QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Predicting the carcinogenicity potential of

1-(2-hydroxyethyl)-1-nitrosourea (CAS 13743-07-2)

by filtering with Ames experimental data

Outlook

- **Background**
- Objectives
- The exercise
- Workflow

Background

- This is a step-by-step presentation designed to take the user through the workflow for filling data gap for carcinogenicity effects by read-across based on an analogue approach. The aim of this data gap filling is to illustrate how the initial set of identified analogues could be filtered based on measured AMES data.

Outlook

- Background
- **Objectives**
- The exercise
- Workflow

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Identify analogues of the target chemical.
- Retrieve experimental results available for those analogues.
- Fill data gaps for carcinogenicity by read across.
- Filtering analogues by measured AMES data.
- Save the prediction.

Outlook

- Background
- Objectives
- **The exercise**
- Workflow

The exercise

- The aim of exercise is to predict the carcinogenicity potential of **1-(2-hydroxyethyl)-1-nitrosourea (CAS 13743-07-2)**, taken as a "target" chemical, by filtering with Ames mutagenicity data.
- The prediction will be accomplished by collecting set analogues considered to be in the same category as the target molecule.
- The category will be defined based on structural similarity of all the chemicals in the category with respect to the Organic functional group profiler.
- The experimental data for the target chemical and identified analogues will be collected from databases including carcinogenicity data.
- Read across will be applied based on analogue approach.
- Analogues will be filtered based on two approaches:
  - Data filter – filtering by AMES experimental data
  - Mechanism based - DNA and Protein binding profilers
- Finally the obtained prediction result will be saved.

Cancerogenicity
Background

- Cancers are a large family of diseases that involve abnormal cell growth with the potential to invade or spread to other parts of the body [1,2].
- Classically, cancer has been viewed as a set of diseases that are driven by progressive genetic abnormalities that include mutations in tumor-suppressor genes and oncogenes, and chromosomal abnormalities. However, it has become apparent that cancer is also driven by epigenetic alterations [3].
- Epigenetic alterations refer to functionally relevant modifications to the genome that do not involve a change in the nucleotide sequence. Examples of such modifications are changes in DNA methylation (hypermethylation and hypomethylation) and histone modification [4] and changes in chromosomal architecture (caused by inappropriate expression of proteins such as HMGA2 or HMGA1).
- Carcinogenicity is a complex, multistep process, conditioned by DNA and Protein alterations[5].
- Basically chemical carcinogens are classified as: into two types carcinogens:
  - Genotoxic, and
  - Nongenotoxic carcinogens
- The analysis continues with assessment of the carcinogenicity of the target chemical taking into account DNA and protein interactions.

1. Cancer Fact sheet N° 297". *World Health Organization*. February 2014. Retrieved 10 June 2014.
2. Cancer - Signs and symptoms". *NHS Choices*. Retrieved 10 June 2014.
3. Baylin SB, Ohm JE (February 2006). "Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?". *Nature Reviews Cancer* **6** (2): 107–16. doi:10.1038/nrc1799. PMID 16491070
4. Kanwal, R; Gupta, S (2012). "Epigenetic modifications in cancer". *Clinical Genetics* **81** (4): 303–11. doi:10.1111/j.1399-0004.2011.01809.x. PMC 3590802.
5. Baker, S.G., Cappuccio, A., Potter, J.D. Research on early-stage carcinogenesis: Are we approaching paradigm instability. *Journal of Clinical Oncology.* Volume 28, Issue 20, 10 July 2010, Pages 3215-3218

Outlook

- Background
- Objectives
- The exercise
- **Workflow**

Workflow

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

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - **Input**

Chemical 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.

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
- Query Tool

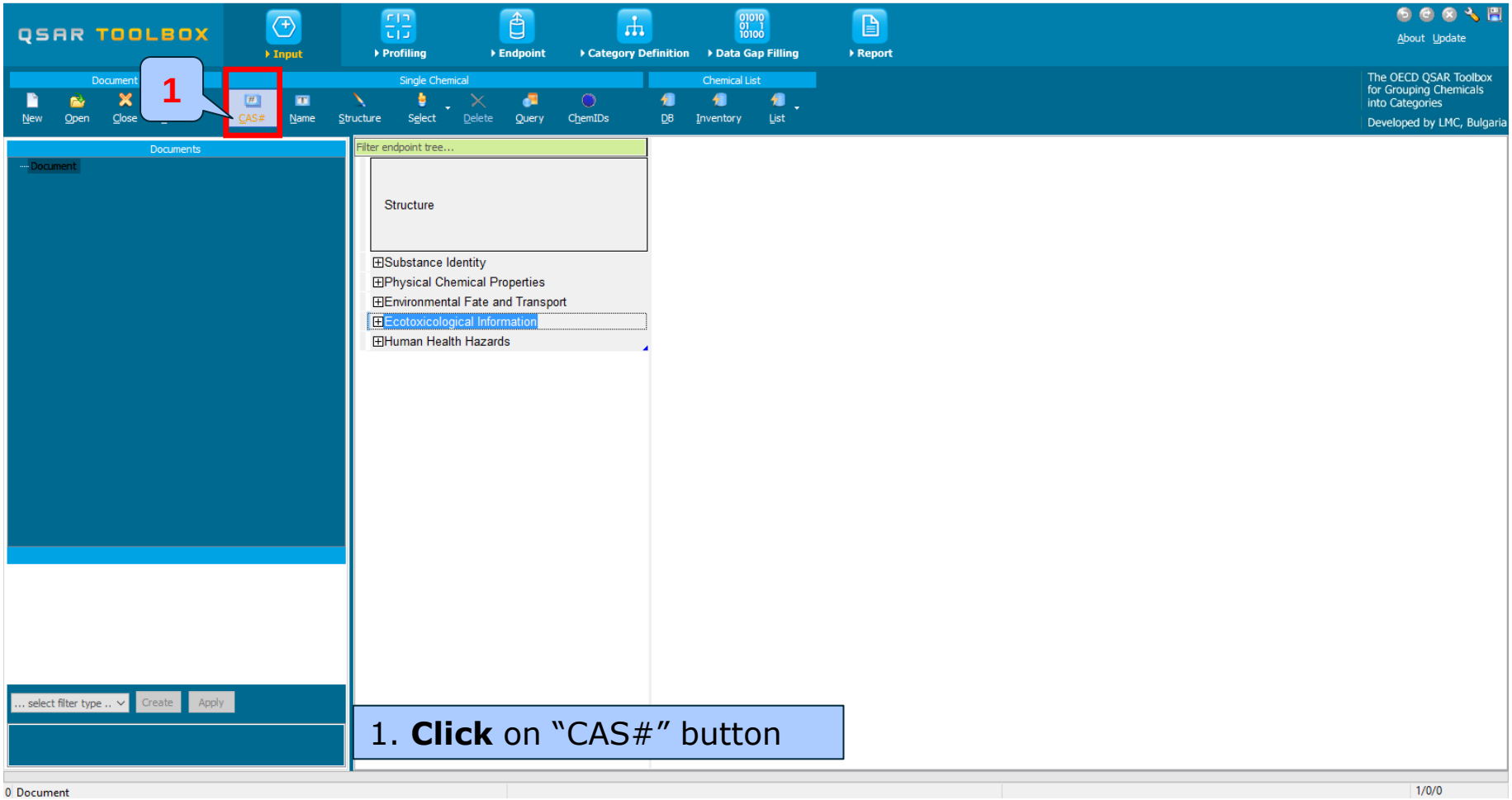
B. Group of chemicals
- User List/Inventory
- Specialized Databases

Getting Started

- Open the Toolbox.
- The six modules in the workflow are seen listed next to "QSAR TOOLBOX".
- Click on "Input" (see next screen shot).

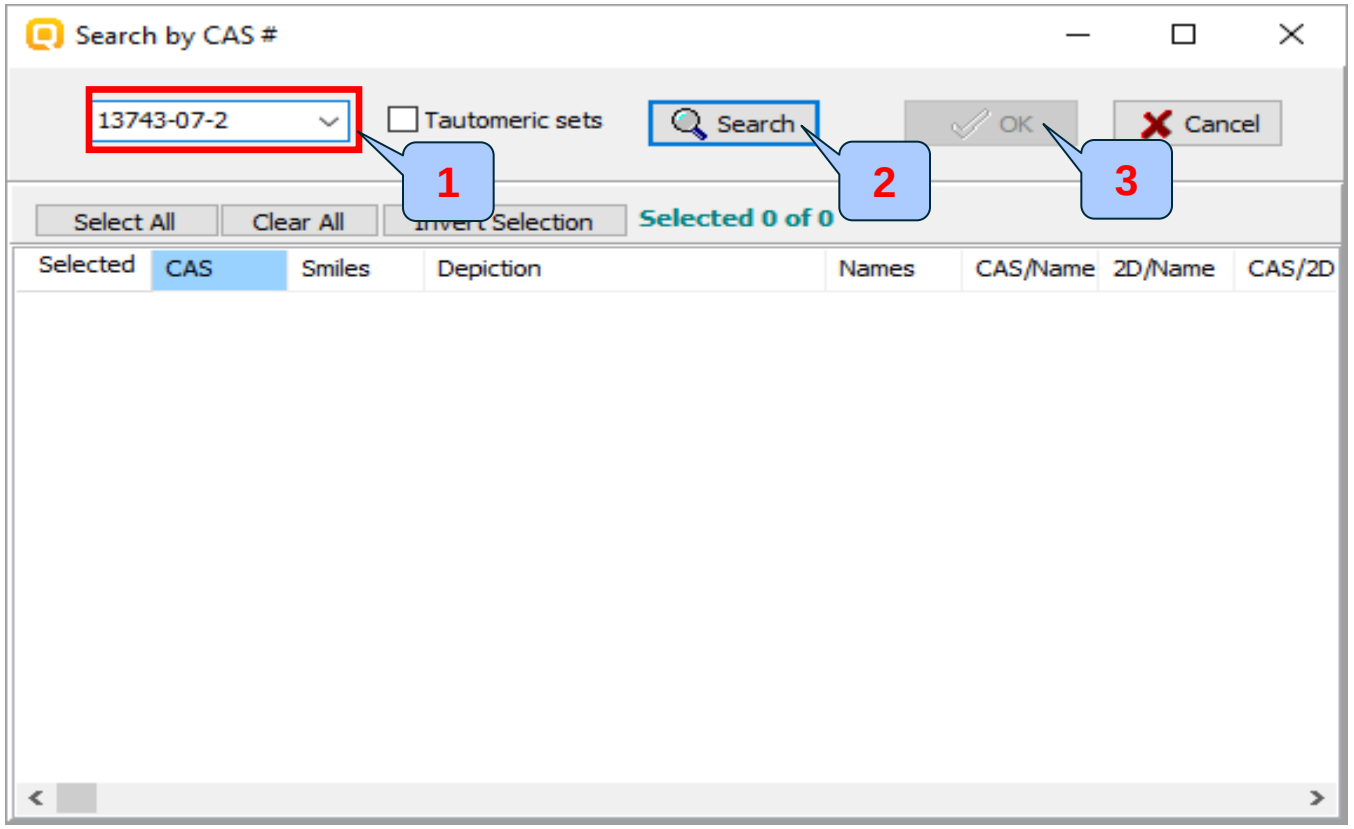
Chemical Input Screen

Input target chemical by CAS#



Chemical Input Screen

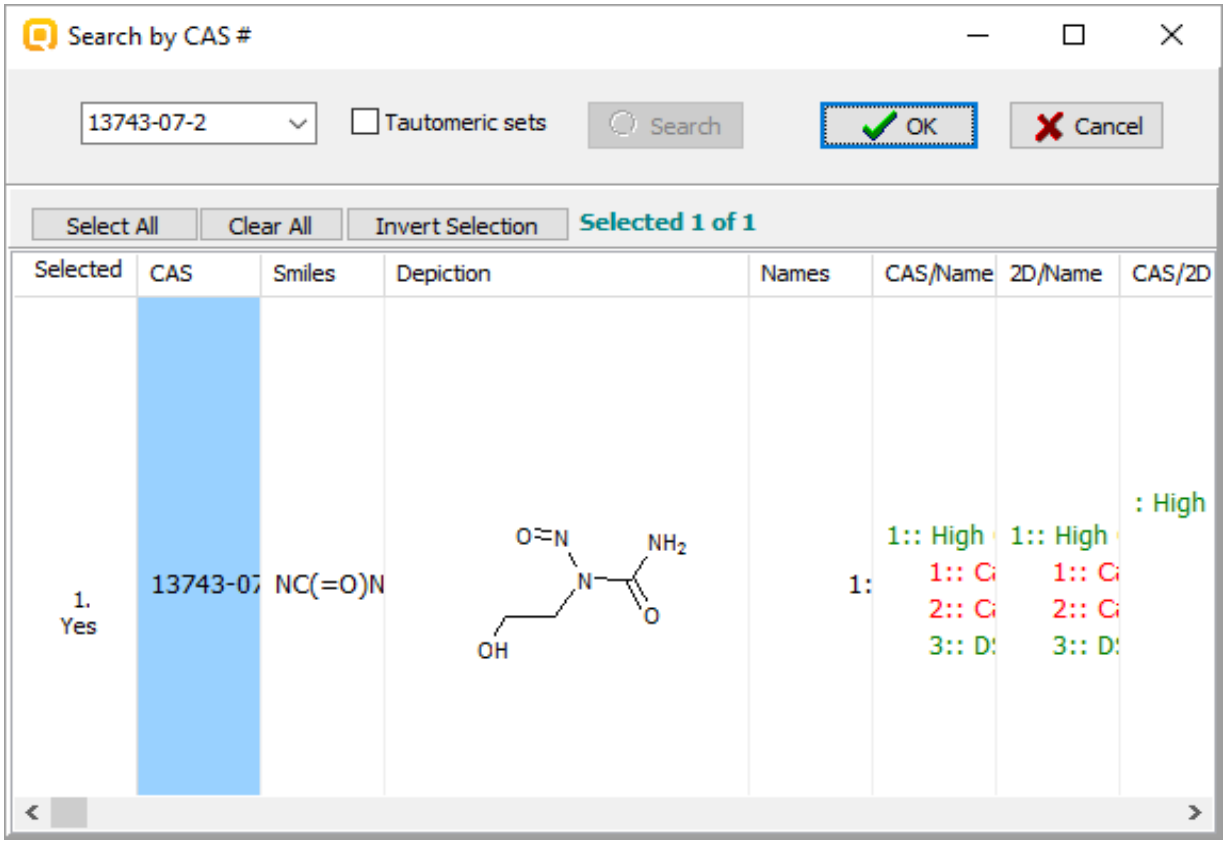
Enter CAS# 13743-07-2



1. **Enter** the CAS# in the blank field; 2. **Click** "Search" button; 3. **Press** "OK"

Chemical Input
Target chemical identity

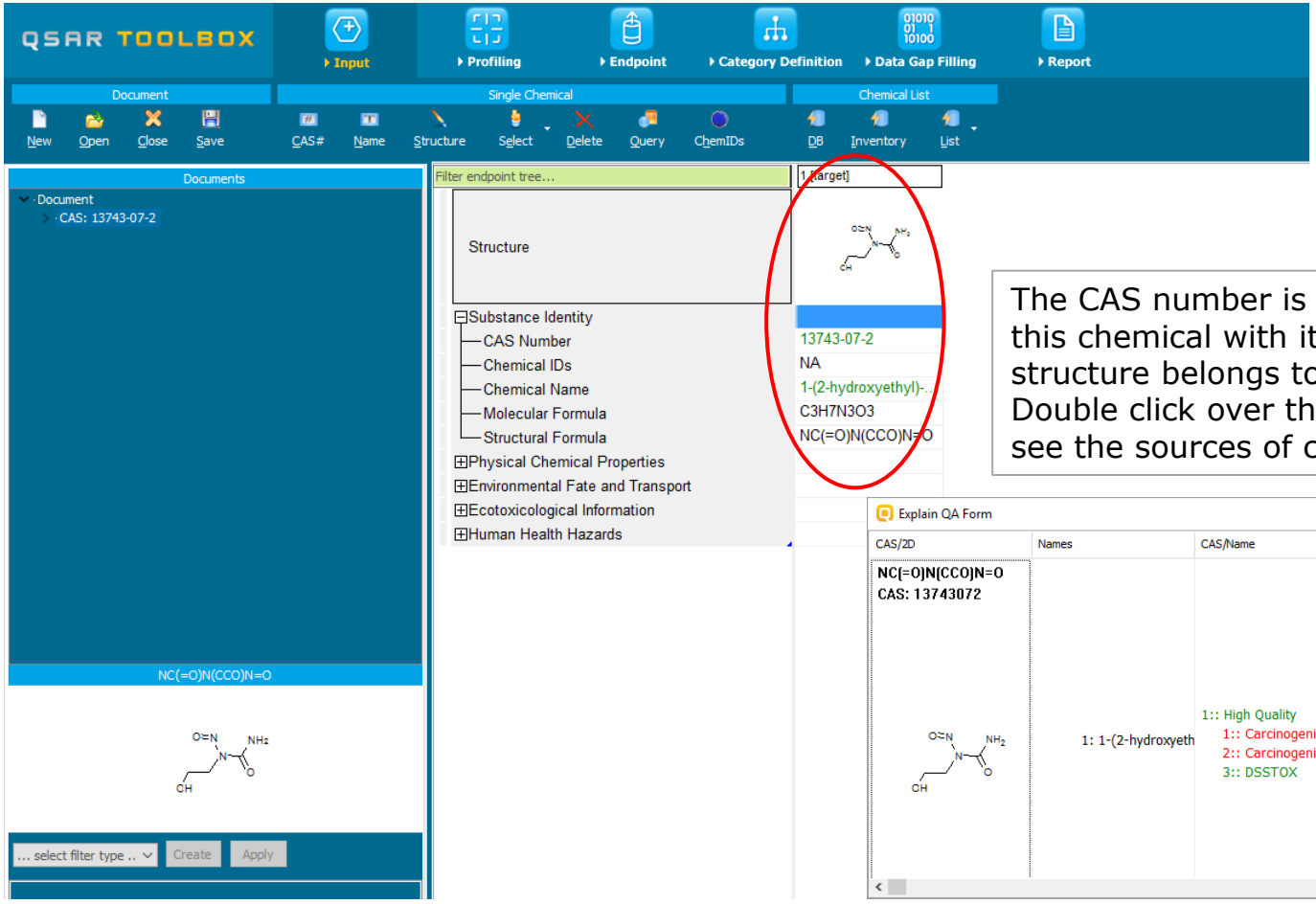
The Toolbox now searches the databases to find out if the CAS# you entered is linked to a molecular structure stored in the Toolbox. It is displayed as a 2-demensional depiction.



Chemical Input
Target chemical identity

- Double click "Substance Identity" displays the chemical identification information.
- The user should note that existing names of the target chemical are presented in different colours. This indicates the reliability of relation CAS-Name-SMILES for the target chemical (see next screen shots).
- The workflow on the first module is now complete, and the user can proceed to the next module.

Chemical Input
Target chemical identity



The CAS number is colored green* because this chemical with its CAS number and 2D structure belongs to high quality inventories. Double click over the cell with CAS number to see the sources of chemical ID

*More details about color legend are provided on next slide

Chemical Input
Chemical identity

The colour code indicates the reliability of the chemical identifier:

- **Green**: There is a high reliability between the identifier and the structure. This colour is applied if the identifier is the same in several quality assured databases.
- **Yellow:** There is only a moderate reliability between the identifier and the structure. The colour is applied if the identifier is the same in several databases for which the quality assurance could not be established.
- **Red**: There is a poor reliability between the identifier and the structure. The colour is applied if the identifier is allocated to different structures in different databases.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - **Profiling**

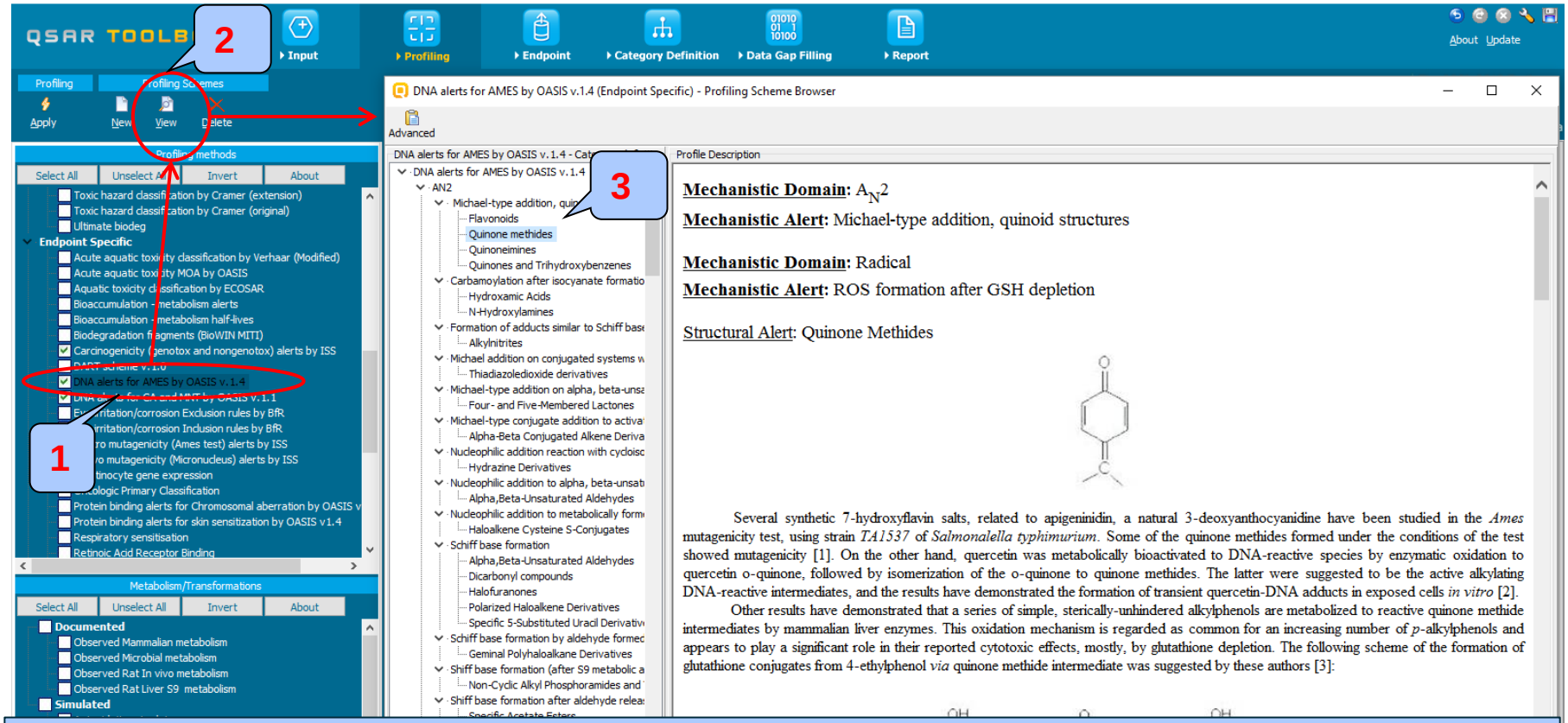
Profiling
Overview

- "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.
- Available information includes likely mechanism(s) of action, as well as observed or simulated metabolites.

Profiling
Side-Bar to Profiling

- For most of the profilers, background information can be retrieved by highlighting one of the profilers and clicking on “View”

Profiling
Side-Bar to Profiling



1. **Highlight** the profiler
2. **Click** View
3. **Click** on category “Quinone methides”. The textual description is displayed on the right

Profiling
Profiling the target chemical

- The outcome of the profiling determines the most appropriate way to search for analogues (detailed information in Manual for getting started (Chapter 4). http://www.oecd.org/dataoecd/58/56/46210452.pdf
- Table 4 - 1 in chapter 4 (Manual for getting started) lists a selection of profilers and their relevance for different endpoints of regulatory relevance.
- In our case study the following endpoint specific and general mechanistic profiling schemes are relevant to Carcinogenicity endpoint:
  - DNA alerts for AMES by OASIS v.1.4
  - DNA alerts for CA and MNT by OASIS v.1.1
  - DNA binding by OECD
  - Protein binding alerts for Chromosomal aberration by OASIS v1.2
  - Carcinogenicity (genotox and nongenotox) alerts by ISS
  - Oncologic Primary Classification
  - Organic functional groups – all four types
- More details about identified analogues is provided on slide 41

Profiling
Profiling the target chemical

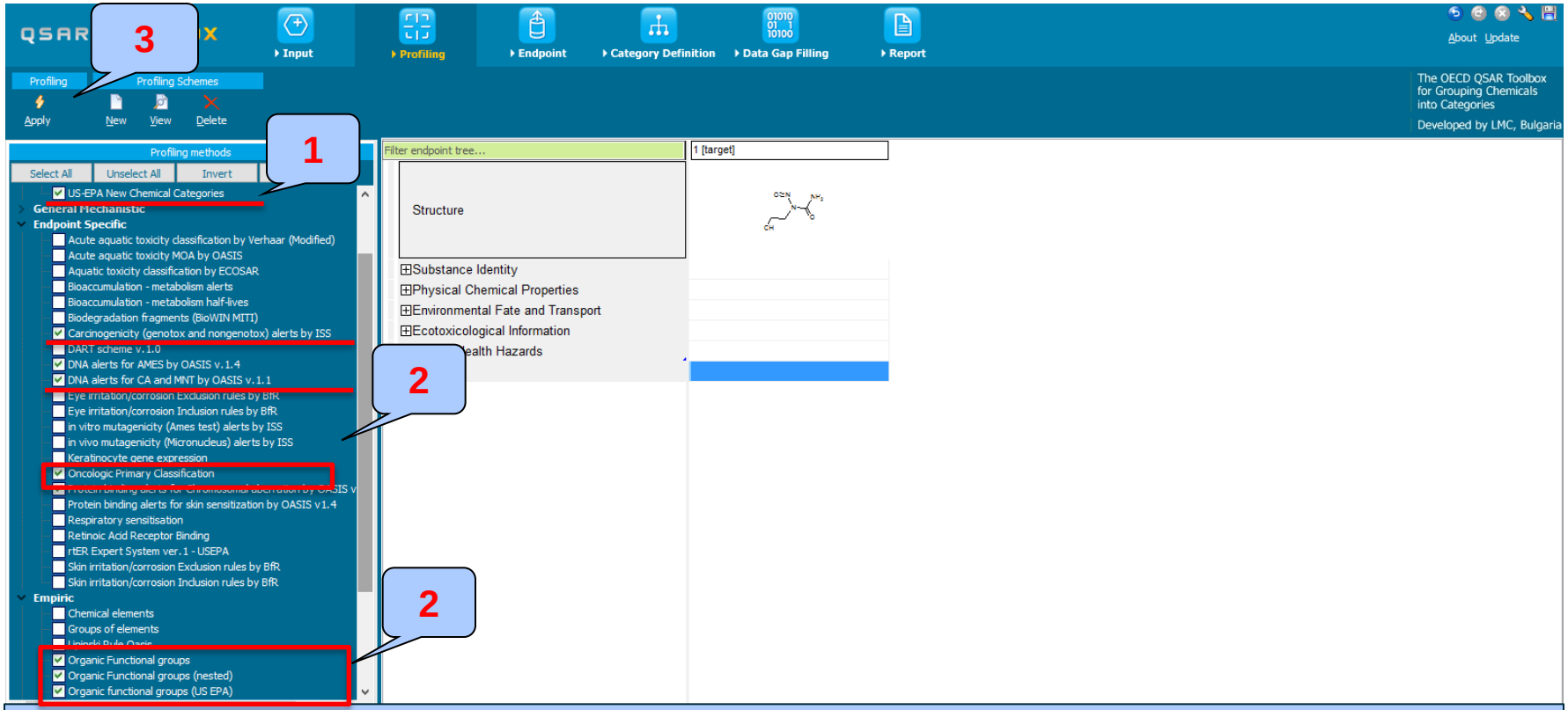
- **Click** in the box next to the name of the profiling methods related to the target endpoint.

- This selects (a **green** check mark appears) or deselects (**green** check mark disappears) profilers.

- For this example, **check** all the profilers mentioned above and **click** on apply (see next screen shot).

Profiling

Profiling the target chemical



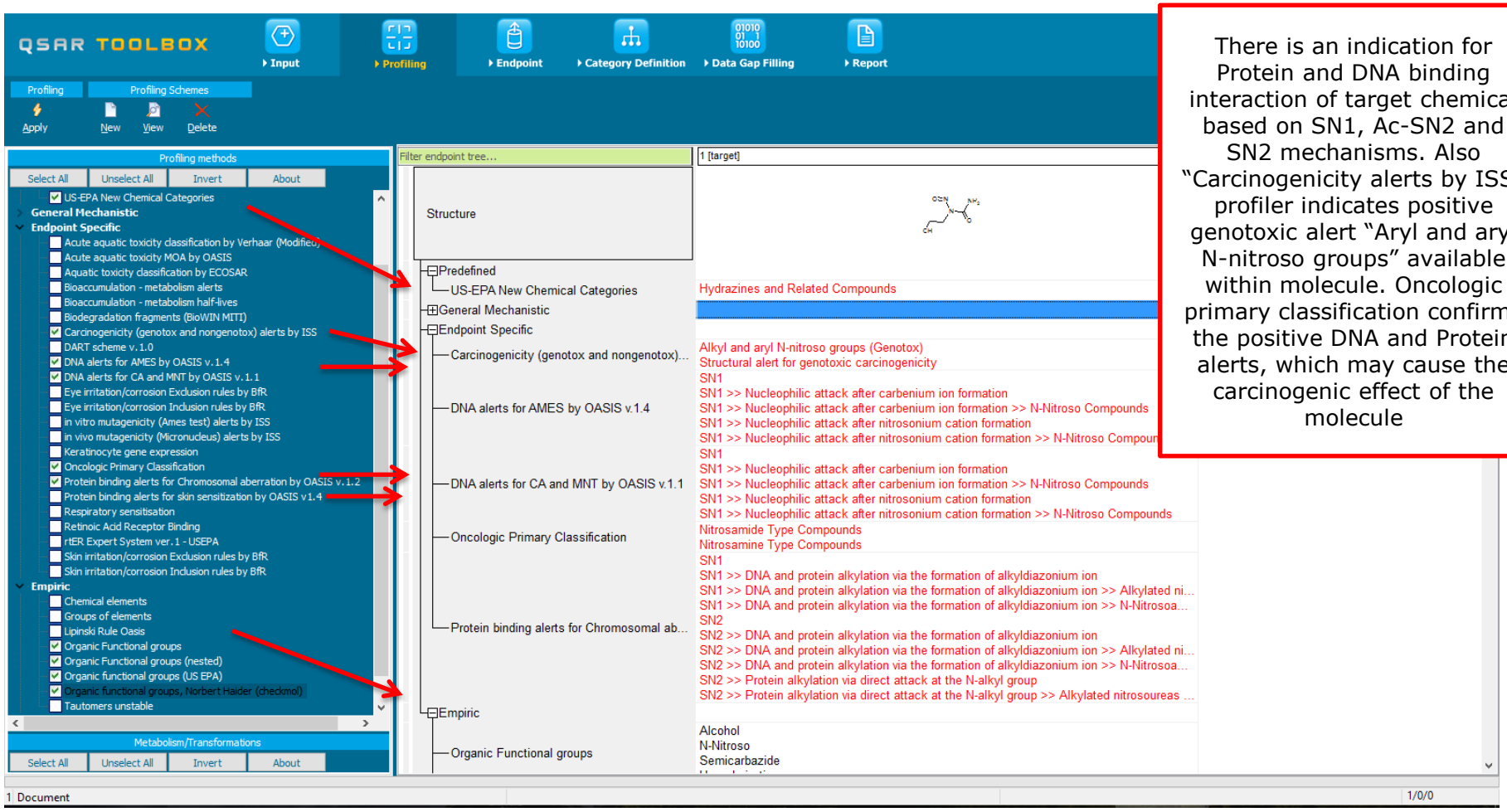
1. **Select** the "US-EPA new Chemical categories" profiler
2. **Select** the endpoint specific profilers associated with target endpoint and mentioned on slide 25
3. **Click** "Apply"

Profiling
Profiling the target chemical

- The profiling will take up to several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appear as a dropdown box under the target chemical (see next screen shot).
- Please pay attention on the outcome of endpoint-specific profilers – Protein and DNA binding by OASIS (see side-bar on carcinogenicity above) and general (endpoint – non specific, OFG) profilers.
- This result will be used to search for suitable analogues in the next steps of the exercise.

Profiling

Profiling the target chemical



There is an indication for Protein and DNA binding interaction of target chemical based on SN1, Ac-SN2 and SN2 mechanisms. Also "Carcinogenicity alerts by ISS" profiler indicates positive genotoxic alert "Aryl and aryl N-nitroso groups" available within molecule. Oncologic primary classification confirms the positive DNA and Protein alerts, which may cause the carcinogenic effect of the molecule

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - **Endpoint**

Endpoint
Overview

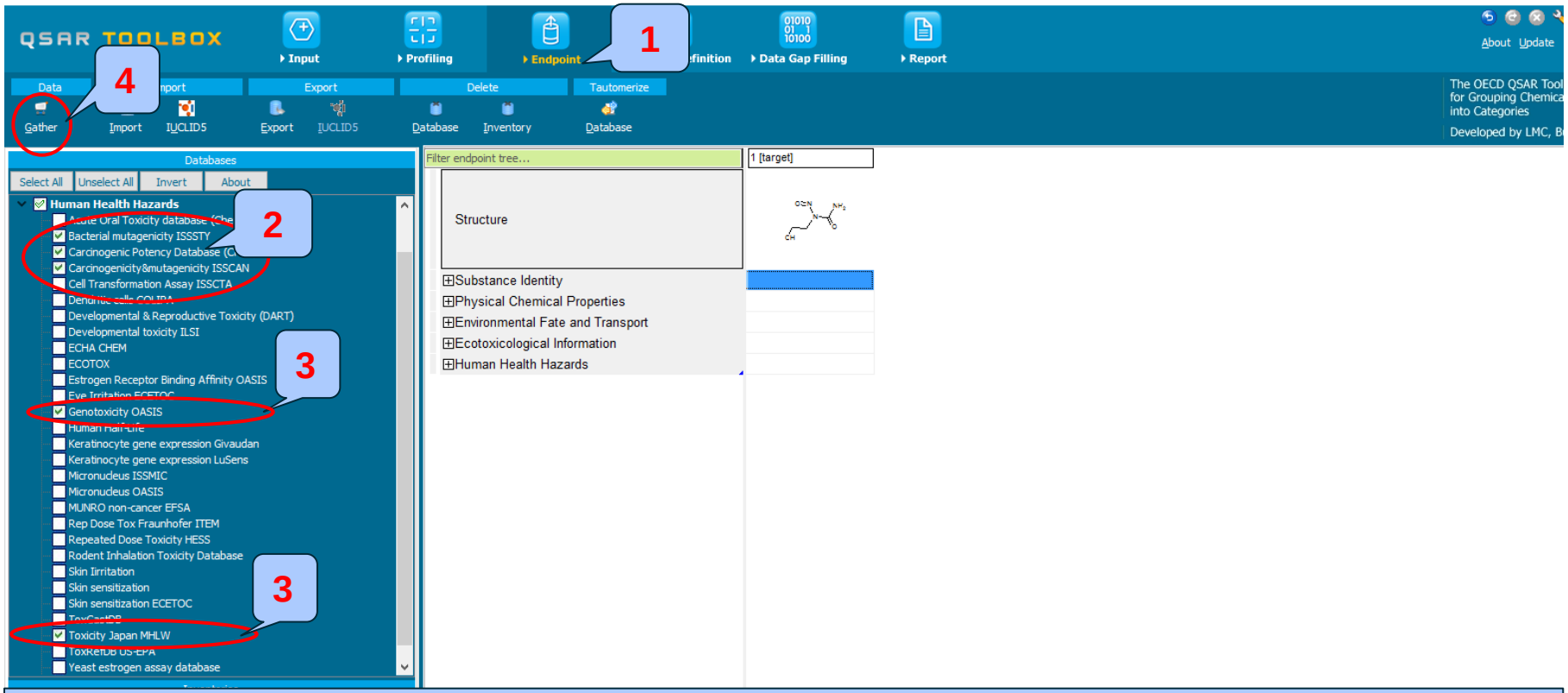
- "Endpoint" refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.

- 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).

Endpoint
Case study

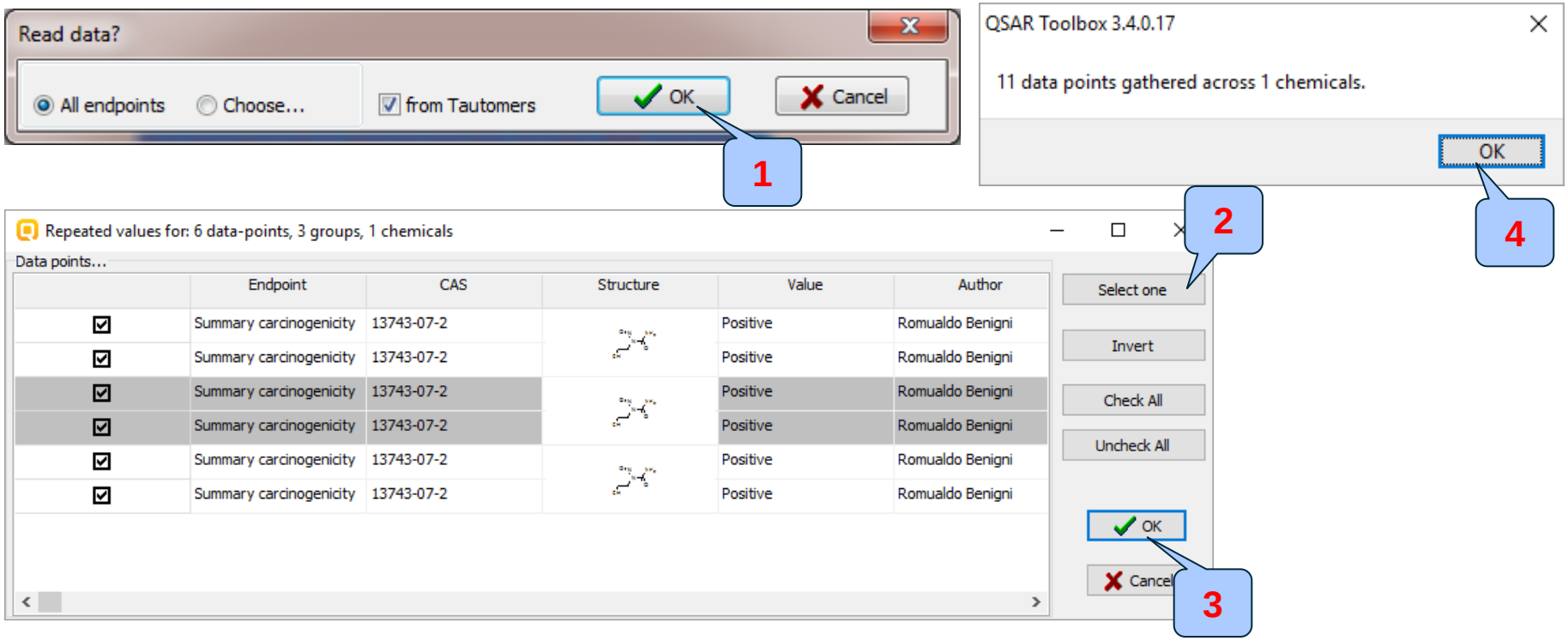
- In this example, we limit our data gathering to the two endpoints: "carcinogenicity" and "Ames mutagenicity". The latter will be used for filtering by Ames data in the subcategorization process. Selection of databases including "Ames mutagenicity" data is needed prior to entering data gap filling and assessing carcinogenicity effect.
- The following "carcinogenicty" and "Ames mutagenicity" databases have been selected and used in this analysis:
  - Bacterial ISSTY
  - Carcinogenicity potency database (CPDB)
  - Carcinogenicity&mutagenicity ISSCAN
  - Genotoxicity OASIS
  - Toxicity Japan MHLW
- Follow the steps:
  - Click on "Endpoint" in the Toolbox workflow.
  - Expand the "Human Health Hazards" section
  - Click on the box to select databases mentioned above
  - Click on "Gather data" (see next screen shot).

Endpoint
Gather data



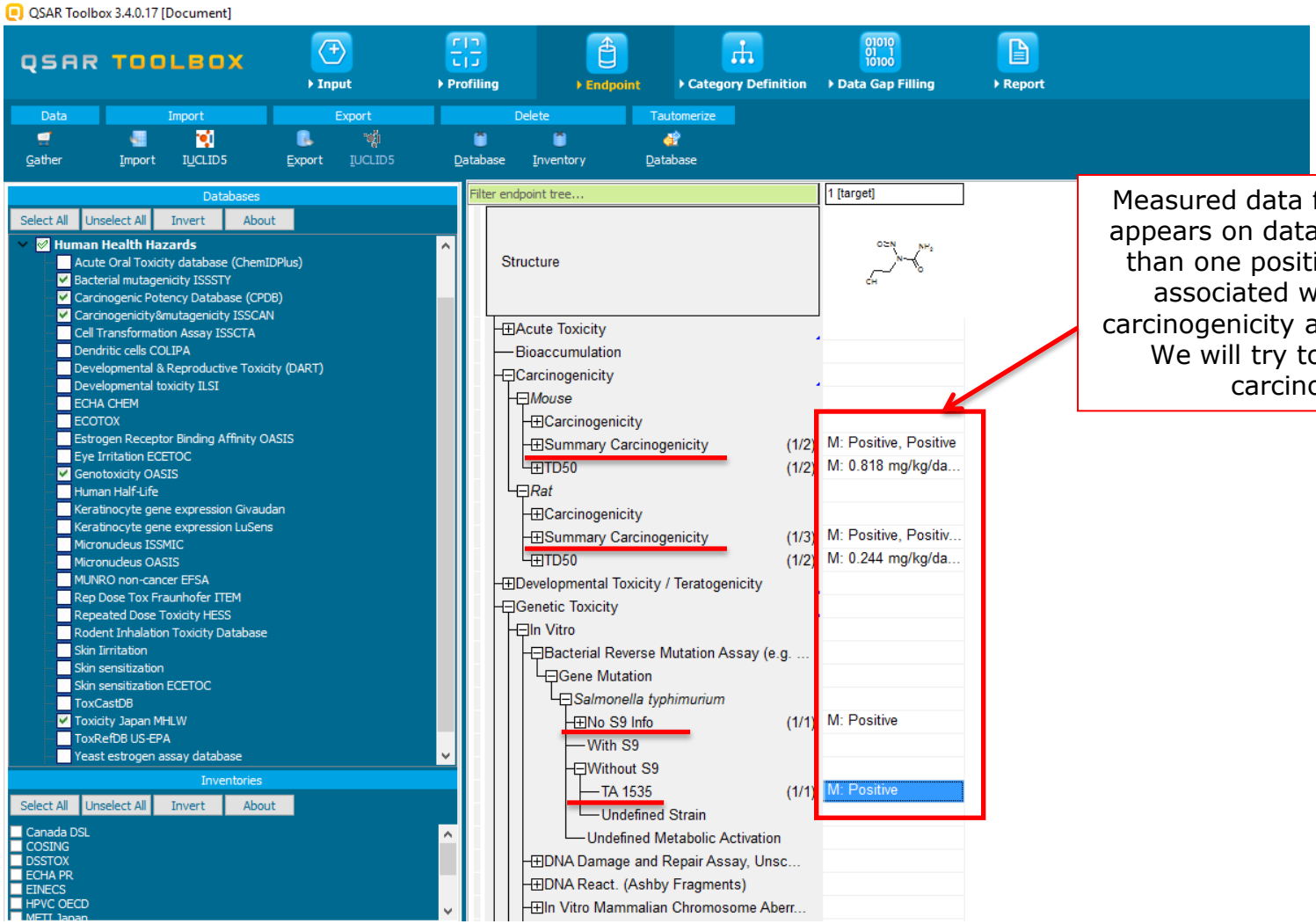
1. **Go** to "Endpoint"
2. **Expand** the "Human Health Hazards" section
3. **Select** databases related to the target endpoint: Carcinogenicity and Ames mutagenicity. This is needed for forthcoming filtering by Ames measured data in the stage of "Data gap filling"
4. **Click** "Gather"

Endpoint
Gather data



1. **Click** "OK" to extract data from database
2. **Click** on "Select one" button
3. **Click** "OK"
4. The message informing for number of gathered data for the target chemical appears. **Click** "OK"

Endpoint
Gather data



Measured data for the target chemical appears on data matrix. There is more than one positive experimental data associated with both endpoints: carcinogenicity and Ames mutagenicity. We will try to reproduce positive carcinogenicity data

Recap

- In the first module, you have entered the target chemical being sure of the correctness of the structure.

- In the second module, you have profiled the target chemical and found that the target chemical may interact with DNA and proteins based on 3 different mechanisms. This could be an indication for positive carcinogenicity effect of the target.

- In the third module, you have found that there are more than one positive carcinogenicity data found for the target structure associated with rat and mouse species. Also positive experimental data was found for Ames mutagenicity. Positive experimental data supports positive DNA and Protein binding alerts found for the target chemical. We will try to reproduce the positive carcinogenicity data by using read-across analysis.

Next actions

- Read-across analysis will be performed for "Summary carcinogenicity" endpoint associated with rat species

- Before proceeding with "Data Gap Filling" module, the user should define a category of similar analogues. The "Organic functional groups" as broad endpoint-non specific grouper will be used for defining similar analogues (see next slides).

- Click on "Category Definition" to move to the next module.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Category definition**

Category Definition
Overview

- 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.
- Several options are available in the Toolbox to assist the user in refining the category definition.

Category Definition
Grouping methods

- The forthcoming 4 slides provide basic information about definition and procedure of "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.
- Detailed information about grouping chemical (Chapter 4) could be found in document "Manual for Getting started" published on OECD website:

  http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm

Basic guidance for category formation and assessment

Usually, a three stages procedure is recommended for building categories for read-across, in Toolbox. The categorization phases could be organized as follows:

- Stage I: Broad and endpoint non-specific primary categorization of chemicals based on their belonging to common chemical classes, predefined categories or being structurally similar
- Stage II: Subcategorization based on mechanisms conditioning the target endpoint thus coming to endpoint specific subset of chemicals reacting by same interaction mechanisms.
- Stage III: Further narrowing down the category based on elimination of chemicals most dissimilar to target one by using additional structure-related profilers

This sequence of stages is not mandatory and depends on the specificity and number of the chemical analogues and target endpoint. Moreover, some of the stages could be skipped if consistency of category members is reached earlier. It is also recommended only primary categorization to be applied in the Category Definition phase of the Toolbox workflow whereas the subcategorization to be applied at Data gap filling phase; thus, one could follow up the effect of subcategorization on the read-across results (having visualization of the endpoint vs. parameter relationship).

The structural similarity is not recommended to be applied as primary categorization. However, often it is needed to be used in the last stage of the subcategorization – for eliminating most dissimilar chemicals. This holds for read-across implementation for any endpoint.

Graphical illustration of suitable categorization phases is shown on next slide

Suitable Categorization/Assessment Phases

Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Broad grouping
Endpoint Non-specific

Repeating Phase I due to Multifunctionality of chemicals

Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- Genotoxicity/carcinogenicity
- DART v1.0
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules
- Repeated dose profiler (NITE)

Subcategorization
Endpoint Specific

Metabolism accounted for

Phase III. Eliminating dissimilar chemicals

Apply Phase I – for structural dissimilarity
Filter by test conditions – for Biological dissimilarity

Category Definition

Grouping methods – phase I

Suitable Categorization/Assessment Phases

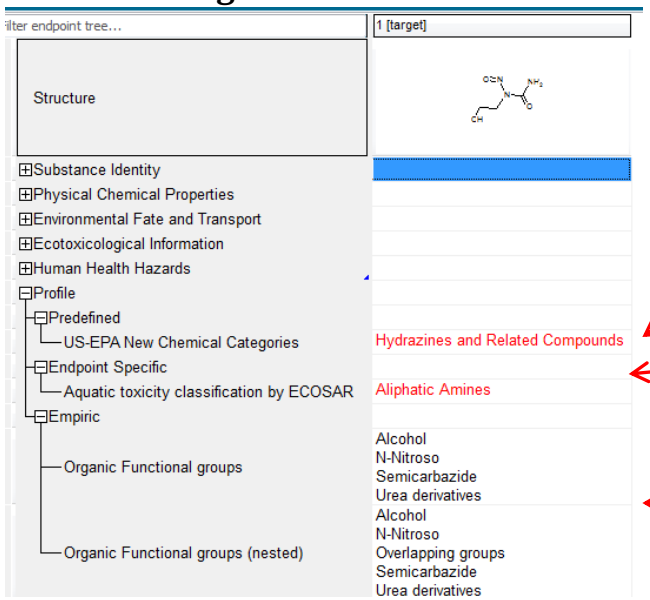
Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Broad grouping
Endpoint Non-specific

Each of the above grouping method is applied to the target chemical and number of identified analogue is provided below. The structural features met in the target chemical regarding each of the applied grouping method are provided on next slide. The aim is to illustrate why OFG is used for categorization purpose.

Phase I categorization in Toolbox



520 analogues are identified.

1687 analogues are identified

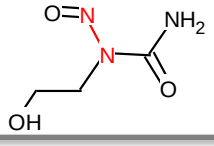
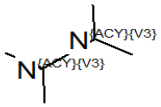
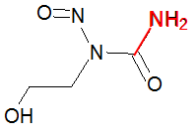
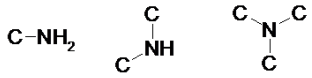
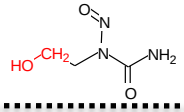
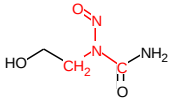
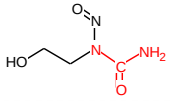
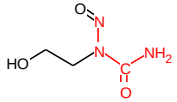
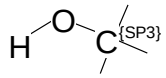
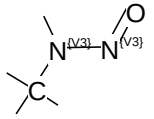
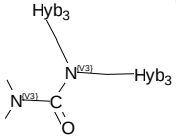
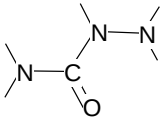
11 analogues are identified (in case all categories are preserved)
53 analogues are identified (in case N-nitroso and Urea are preserved)

Structural similarity, Dice ACF, 50%

48 analogues are identified

Category Definition

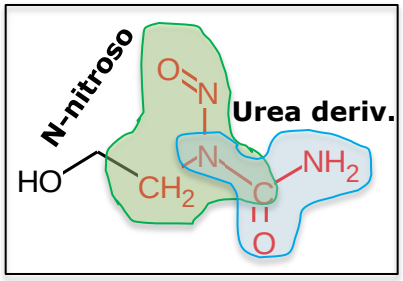
Phase I grouping methods – definitions

Category used for grouping (phase I)	Structural feature met in the chemical structure - highlighted in red	Boundary(ies) coding the rule	Comment
US-EPA (New Chemical categories)		 **AND** MW= 133 Da	**The chemical meet the structural and parametric criteria of the category**
ECOSAR			**The chemical meets the structural criteria of the category**
Organic functional groups (OFG)	**Alcohol**  **N-nitroso**  **Urea derivatives**  **Semicarbazide** 	   	**The chemical meet the structural criteria of each of these four OFG categories: Alcohol, N-nitroso, Urea and Semicarbazide**

Category Definition

Grouping methods - recap

- Based on above recommendations and basic guidance for grouping chemicals explained on the previous slides, phase I grouping methods are compared each other and illustrated on slide 43
- Definition of structural fragments met in the target chemical regarding different categorization groups are provided on slide 44
- Analysis of structural fragments shows that the most appropriate group for identifying analogues is OFG, when two out of 4 structural fragments are combined: "N-Nitroso" and "Urea derivatives".

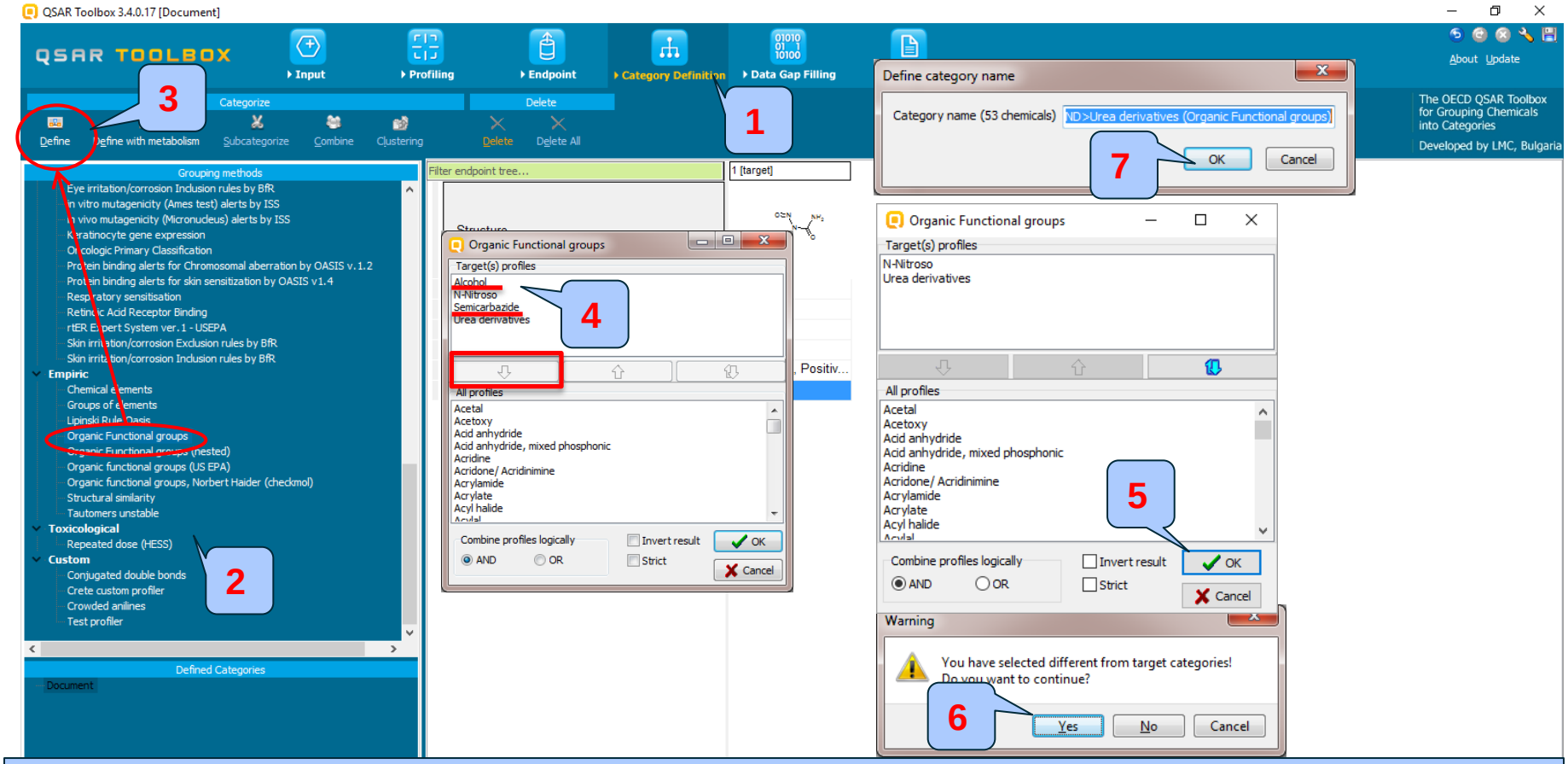


The two alerting categories mentioned above are used for categorization, because:

- They cover basic functionalities, which may explain the positive DNA and protein interaction. Alcohol group is an inert group and usually does not lead to positive genotoxic effect. Hence, no carcinogenic effect will be expected from this group and is omitted from the selection. "Semicarbazide" group is an alternative group, which could be used in the group selection instead of "Urea derivatives".
- US-EPA and ECOSAR grouping methods are not used for categorization, because they are broad and do not cover the two basic functionalities, which could explain the positive carcinogenicity effect.

Category Definition

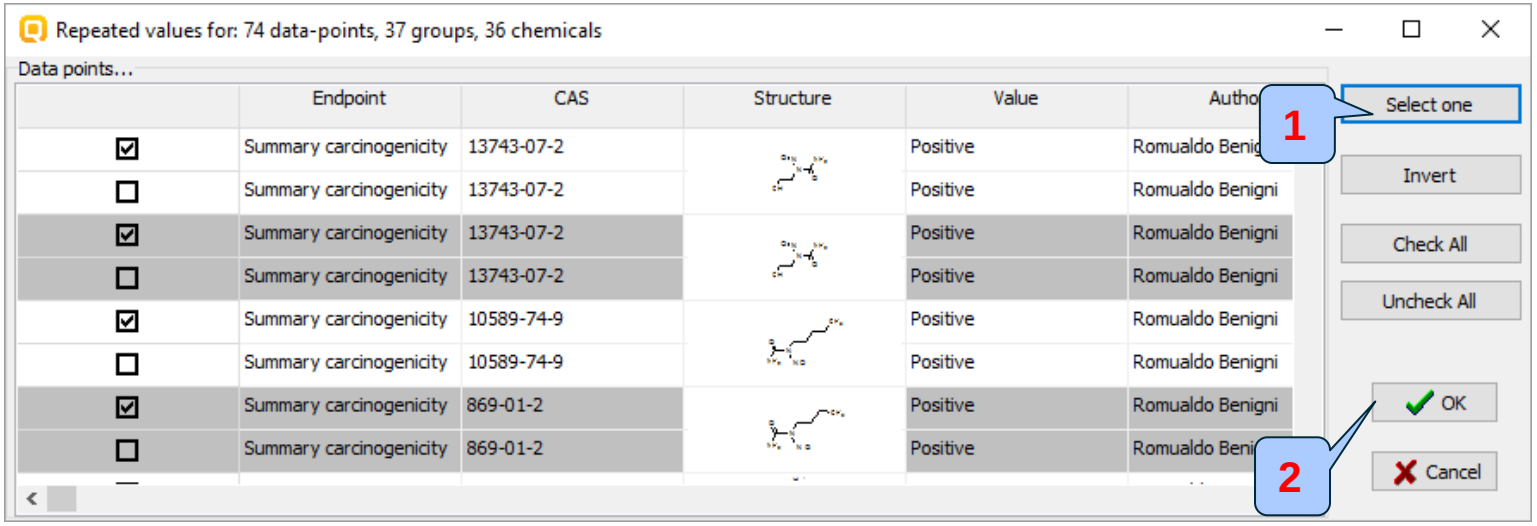
Defining category by OFG

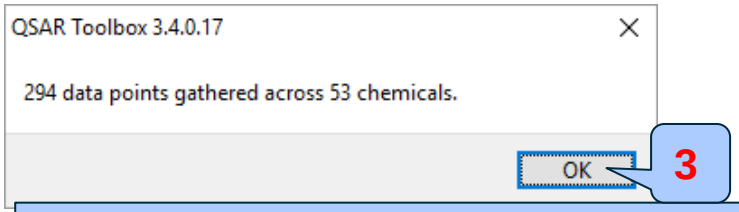


1. **Go** to "Category definition" 2. **Highlight** the "Organic functional group"
3. **Click** "Define" 4. **Click** over the category "Alcohol" and "Semicarbazide" and **remove** it from the selection by clicking on ⇩ button
5.**Click** "OK" 6. **Click** "Yes" on the appeared warning message. 53 analogues were identified.
7.**Click** "OK"

Category Definition
Read data for Analogues

Redundancy table appears informing for repeating data values for same chemical presented in different databases.

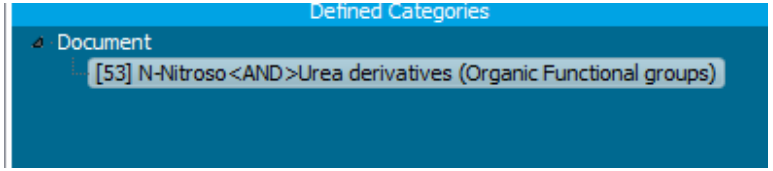




In our case we use single data value: 1. **Click** "Select one" button, then 2. **Click** "OK"; 3. 276 data points are gathered for the identified 53 analogues. **Click** "OK"

Category Definition
Defining category by OFG

- The data is automatically collected.
- Based on the defined category ("N-Nitroso" and "Urea derivatives") 53 analogues have been identified
- The name of the category appears in the "Defined Categories" window, indicating the number of substances belonging to the category.

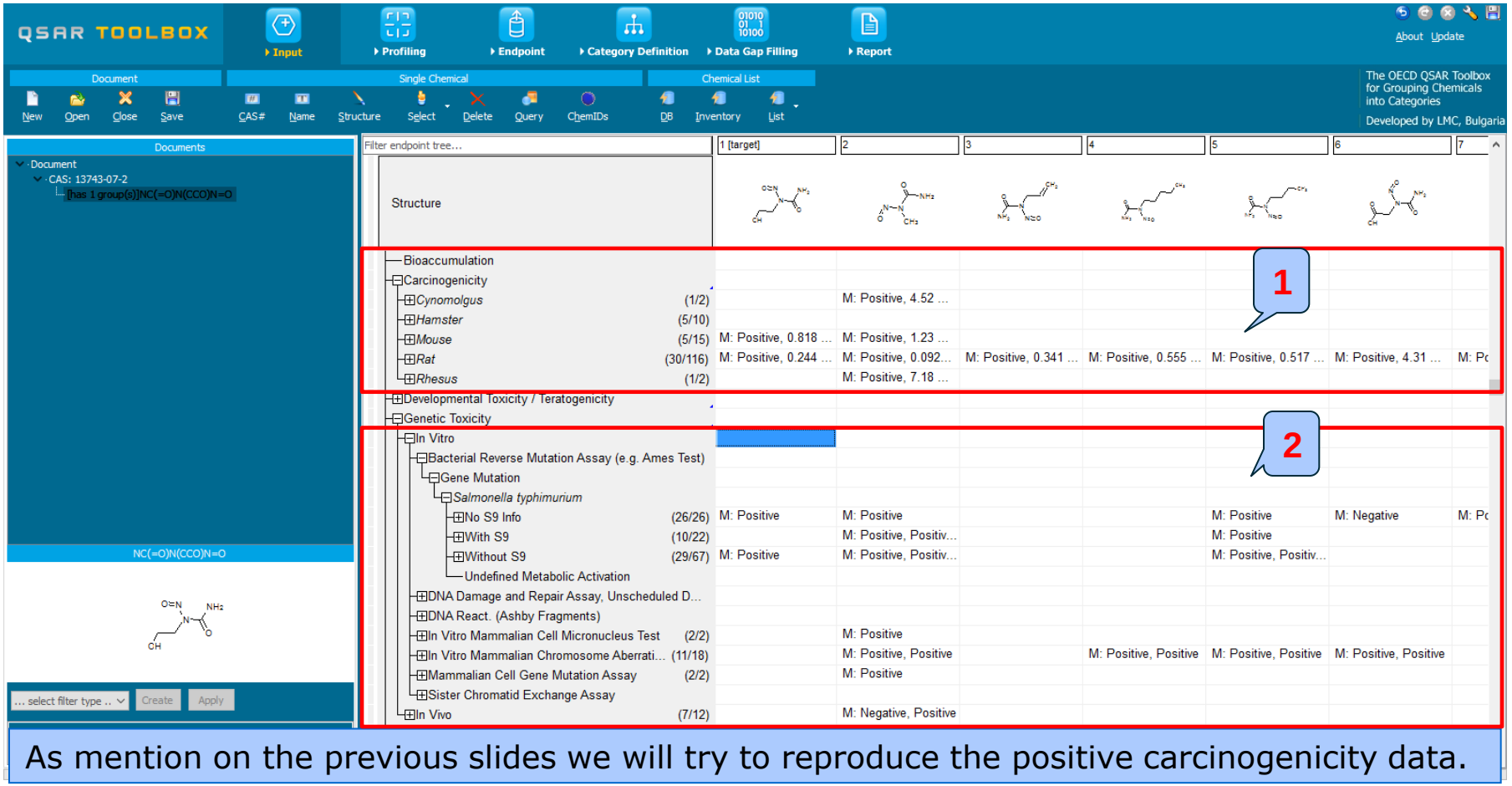


- In other words, these 53 compounds along with the target chemical form a category, which will be used for data gap filling (see next slide).

Category Definition

Defining category by OFG

The experimental carcinogenicity (1) and genotoxicity (2) data for the analogues appears on datamatrix



As mention on the previous slides we will try to reproduce the positive carcinogenicity data.

Outlook

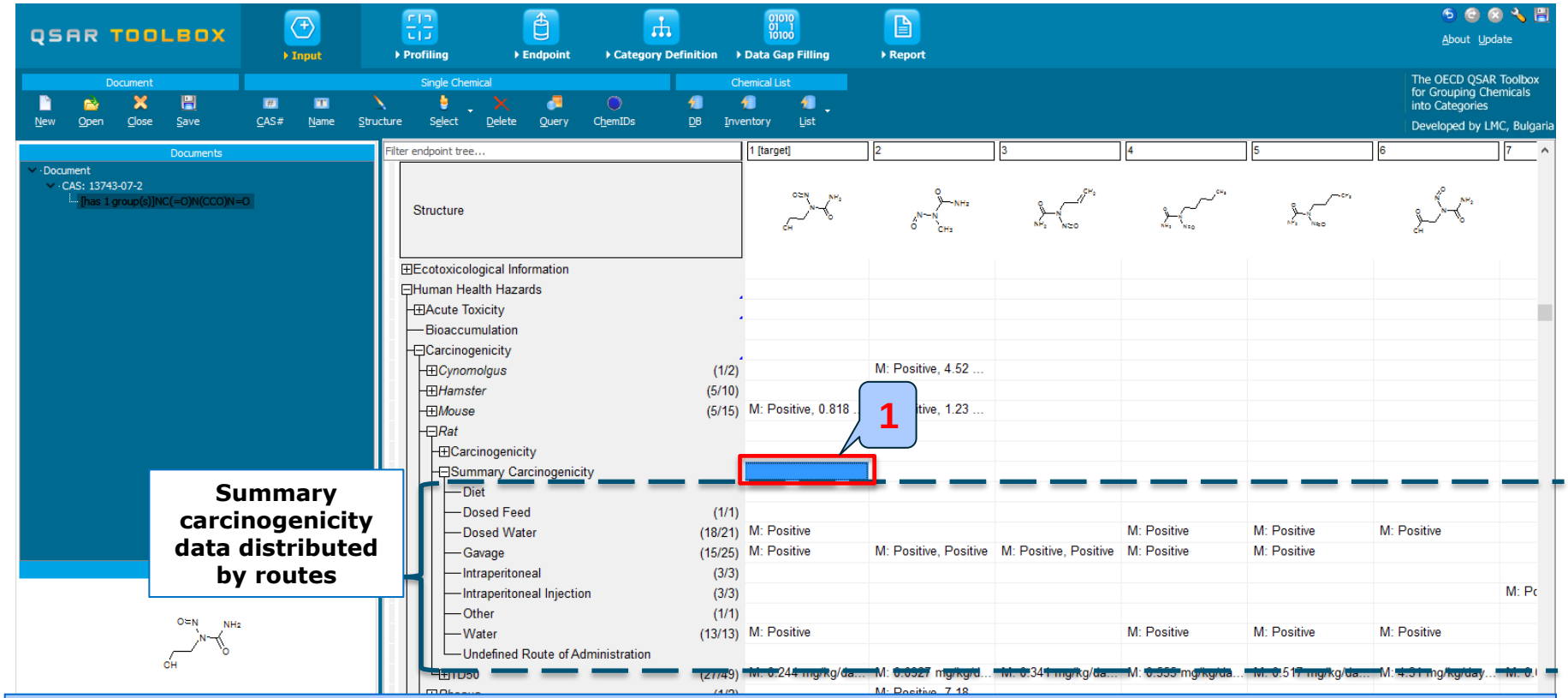
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Category definition
  - **Data gap filling**
    - **Navigation to target endpoint**

Data gap filling
Navigation through the endpoint tree

- Before proceeding with the "Data Gap Filling" module, the user should navigate through the endpoint tree and find the specific gap that will be filled in.
- The user can navigate through the data tree by closing or opening the nodes of the endpoint tree.
- Double-click on the node next to Human Health Hazards then Carcinogenicity, followed by Rat, Summary carcinogenicity.
- In this example, all data associated with different route of administration is used in read-across analysis. In this respect the user should enter gap filling on the level of "Summary carcinogenicity" (see next screen shot).

Data gap filling

Navigation through the endpoint tree



1. **Point** the mouse cursor on the cell of "Summary carcinogenicity" endpoint in order to combine data with different routes. This is the target endpoint. Carcinogenicity experimental data distributed by different "Route of administration" is displayed beneath.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Category definition
  - **Data gap filling**
    - Navigation to target endpoint
    - **Read-across analysis**

Data Gap Filling
Overview

- "Data Gap Filling" module gives 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 sensitisation 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 this example, we use read-across approach.

Data Gap Filling
Interpreting Read-across

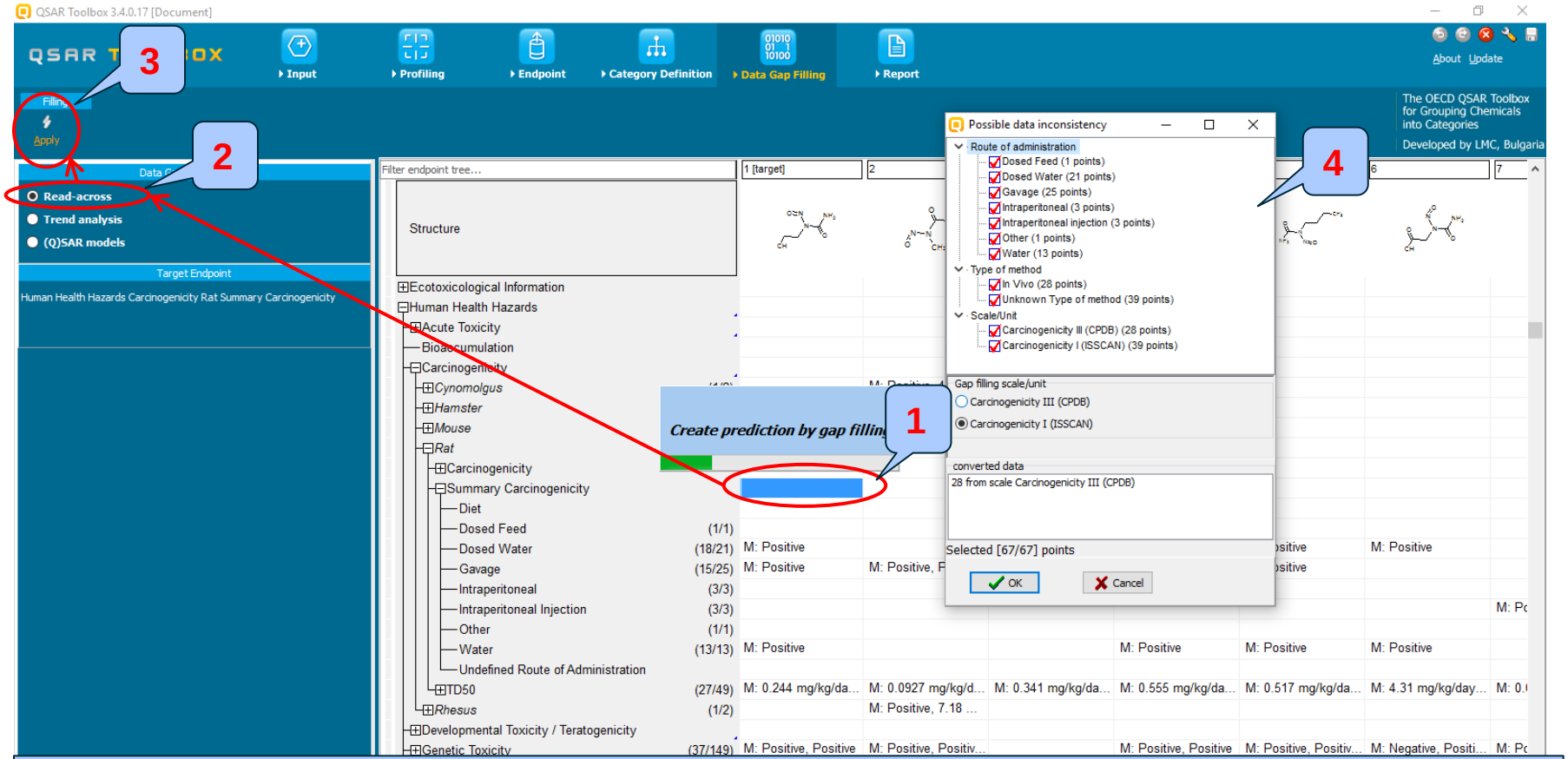
- Read-across prediction is based on an initial group of analogues having "N-Nitroso" and "Urea derivatives" categories defined by Organic functional groups. A group of 53 analogues along with the target chemical is identified.

- Summary carcinogenicity data associated with rat species is available for 30 out of 53 analogue chemicals. This data will be used for filling data gap.

- As mention in the exercise of this tutorial (slide #7) the aim is to predict the carcinogenicity effect of the target by filtering with Ames mutagenicity data. As noted earlier carcinogenicity effect is a complex endpoint depending on DNA and Protein alterations. Based on this the initial group of analogues is further refined using the subsequent subcategorizations starting with:
  - Filter by AMES observed data
  - DNA alerts for Ames, MN, CA by OASIS v.1.4
  - Protein binding alerts for chromosomal aberration by OASIS v1.2

See next screen shots

Data Gap Filling
Apply Read across



1. **Click** on the cell corresponding to "Summary carcinogenicity" node of endpoint tree; 2. **Select** "Read-across"; 3. **Click** "Apply" 4. "Possible inconsistency window" appears indicating that more than one unit/scale is mixed. For more details see next slide.

Data Gap Filling
Scale definition

- Carcinogenicity is a "qualitative" endpoint for which the results are presented with categorical data (for example: positive; equivocal; negative).
- Summary carcinogenicity data of the chemicals originate from different databases, coded with different names (for example: data from ISSCAN are: *Positive, Weakly positive, Equivocal and Negative*; data from CPDB database are: *Positive, Negative, and unspecified,* data from ECHA CHEM database came with different data interpretation: *Yes, no effect, not examined etc.*)
- The main purpose of the scales is to unify all data available in the Toolbox databases for a certain endpoint. To this aim, scale conversion of one data to another data type is implemented. Usually more informative scale is transferred to less informative scale.
- Scale definition and conversions associated with summary carcinogenicity data is presented on next three slides (only for illustration)

Data Gap Filling
Summary carcinogenicity - scale definition

- Summary carcinogenicity data is presented in four different scales:
  - Carcinogenicity I (ISSCAN) including the following scale members:
    - *Positive,*
    - *Equivocal and*
    - *Negative*
  - Carcinogenicity II (ISSCAN) including the following scale members:
    - *Positive,*
    - *Weakly positive*
    - *Equivocal and*
    - *Negative*

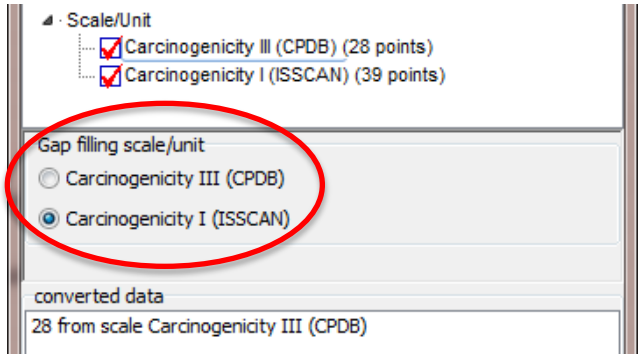
Data Gap Filling
Summary carcinogenicity - scale definition
(continued from previous slide)

- Summary carcinogenicity data is presented in four different scales:
  - Carcinogenicity III (CPDB) including the following scale members:
    - *Positive,*
    - *Negative, and*
    - *Unspecified*
  - Carcinogenicity IUCLID (data coming from ECHA Chem DB) including the following scale members:
    - *Yes*
    - *no effect,*
    - *not examined etc.*

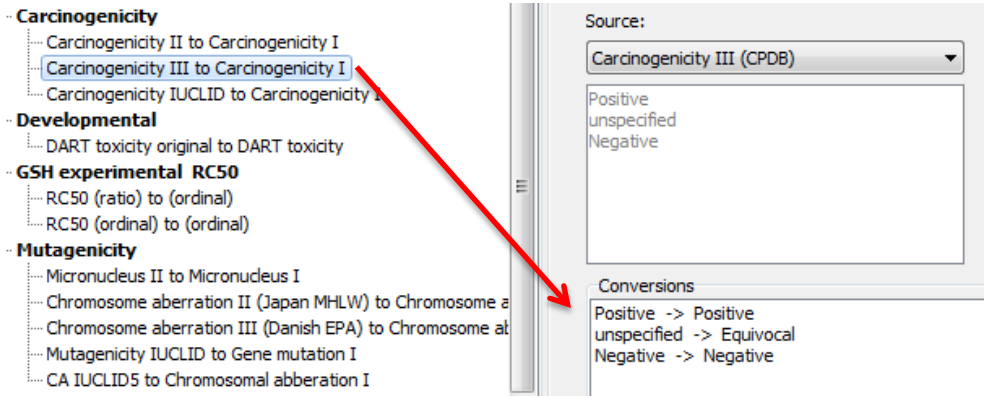
Data Gap Filling

Summary carcinogenicity - scale conversion

Summary carcinogenicity scales used in the current read-across analysis



Implemented scale conversions

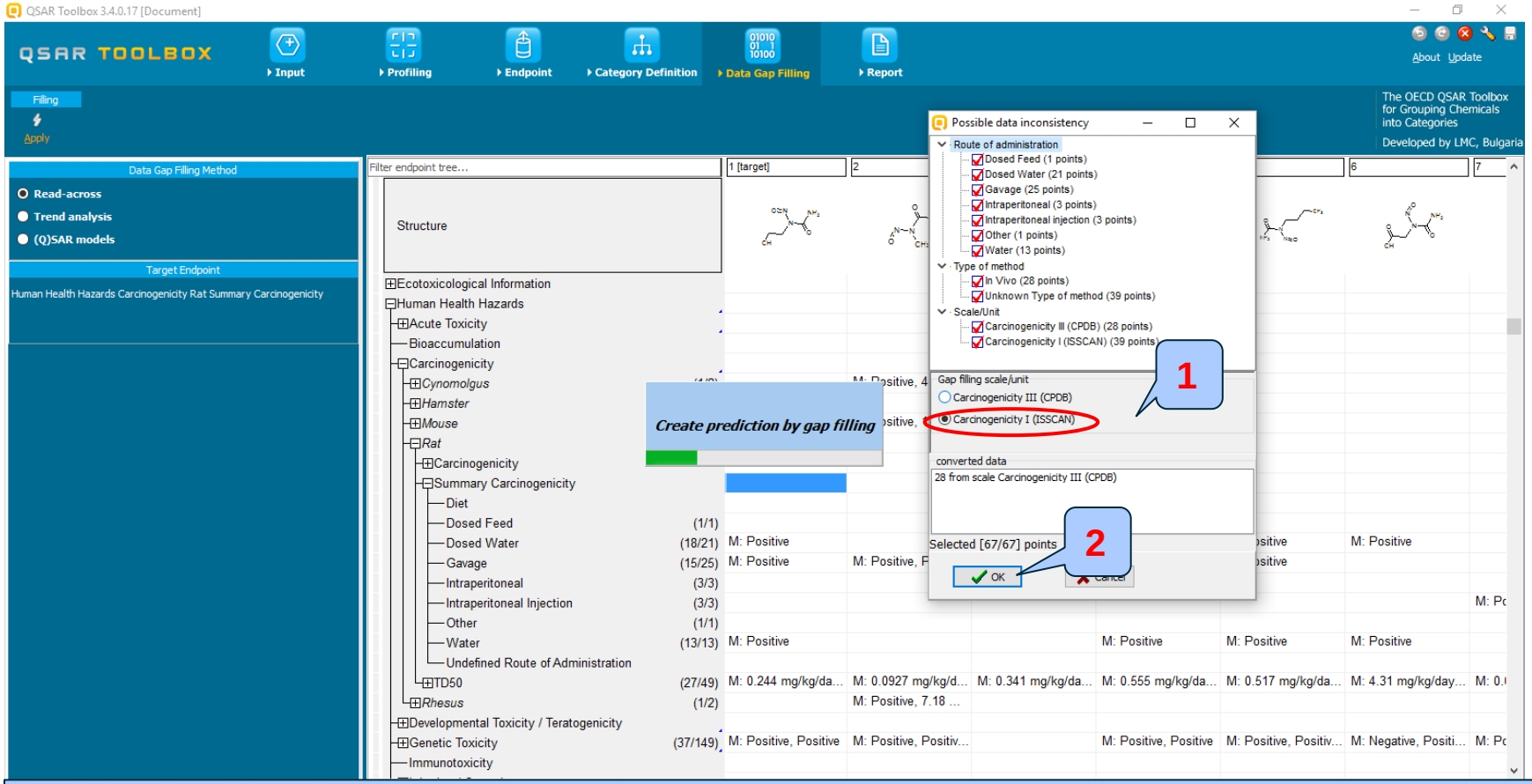


- Scales "Carcinogenicity II, III and Carcinogenicity IUCLID" mentioned on the previous two slides are converted to scale "Carcinogenicity I". As an example, conversion of "Carcinogenicity III" to "Carcinogenicity I" is provided on the right snapshot. The conversion consists of:
  - Positive to Positive
  - Unspecified to Equivocal
  - Negative to Negative

- The data used in current read-across analysis is associated with two scales: Carcinogenicity III and I (see left snapshot). As mentioned above a conversion of III to I has been available. This is the reason the scale "Carcinogenicity I (ISSCAN)" to be used in further read-across analysis.

Data Gap Filling
Apply Read across

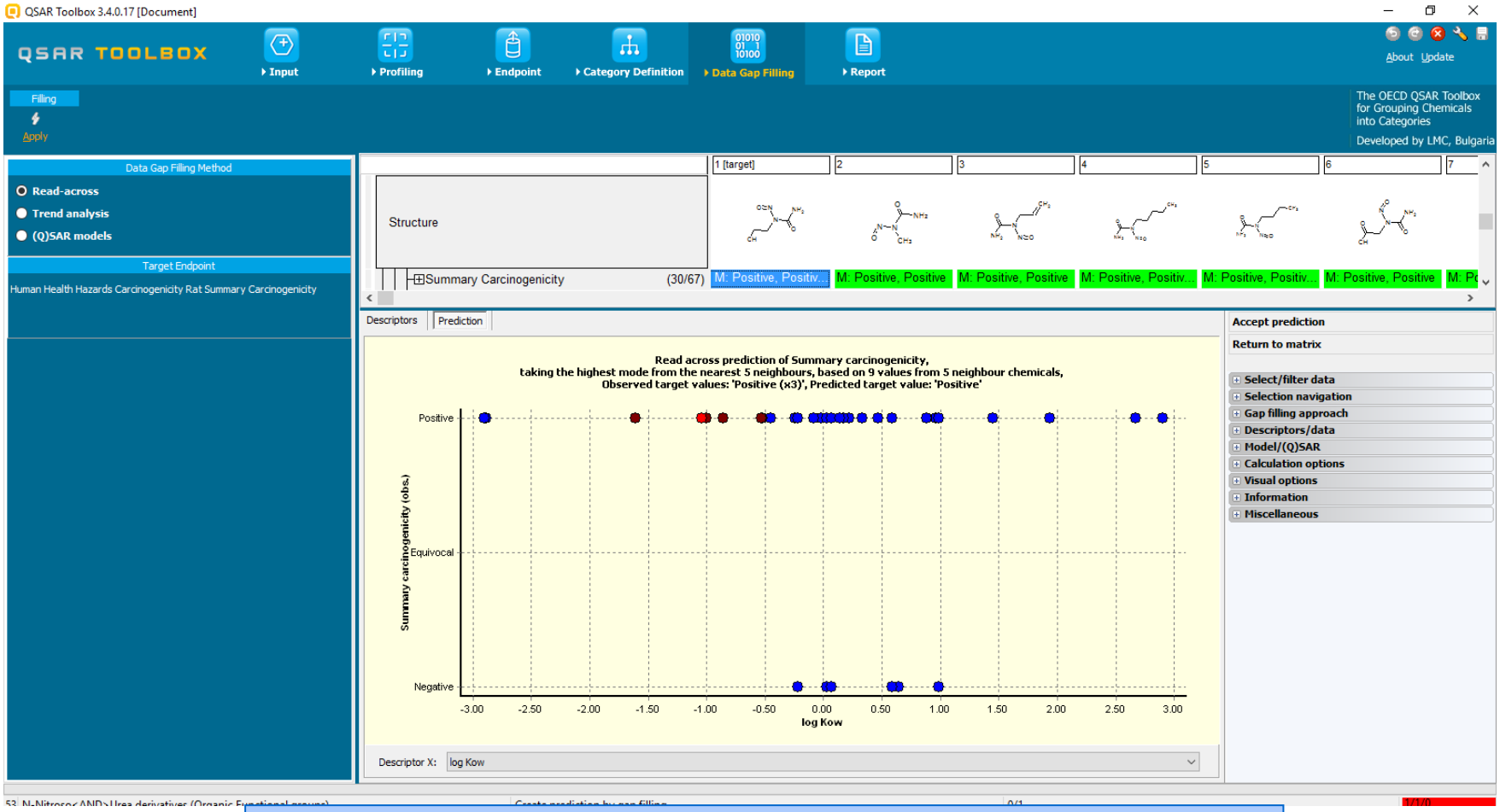
Back to our example



1. **Select** "Carcinogenicity I (ISSCAN)" scale 2. **Click** "OK"

Data Gap Filling

Read across applied for summary carcinogenicity



Initial data gap filling graph without any subcategorization steps.

Data Gap Filling
Subcategorizations

In this example, the following subcategorizations are applied in order to eliminate dissimilar analogues (phase II, see slides 41-42):

- Filter by AMES observed data
- Oncologic Primary classification ("Protein binding alerts for chromosomal aberration by OASIS v1.2" profiler as an alternative subcategorization)
- DNA alerts for AMES by OASIS v.1.4
- DNA alerts for CA and MNT by OASIS v.1.1

See next screen shots

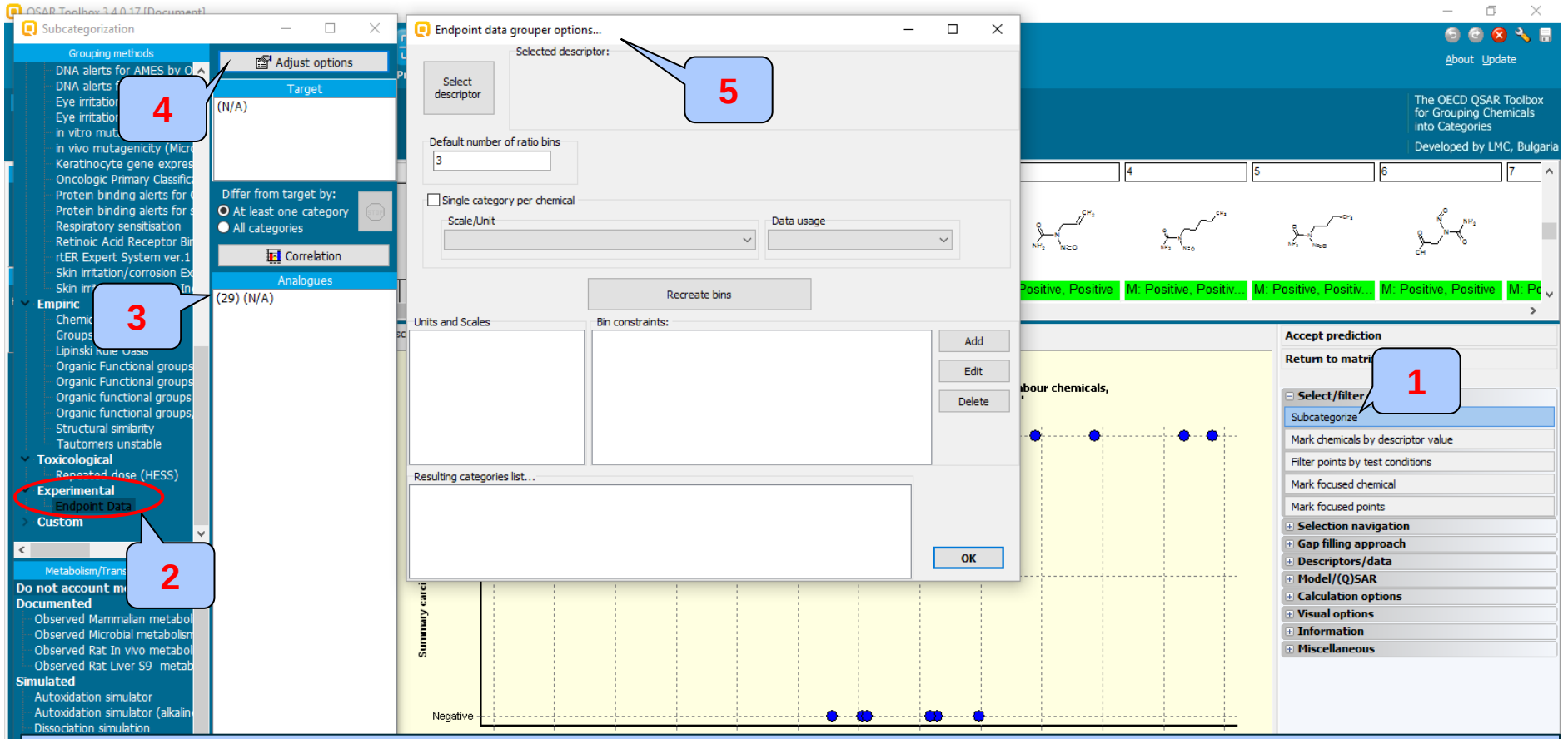
Data Gap Filling

Filter by AMES mutagenicity data – subcategorization 1 (new functionality)

- This is a new functionality introduced in Toolbox 3.3.
- It is a functionality within the data gap filling module allowing to filter chemicals by measured data.
- This functionality separates analogues in bins based on observed data.
- Domain of the obtained read-across prediction takes into account filtering by measured data.
- Filtering by data is a multi-step procedure illustrated in the forthcoming sequence of screenshots.

Data Gap Filling

Filter by AMES mutagenicity data – subcategorization 1 (new functionality)

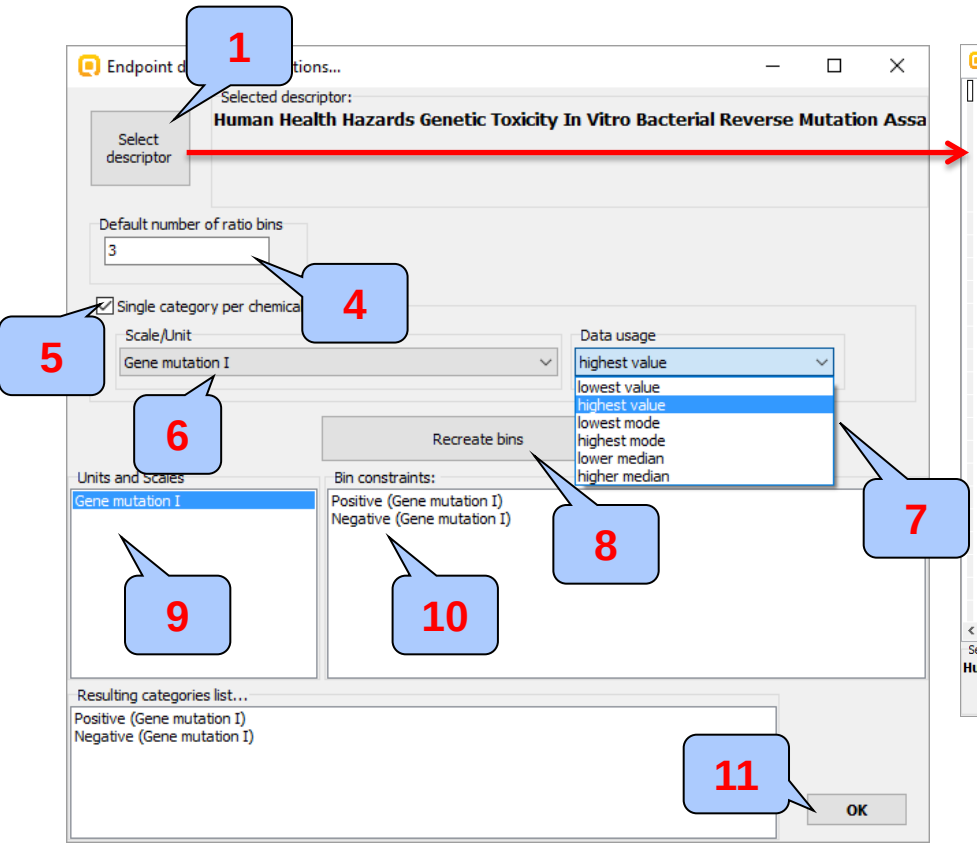


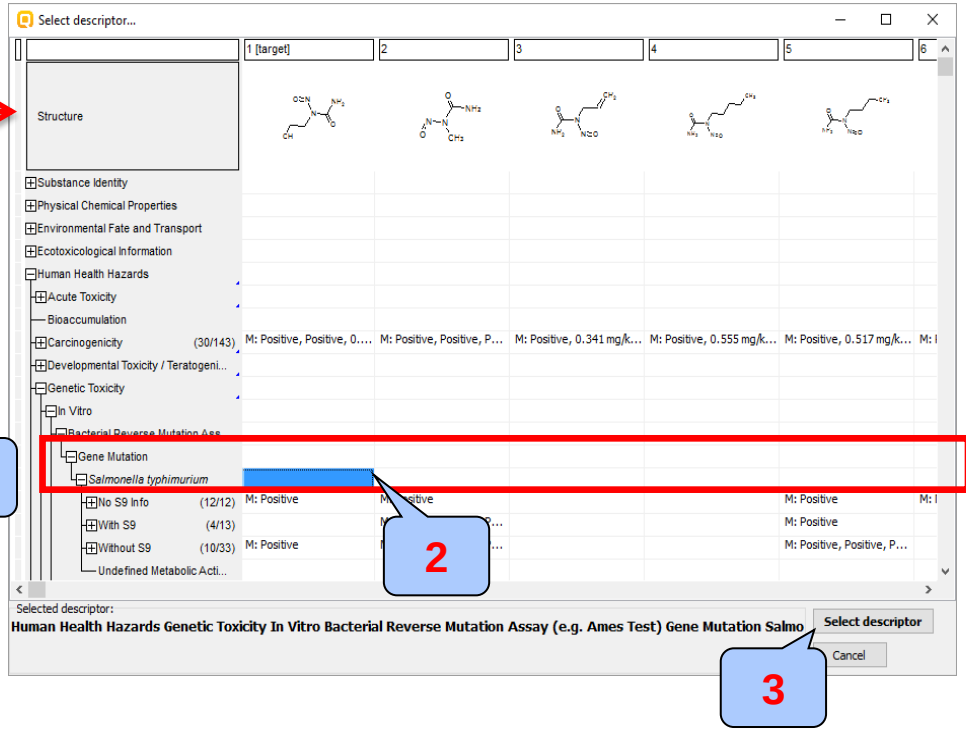
1. **Open** "Subcategorize" panel. The new functionality is implemented as a sub-node of new endpoint tree node called "Experimental"
2. **Click** on "Endpoint data" node 3. Analogues are labeled as N/A, because no observed data has been selected. Follow the steps
4. **Click** on "Adjust options" button 5. "Endpoint data grouper" window appears.

More details about this window are provided on next slide

Data Gap Filling

Filter by AMES mutagenicity data – subcategorization 1

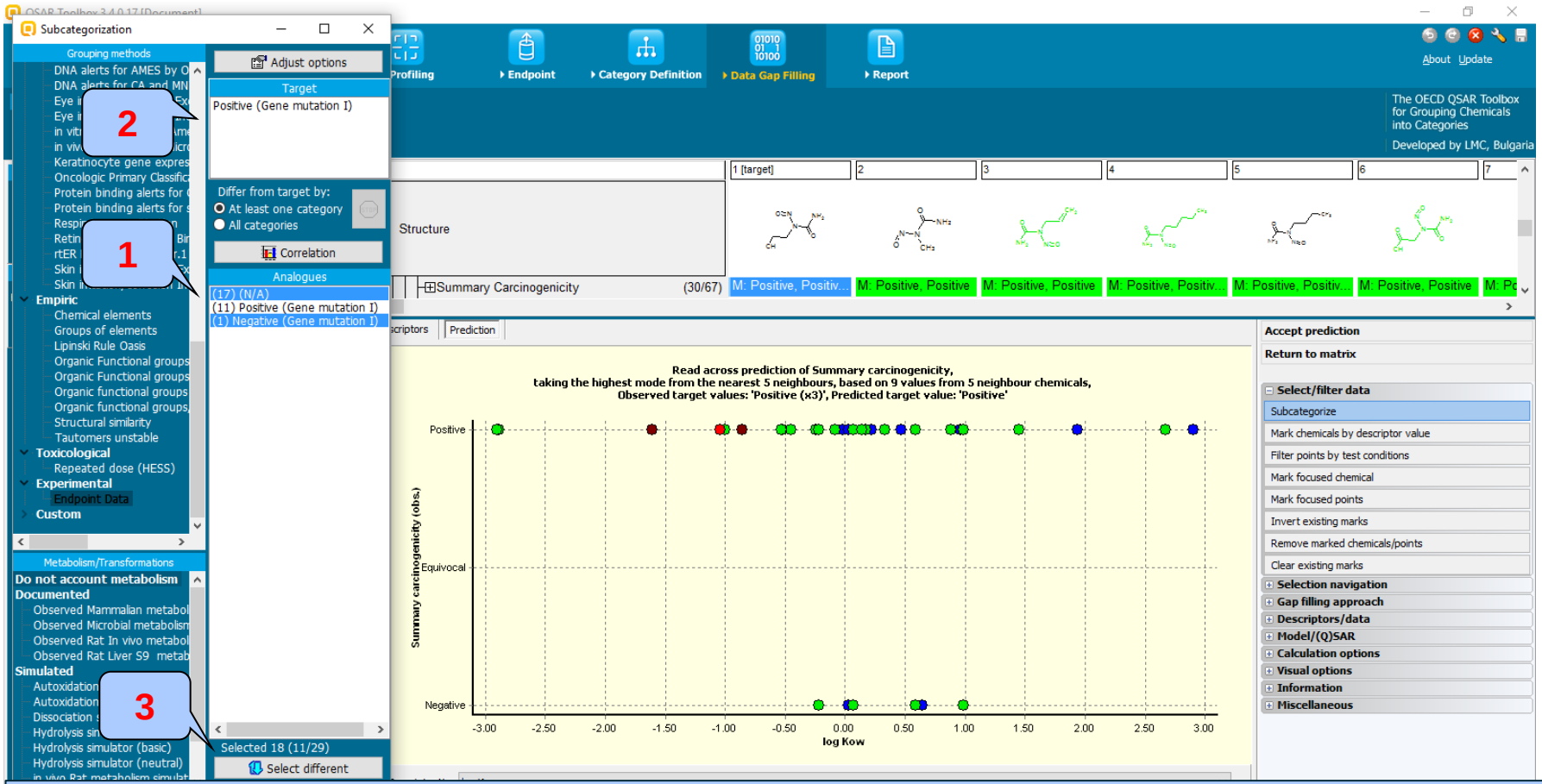




1. "Select descriptor" button evoke an additional window from where the user could select endpoint (observed data) used further for filtering;
2. In our case we mixed up gene mutation data with and without S9;
3. Once the endpoint of interest is selected **click on** "Select descriptor" button;
4. Data are distributed into 3 bins by default. Number of bins could be changed;
5. "Single category per chemical" produces a single value per chemical when multiple values of single unit/scale are present;
6. List of used scales
7. Combo-box list with different data usage options. In our case we use highest values, in case worst case scenario is played
8. **Click** "Recreate bins" to finish the initialization process.
9. Units and scales used
10. A panel with bin constrains of the selected scale
11. **Click** OK to finalize the correlation settings

Data Gap Filling

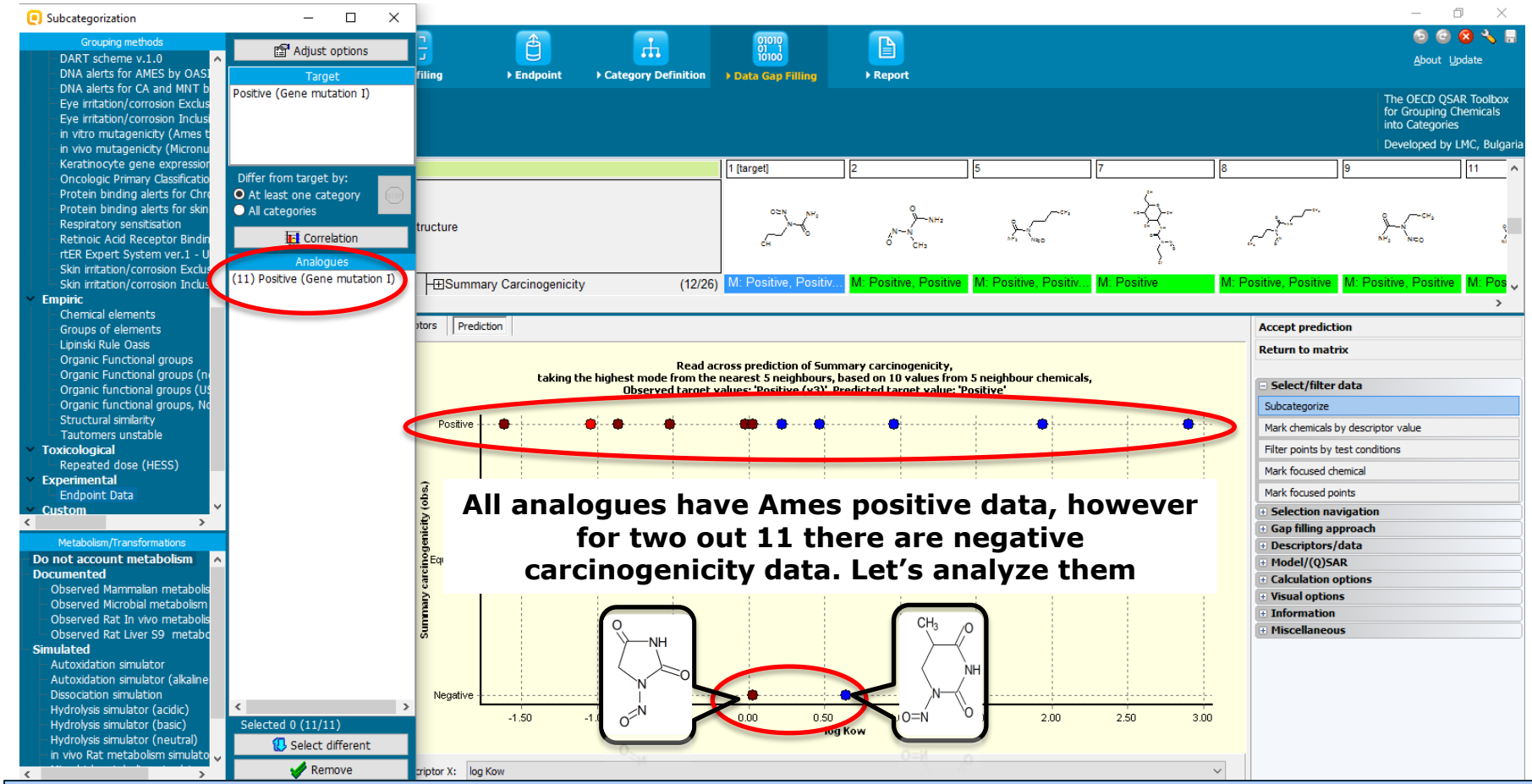
Filter by AMES mutagenicity data – subcategorization 1 (new functionality)



After click the OK button on the previous panel where we have filter the chemicals by Ames data the set of analogues have been distributed into 3 bins (default bins in the Endpoint data grouper window) depending on Ames experimental data (1). The target chemical has positive experimental data (2). As seen from the graph the negative (1 analogue)analogue and analogues without Ames data, labeled as N/A (17 analogues) are distributed into positive and negative level of the graph (green dots). The prediction is apparently positive. In this respect analogues with negative and absence of Ames data are eliminated from the graph. The two bins of analogues (with negative Ames and absence of data) are selected by default and eliminated from the graph by clicking on "**Remove** button"(3).

Data Gap Filling

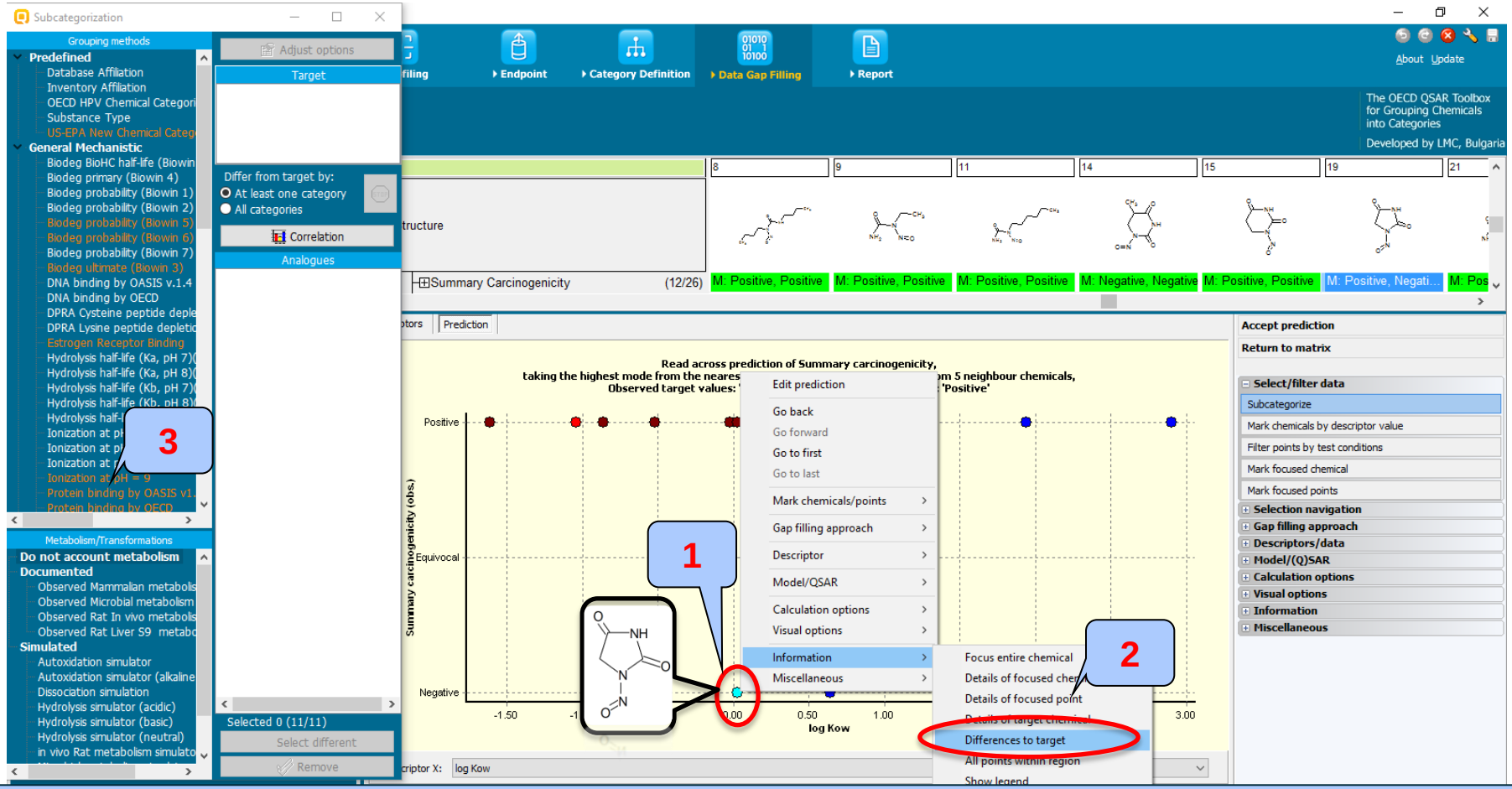
Filter by AMES mutagenicity data – subcategorization 1 (new functionality)



After analogue elimination, the chemicals with positive Ames data remained only. However, as seen in the set of positive analogues there are two chemicals with negative Carcinogenicity data. The next step is to see why these analogues are negative. See next two slides.

Data Gap Filling

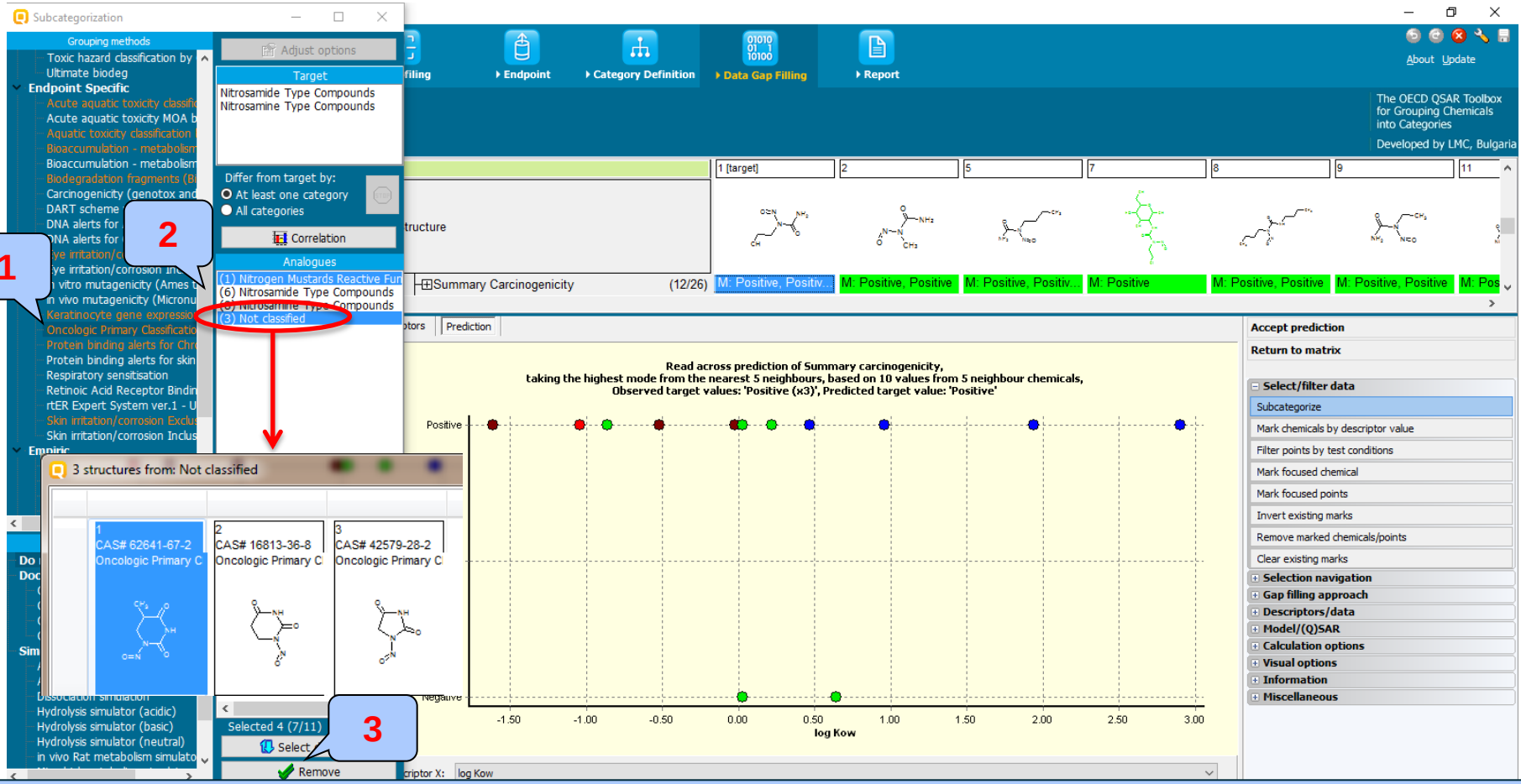
Oncologic primary classification – subcategorization 2



In order to see why the outlier chemical is difference to target follow the steps: 1. **Right click** over the outlier chemical 2. **Select** "Information" and then "Difference to target". The profilers according to which the analogue differ from the target are highlighted in red (3). The most appropriate profiler for subcategorization in case of carcinogenicity assessment is "Oncologic primary classification" and "Protein binding alerts for CA". Let's check both of them separately (see next few slides)

Data Gap Filling

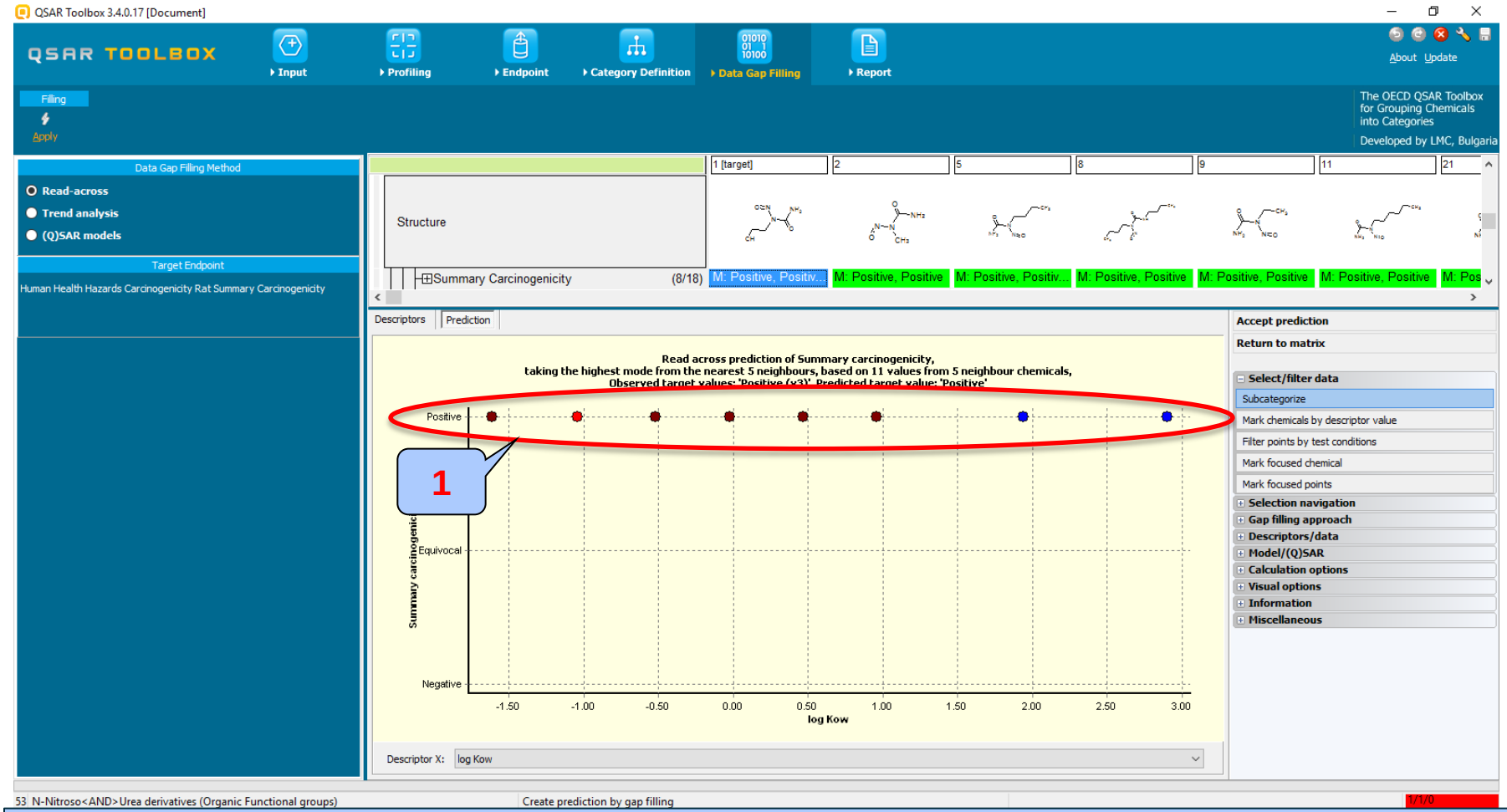
Oncologic primary classification – subcategorization 2



1. **Select** "Oncologic Primary classification".
2. As seen there are 4 analogues with different classification than those of the target. 3 out of 4 are categorized as "Not classified" and 1 outlier is categorized as "Nitrogen mustard" category", which do not belongs to the set of categories of the target chemical. Based on this different classification, these four outlying chemicals could be eliminated from the group.
3. **Click** "Remove".

Data Gap Filling

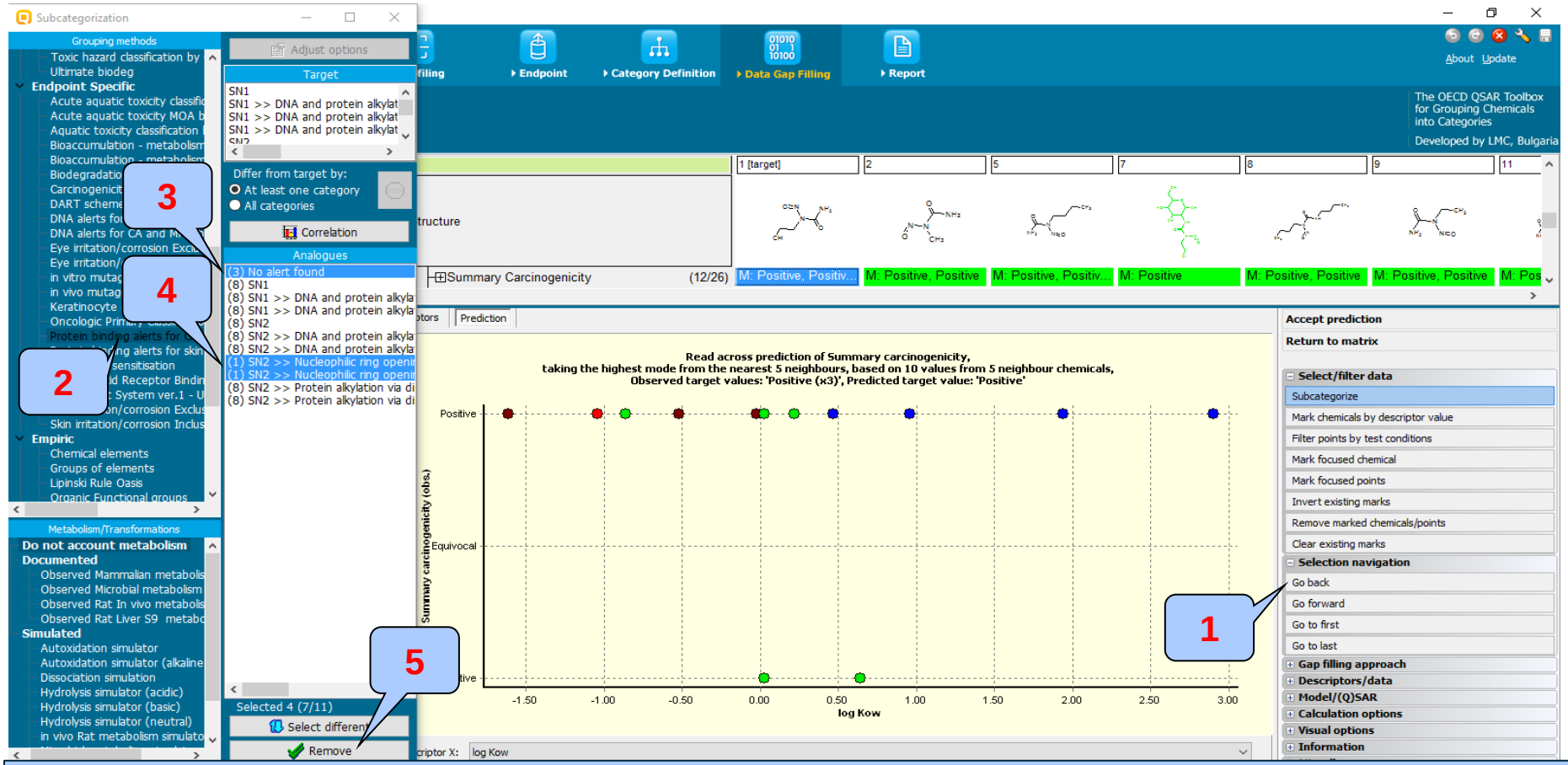
Oncologic primary classification – subcategorization 2



1. All the analogue are positive and similar with respect to **"Oncologic primary classification".**

Data Gap Filling

Protein binding alerts for chromosomal aberration by OASIS v1.2 – (alternative subcategorization)



This slide illustrates an alternative to the "Oncologic" subcategorization.

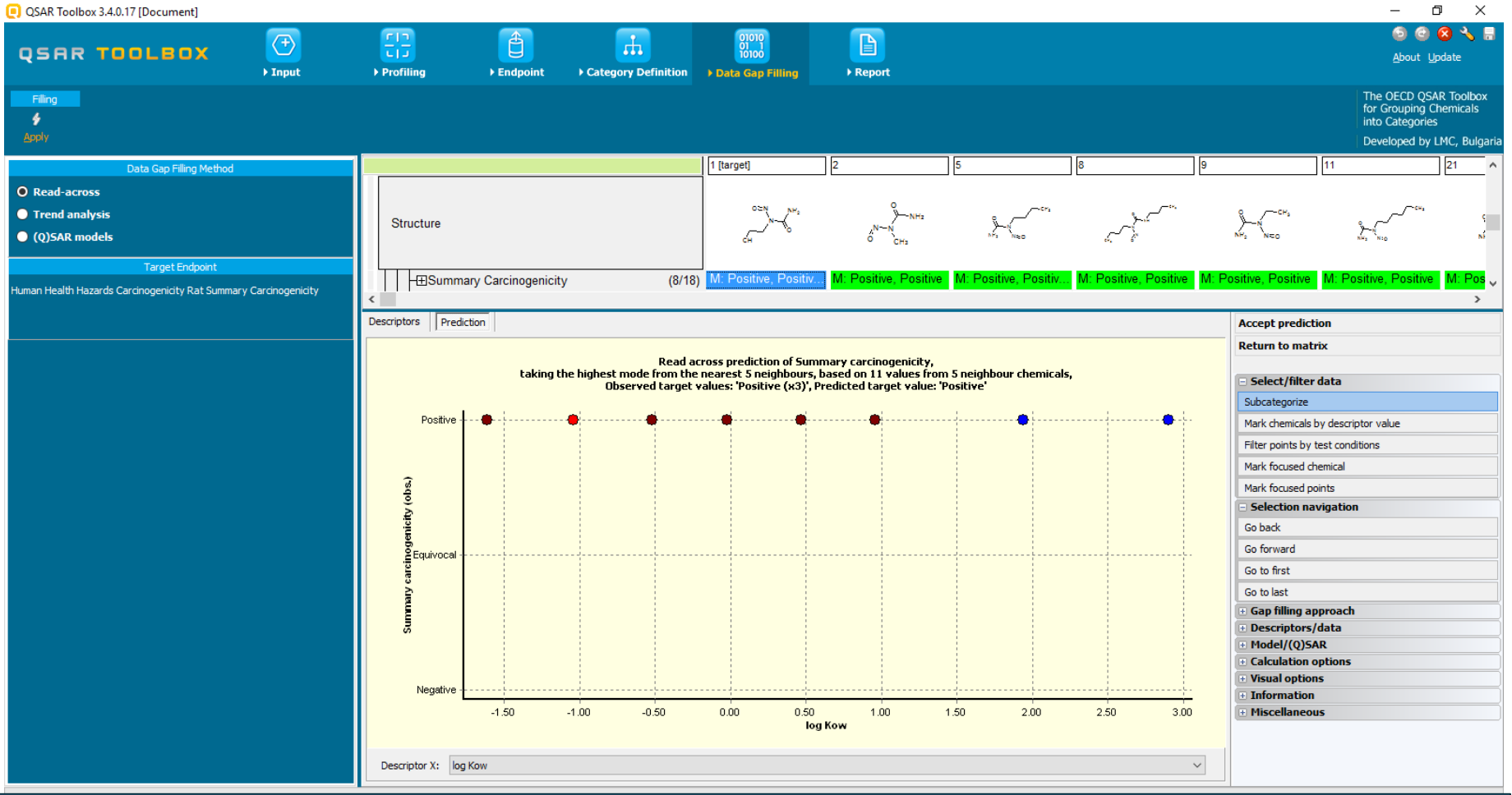
1. **Open** "Select navigation" panel and **click** "Go back" 2. **Select** "Protein binding alerts for CA by OASIS v1.2"

3. As seen 3* out of all analogues do not have "protein binding alert". They could be eliminated from the graph.

4. There is one additional chemical which has Protein binding mechanism different than the target chemical. 5. **Click** "Remove" to eliminate dissimilar chemicals

*3 – four dots are highlighted in green, because for one of the chemicals there is positive and negative experimental data.

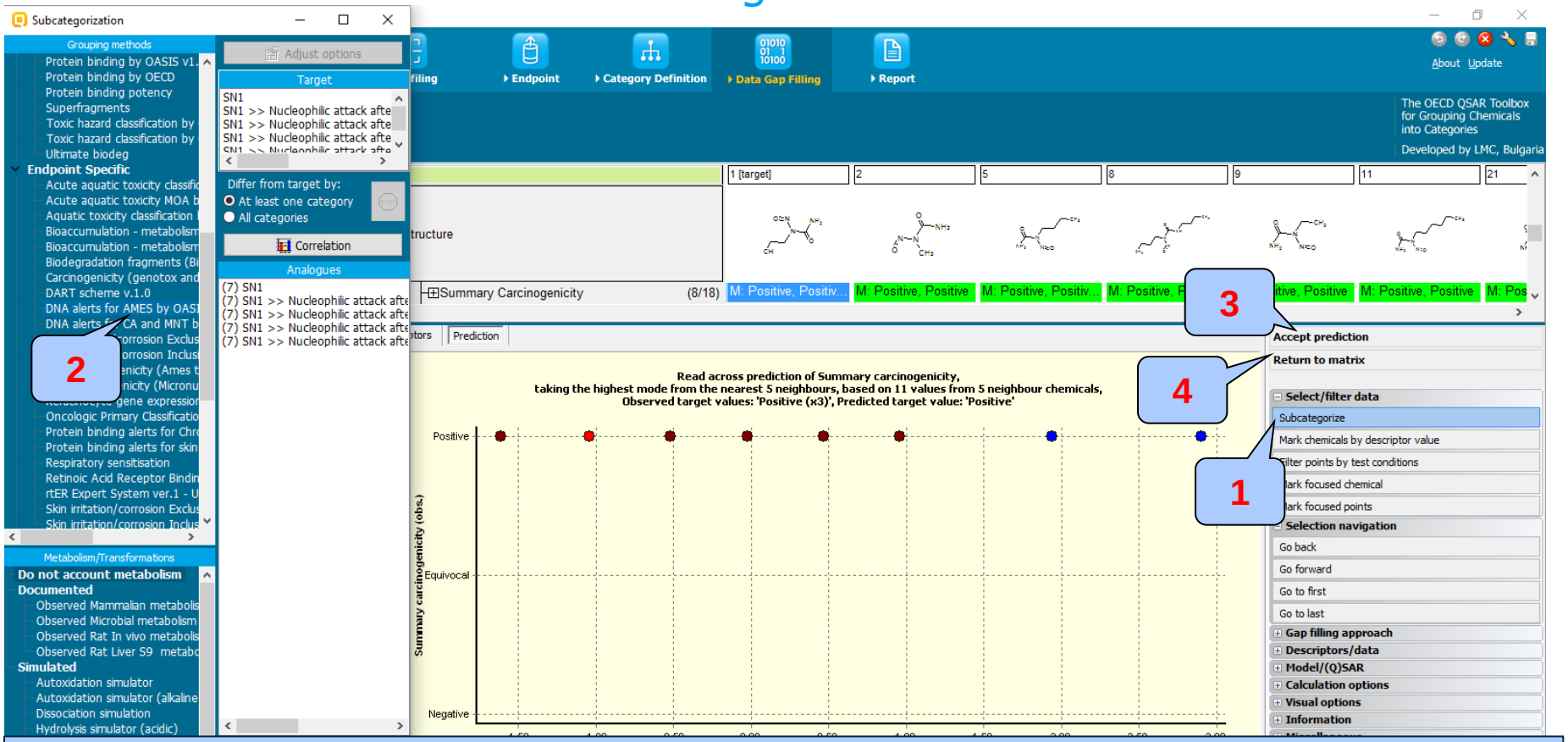
Data Gap Filling
Results after subcategorization 2



Same prediction is obtained after subcategorization by "Protein binding" profiler.

Data Gap Filling

DNA binding alerts for Ames, MN, CA by OASIS v1.3 – subcategorization 3



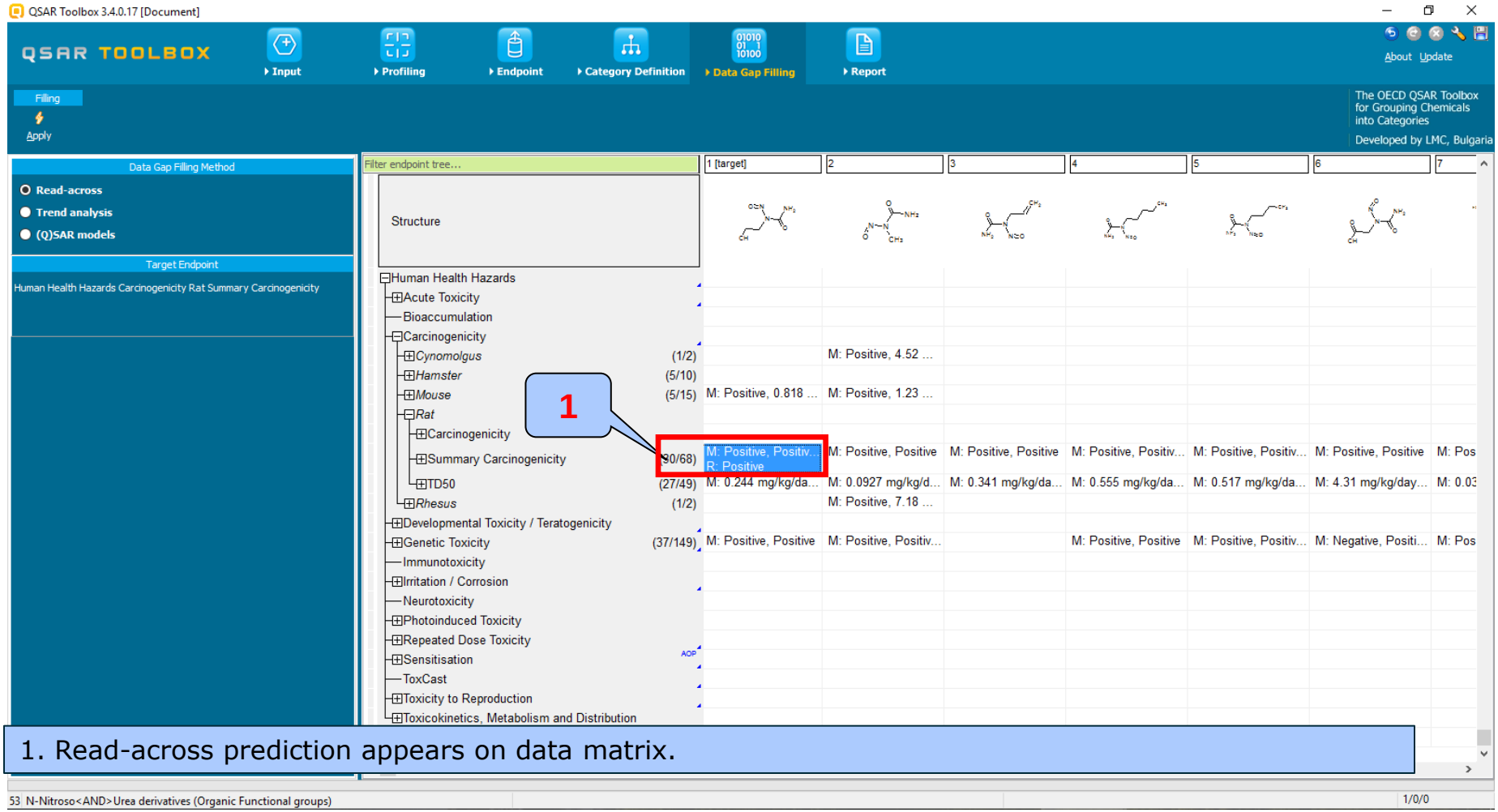
Next subcategorization, is by "DNA alerts for Ames, by OASIS v1.4" because as mentioned before "Carcinogenicity effect" is conditioned by DNA and Protein interactions. Follow the steps

1. **Click** "Subcategorize" 2. **Select** "DNA alerts for Ames, MN…".

As seen all the analogues have same DNA mechanism (SN1, Nucleophilic attack, N-nitroso compounds) as the target chemical. So the prediction is reliable enough, in domain and could be accepted.

3. **Click** "Accept prediction" 4. **Click** "Return to matrix"

Data Gap Filling
Read-across results



1. Read-across prediction appears on data matrix.

Recap

- This example illustrates the read-across prediction for carcinogenicity of chemical **(1-(2-hydroxyethyl)-1-nitrosourea)** when filter by Ames experimental data is applied to the initial group of analogues.
- All the analogues were identified based on Organic functional group classification. Two out of four most important functional groups were used for searching analogues: "N-Nitroso" and "Urea derivatives".
- Data gap filling was based on read-across approach due to "qualitative" character of experimental data (positive, negative).
- Read-across analysis was applied for analogues with "Summary carcinogenicity" data associated with rat species.
- A set of 3 subcategorizations were applied for refining the initial group of analogues starting with filtering all the analogues by Ames experimental data.
- Refinement of category continues with two mechanism-based subcategorizations:
  - Oncologic Primary classification (Protein binding alerts for Chromosomal aberration by OASIS v1.2 as an alternative subcategorization)
  - DNA alerts for Ames by OASIS v1.4
- Results of subcategorization shows positive read-across prediction for the target based on positive Ames mutagenicity analogues supported by positive DNA and Protein binding alerts.
- Hence one could conclude that the target chemical may elicit carcinogenicity effect in rats.

Outlook

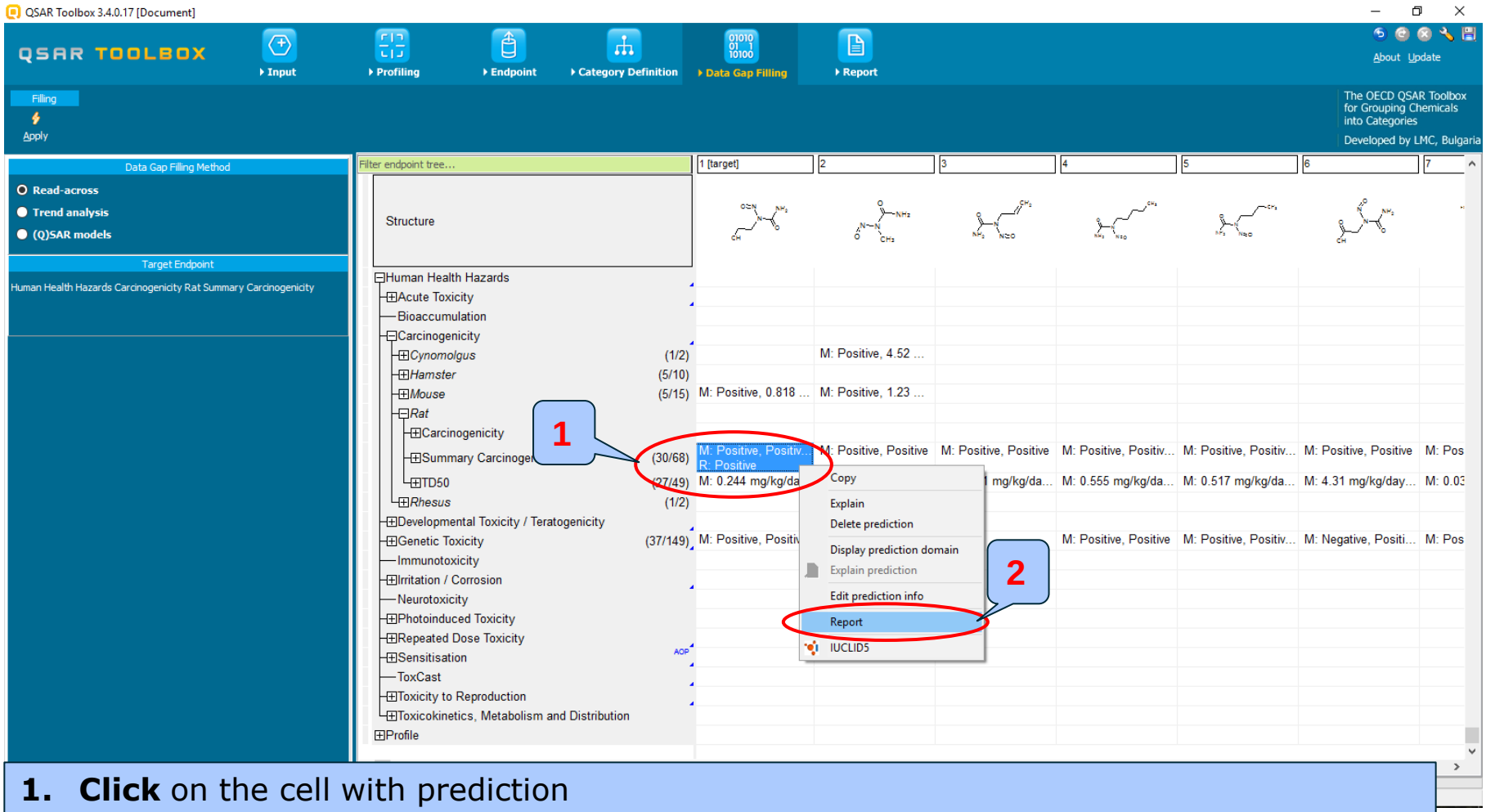
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Category definition
  - Data gap filling
  - **Reporting**
  - Save the prediction

Report

- The report module allows generation of report on the predictions obtained with the Toolbox. This 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.

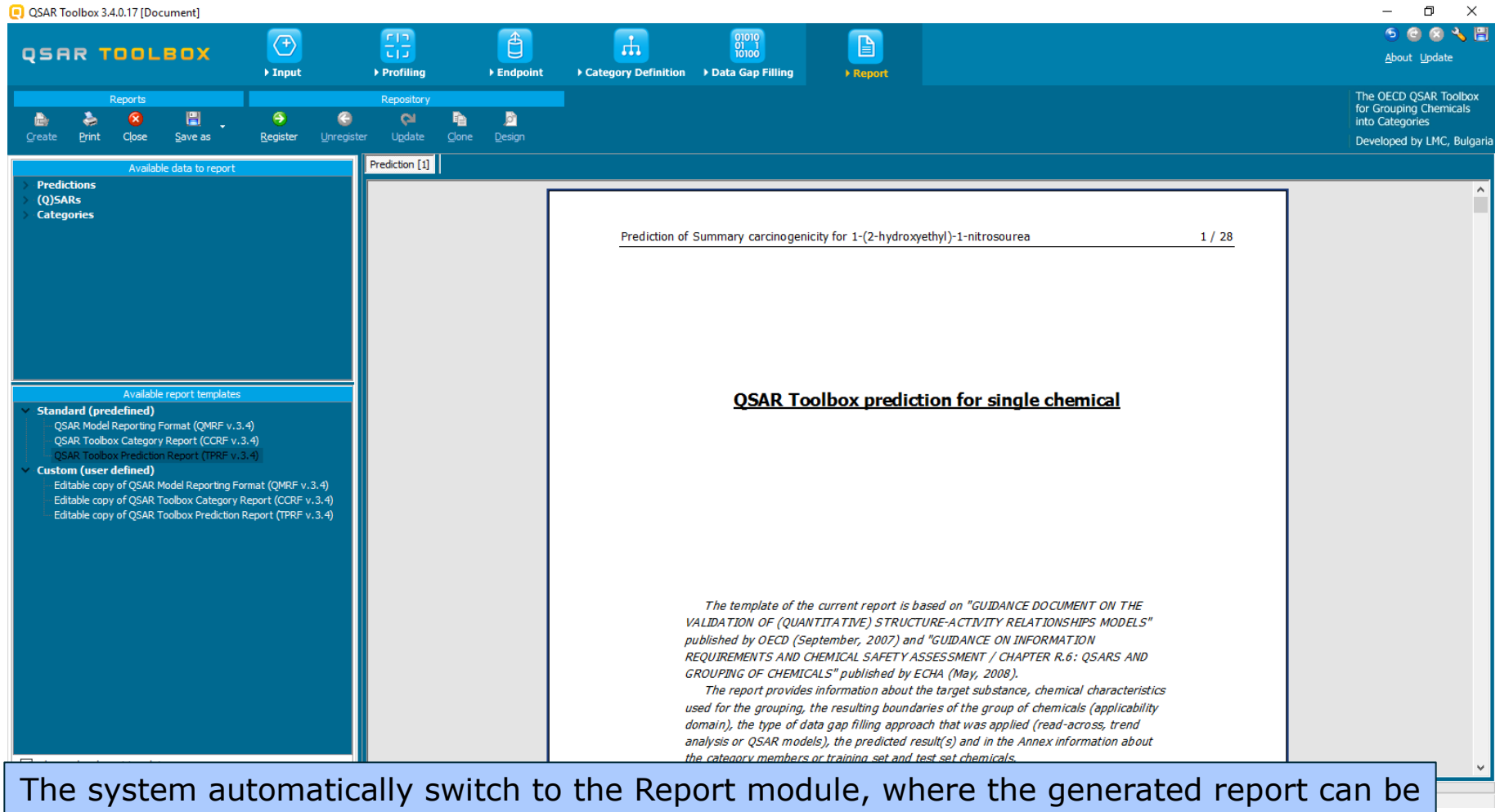
- Generating the report is shown on next screenshots.

Report



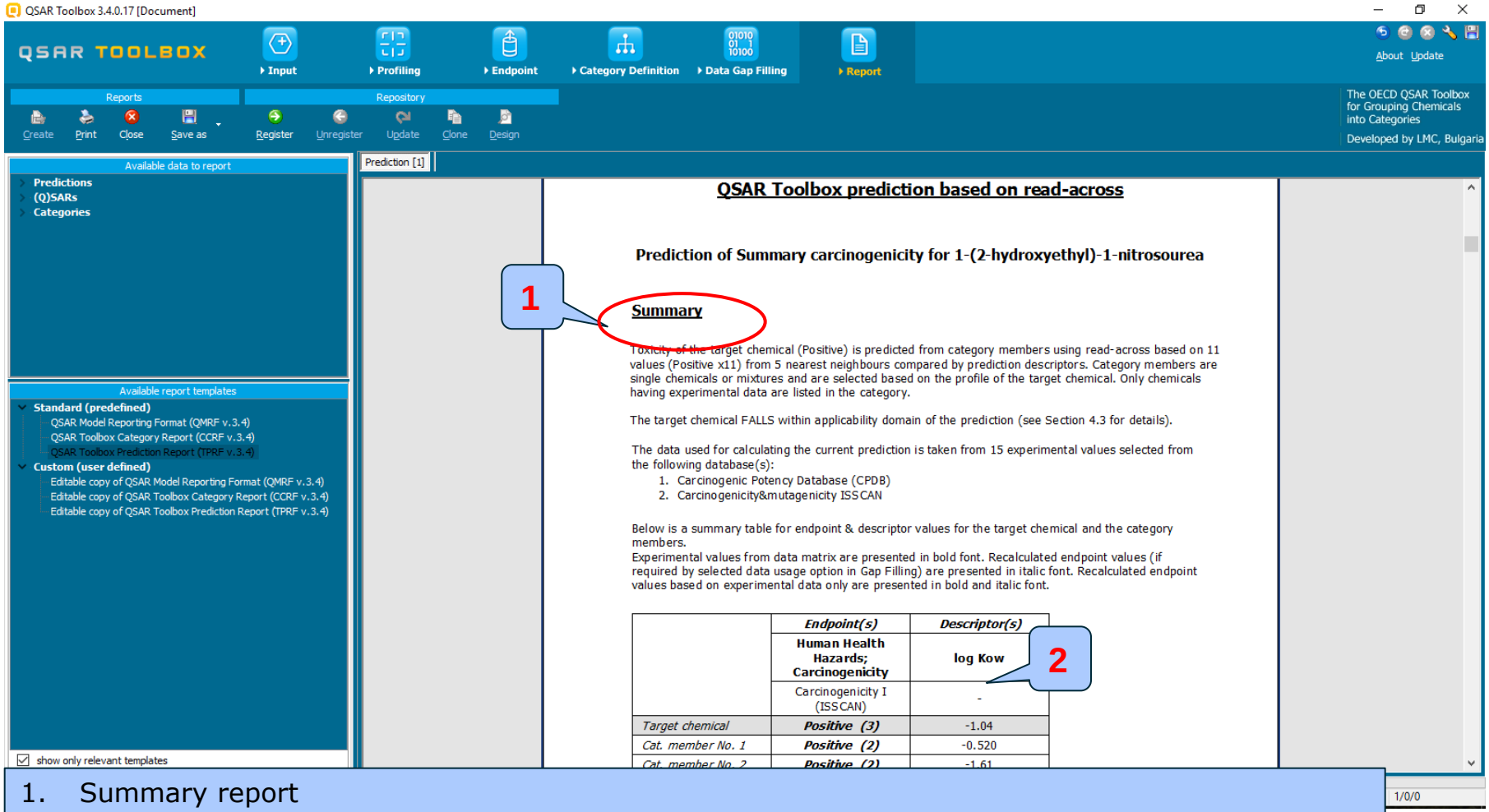
1. **Click** on the cell with prediction
2. Perform **Right click** and **Select** Report

Report



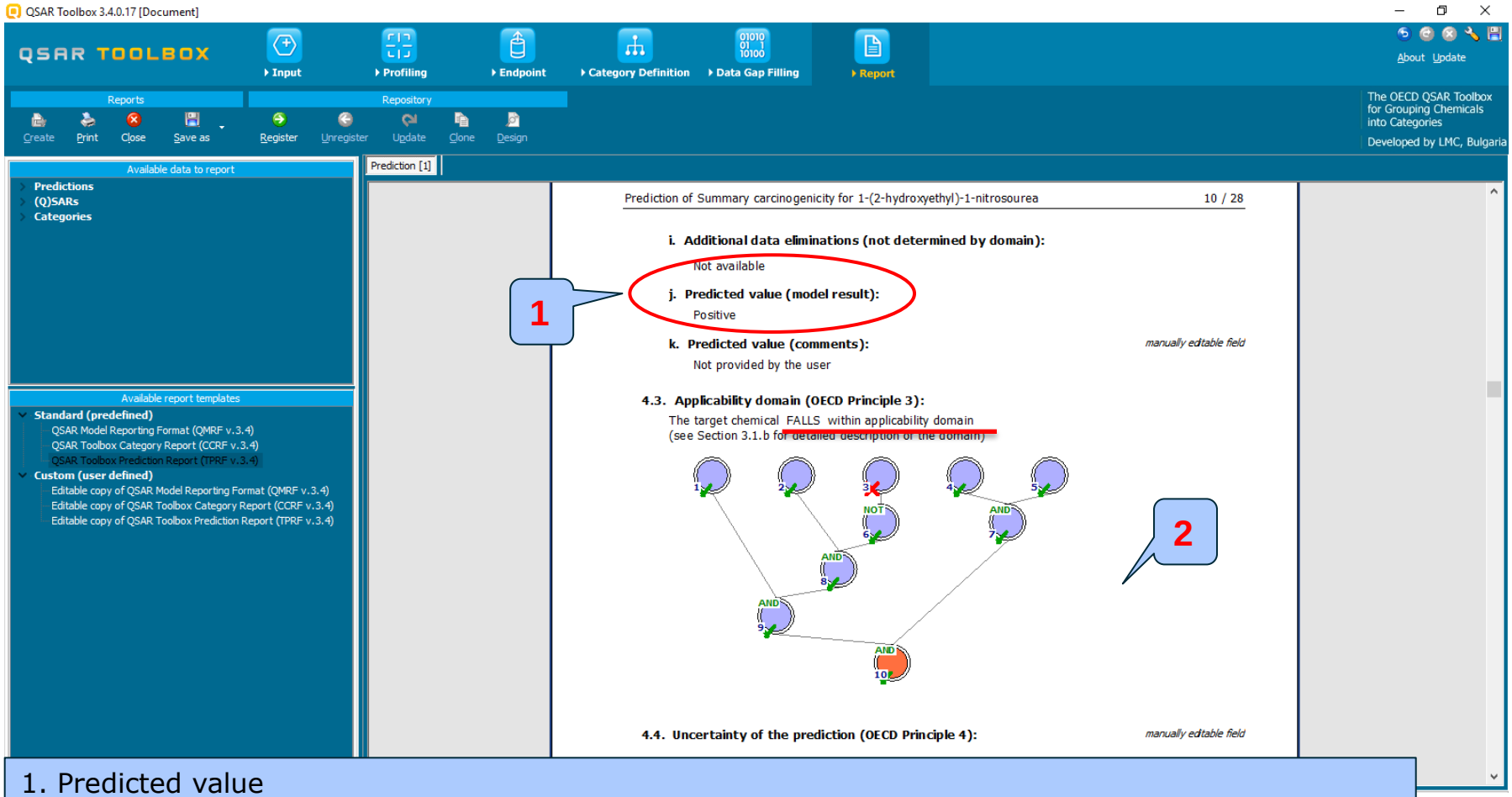
The system automatically switch to the Report module, where the generated report can be seen.

Report



1. Summary report
2. Table with for endpoint and descriptors of the target and chemicals within the category

Report



1. Predicted value
2. Applicability domain

The target chemical is "In domain".

Outlook

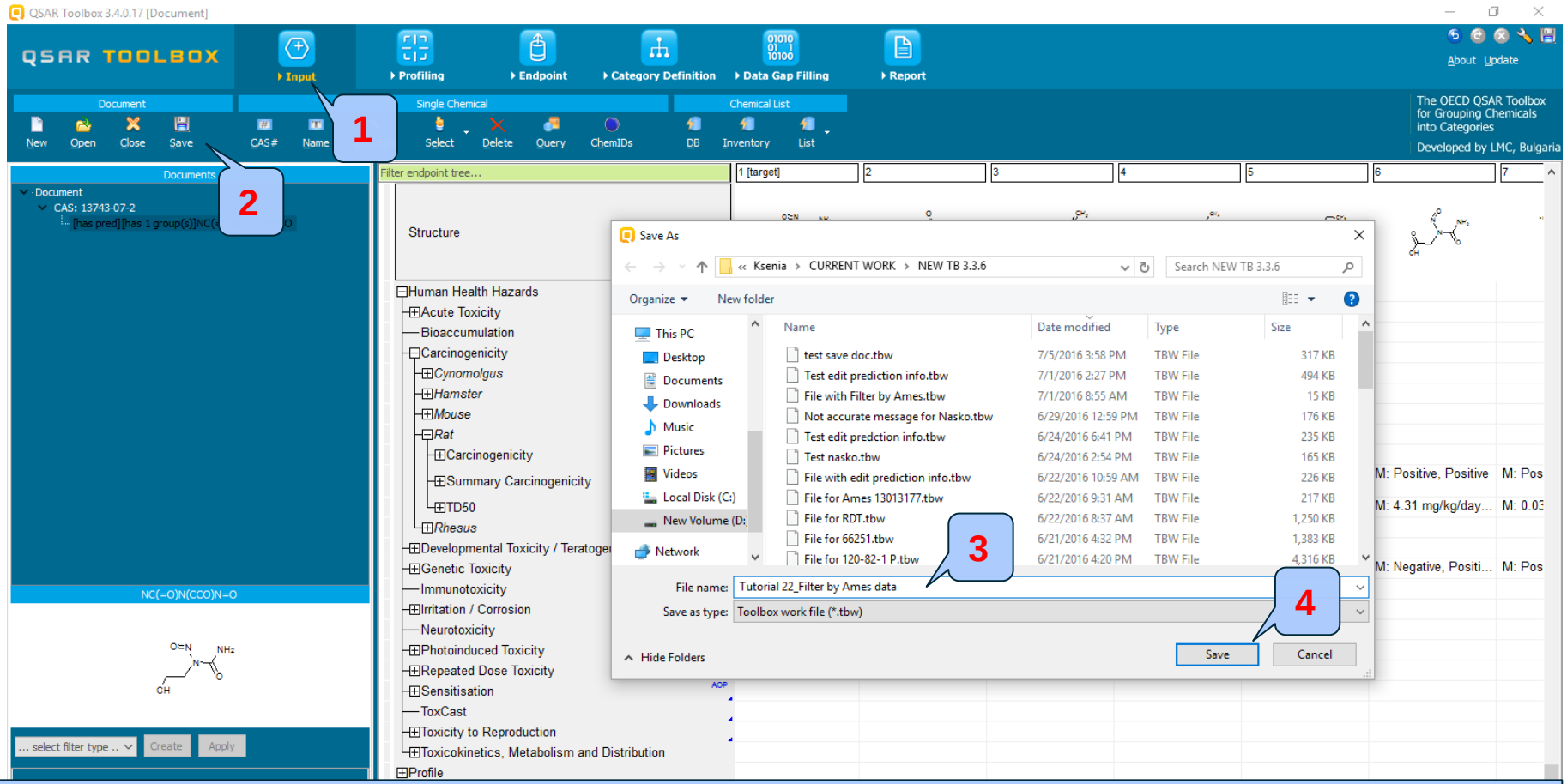
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Category definition
  - Data gap filling
  - Reporting
  - **Save the prediction**

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).

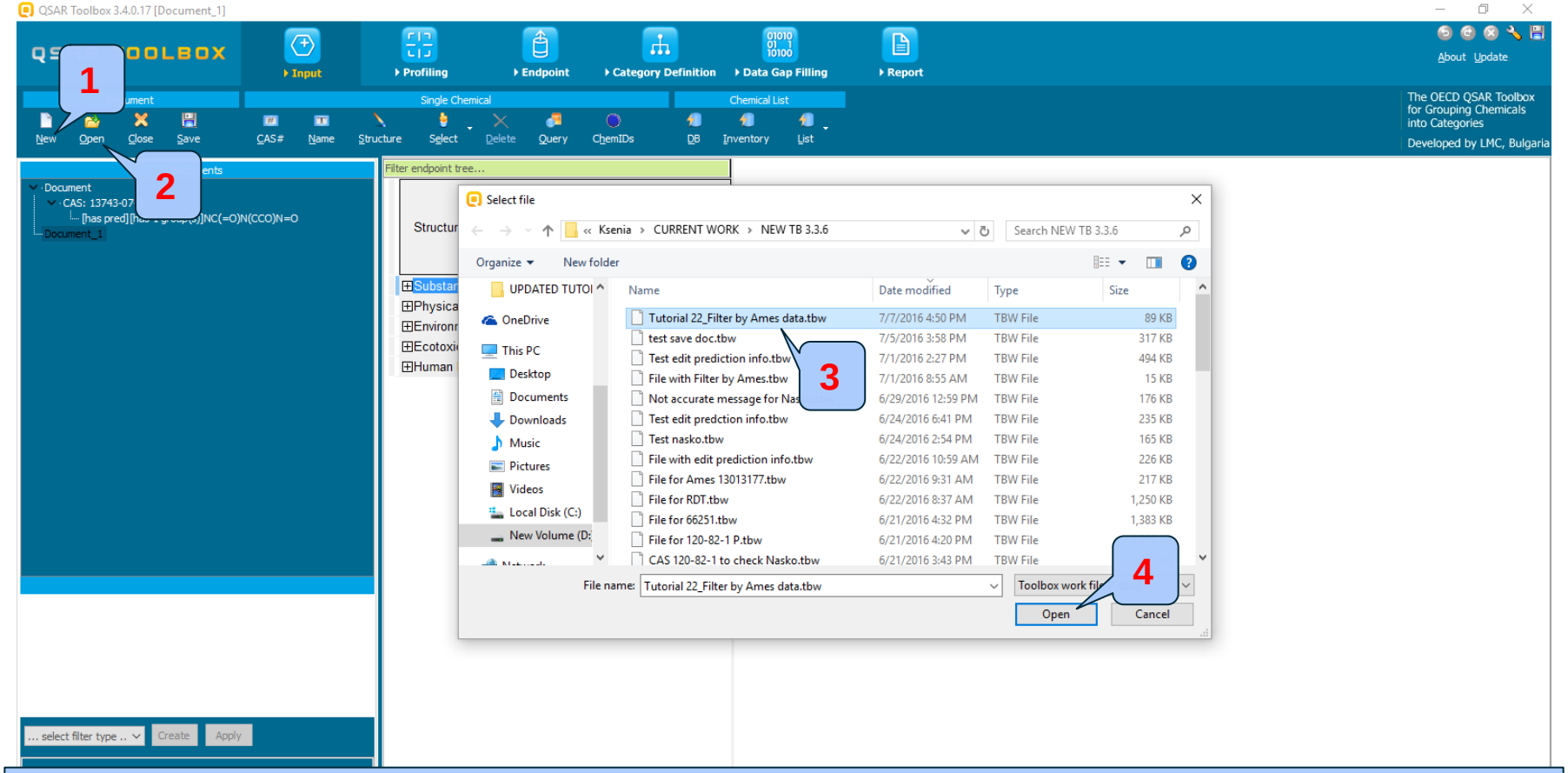
- Saving/Loading the file with TB prediction is shown on next screenshots.

Saving the prediction result



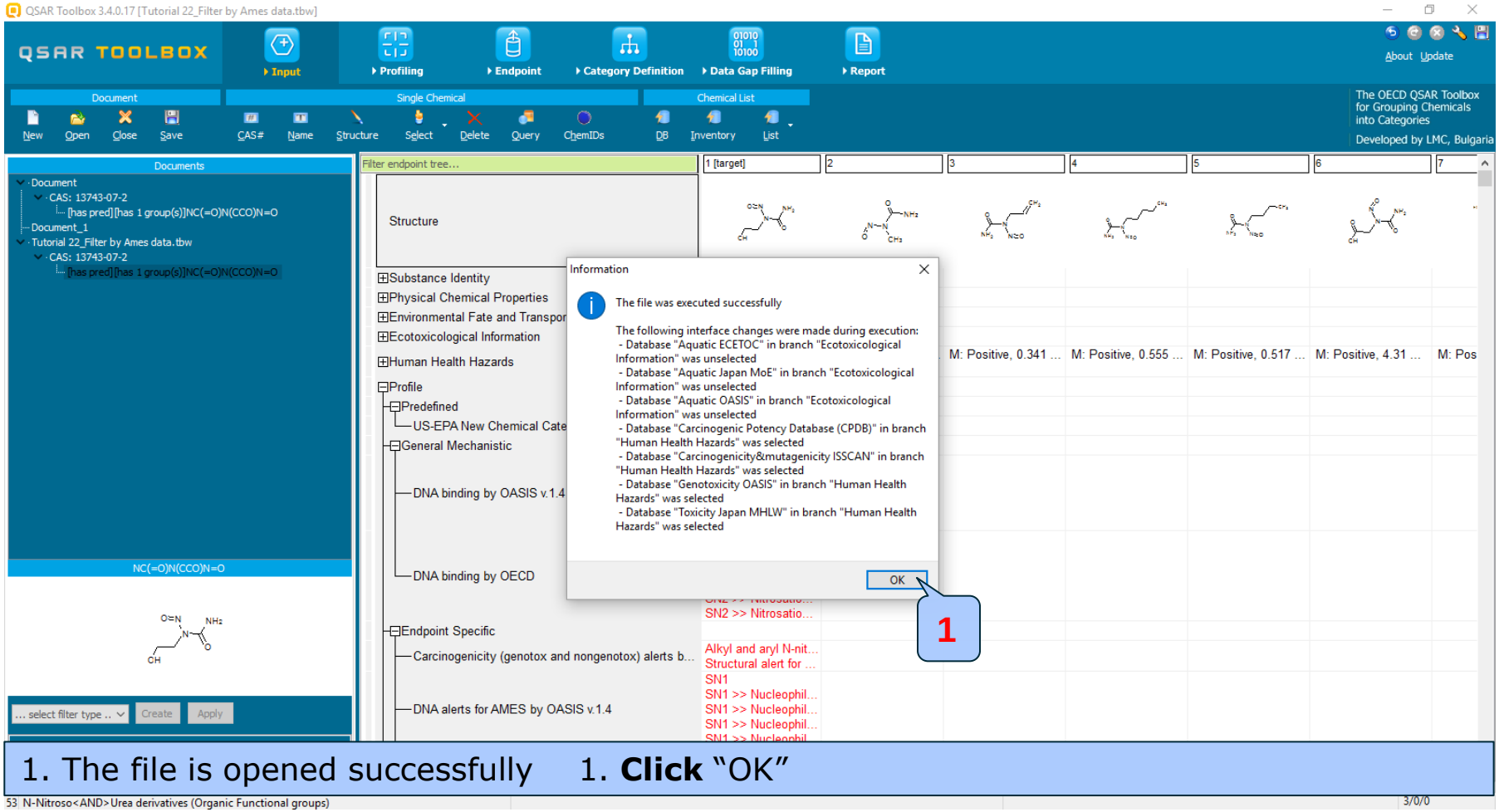
1. **Go** to "Input" section 2.**Click** on "Save" button 3. **Define** name of the file; 4. **Click** "Save" button

Open saved file



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

Open saved file



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