap Filling (Ames with S9)

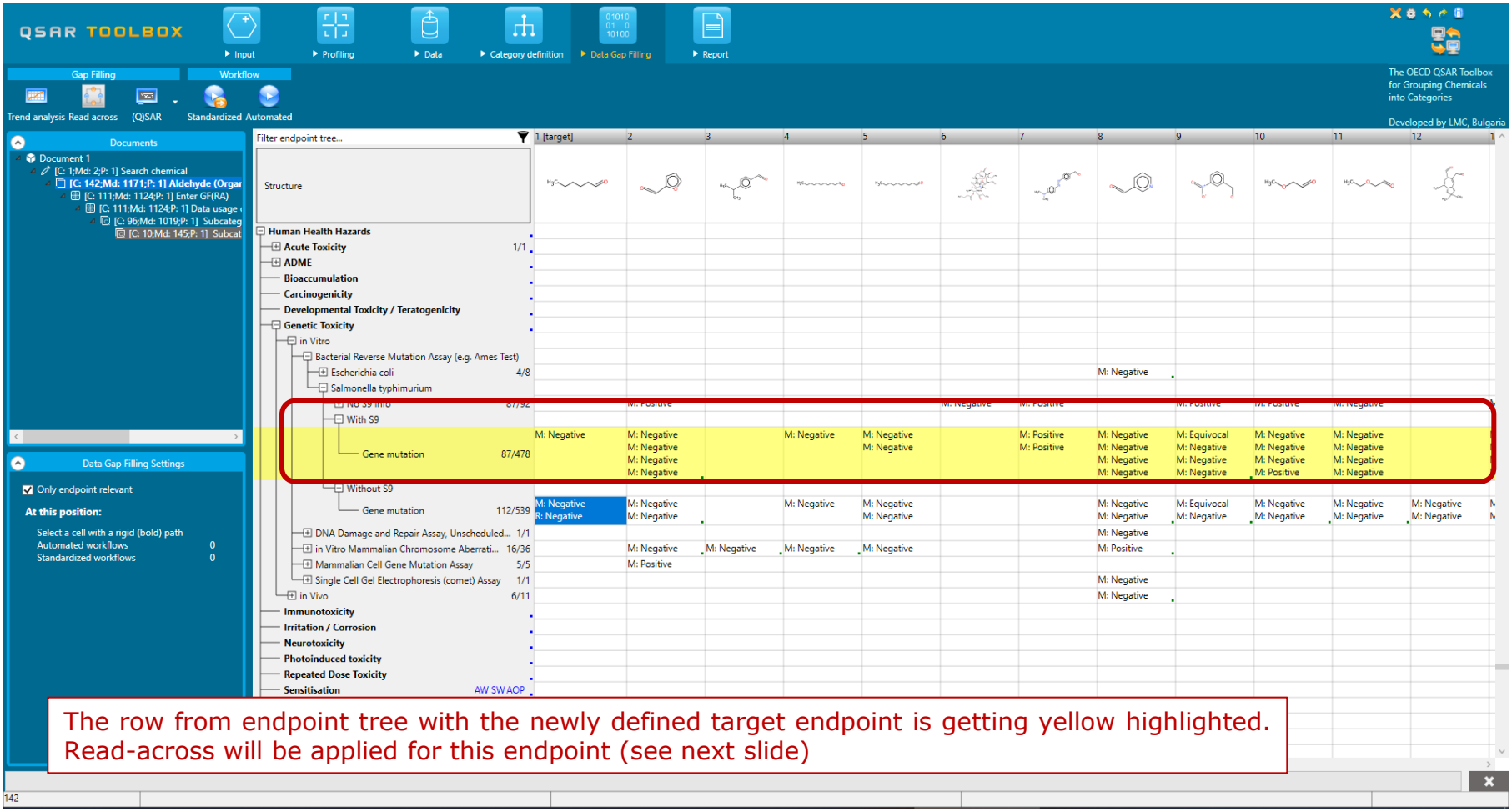
Define the target endpoint



1. A window with the automatically fulfilled fields appears. 2. The only difference with respect to the previous endpoint is field **"Metabolic activation"**. In this case **"With S9"** is preselected. Keep it as it is; 3. Click **Finish** button.

Data Gap Filling (Ames with S9)

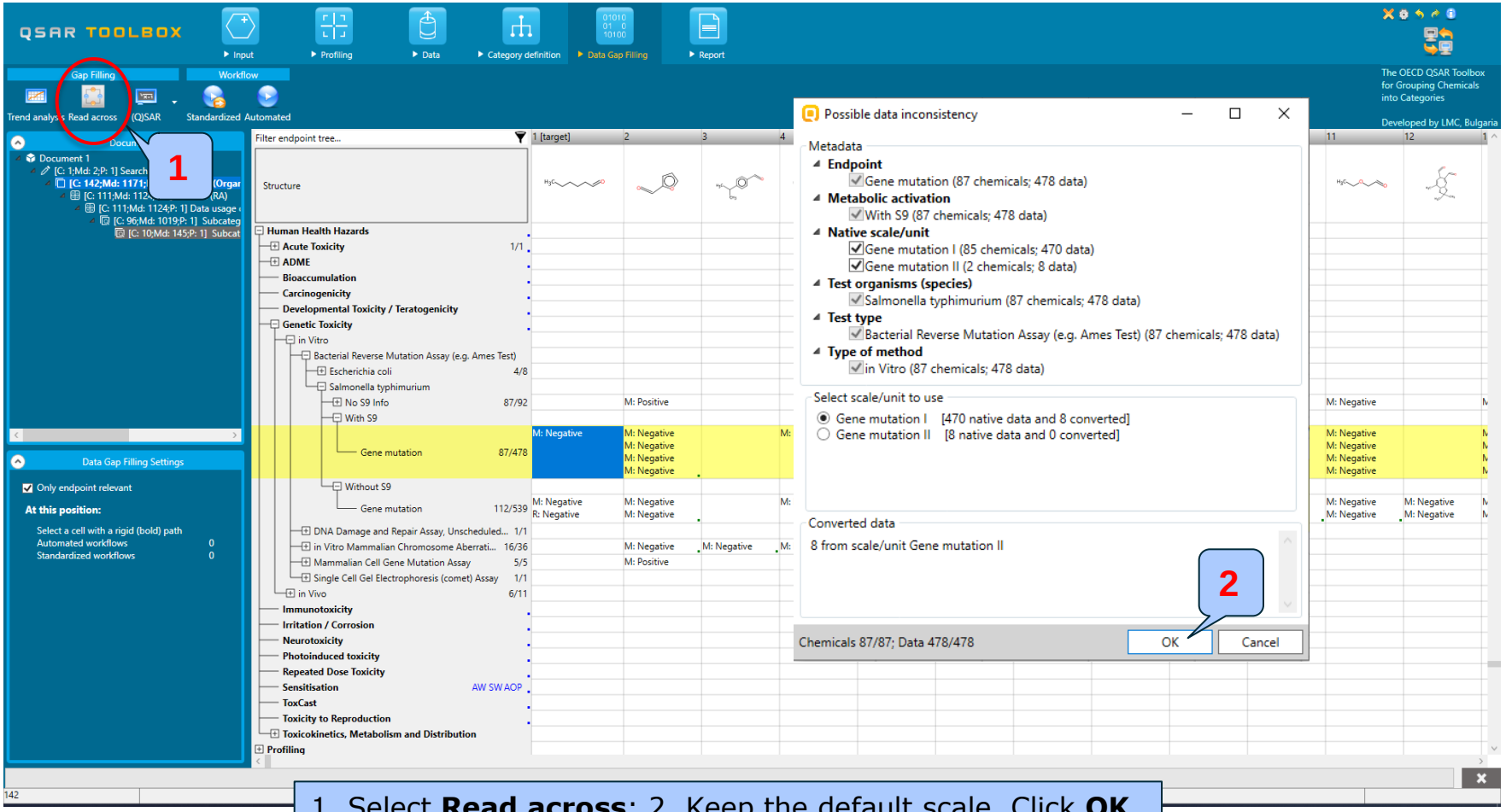
Define the target endpoint



The row from endpoint tree with the newly defined target endpoint is getting yellow highlighted. Read-across will be applied for this endpoint (see next slide)

Data Gap Filling (Ames with S9)

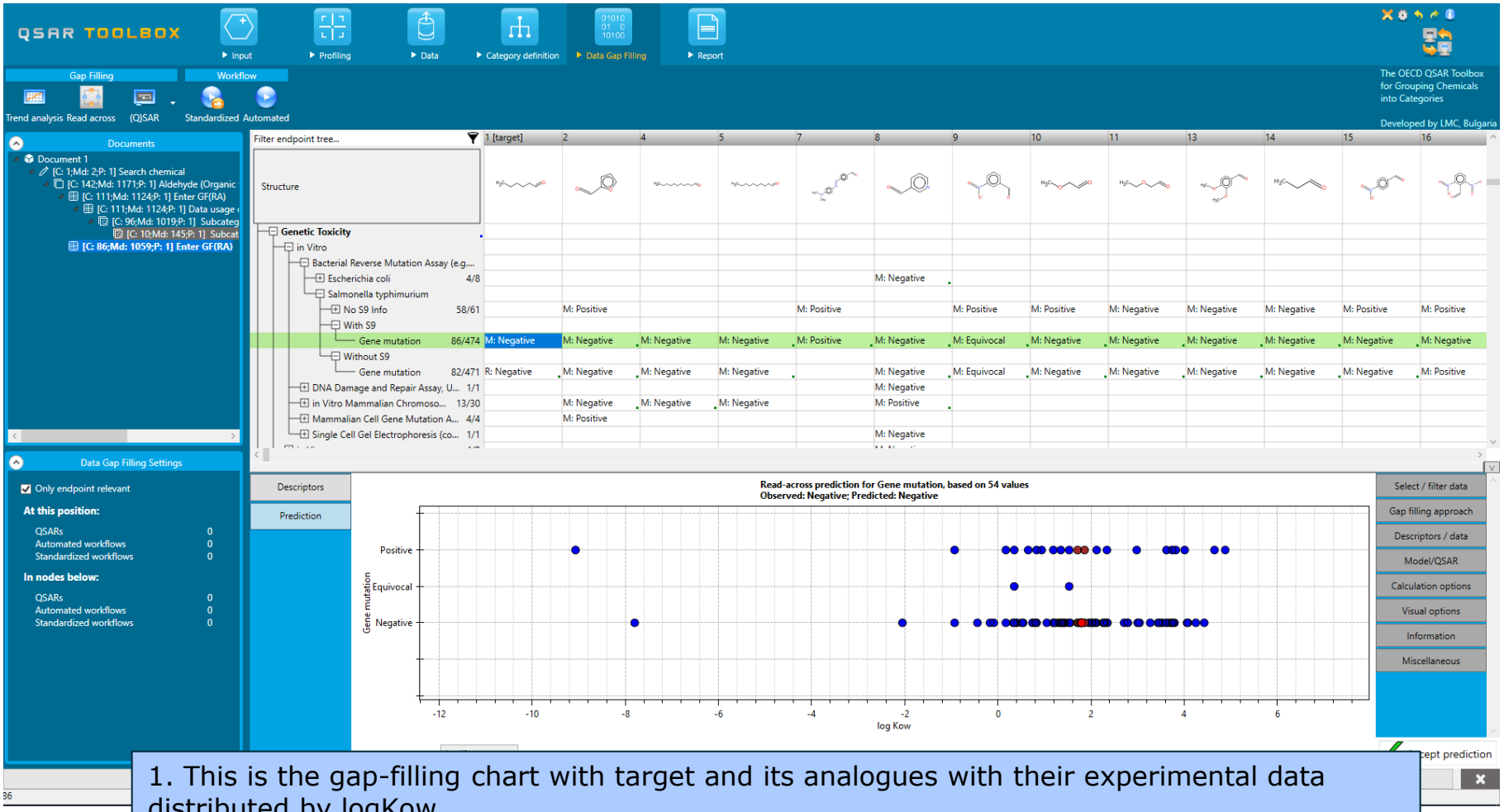
Apply Read-across



1. Select **Read across**; 2. Keep the default scale. Click **OK**.

Data Gap Filling (Ames with S9)

Apply Read-across



1. This is the gap-filling chart with target and its analogues with their experimental data distributed by logKow.

Data Gap Filling (Ames with S9)
Results of Read across

- As while predicting Ames mutagenicity without S9, before accepting the estimated result for the target chemical by read-across, the user should refined the category by subcategorization.
- Subcategorization refers to the process of applying additional profilers to the previously defined category. This aim to identify the chemicals which have differing profiling results and eventually eliminating these chemicals from the category.
- In this example, we are going to use several different profilers to repeatedly subcategorise the data set.

Data Gap Filling (Ames with S9)
Side Bar of Subcategorization

The analogues which are dissimilar to the target chemical with respect to:

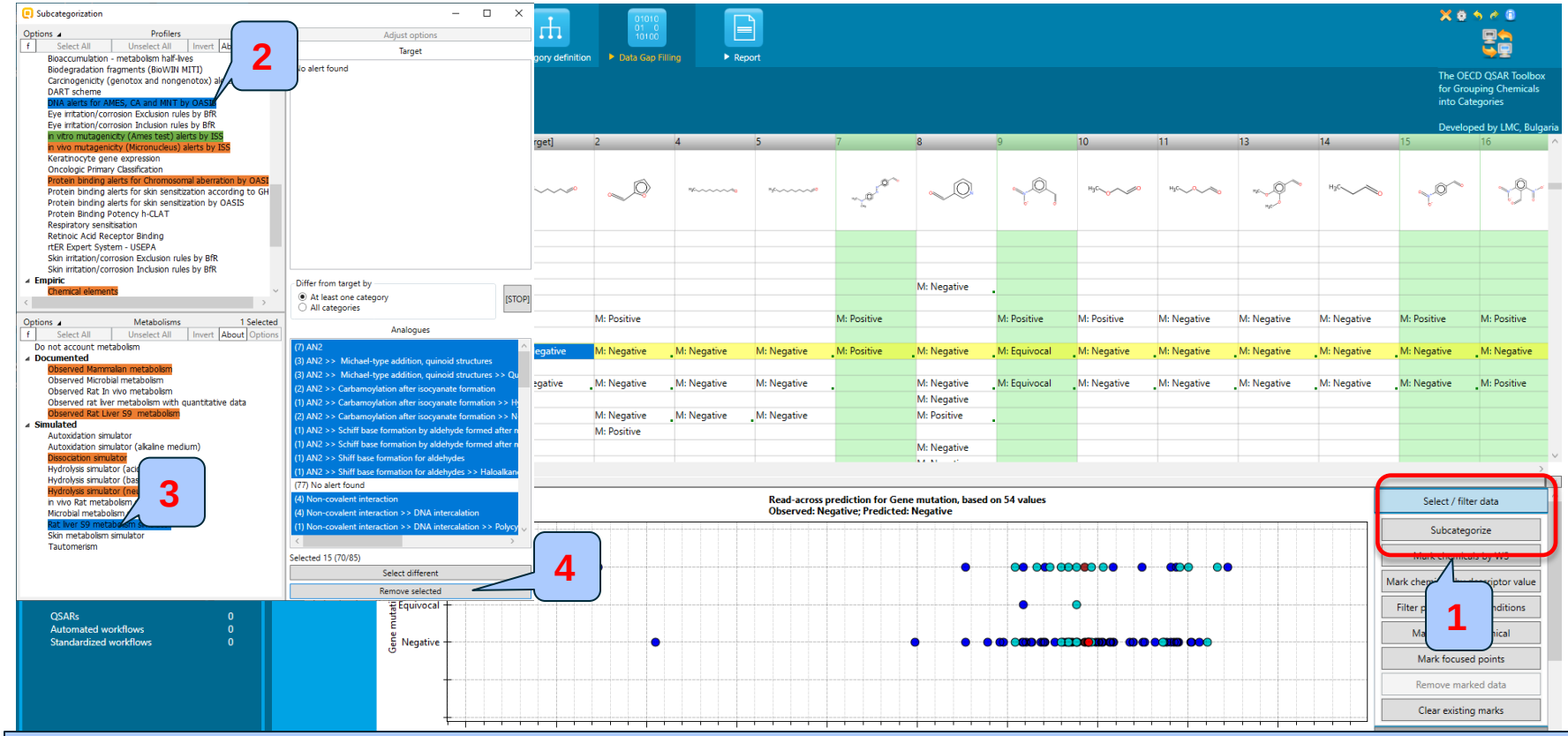
- DNA alerts for AMES, CA and MNT by OASIS (endpoint specific) taking into account in vitro S9 metabolism – The categorization based on this profiler identifies analogues having same DNA binding alerts as the target after metabolic activation

- Organic functional groups (US-EPA) – The categorization based on this profiler identifies analogues having the same organic functional groups

can be removed from the initial list of analogues previously defined by primary group according to OFG (US-EPA) *(see slides 49-52).*

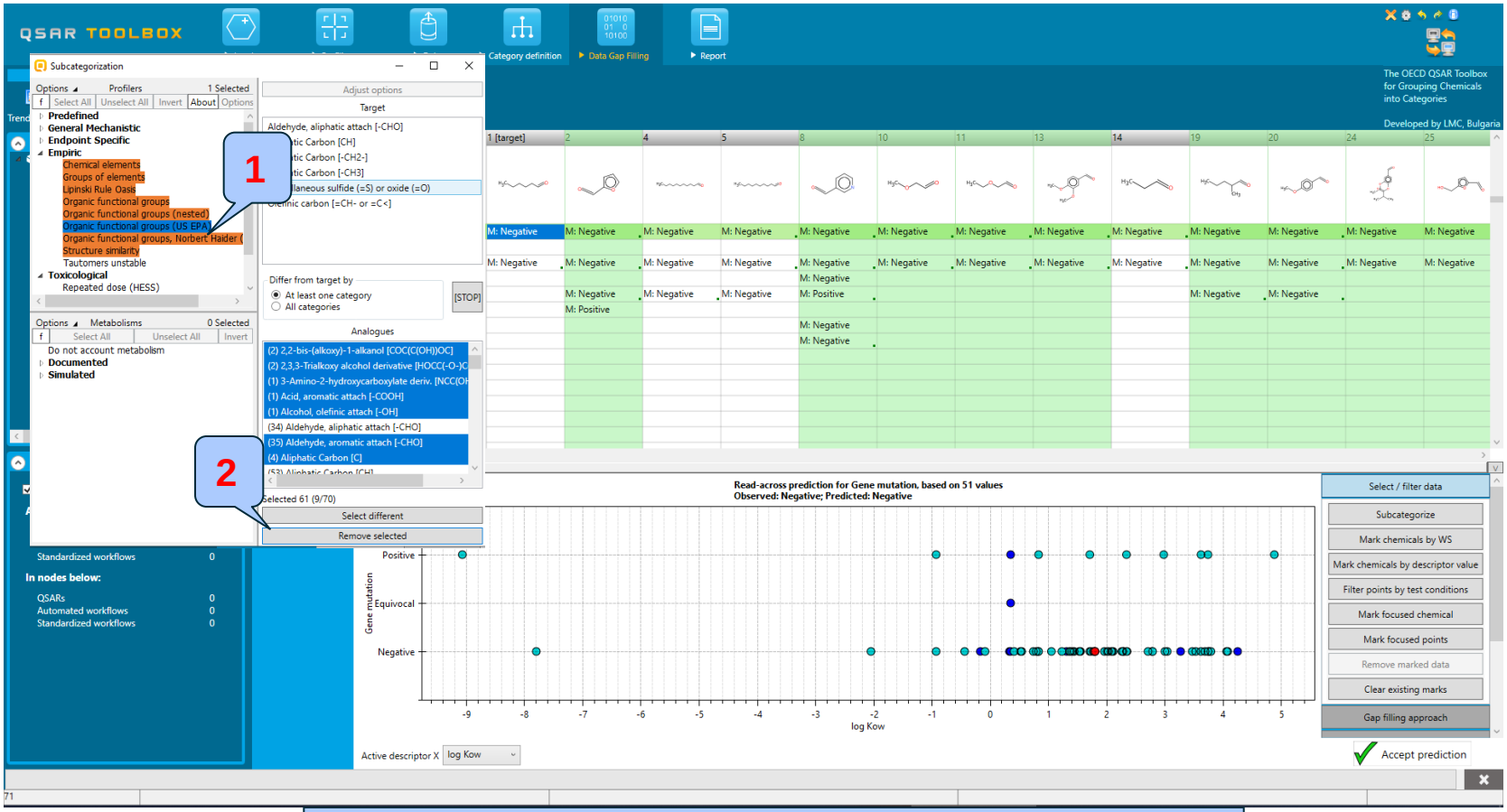
Data Gap Filling (Ames with S9)

Subcategorization by DNA binding alerts taking into account Rat liver S9 metabolism



Before proceeding with subcategorization, worst-case scenario is applied. **Repeat steps from slide 61**.
Once ready with data usage please 1. Go to **Select/Filter data** >>**Subcategorize** 2. Select **DNA alerts from AMES, CA and MNT by OASIS** profiler; 3. Combine it with the **Rat liver S9 metabolism** simulator; 4. Click **Remove selected**

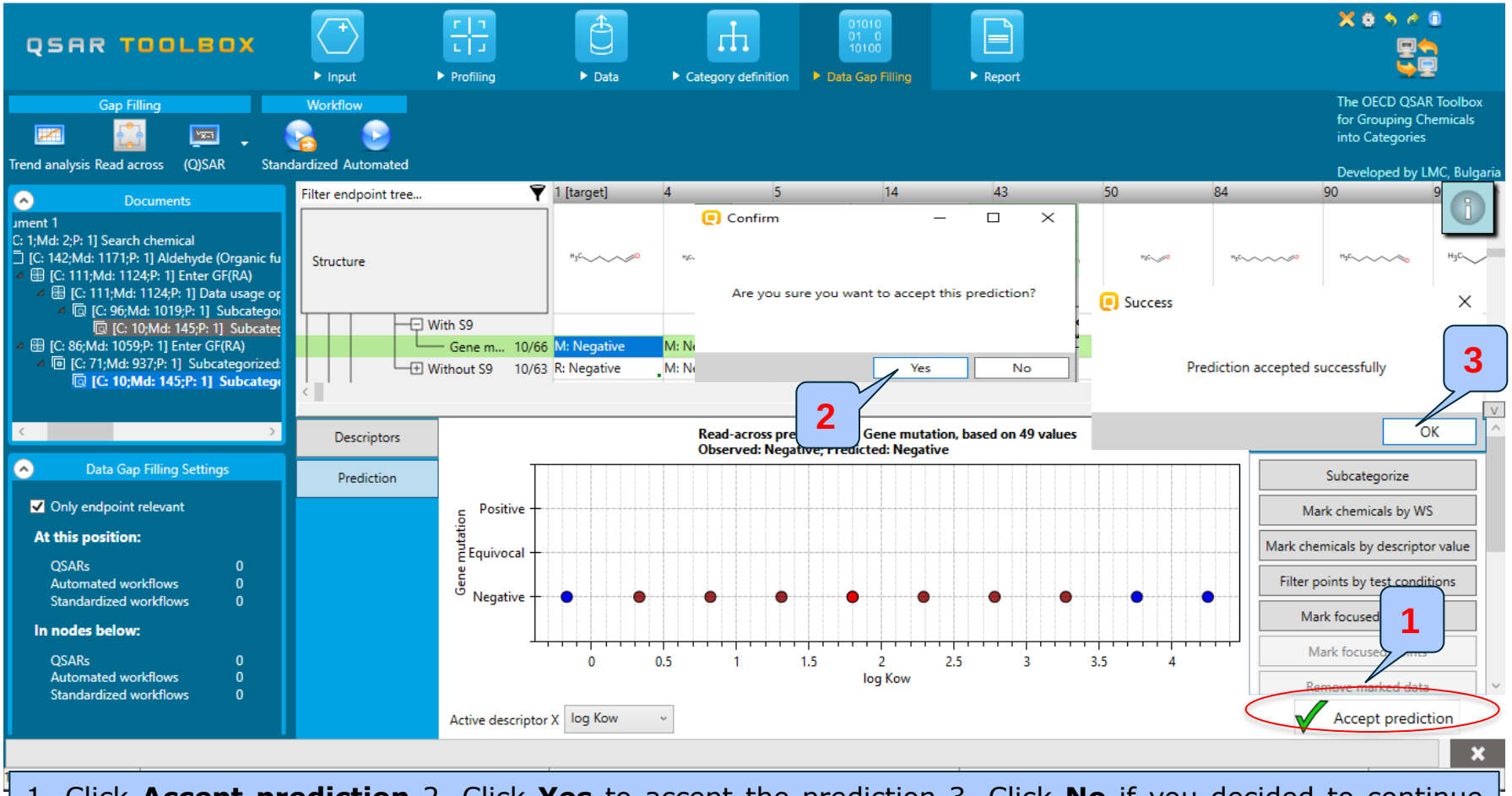
Data Gap Filling (Ames with S9)
Subcategorization by OFG (US-EPA)



1. Select **OFG (US-EPA)**; 2. Click **Remove selected**

Data Gap Filling (Ames with S9)

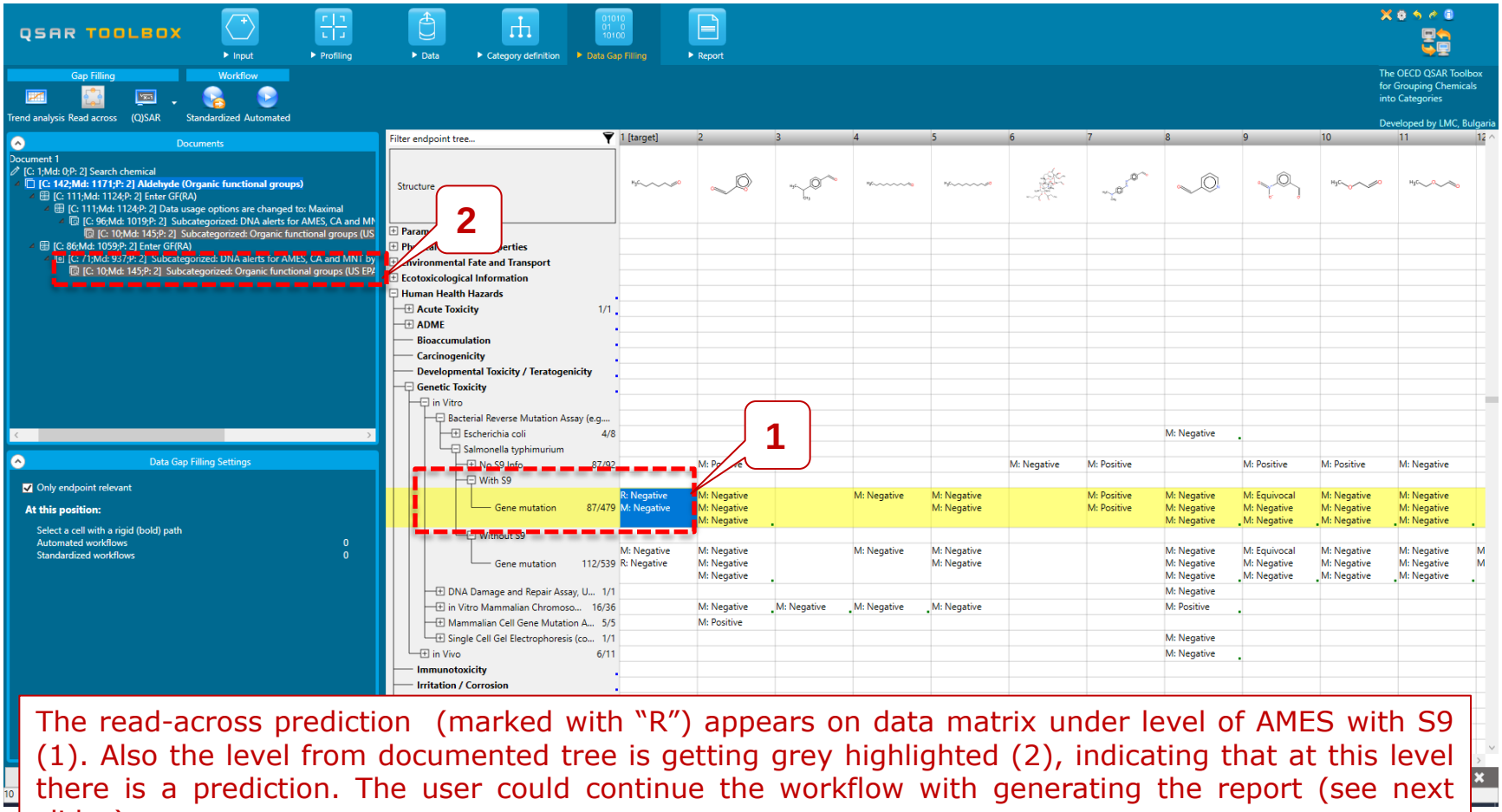
Result of read-across



1. Click **Accept prediction** 2. Click **Yes** to accept the prediction 3. Click **No** if you decided to continue with Subcategorization; 3. Click **OK** on the appeared message

Data Gap Filling (Ames with S9)

Result of read-across



The read-across prediction (marked with "R") appears on data matrix under level of AMES with S9 (1). Also the level from documented tree is getting grey highlighted (2), indicating that at this level there is a prediction. The user could continue the workflow with generating the report (see next slides)

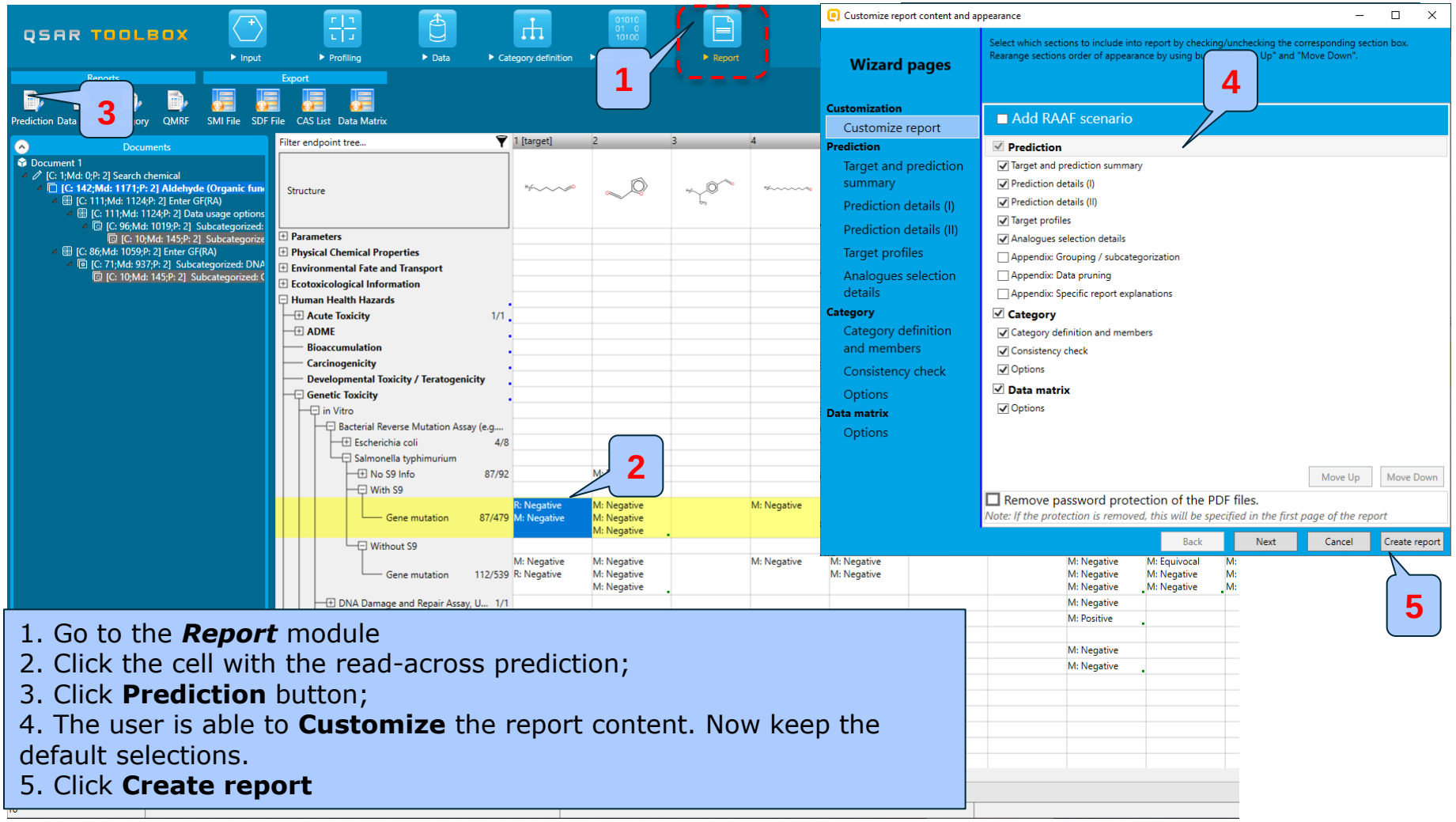
Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - Data
  - Category definition
  - Data Gap Filling
  - **Report**

Report
Overview

- The Report module allows you to generate a report on the predictions performed within the Toolbox.
- This module contains a predefined report template with automatically populated sections as well as manually editable sections, where the users could add some additional/custom information.
- The Prediction report generates three files: 1) *Prediction report*; 2) *Category report* and 3) *Data matrix .*
- The generated reports can then be saved or just opened.

Report
Generate Report

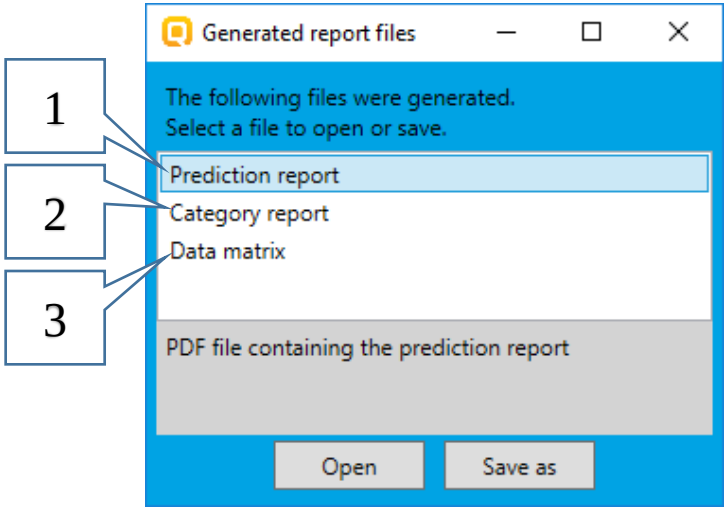


1. Go to the ***Report*** module
2. Click the cell with the read-across prediction;
3. Click **Prediction** button;
4. The user is able to **Customize** the report content. Now keep the default selections.
5. Click **Create report**

Report
Generation

After clicking *Create report* button, *Generated report files* window appears. It contains three type of files:

1) **Prediction report** - a PDF file containing the prediction information related to the target.
2) **Category report** - a PDF file containing information for the consistency of the final category (target plus used analogues)
3) **Data matrix** - a MS Excel file containing chemicals used for prediction along with their data for selected parameters, profiles and endpoint tree positions.



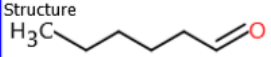
Report
Generated report files

Prediction report

QSAR Toolbox prediction for single chemical

Date: 3 Apr 2020
Author(s):
Contact details:

Target information

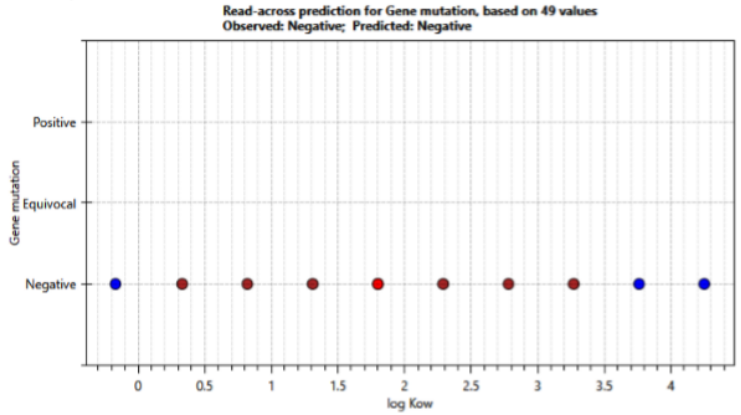
Structural information	Numerical identifiers	Chemical names
SMILES: CCCCCC=O Structure 	CAS#: 66-25-1 Other: EC Number:2006245	hexaldehyde Hexanal Hexylaldehyde

Prediction summary

Predicted endpoint: Gene mutation; No effect specified; Salmonella typhimurium; No duration specified; No guideline specified
Predicted value: Negative
Unit/scale: Gene mutation I
Data gap filling method: Read-across analysis
Summary: *manually editable field*
Not provided by the user

Prediction details (I)

Predicted value: Negative
Predicted endpoint (OECD Principle 1 - Defined endpoint): Human Health Hazards -> Genetic Toxicity -> in Vitro -> Bacterial Reverse Mutation Assay (e.g. Ames Test) -> Gene mutation -> Salmonella typhimurium -> With S9
Prediction plot:

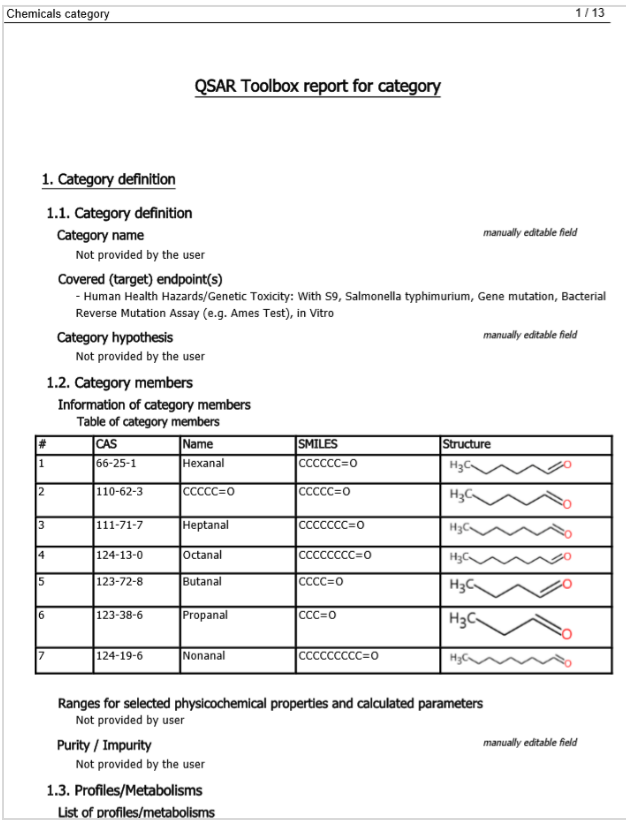


Calculation approach (OECD principle 2 - Unambiguous algorithm): takes the highest mode value from the 6 nearest neighbours
Active descriptor: log Kow (calculated)
Data usage: All values*
*When multiple values are available for the same chemical, all of them are taken individually in prediction calculations

Report

Generated report files

Category report



QSAR Toolbox report for category

1. Category definition

1.1. Category definition

Category name *manually editable field*

Not provided by the user

Covered (target) endpoint(s)

- Human Health Hazards/Genetic Toxicity: With S9, Salmonella typhimurium, Gene mutation, Bacterial Reverse Mutation Assay (e.g. Ames Test), in Vitro

Category hypothesis *manually editable field*

Not provided by the user

1.2. Category members

Information of category members

Table of category members

#	CAS	Name	SMILES	Structure
1	66-25-1	Hexanal	CCCCCC=O	
2	110-62-3	CCCCC=O	CCCCC=O	
3	111-71-7	Heptanal	CCCCCCC=O	
4	124-13-0	Octanal	CCCCCCCC=O	
5	123-72-8	Butanal	CCCC=O	
6	123-38-6	Propanal	CCC=O	
7	124-19-6	Nonanal	CCCCCCCCC=O	

Ranges for selected physicochemical properties and calculated parameters

Not provided by user

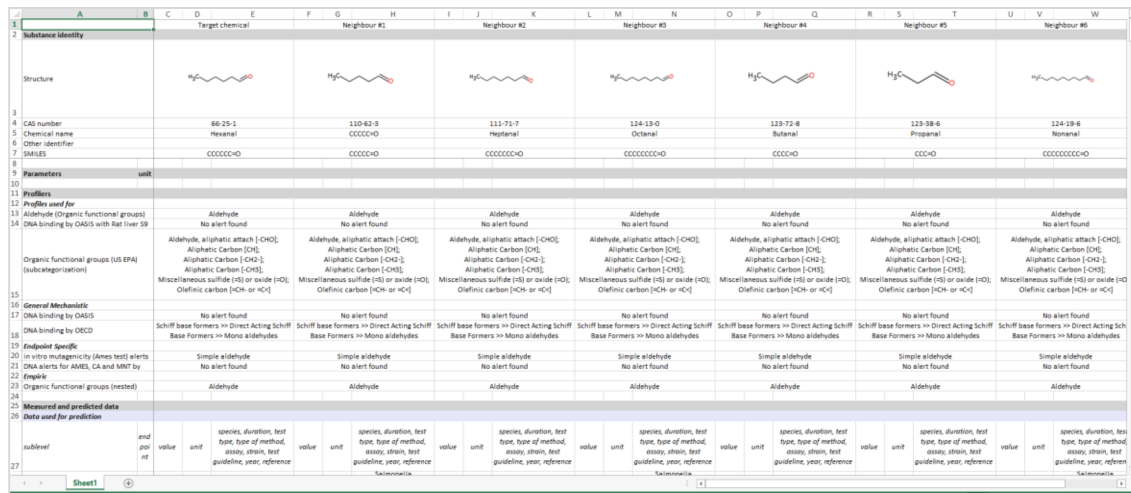
Purity / Impurity *manually editable field*

Not provided by the user

1.3. Profiles/Metabolisms

List of profiles/metabolisms

Data matrix report



		Target chemical	Neighbour #1	Neighbour #2	Neighbour #3	Neighbour #4	Neighbour #5	Neighbour #6
Substance identity								
Structure								
CAS number		66-25-1	110-62-3	111-71-7	124-13-0	123-72-8	123-38-6	124-19-6
Chemical name		Hexanal	CCCCC=O	Heptanal	Octanal	Butanal	Propanal	Nonanal
Other identifier								
SMILES		CCCCCC=O	CCCCC=O	CCCCCCC=O	CCCCCCCC=O	CCCC=O	CCC=O	CCCCCCCCC=O
Parameters	unit							
Profilers								
Profiles used for								
Aldehyde (Organic functional groups)		Aldehyde	Aldehyde	Aldehyde	Aldehyde	Aldehyde	Aldehyde	Aldehyde
DNA binding by OASIS with Rat liver S9		No alert found	No alert found	No alert found	No alert found	No alert found	No alert found	No alert found
Organic functional groups (US EPA) (subcategorization)		Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]	Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]	Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]	Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]	Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]	Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]	Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (=S) or oxide (=O; Olefinic carbon [=CH- or =C<]
General Mechanistic								
DNA binding by OASIS		No alert found	No alert found	No alert found	No alert found	No alert found	No alert found	No alert found
DNA binding by OECD		Schiff base formers >> Direct Acting Schiff Base Formers >> Mono aldehydes	Schiff base formers >> Direct Acting Schiff Base Formers >> Mono aldehydes	Schiff base formers >> Direct Acting Schiff Base Formers >> Mono aldehydes	Schiff base formers >> Direct Acting Schiff Base Formers >> Mono aldehydes	Schiff base formers >> Direct Acting Schiff Base Formers >> Mono aldehydes	Schiff base formers >> Direct Acting Schiff Base Formers >> Mono aldehydes	Schiff base formers >> Direct Acting Sch Base Formers >> Mono aldehydes
Endpoint Specific								
in vitro mutagenicity (Ames test) alerts		Simple aldehyde	Simple aldehyde	Simple aldehyde	Simple aldehyde	Simple aldehyde	Simple aldehyde	Simple aldehyde
DNA alerts for AMES, CA and MNT by		No alert found	No alert found	No alert found	No alert found	No alert found	No alert found	No alert found
Empiric								
Organic functional groups (nested)		Aldehyde	Aldehyde	Aldehyde	Aldehyde	Aldehyde	Aldehyde	Aldehyde
Measured and predicted data								
Data used for prediction								
sublevel	*end point*	*value / unit / species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value / unit / species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value / unit / species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value / unit / species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value / unit / species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value / unit / species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value / unit / species, duration, test type, type of method, assay, strain, test guideline, year, reference*

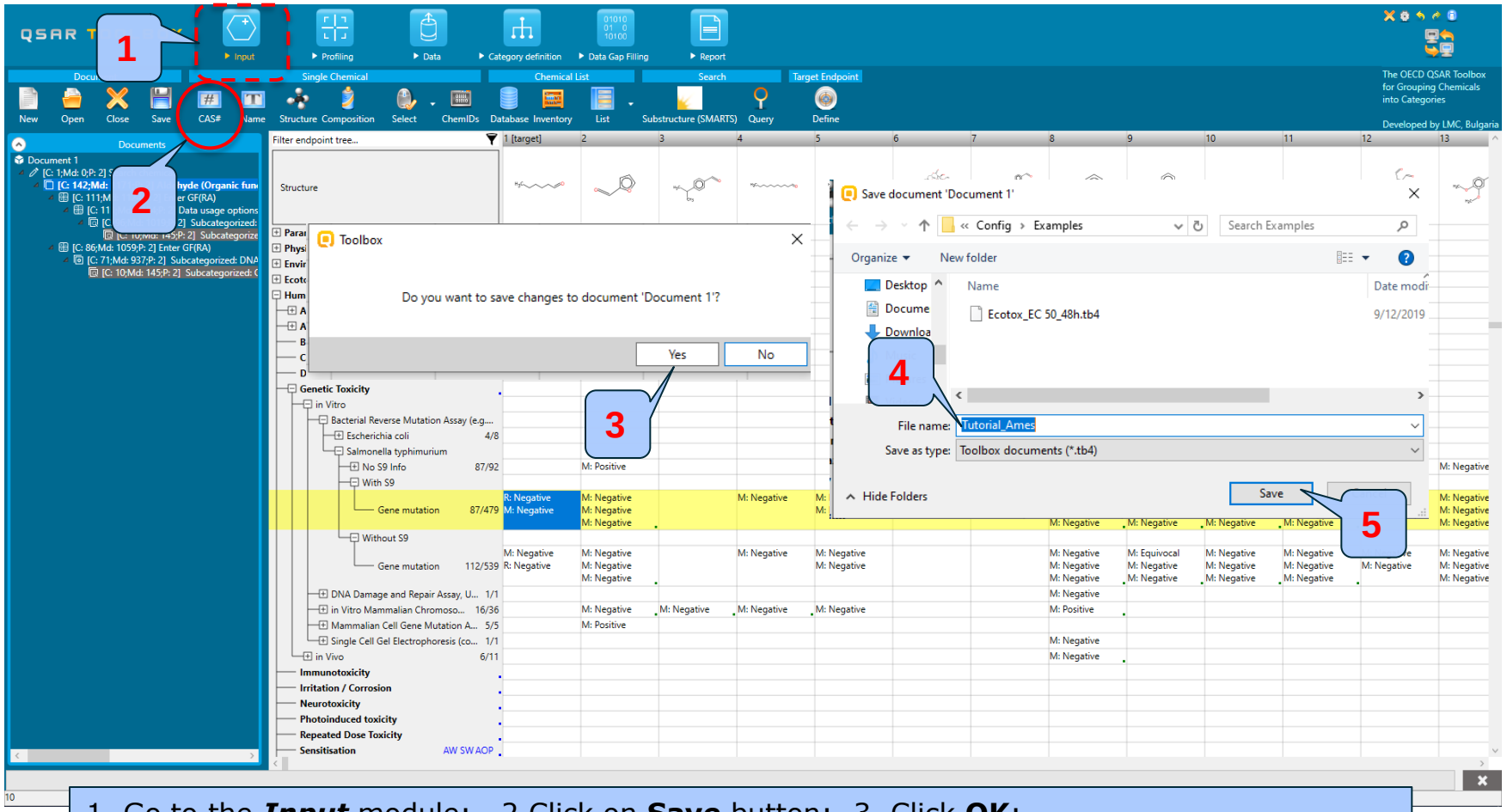
Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read-across
- The exercise
- Workflow of the exercise
- **Save the prediction**

Saving the prediction result

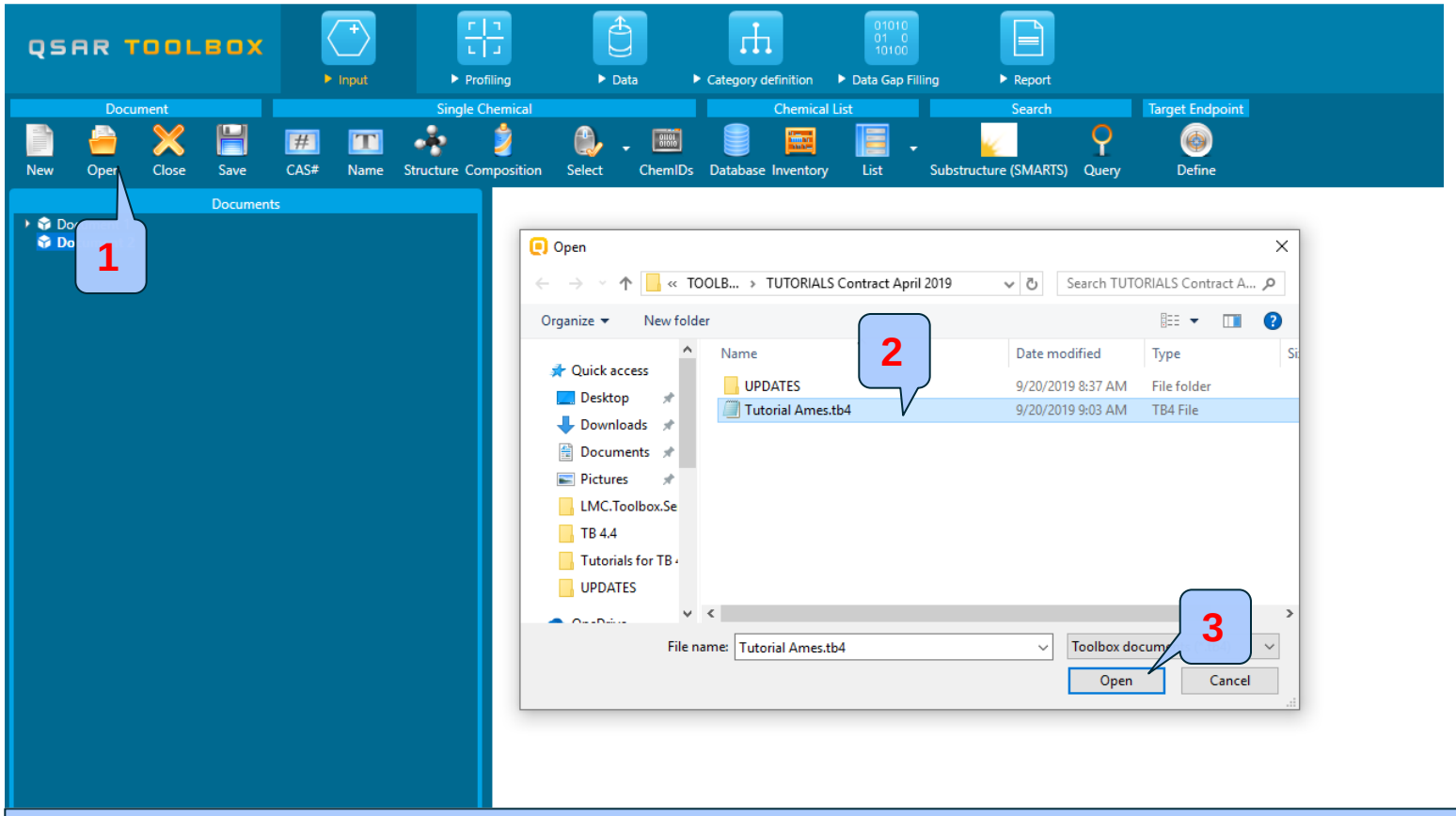
- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc., on the same computer.
- The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox 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 the ***Input*** module; 2.Click on **Save** button; 3. Click **OK**;
4. Define name of the file; 5. Click **Save** button (Message "File saved successfully" appears)

Open saved file



1. Click **Open**; 2. Find and select file; 3. Click **Open**. The saved document will be opened. Finally, a message confirming the successful loading the file will appear.

Congratulations

- By now you should feel comfortable with the six basic modules of the Toolbox and how they form the work flow of the Toolbox.
- In this tutorial you have now been introduced to several additional functions in the Toolbox, especially using different profilers in subcategorizing the category of the target chemical.
- Remember - proficiency only comes with practice.