QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.1

Example for predicting skin sensitisation potential of (*2E,6Z*)-2,6-nonadien-1-ol accounting for skin metabolism

Outlook

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

Background

- This is a step-by-step presentation designed to take the user through the Toolbox workflow for filling data gap for skin sensitization of trans-2,cis-6-nonadienol accounting for its skin metabolism

Outlook

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

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Simulating skin metabolism of target chemical
- Identify analogues for a selected active metabolite
- Filling data gaps for active metabolites by read across
- Assign prediction of metabolite to the parent chemical

Outlook

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

The Exercise

- In this exercise we will predict the skin sensitization potential for target chemical (**trans-2,cis-6-nonadienol**) [**CAS # 28069-72-9**].
- Profile the target chemical and identify no Protein binding alert for target chemical.
- Gather available experimental data for the target chemical and identify positive experimental data.
- Skin metabolism of target chemical will be accounted for.
- Read across prediction for active metabolite will be applied.
- The predicted result of metabolite will be assigned to the target chemical.

Outlook

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

Workflow

- **As you know the Toolbox has 6 modules which are typically used in sequence:**
  - Chemical Input
  - Profiling
  - Endpoint
  - Category Definition
  - Data Gap Filling
  - Report
- **In this example we will use the modules in a different order, tailored to the aims of the example.**

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 Chemicals

User Alternatives for input of Chemical:

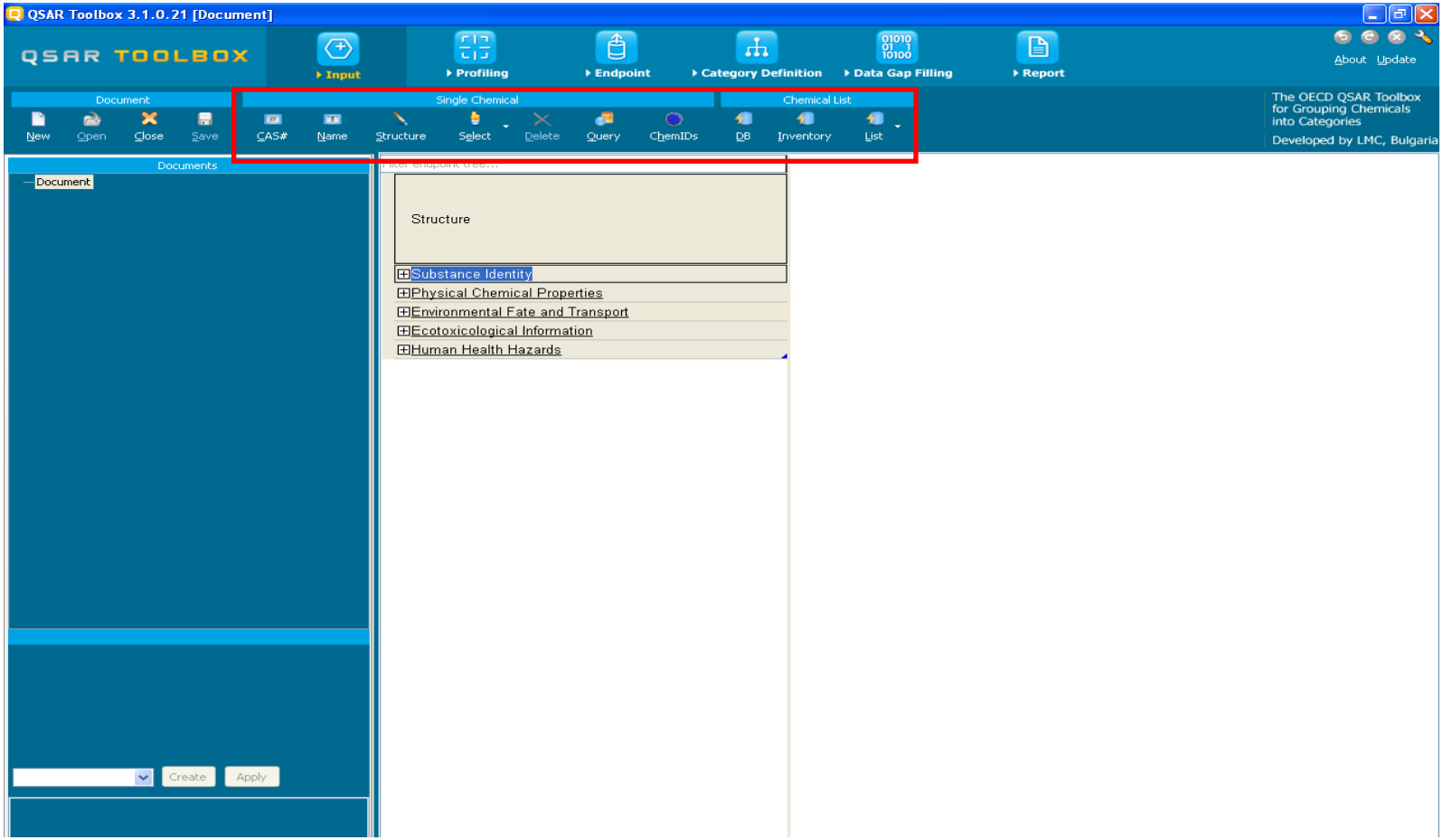
A. Single target chemical
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases
- Chemical IDs such as EC number, ENECS number

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

Chemical Input
Input Screen

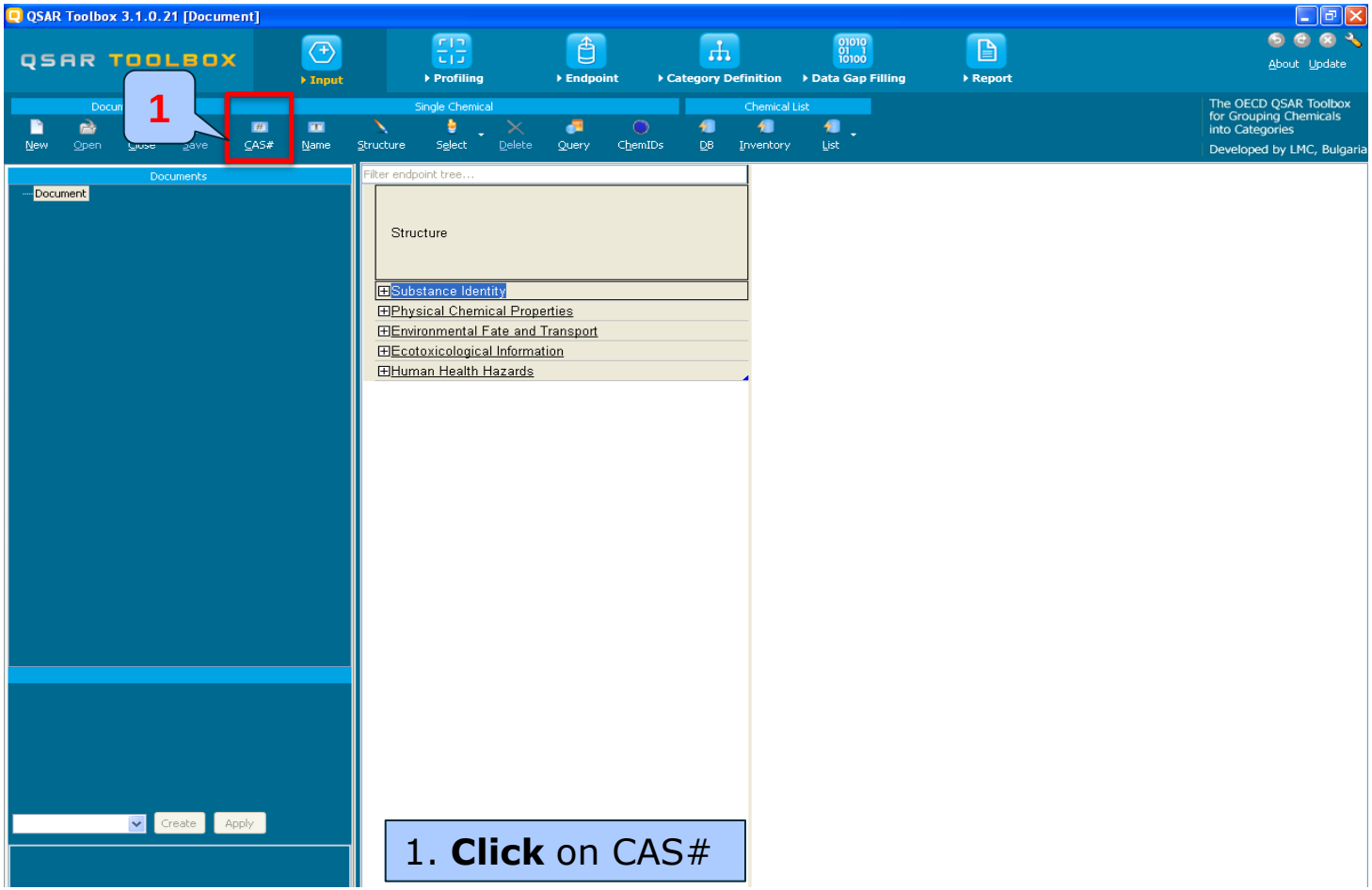
- Open the Toolbox.
- The six modules in the workflow are seen listed next to "QSAR TOOLBOX" title.
- Click on "Input" (see next screen shot)

Chemical Input
Input Screen



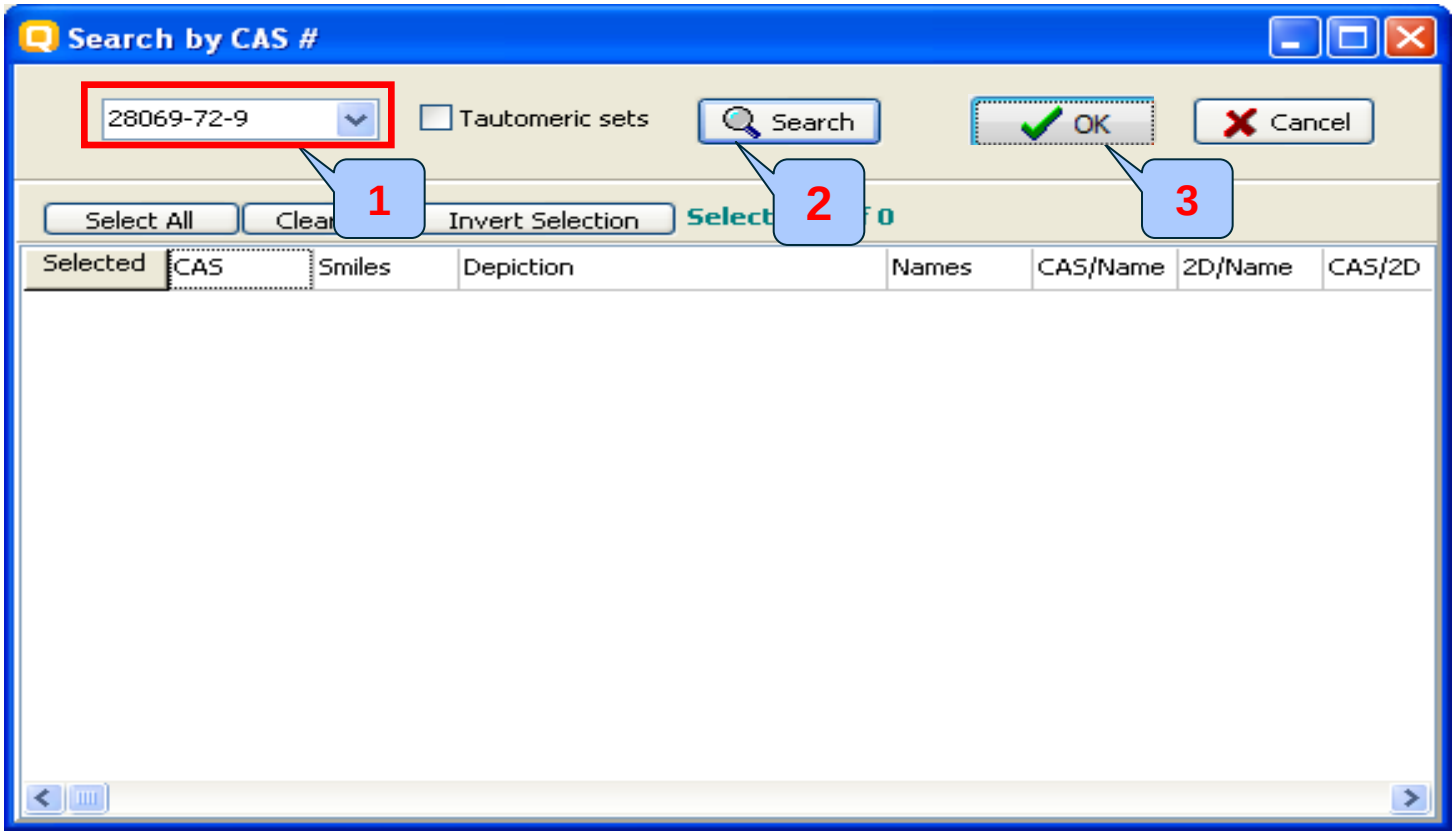
Chemical Input

Input target chemical by CAS#



Chemical Input

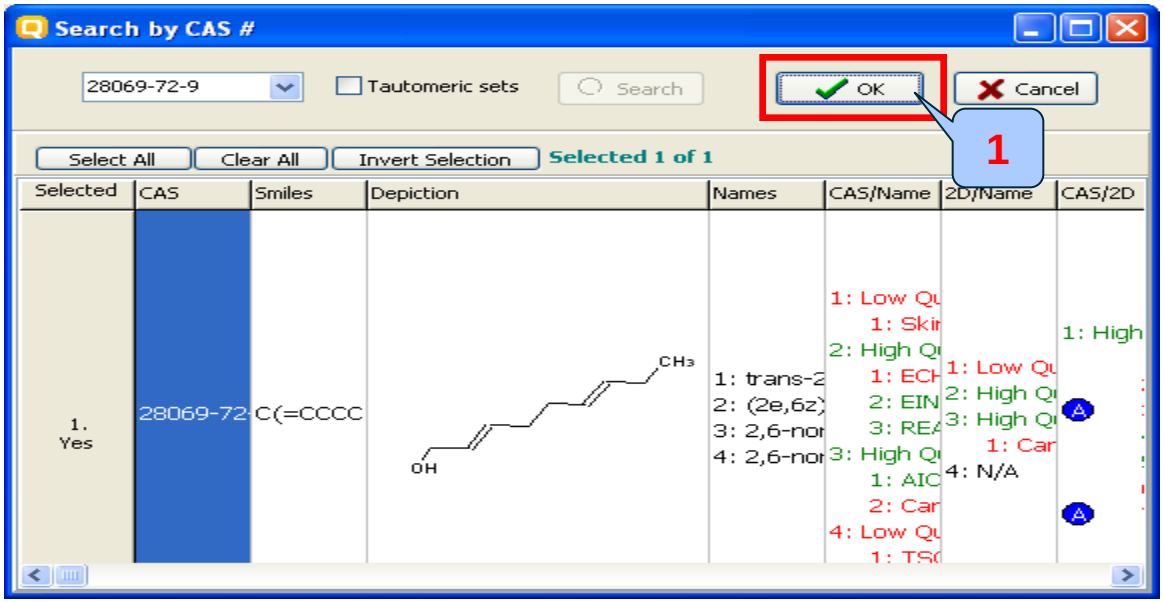
Enter CAS# of trans-2,cis-6-nonadienol



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.

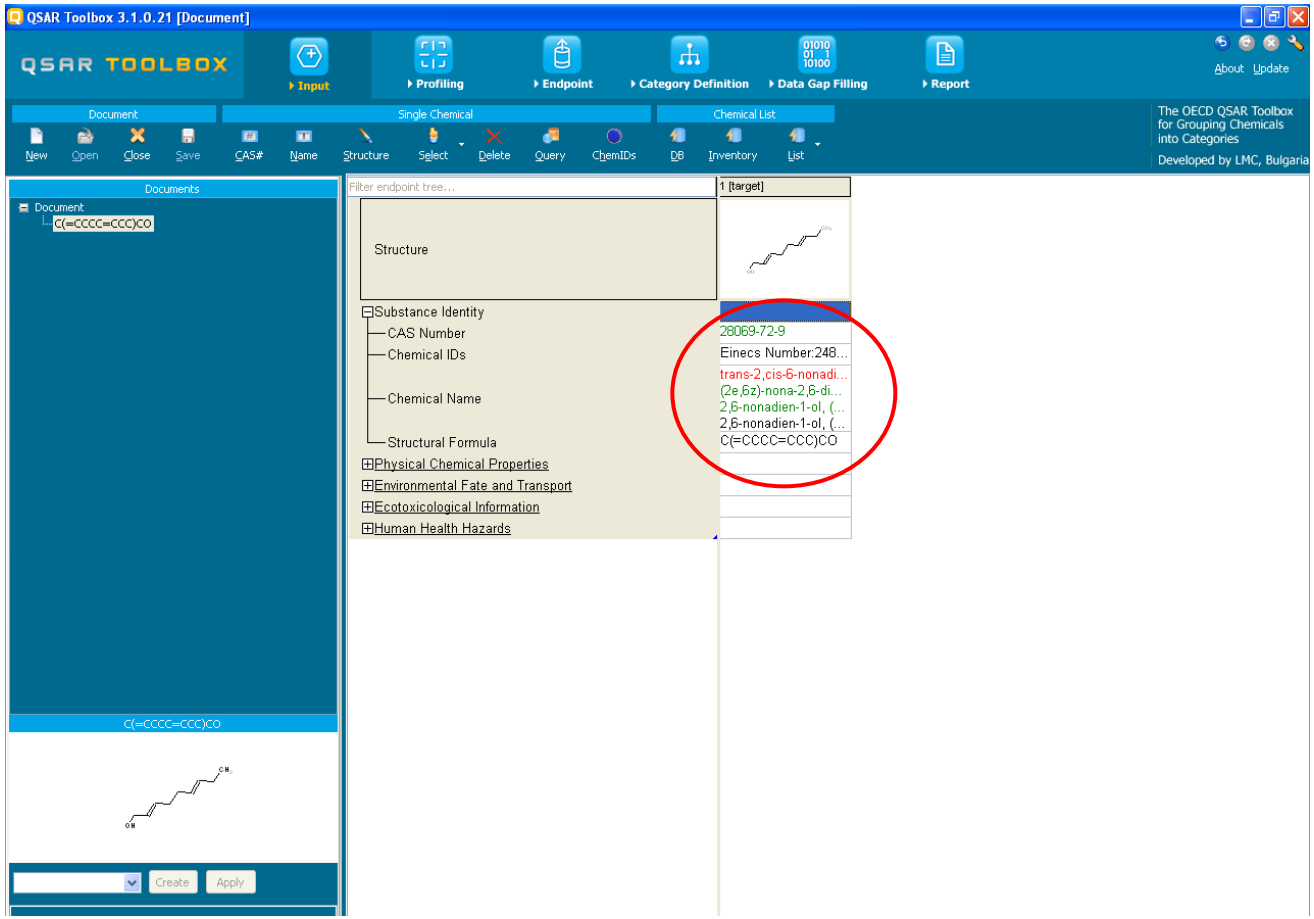


1. **Click** OK to enter the target structure into data matrix

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 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



Chemical Input
Target 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**

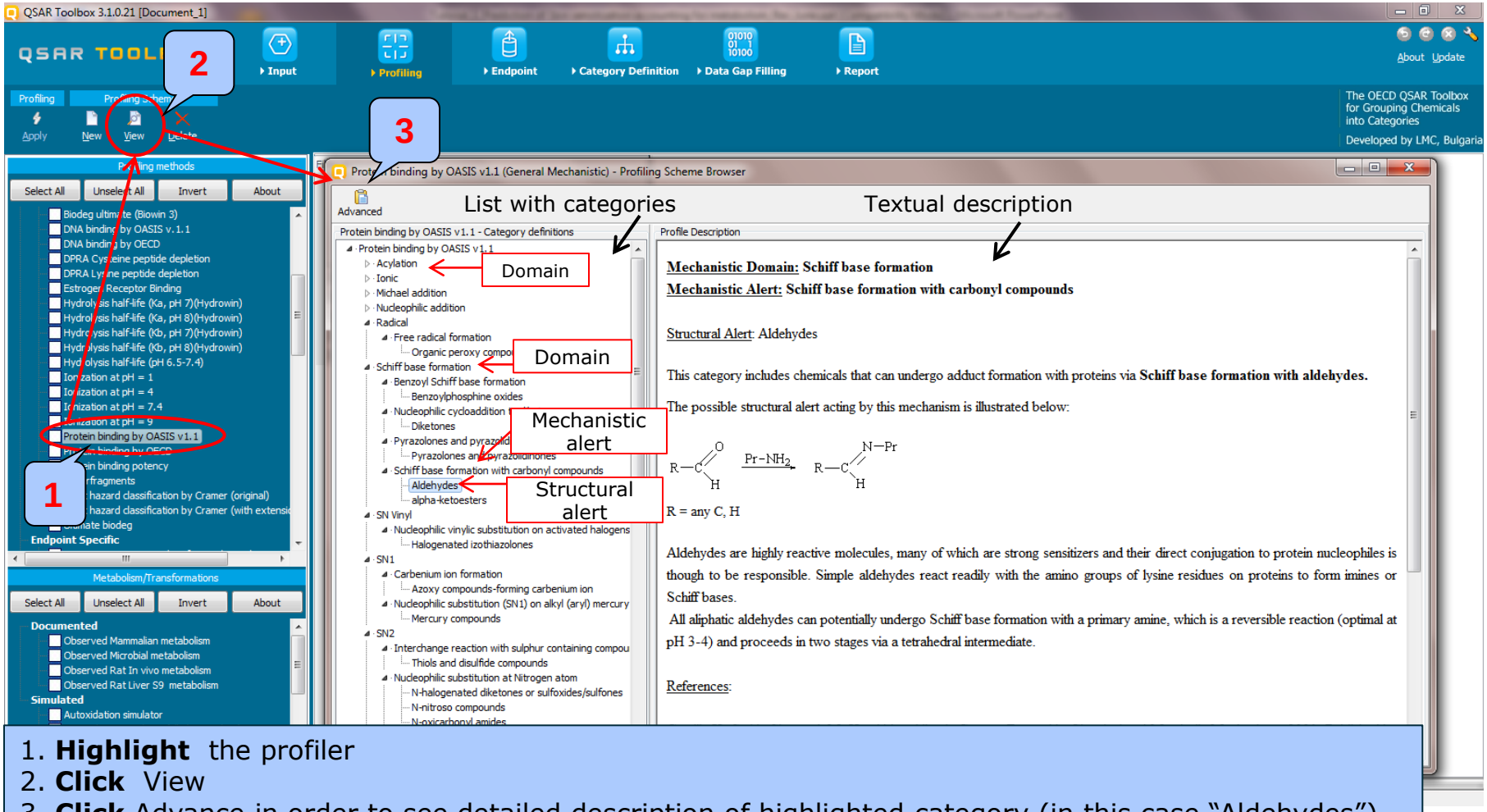
Profiling
Overview

- "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

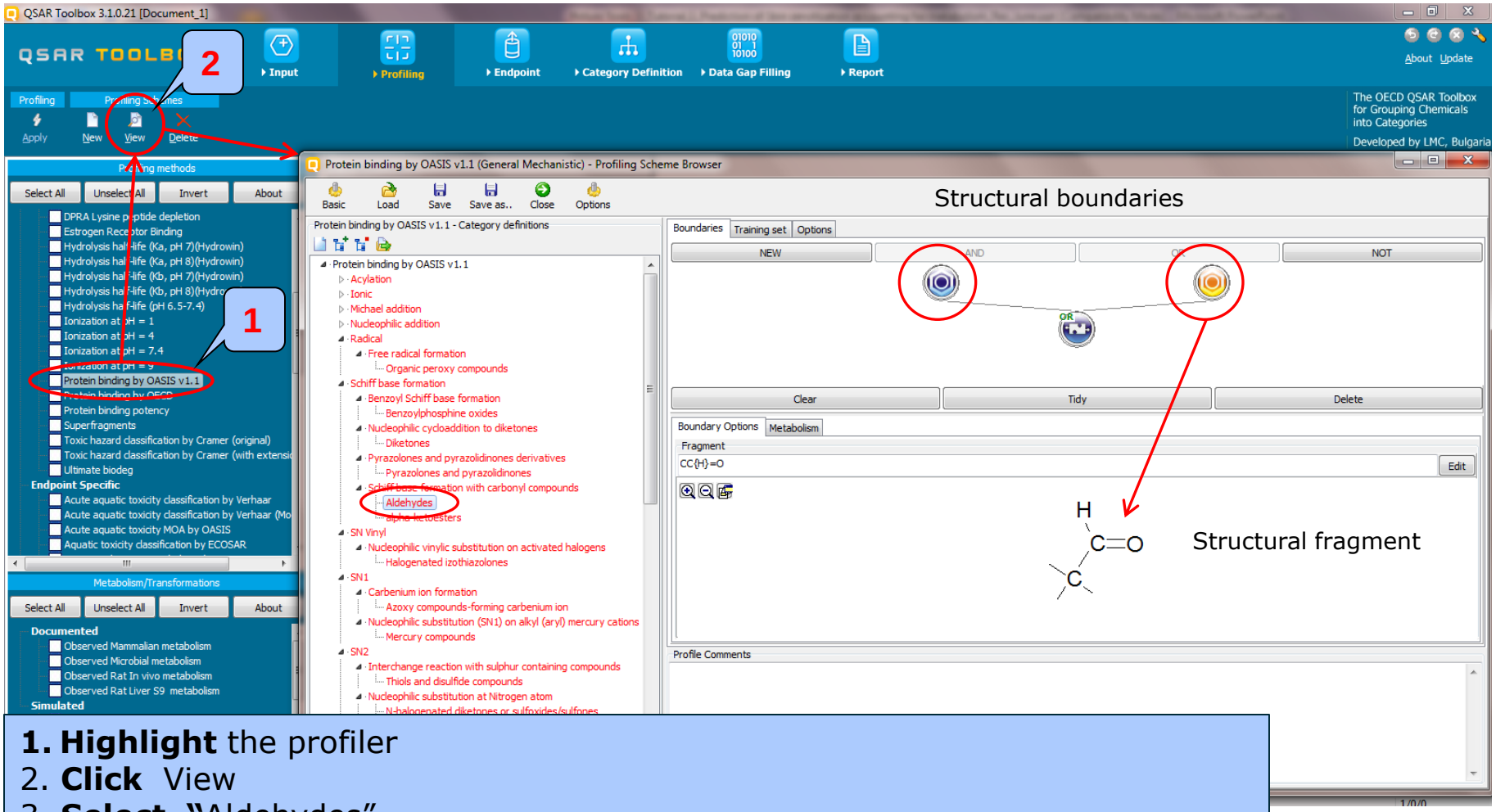
- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, Protein binding by OASIS v1.1 and clicking on "View" (see next screen shot).

Profiling
Side-Bar to Profiling



1. **Highlight** the profiler
2. **Click** View
3. **Click** Advance in order to see detailed description of highlighted category (in this case "Aldehydes")

Profiling
Side-Bar to Profiling



1. **Highlight** the profiler
2. **Click** View
3. **Select** "Aldehydes"

Profiling
Profiling the target chemical

- The outcome of the profiling determines the most appropriate way to search for analogues (detailed information about profilers could be found in "Manual for Getting started" (Chapter 4) published on the OECD website:

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

- Table 4 - 1 in chapter 4 (Manual for getting started) lists a selection of profilers and their relevance for different endpoints of regulatory relevance

- The following profiling schemes are relevant to the **Skin sensitization**:
  - Protein binding by OASIS v.1.1 – general mechanistic
  - Protein binding by OECD – general mechanistic
  - Protein Binding Potency – general mechanistic
  - Protein binding alerts for skin sensitization by OASIS v1.1 – endpoint specific

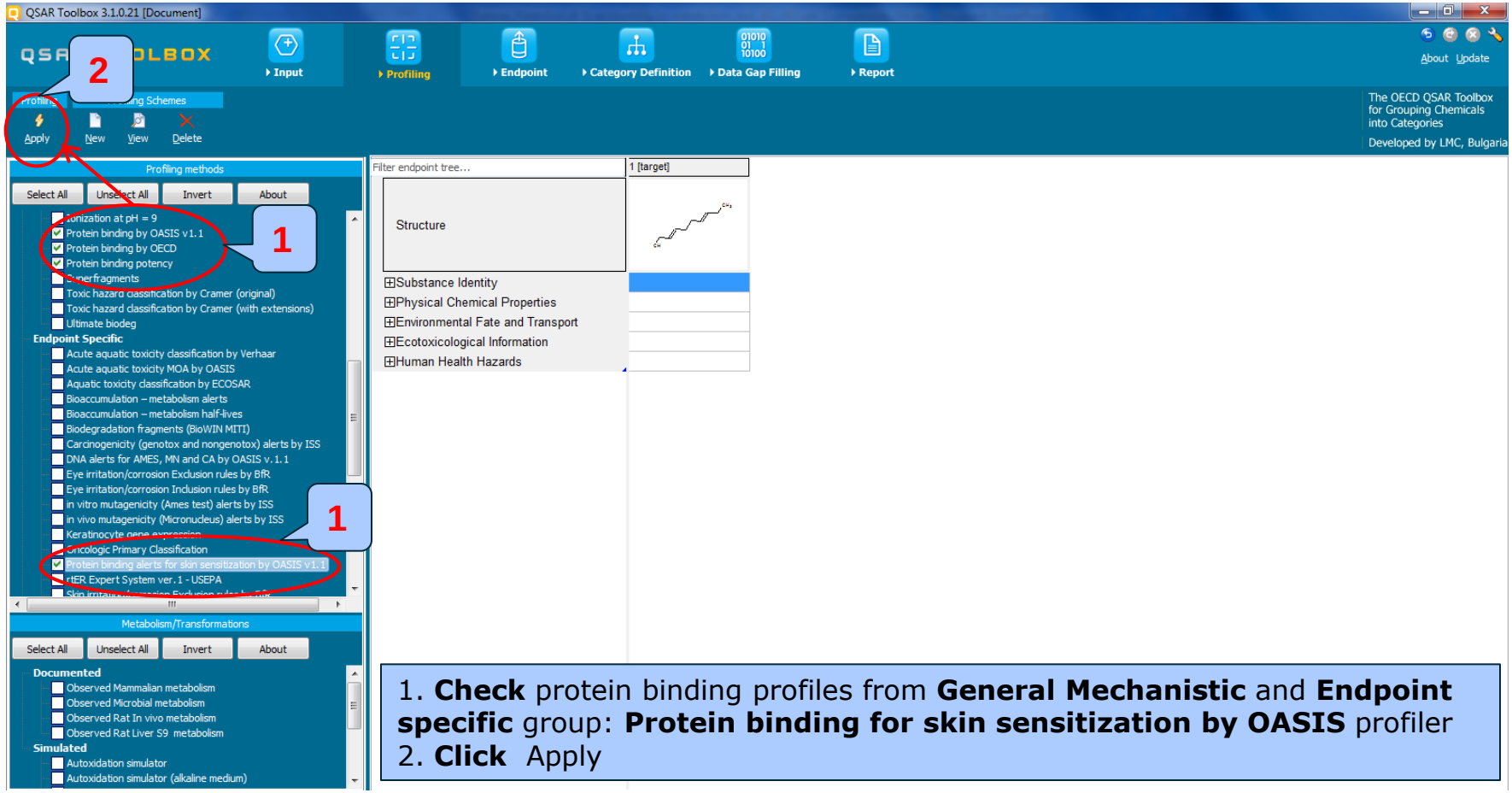
Profiling
Profiling the target chemical

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

- For this example, go through the general and endpoint specific profiling mechanisms and highlight those that apply to skin sensitization(see next screen shot).

Profiling

Profiling the target chemical



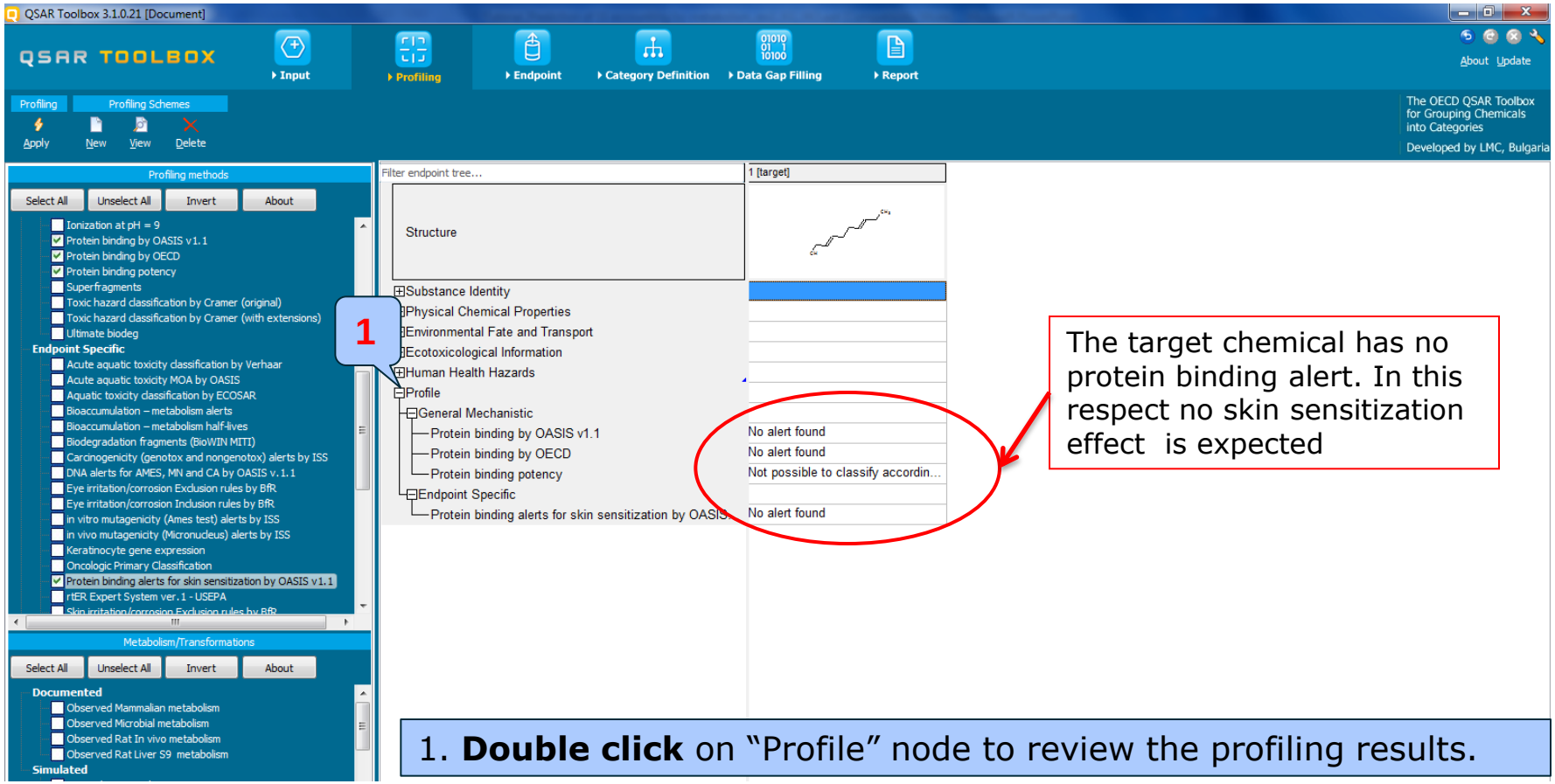
1. **Check** protein binding profiles from **General Mechanistic** and **Endpoint specific** group: **Protein binding for skin sensitization by OASIS** profiler
2. **Click** Apply

Profiling
Profiling the target chemical

- The actual profiling will take up to several seconds depending on the number and type of profilers selected
- The results of profiling automatically appear as a dropdown box under the target chemical (see next screen shot)
- Please note the specific protein-binding profilers
- No protein binding alert has been found for the test compound (trans-2,cis-6-nonadienol).

Profiling

Profiling the target chemical



Outlook

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

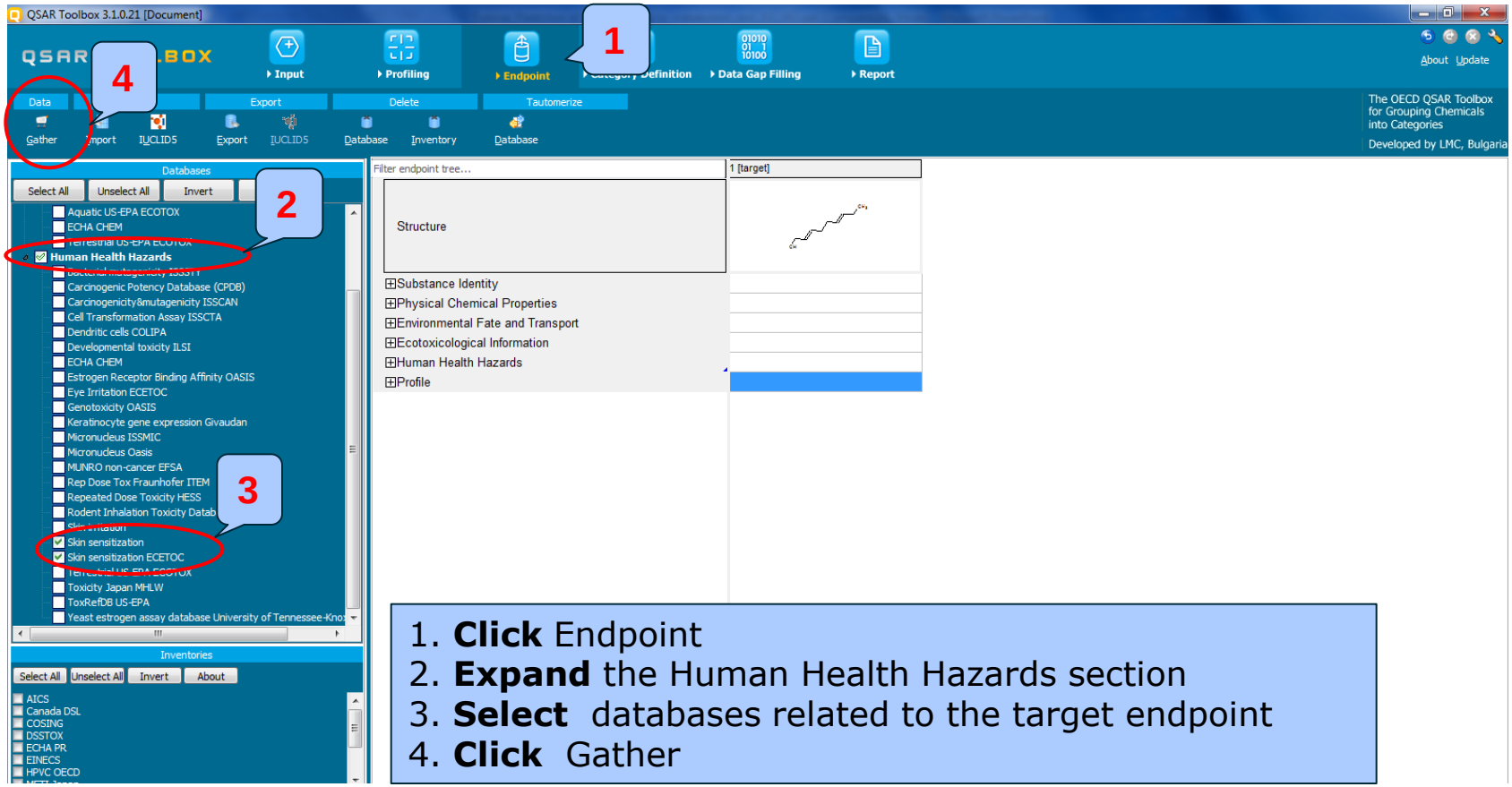
Endpoint
Overview

- "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 a single toxicity endpoint (skin sensitization).
- In this example, we collect data from the databases containing experimental results for Skin sensitisation (Skin sensitisation and Skin sensitisation ECETOC).
- Click on "Endpoint" in the Toolbox workflow.
- Expand the "Human Health Hazards" section
- Click on the box to select the relevant databases.
- Click on "Gather data" (see next screen shot).

Endpoint
Gather data



1. **Click** Endpoint
2. **Expand** the Human Health Hazards section
3. **Select** databases related to the target endpoint
4. **Click** Gather

Endpoint
Gather data

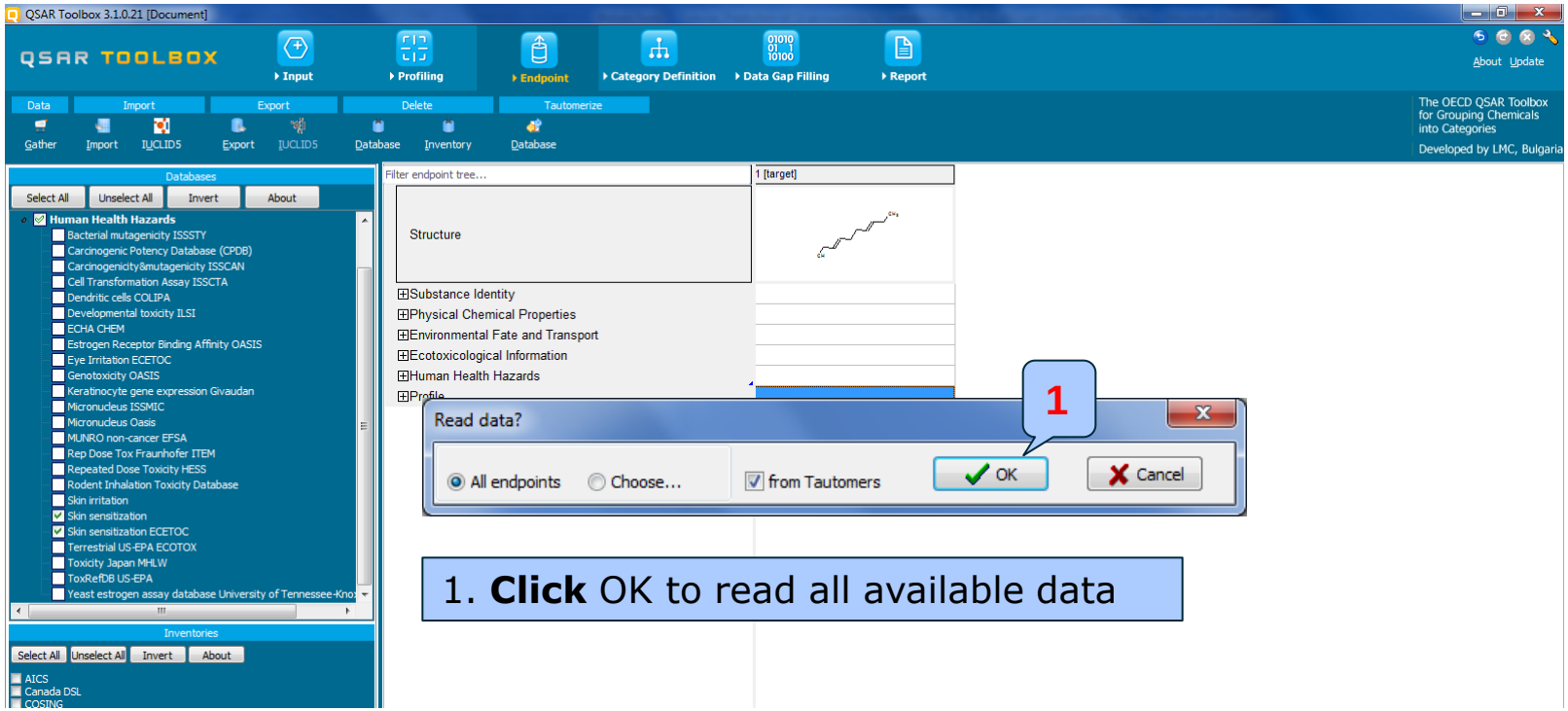
- Toxicity information on the target chemical is electronically collected from the selected dataset(s)

- It should be kept in mind that the search for data and analogues is performed only among the chemicals which are listed in the selected databases, which in this example are **Skin sensitization** and **Skin sensitization ECETOC**

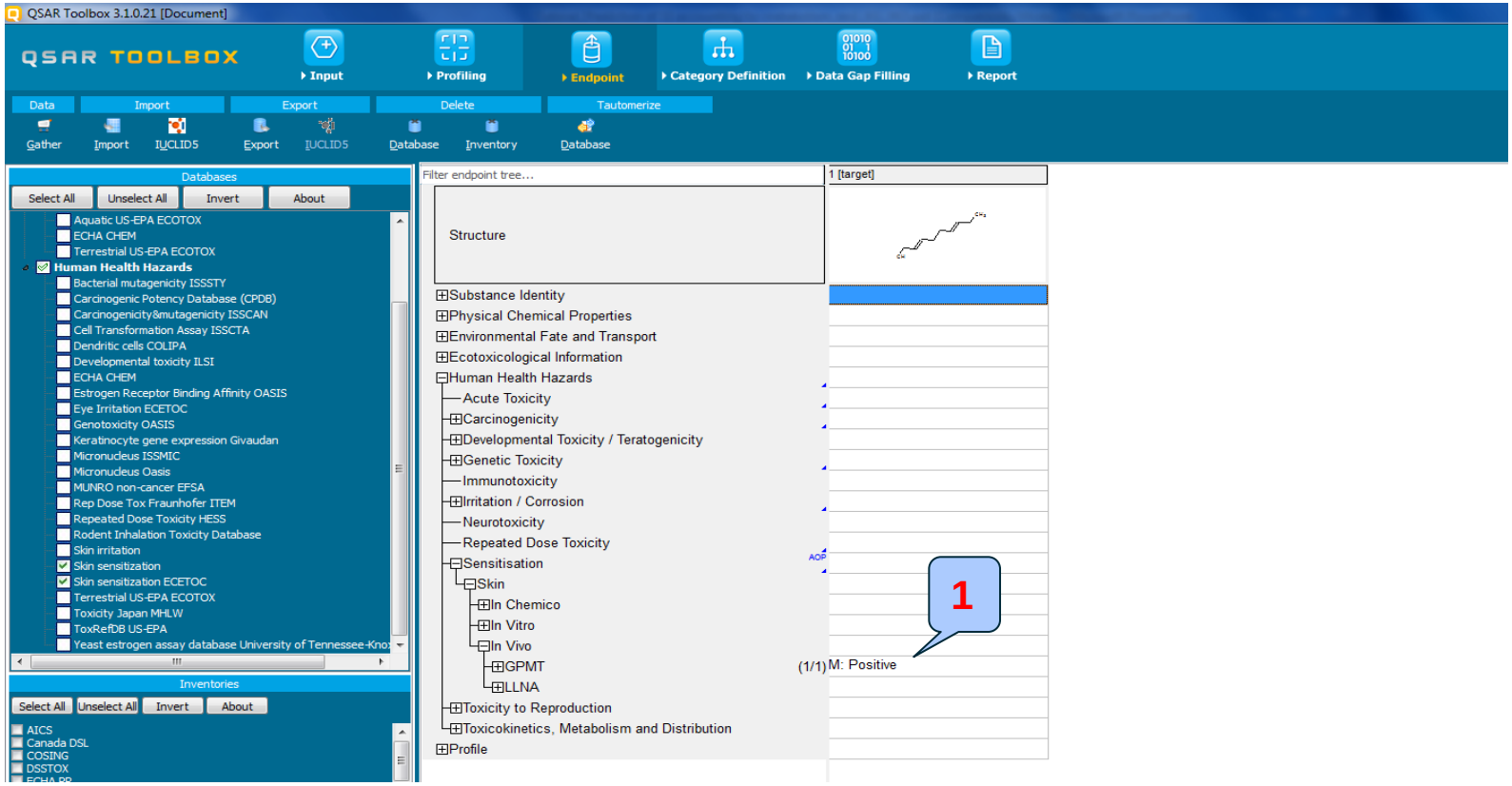
- In this example, there is Positive experimental data for the target chemical(see next screen shots)

Endpoint
Gather data

Toxicity information on the target chemical is electronically collected from the selected datasets. A window with "Read data?" appears. Now the user could choose to collect "all" or "endpoint specific" data.

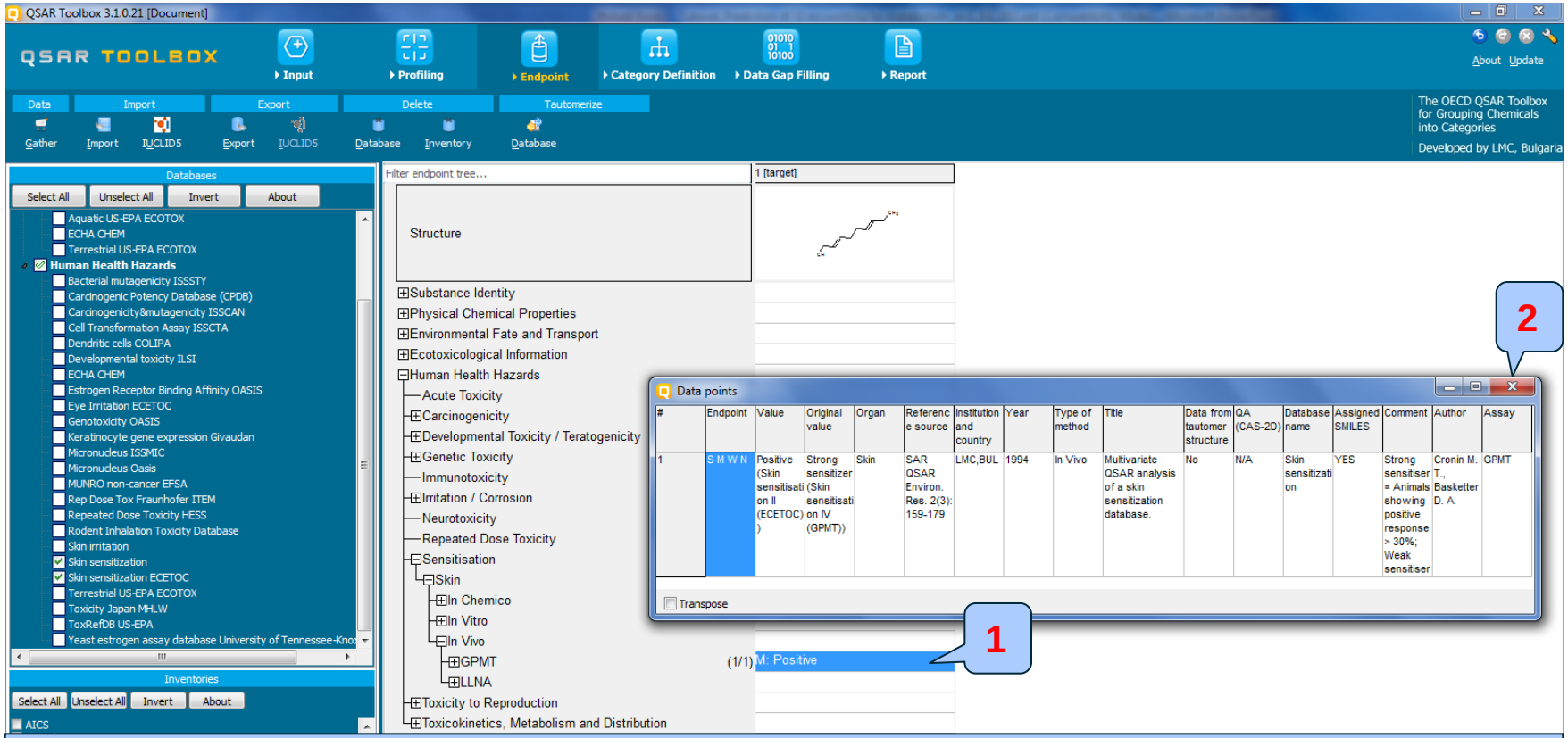


Endpoint
Gather data



1. Available experimental data appears on datamatrix.

Endpoint
Gather data



1. **Double-click** on the cell displays metadata information for the observed data
2. **Click** on the X to close the window

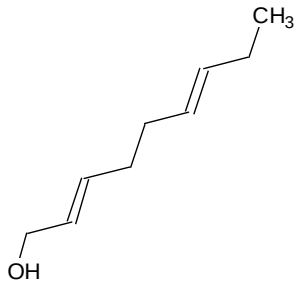
Recap

- The first module, which introduces the target chemical, ensure correctness of the structure
- The second module shows that there is no protein binding alert for target chemical
- In the third module, you have found that the target chemical has positive skin sensitization data
- The positive experimental data could be due to skin metabolism
- The study continues with accounting for skin metabolism of target chemical(see next slides).

Recap

Parent

trans-2,cis-6-nonadienol
CAS# 28069-72-9



Protein binding alert
(Protein binding by OASIS)

No alert found

Profiling Result

No alert found (no skin sensitization effect expected)

GPMT
Strong sensitizer
(experimental data)

Skin metabolism

Outlook

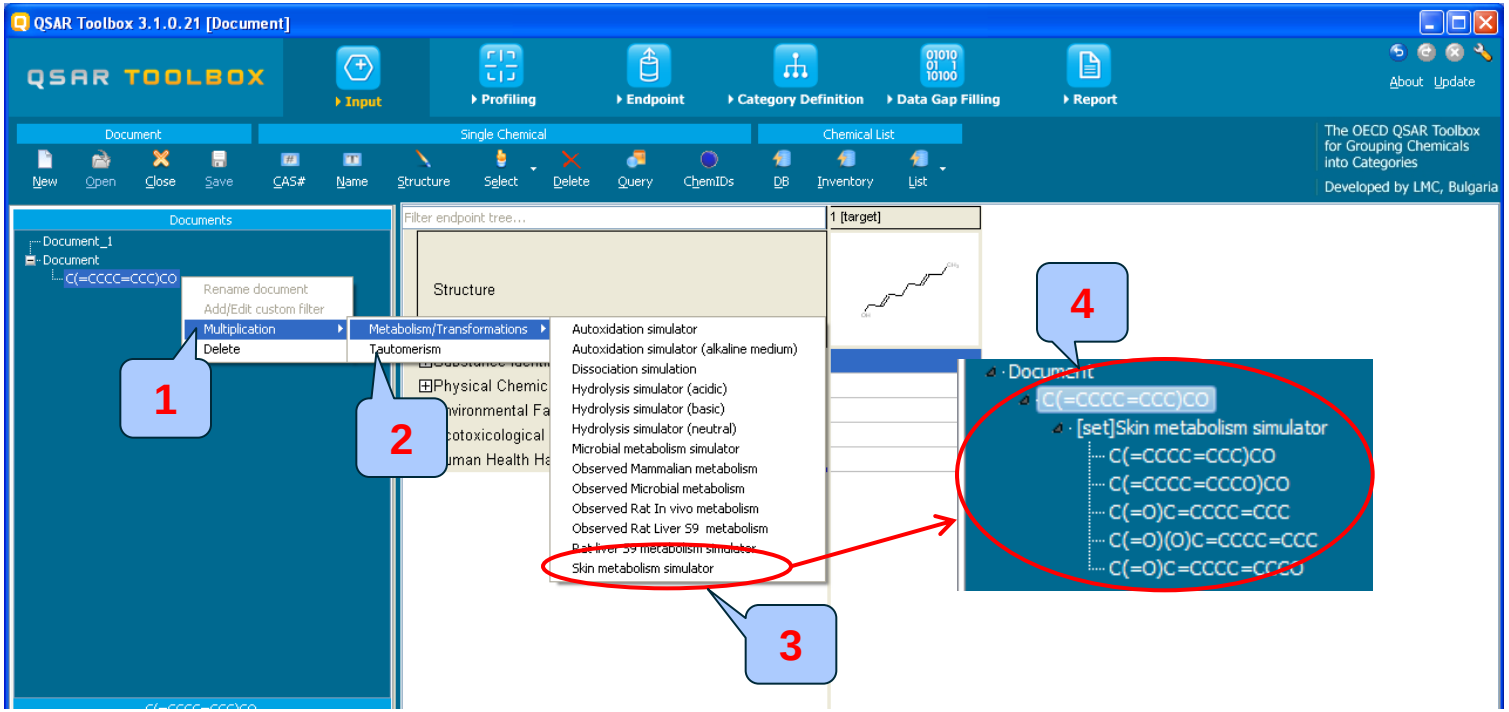
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Handling of skin metabolism**

Handling of skin metabolism of target chemical

- Multiplying target chemical by skin metabolism simulator
- Multiplying target chemical by skin metabolism simulator is accomplished in section **Input**
- The generated metabolites appear in tree like form(see next screen shot)

Handling of skin metabolism of target chemical

Multiplication of target chemical



1. **Click** on the SMILES of the target chemical and perform right click on it, then
2. **Select** Multiplication-Metabolism/Transformations
3. **Select** Skin metabolism simulator
4. Generated metabolites appear in tree like form

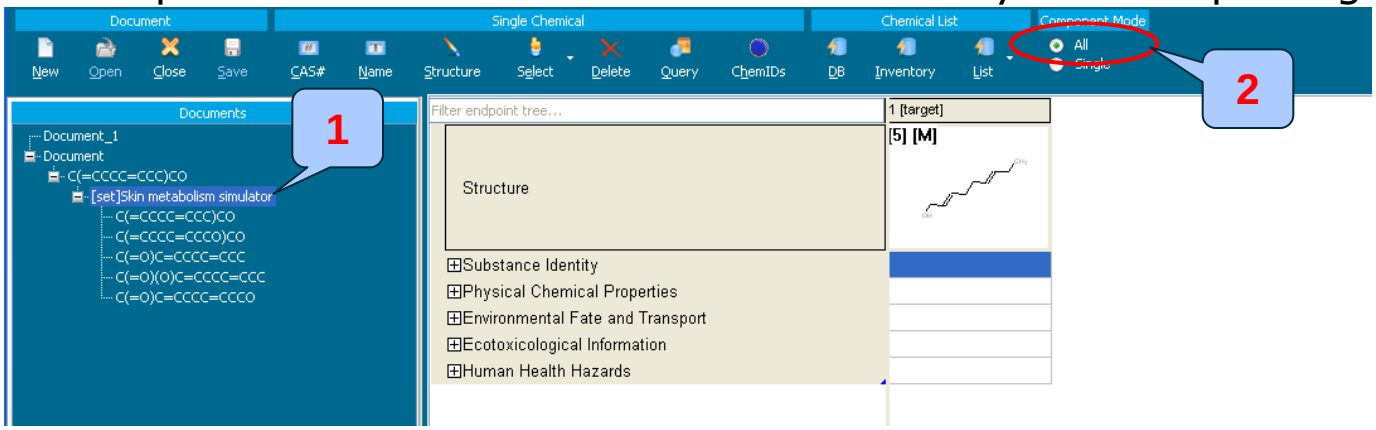
Handling of skin metabolism of target chemical
Visualization of modeling modes

- Two component modes are implemented:
  - Set Mode - all metabolites are analyzed as a package
  - Individual Component Mode - each metabolite is analyzed individually(see next screen shot)

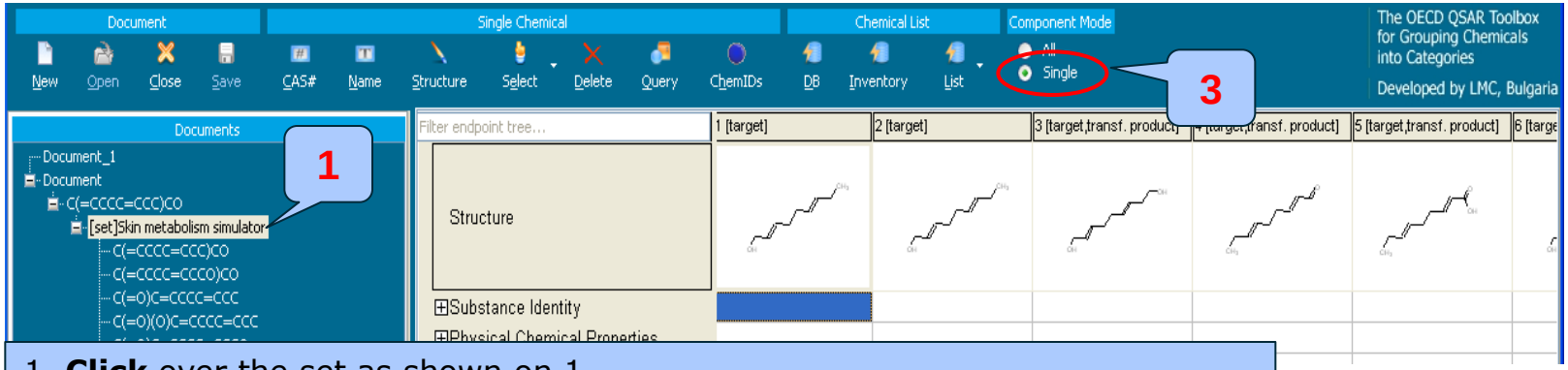
Handling of skin metabolism of target chemical

Visualization of modeling modes

- All Component Mode – all metabolites are analyzed as a package



- Single Mode – each metabolite is analyzed individually



1. **Click** over the set as shown on 1
2. **All** component mode – select All **(2)**
3. **Single** component mode – select Single **(3)**

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Handling of skin metabolism of target chemical**
    - **Profiling set of metabolites**

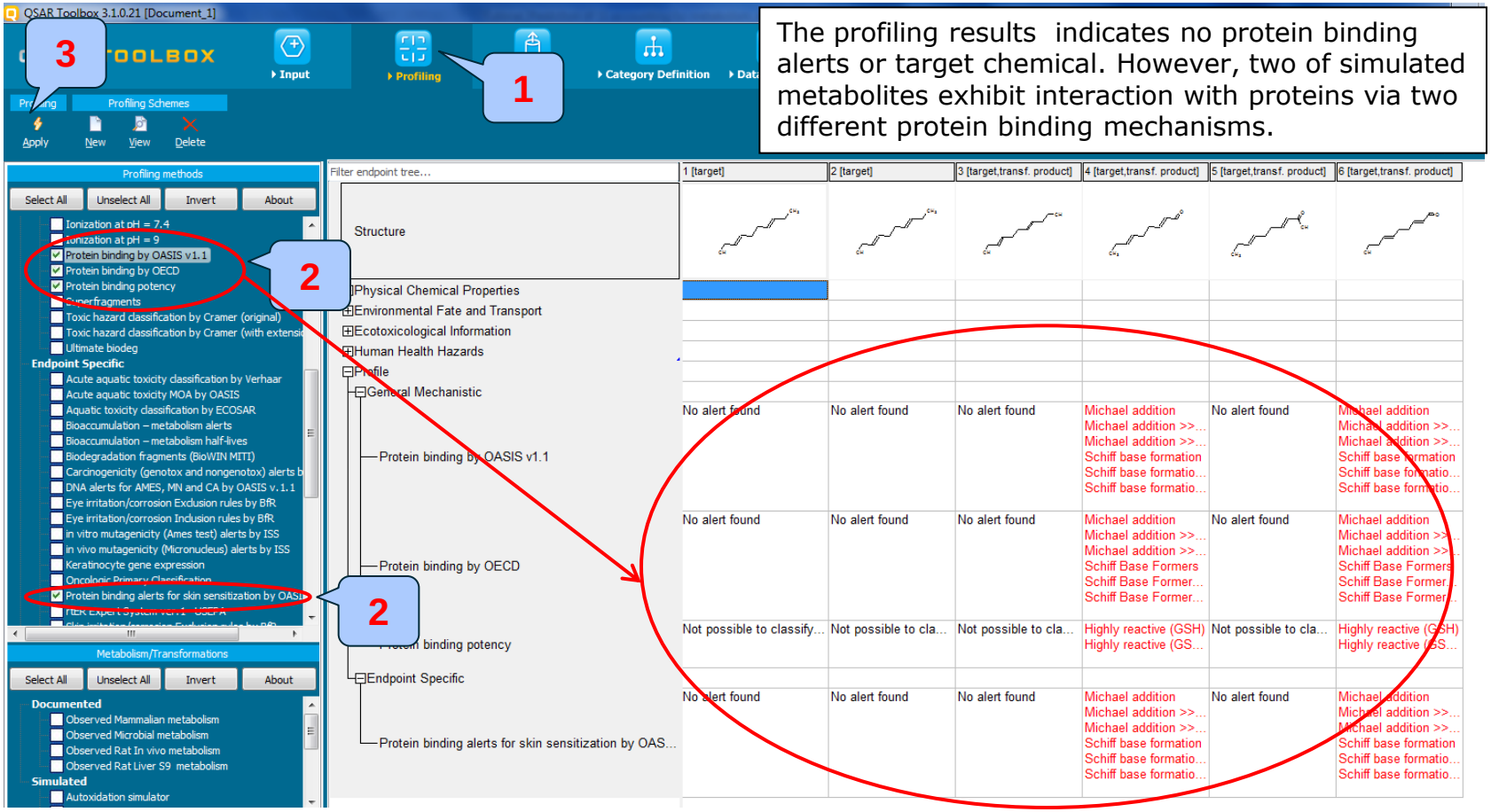
Handling of skin metabolism of target chemical

Profiling set of metabolites

- This module identifies profilers of target chemical and its metabolites
- Protein binding profiles related to skin sensitization are applied on package of target and metabolites
- Profiling results of target and metabolites are illustrated in Single Component mode
- Click on "Profiling" to go to the required module
- Apply Protein binding profilers(see next screen shot)

Handling of skin metabolism of target chemical

Protein binding result of parent and metabolites



The profiling results indicates no protein binding alerts or target chemical. However, two of simulated metabolites exhibit interaction with proteins via two different protein binding mechanisms.

1. **Go** to Profiling 2. **Check** the profilers related to the target endpoints 3. **Click** Apply

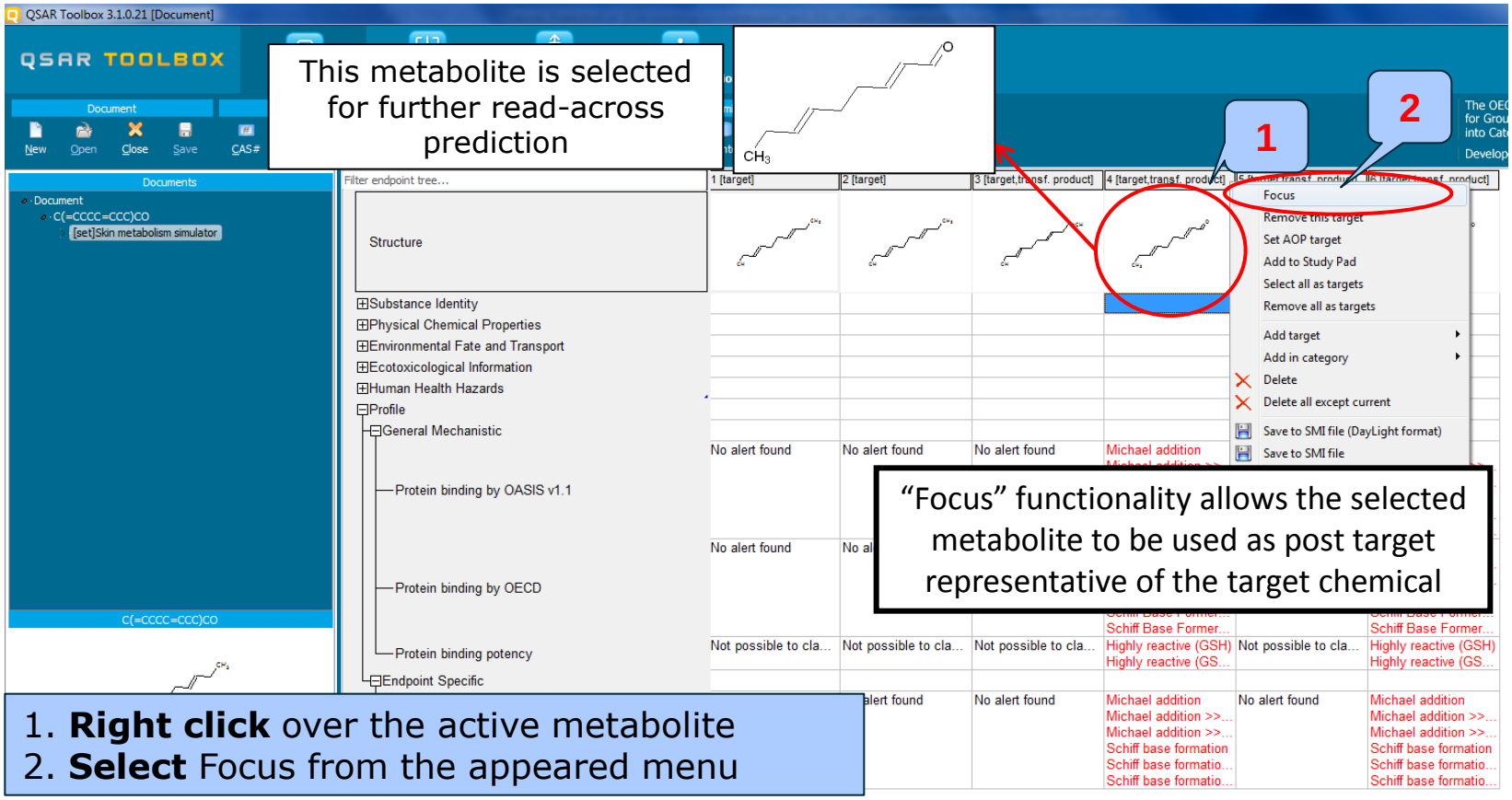
Handling of skin metabolism of target chemical
Recap

- The profiling results indicates no protein binding for target chemical
- Two of simulated skin metabolites have positive protein binding alerts
- One of the reactive metabolites is used for further read across analysis
- The next two parts of the exercise will focus one of the reactive metabolites and find similar analogues of the reactive one (see next screenshot).

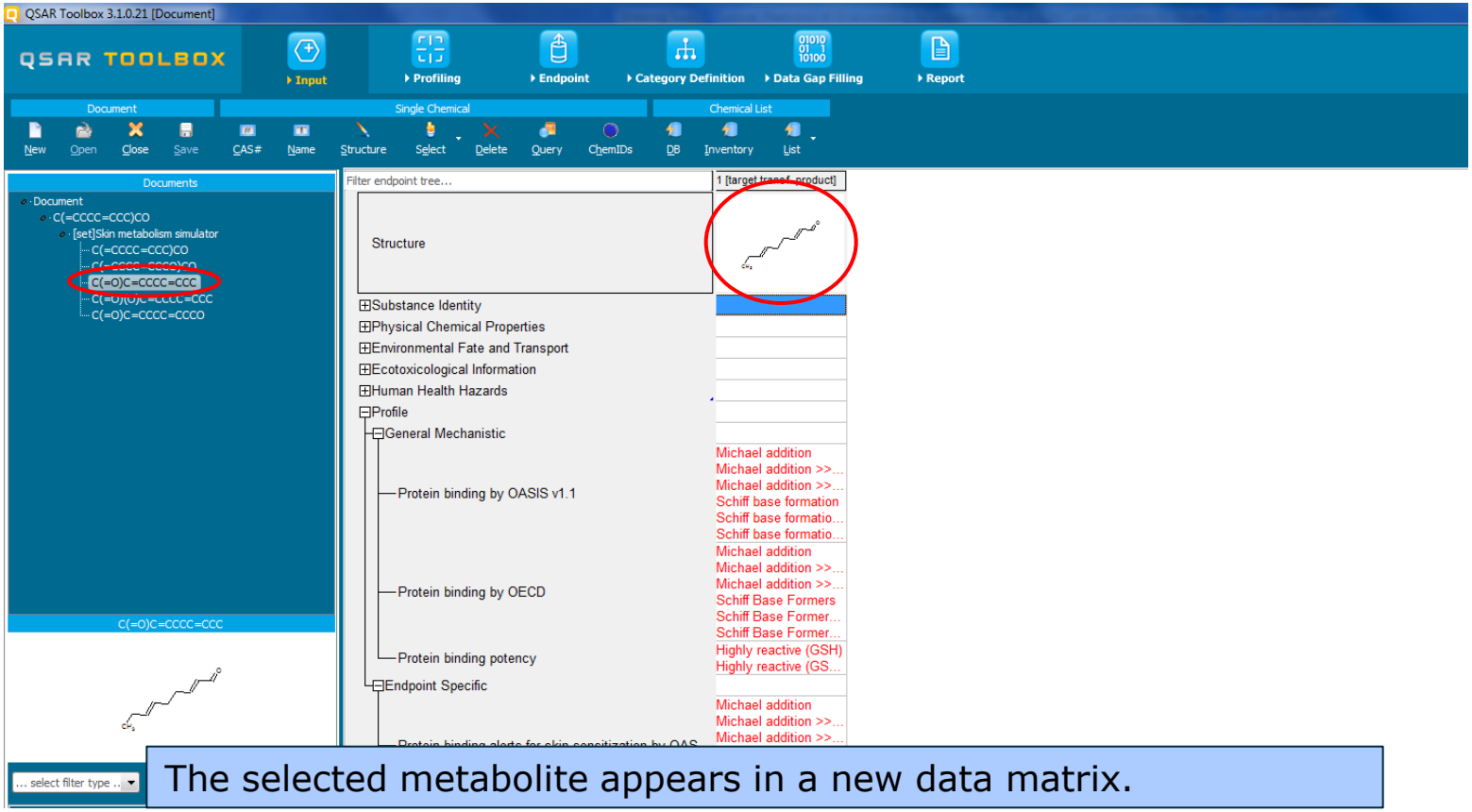
Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Handling of skin metabolism of target chemical**
    - Profiling set of metabolites
    - **Focus of active metabolite**

Handling of skin metabolism of target chemical
Focus of active metabolite



Handling of skin metabolism of target chemical
Focus of active metabolite



The selected metabolite appears in a new data matrix.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Handling of skin metabolism of target chemical**
    - Profiling set of metabolites
    - Focus of active metabolite
    - **Defining category for active metabolite**

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.

- 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

Suitable categorization phases:

1. Structure-related profilers
2. Endpoint specific profilers (for sub-cat)
3. Additional structure-related profilers, if needed to eliminate dissimilar chemicals (to increase the consistency of category) (e.g. chemical elements)

Performing categorization:

1. The categorization phases should be applied successively
2. The application order of the phases depend on the specificity of the data gap filling
3. More categories of same Phase could be used in forming categories
4. Some of the phases could be skipped if consistency of category members is reached

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
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules

Subcategorization
Endpoint Specific

Metabolism accounted for

Phase III. Eliminating dissimilar chemicals

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

Subcategorization
Endpoint Specific

Handling of skin metabolism of target chemical
Category definition for active metabolite

- In this exercise, the reactive metabolite is classified as: Aldehyde by US-EPA New chemical category (phase I)

- The identified Protein binding profiler of the reactive metabolite is: Michael addition >> a,b-unsaturated carbonyl compounds >>Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes (phase II)

- In order to expand the initial group of identified analogues the US-EPA New chemical category is used for categorization purposes (phase I).

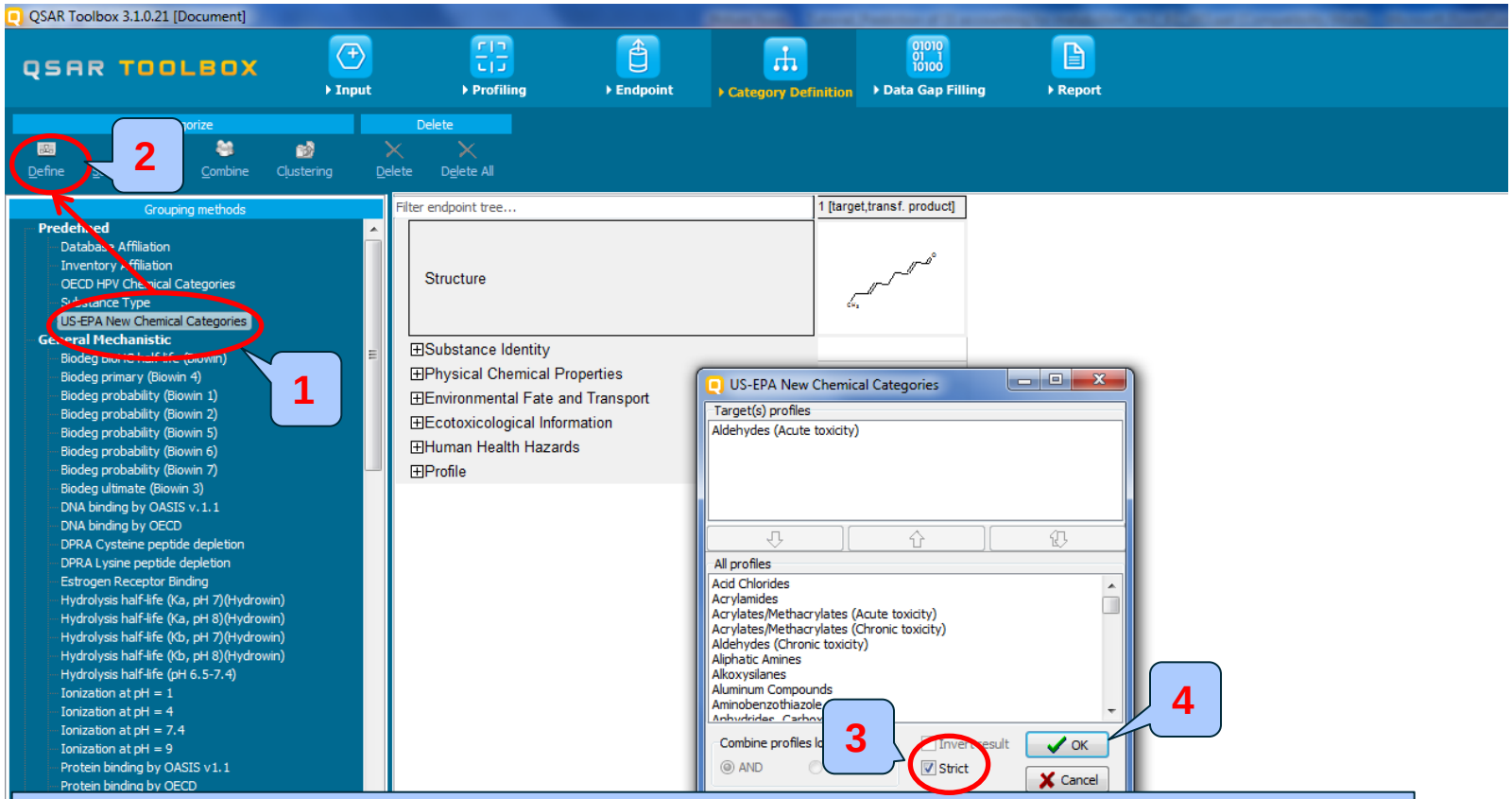
- Searching for similar analogues is accomplished using the two skin sensitization databases: Skin sensitization and Skin sensitization ECETOC

Handling of skin metabolism of target chemical
Defining US-EPA category

- The category US-EPA New chemical category (strict) is used
- **Strict** functionality means that the software will identify analogues having ONLY the categories of the target (i.e aldehydes) and will exclude the analogues having any other categories
- Select US-EPA New Chemical category
- Click Define (see next screen shots)

Handling of skin metabolism of target chemical

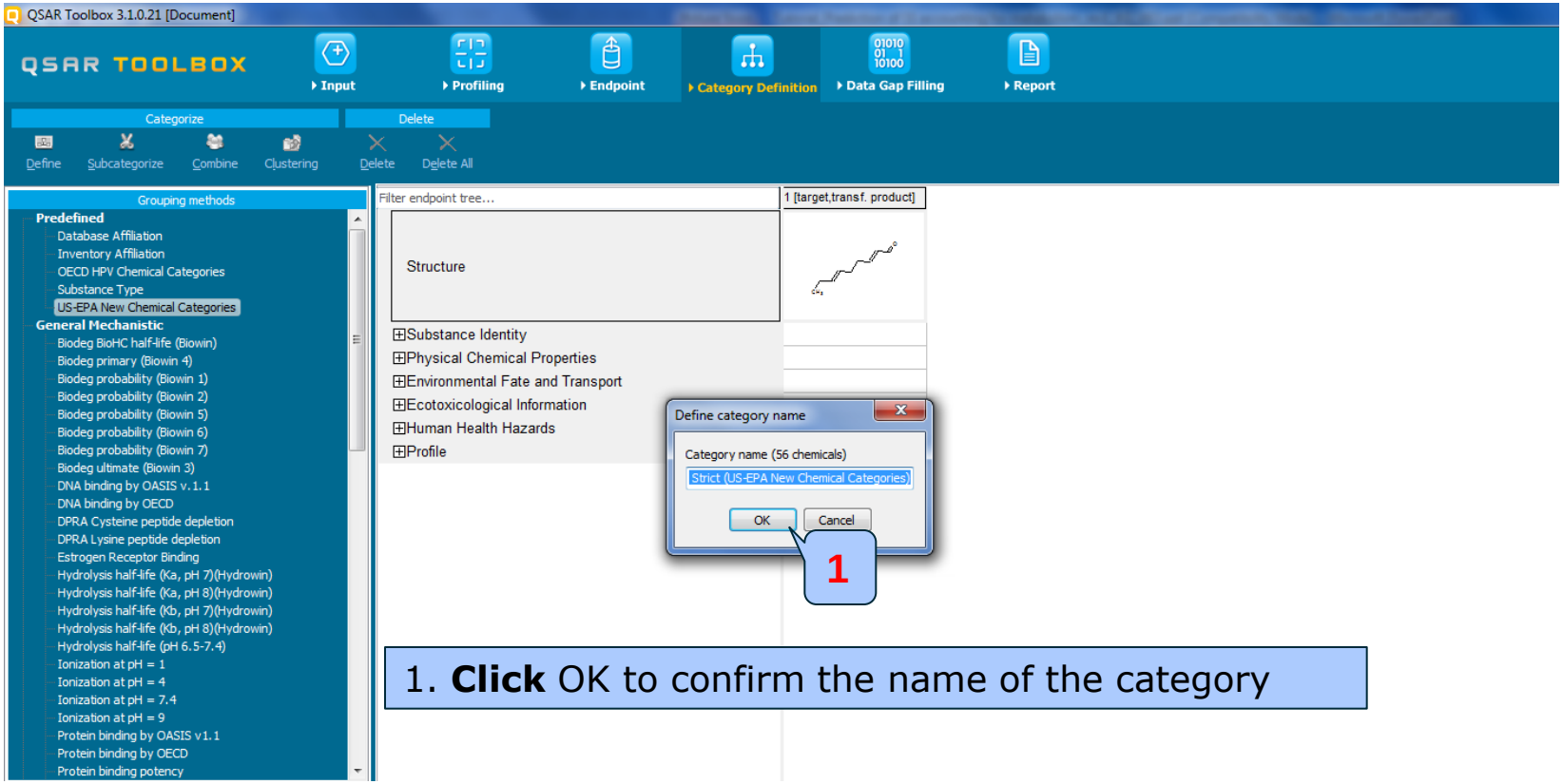
Defining US-EPA category



1. **Highlight** "US-EPA New Chemical Categories" 2. **Click** Define 3. **Select** Strict 4. **Click** OK to confirm the category **Aldehydes (Acute toxicity)** defined by US-EPA category.

Handling of skin metabolism of target chemical

Defining US-EPA category



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

Handling of skin metabolism of target chemical
Category analogues

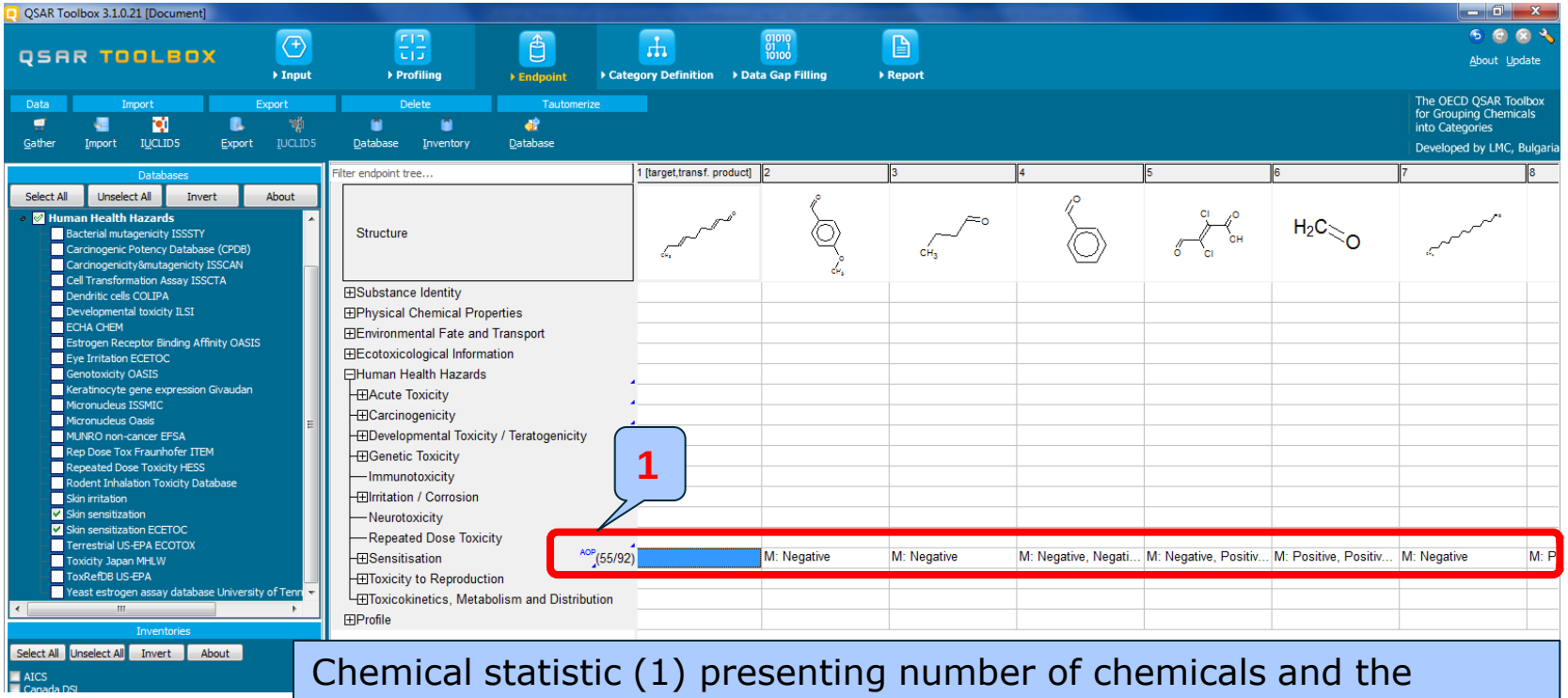
- The Toolbox now identifies all chemicals corresponding to *Aldehydes(Acute toxicity)* by US-EPA listed in the skin sensitization databases.

- 56 analogues including the target chemical are identified; they form a mechanistic category named "**Aldehydes (Acute toxicity)**", which will be used for further data gap filling.

- The experimental data for analogues in the category appears on datamatrix

Handling of skin metabolism of target chemical
Summary information for Analogues

The experimental results for the analogues appeared on datamatrix



Chemical statistic (1) presenting number of chemicals and the available experimental data

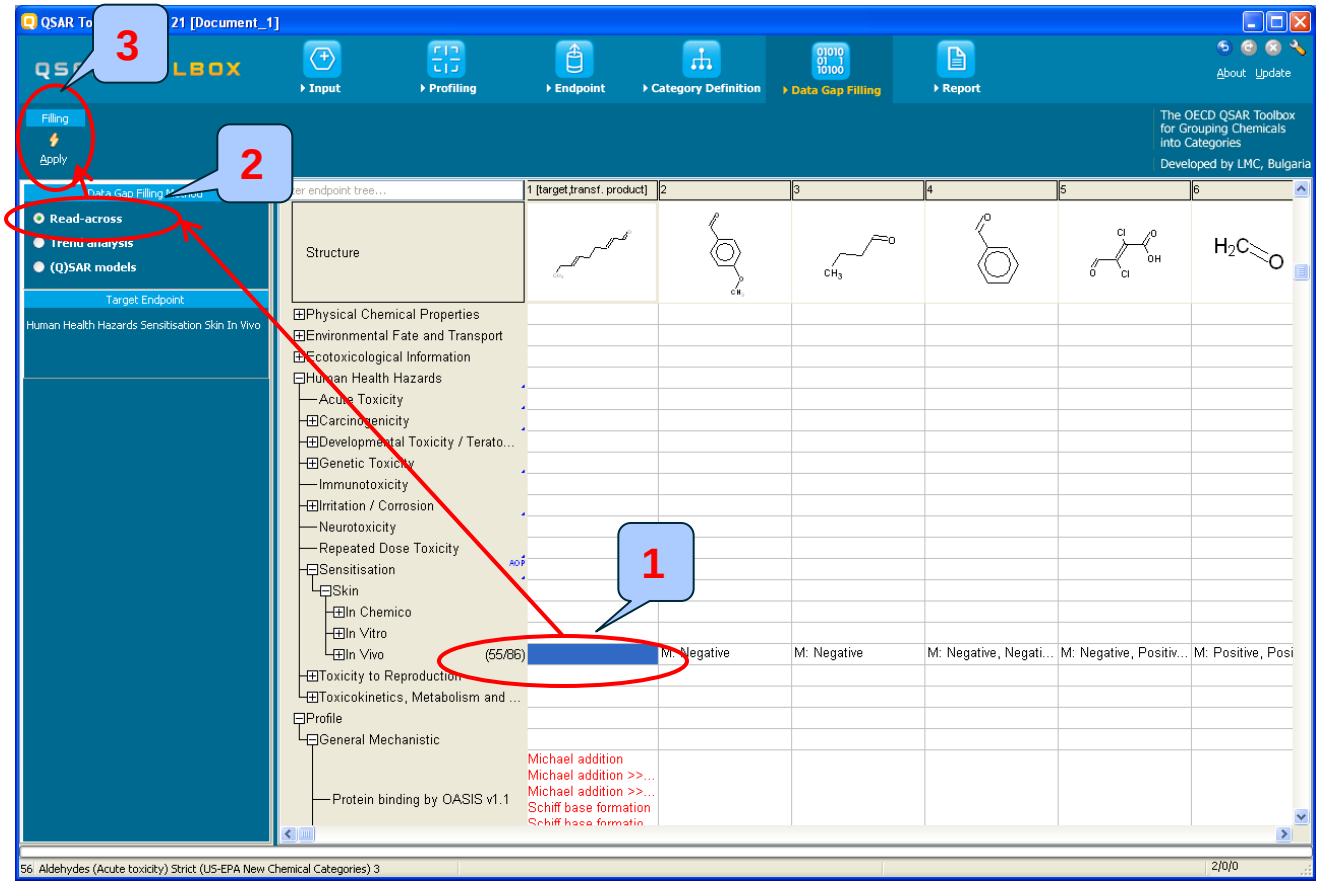
Recap

- In this case "US-EPA New-category Aldehydes(strict) is used for categorization purposes.
- The defined category consist of 56 analogues along with the target chemical
- The available experimental data for these 56 analogues have been collected from two skin sensitization databases.
- But before the user can proceed with the "Filling Data Gap" module, he/she should navigate through the endpoint tree and find the specific gap that will be filled in (in this case Human Health Hazards#Sensitisation#Skin#In Vivo).
- In this case we mixed assays and endpoints (see slides #62-64)

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Handling of skin metabolism of target chemical**
    - Profiling set of metabolites
    - Focus of active metabolite
    - Category definition for active metabolite
    - **Data gap filling**

Data gap filling
Apply Read across

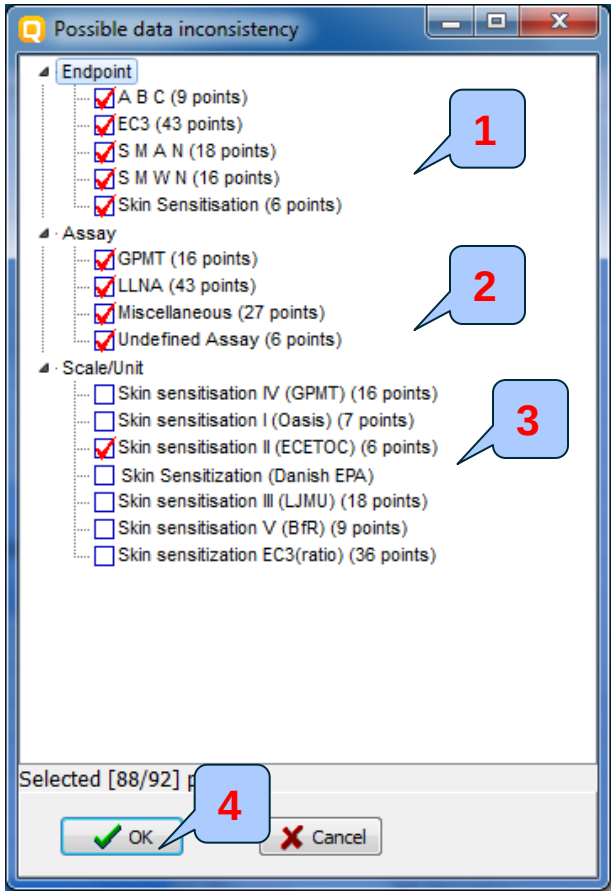


1. **Click** on the cell corresponding to "Sensitisation/Skin/In vivo" for the target chemical(active metabolite) 2. **Select** Read-across 3. **Click** Apply

Data gap filling for active metabolite
Scale definition

- Skin sensitisation is "qualitative" endpoint for which the results are presented with categorical data (for example: positive; negative; weak sensitizer; strong sensitizer ,etc).
- Skin sensitisation potential of the chemicals came from different authors coded with different names (for example: data from John Moores University of Liverpool are: *Strongly sensitizing, Moderately sensitizing etc*.; data from European centre for Ecotoxicology and Toxicology of chemicals are: *Positive, Negative, and Equivocal*).
- The main purpose of the scales is to unify all data available in the Toolbox databases for a certain endpoint.
- The default scale for Skin Sensitisation is "Skin Sensitisation ECETOC". It converts all skin data into: Positive, Negative, and Equivocal.

Data gap filling for active metabolite
Scale definition

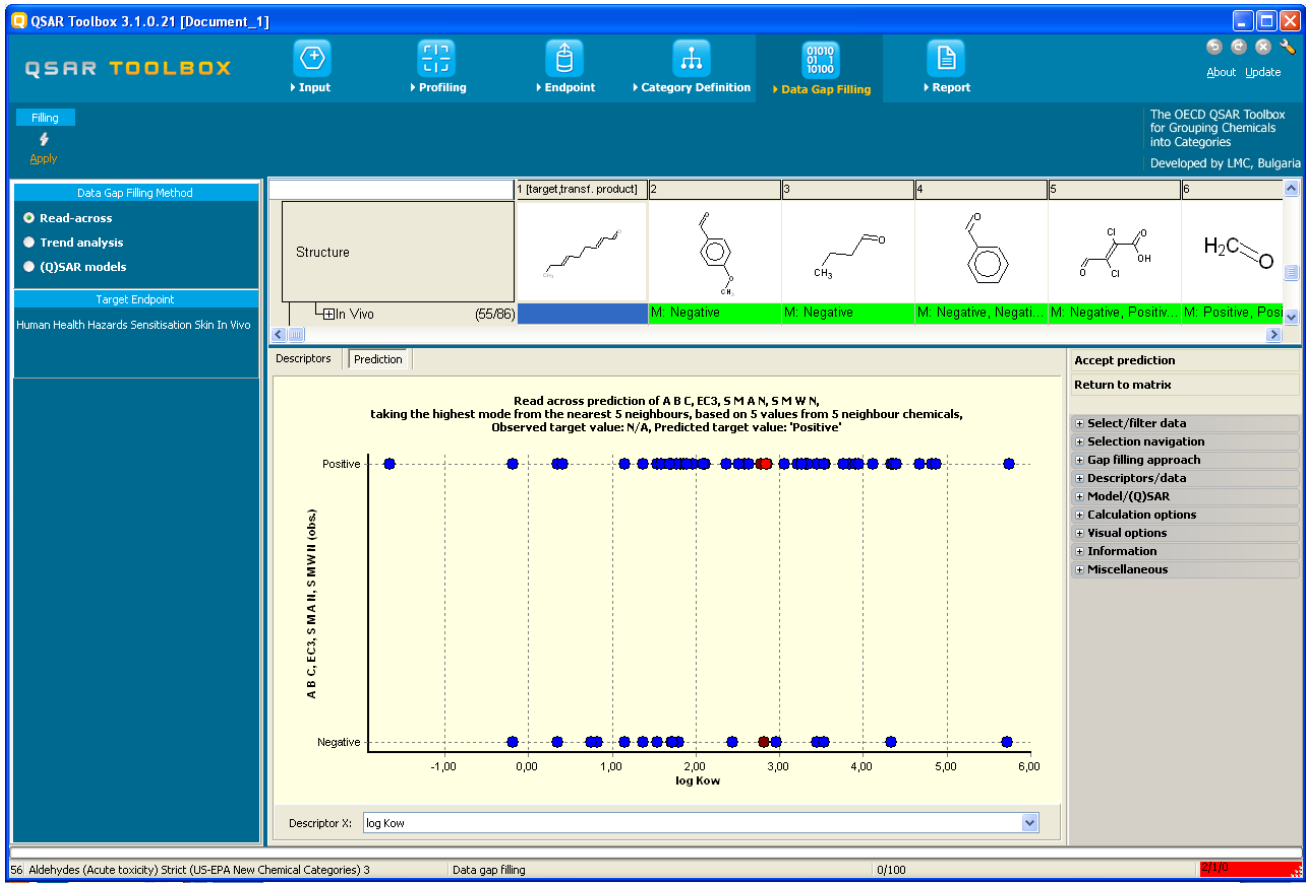


1. In this case all endpoints (1) and assays (2) related to skin sensitization are taken into account
3. Default scale is Skin sensitization II (ECETOC)
4. **Click** OK to enter data gap filling

Data gap filling for active metabolite
Read-across

- The resulting plot is experimental results of all analogues (Y axis) according to a descriptor (X axis) with the default descriptor of log Kow (see next screen shot).
- The **RED** dot represents predicted results for the target chemical .
- The **BROWN** dots represent the experimental results available for the analogues that are used for the read-across.
- The **BLUE** dot represent the experimental results available for the analogues but not used for read-across.

Data gap filling for active metabolite
Read-across



Data gap filling for active metabolite
Subcategorization

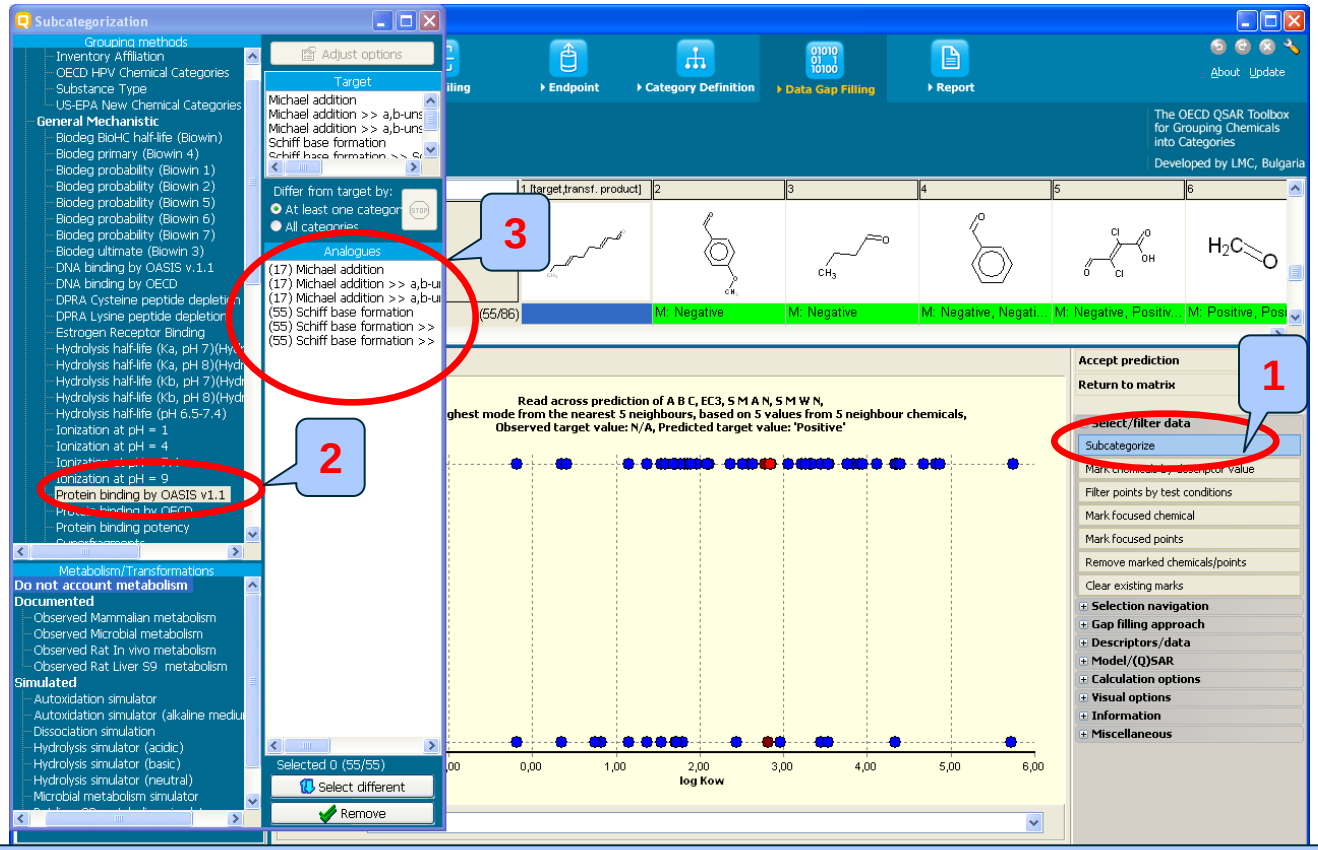
- After the available data has been retrieved, the user can then further subcategorize the results according to the following endpoint-specific subcategorizations (phase II, slide #55):

  - Protein binding by Oasis v1.1.
  - Protein binding by OECD
  - Protein binding potency

- These steps are summarized in the next screen shots.

Data gap filling for active metabolite

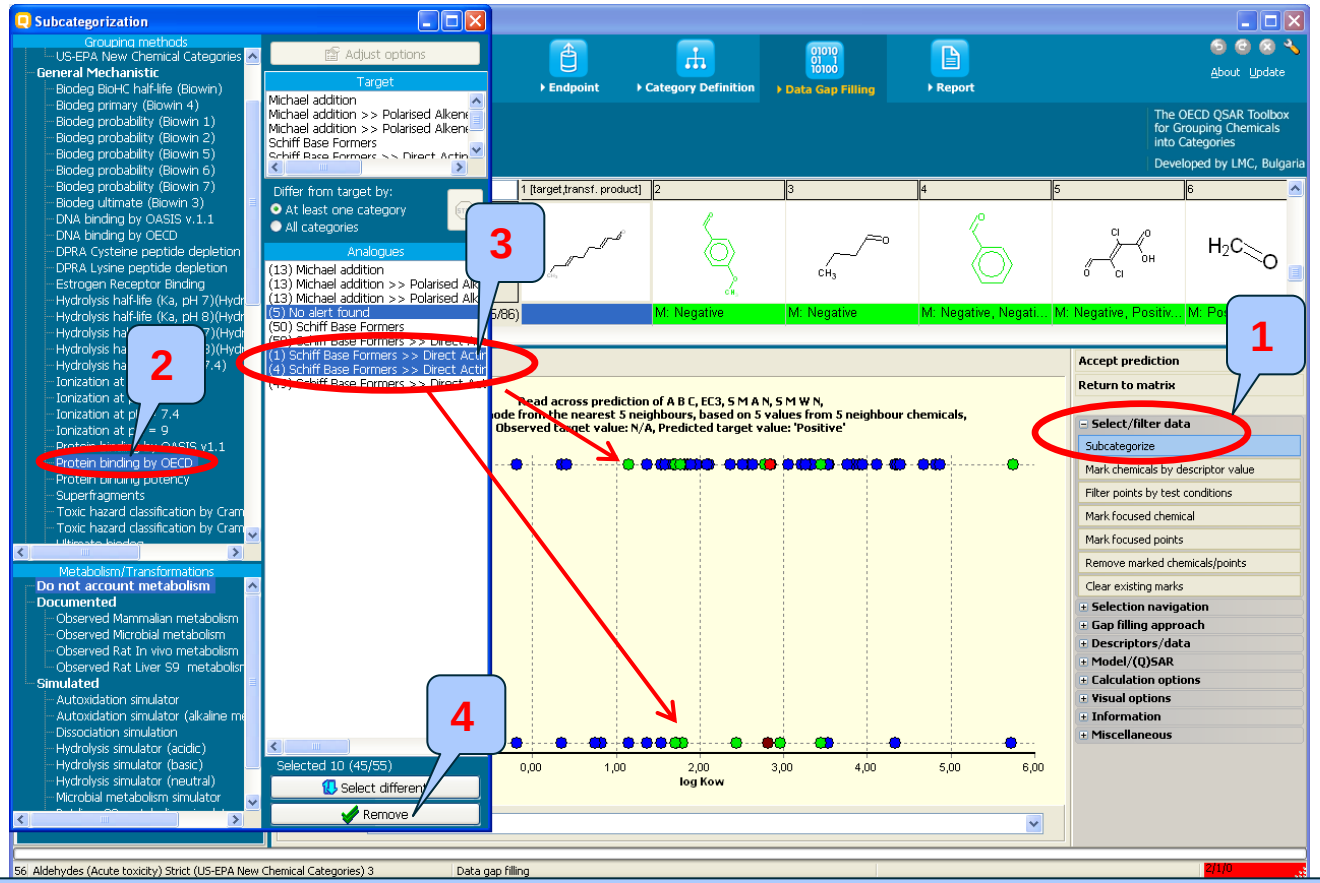
Subcategorization 1: Protein binding by OASIS v.1.



1. **Select filter data/subcategorize** 2. **Select** Protein binding by OASIS v1.1. Note all analogues are in the same category (3) as the target chemical, so no further action is required.

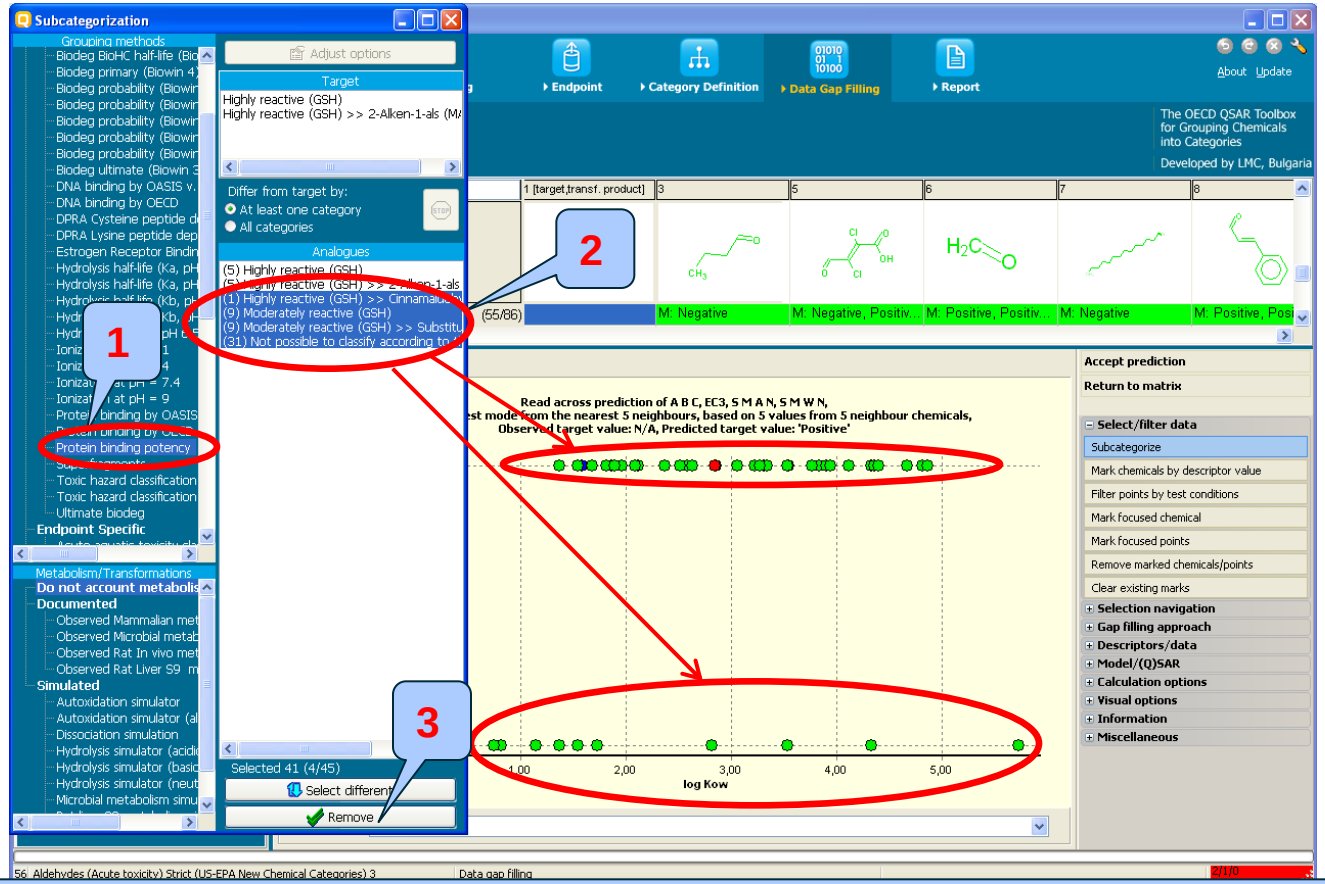
Data gap filling for active metabolite

Subcategorization 2: Protein binding by OECD



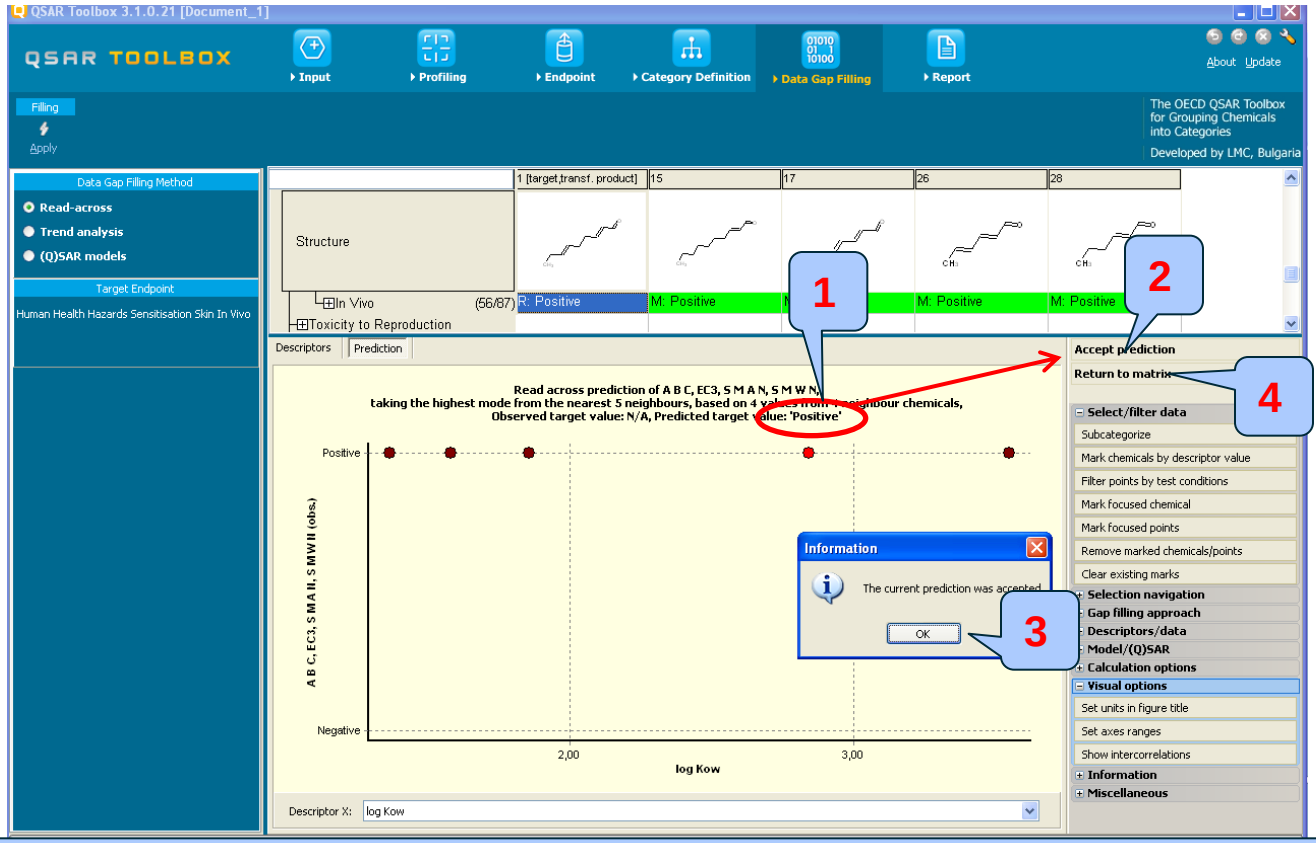
1. **Open** Select/filter data/Subcategorize 2. **Select** Protein binding by OECD 3. Green dots represent analogues having different Protein mechanism of interaction 4. **Remove** dissimilar chemicals

Data gap filling for active metabolite
Subcategorization 3: Protein binding potency



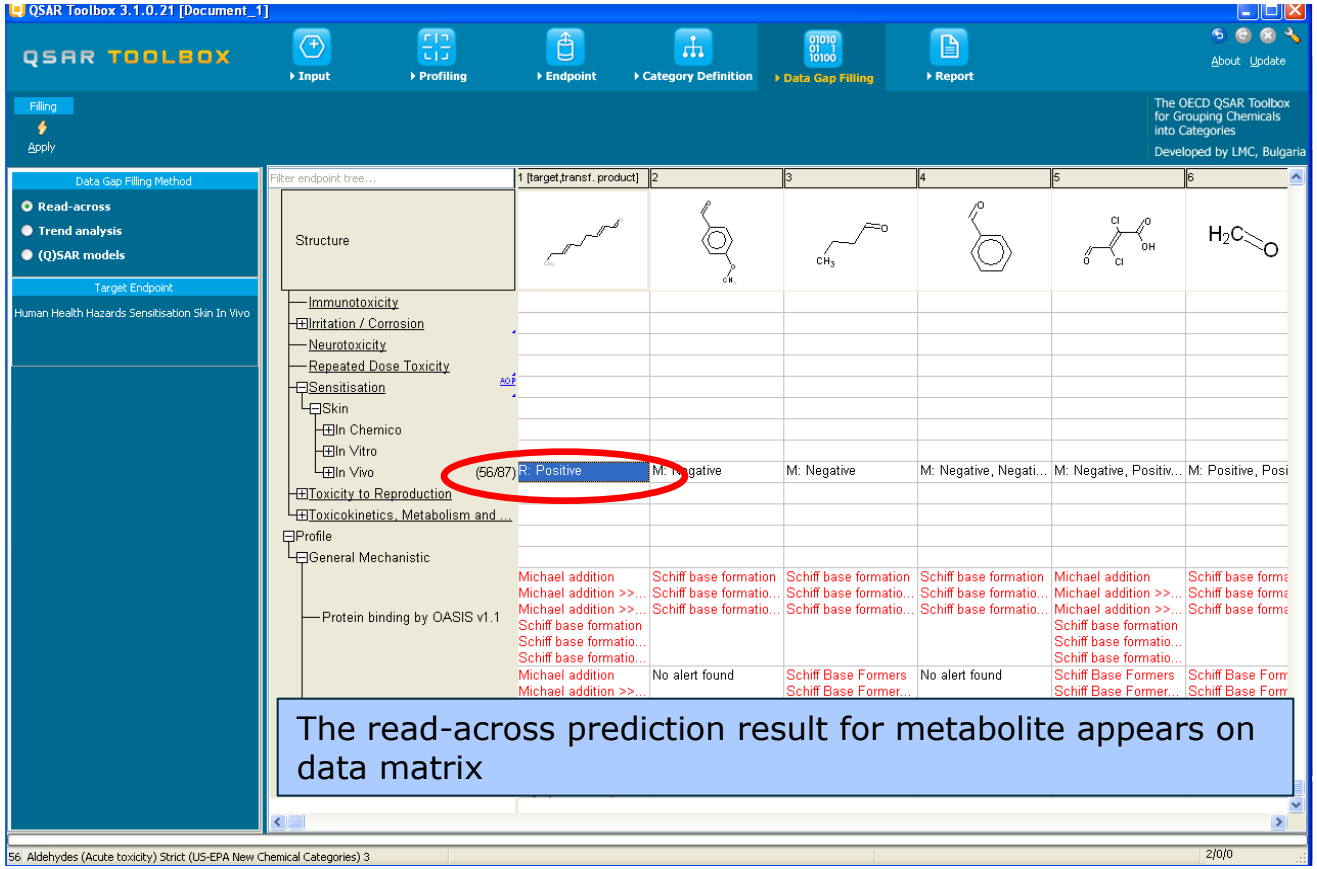
1. **Select** Protein binding potency 2. Green dots represent analogues belonging to different subcategories 3. **Remove** dissimilar chemicals

Data gap filling for active metabolite
Results after subcategorization



1. The predicted result is **positive** 2. **Accep**t prediction 3. **Click** OK 4. **Return to matrix**

Data gap filling for active metabolite
Read-across prediction



Data gap filling for active metabolite
Interpreting Read-across

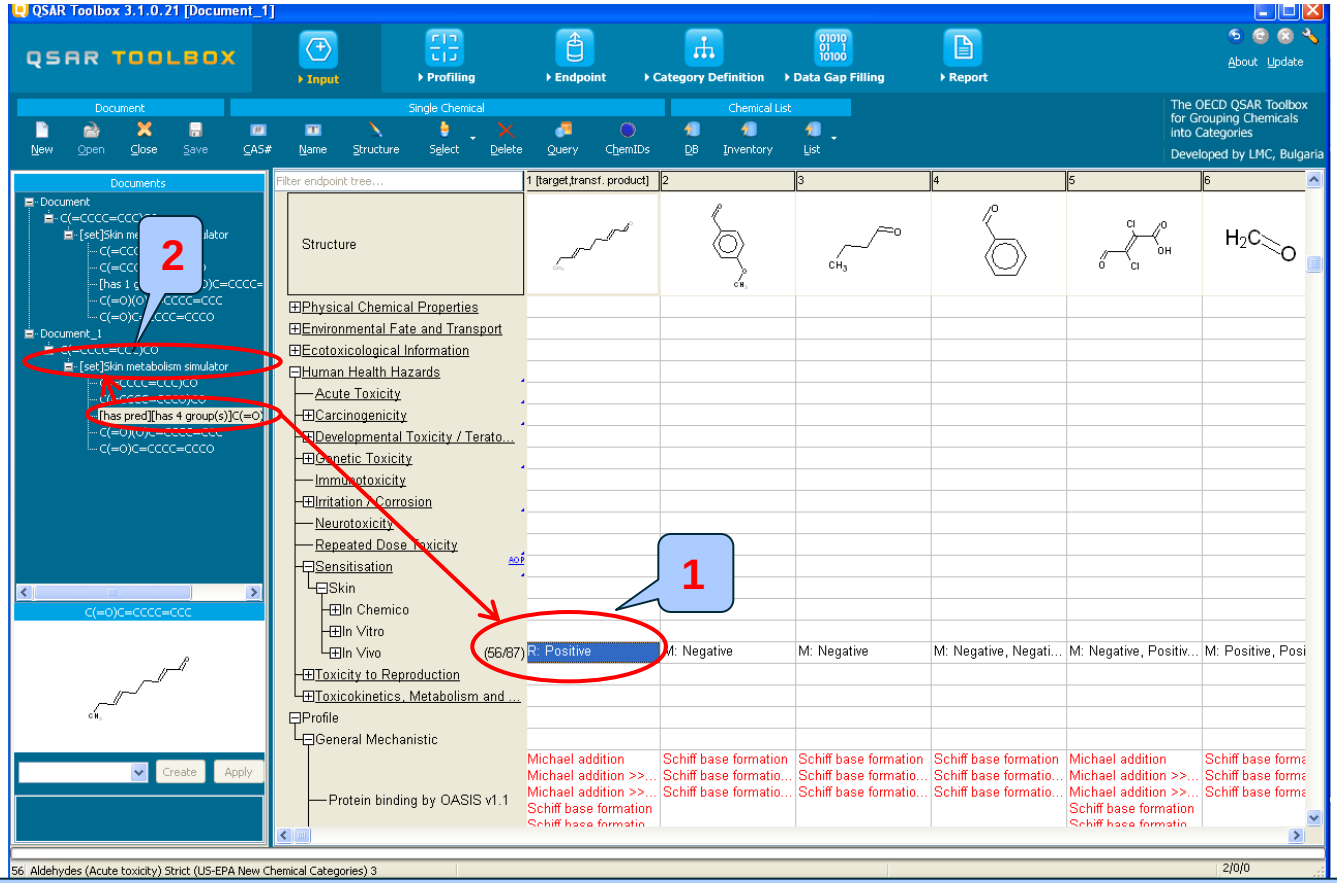
- In this example, all analogues have same protein binding alerts
- All analogues exhibit positive skin sensitization
- The same positive sensitising potential is therefore predicted for the target (i.e. active metabolite).
- The prediction of metabolite is further transferred to the parent chemical using Independent MOA (see next screen shots)

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - **Handling of skin metabolism of target chemical**
    - Profiling set of metabolites
    - Focus of active metabolite
    - Category definition for active metabolite
    - Data gap filling
  - **Assigning prediction of active metabolite to parent**

Handling skin metabolism of target chemical

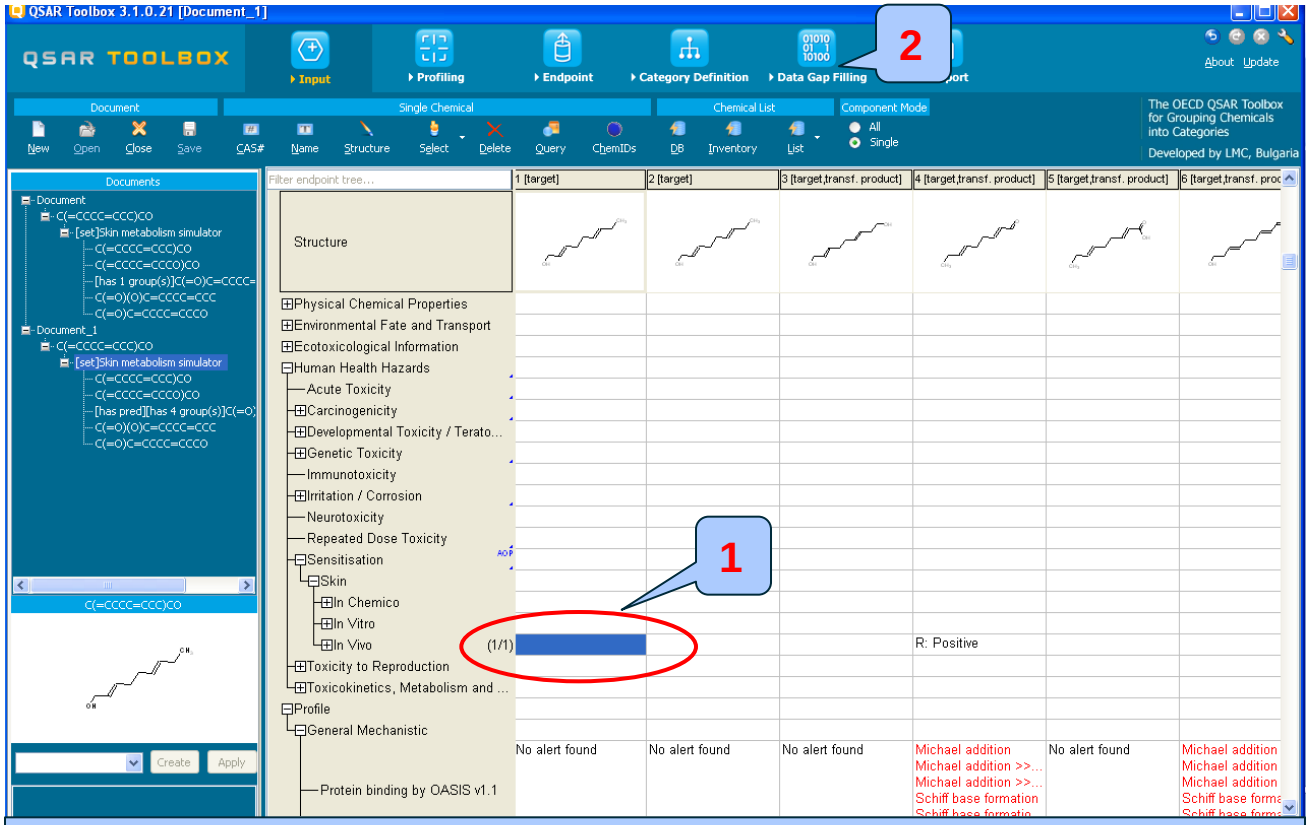
Assigning data to parent chemical



1.The read-across prediction for the metabolite is positive then 2. **Select** datamatrix of the target chemical: click over the node "[set] Skin metabolism simulator"

Handling skin metabolism of target chemical

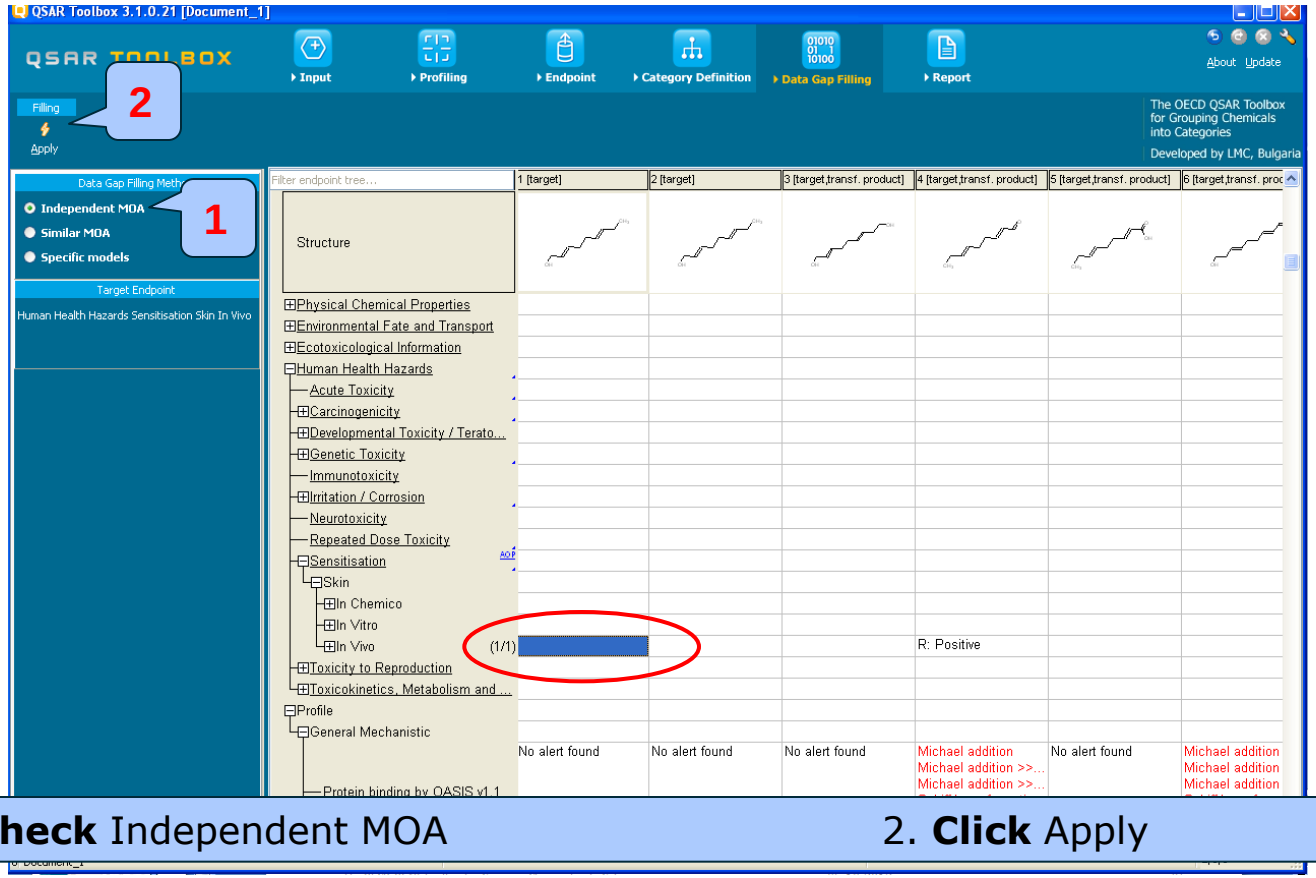
Assigning data to parent chemical



1. **Select** cell corresponding to skin sensitization 2. **Select** Data Gap Filling

Handling skin metabolism of target chemical

Assigning data to parent chemical



1. **Check** Independent MOA

2. **Click** Apply

Handling skin metabolism of target chemical

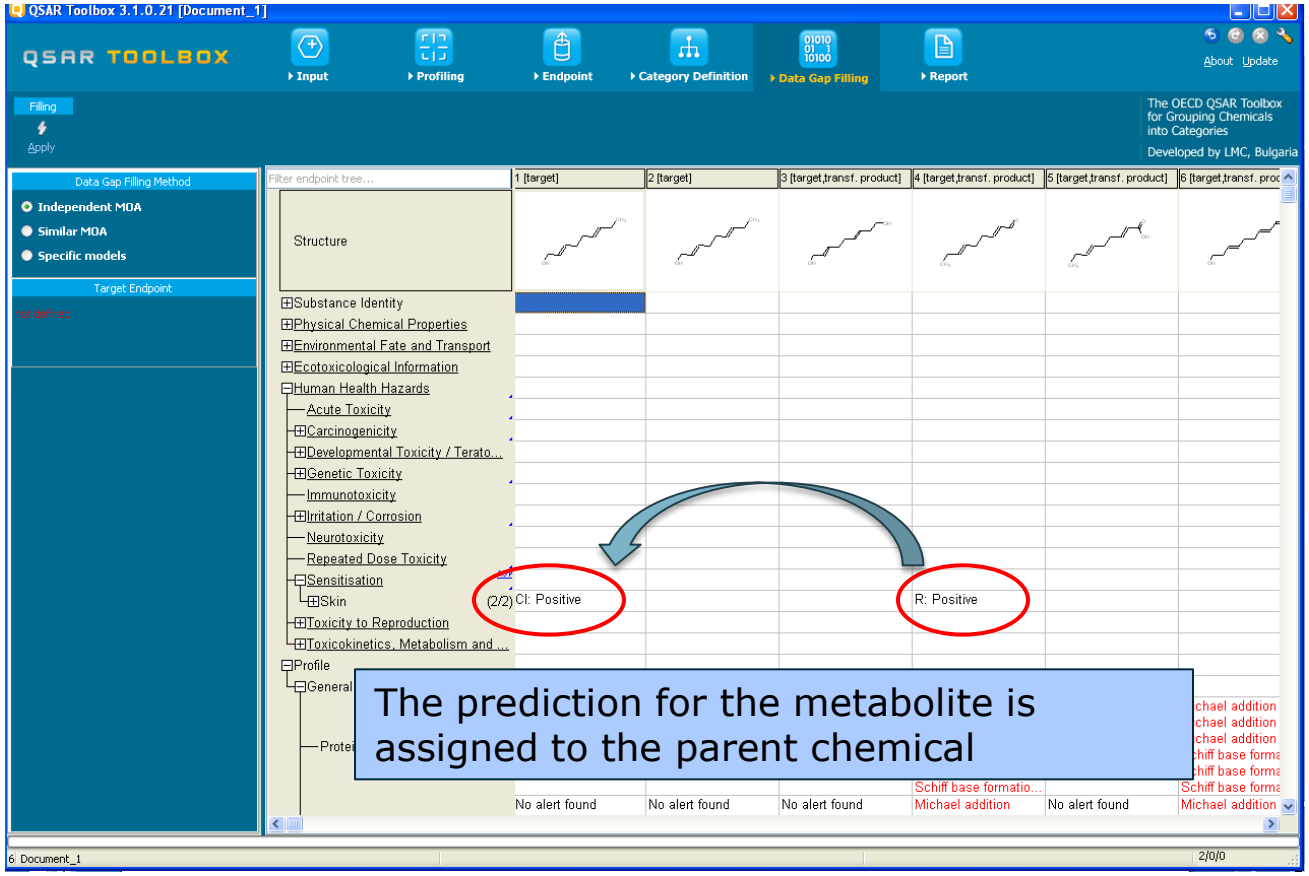
Assigning data to parent chemical

- The following actions (steps) are used for assigning data to parent chemical:
  - Accept prediction
  - Return to matrix

- Final prediction for the parent compound labeled as CI (Component based Independent mode) is positive (see next screen shot)

Handling skin metabolism of target chemical

Assigning data to parent chemical



Recap

- The target chemical **trans-2,cis-6-nonadienol** has been entered into the system.
- It has been profiled by Protein binding profilers; no protein binding has been found for target chemical.
- Positive experimental data has been retrieved for target chemical.
- Skin metabolism of target chemical is investigated. Two of simulated skin metabolites have positive protein binding alerts.
- These metabolites have similar protein binding alert: **α,β-unsaturated aldehydes.** One of the reactive metabolites is used for further read across analysis.
- No experimental data for the selected metabolite has been found, so the category of similar analogues has been defined.
- The initial group of analogues is defined by US-EPA New Chemical categories.
- 56 analogues including the target chemical are identified; they form a mechanistic category "**Aldehydes (Acute toxicity)**", which will be used for gap filling.
- Read-across is used for data gap filling.
- Protein binding categories have been used for refining the initial category.
- Positive skin sensitization has been predicted for the active metabolite.
- Positive prediction for reactive metabolite has been transferred to the parent chemical using Independent MOA.

Outlook

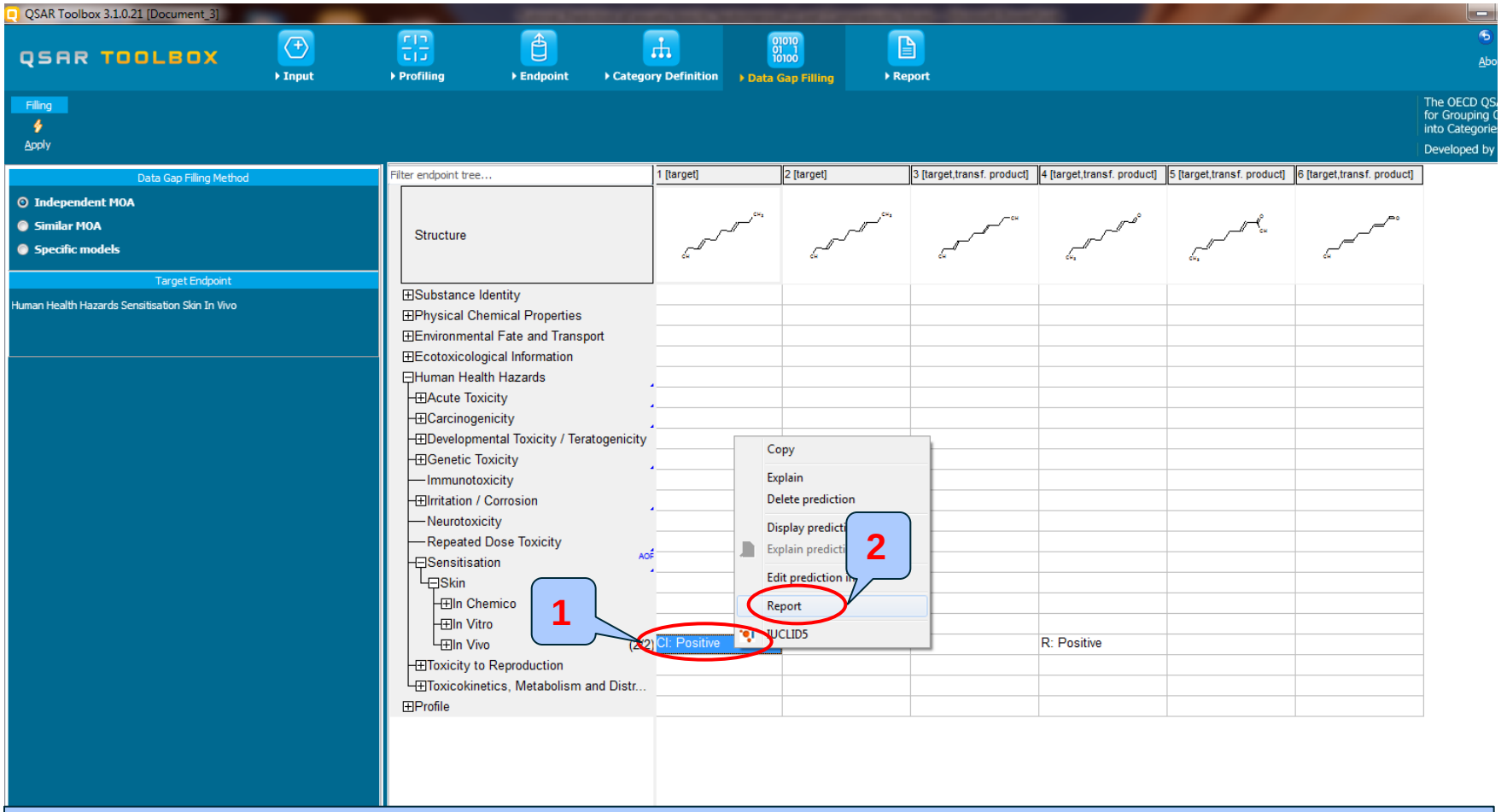
- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Endpoint
  - Handling of skin metabolism of target chemical
  - **Report**

Report

- Remember the report module allows you to generate a report on the predictions performed 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.

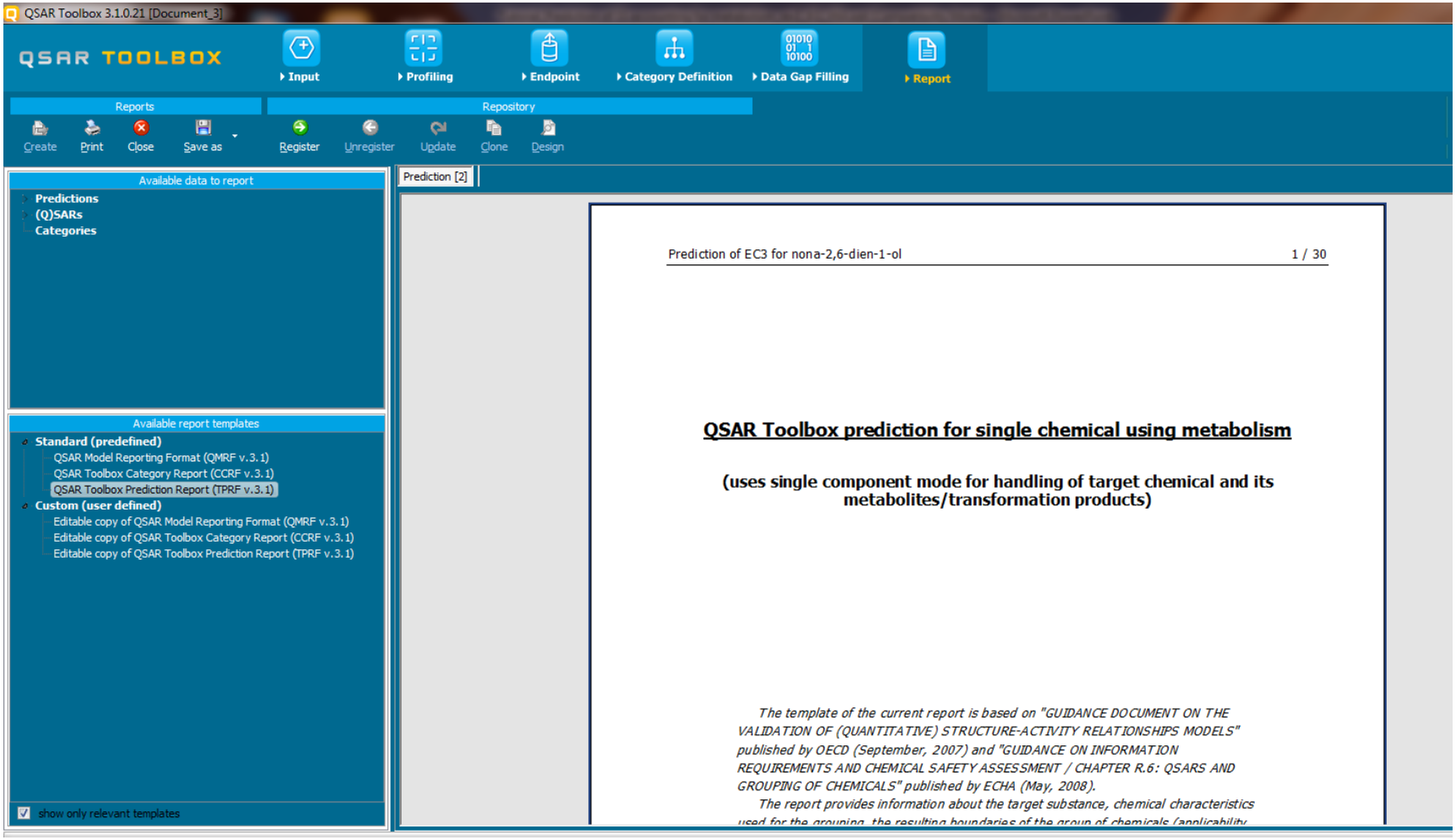
- Generating the report is shown on next screenshots

Report

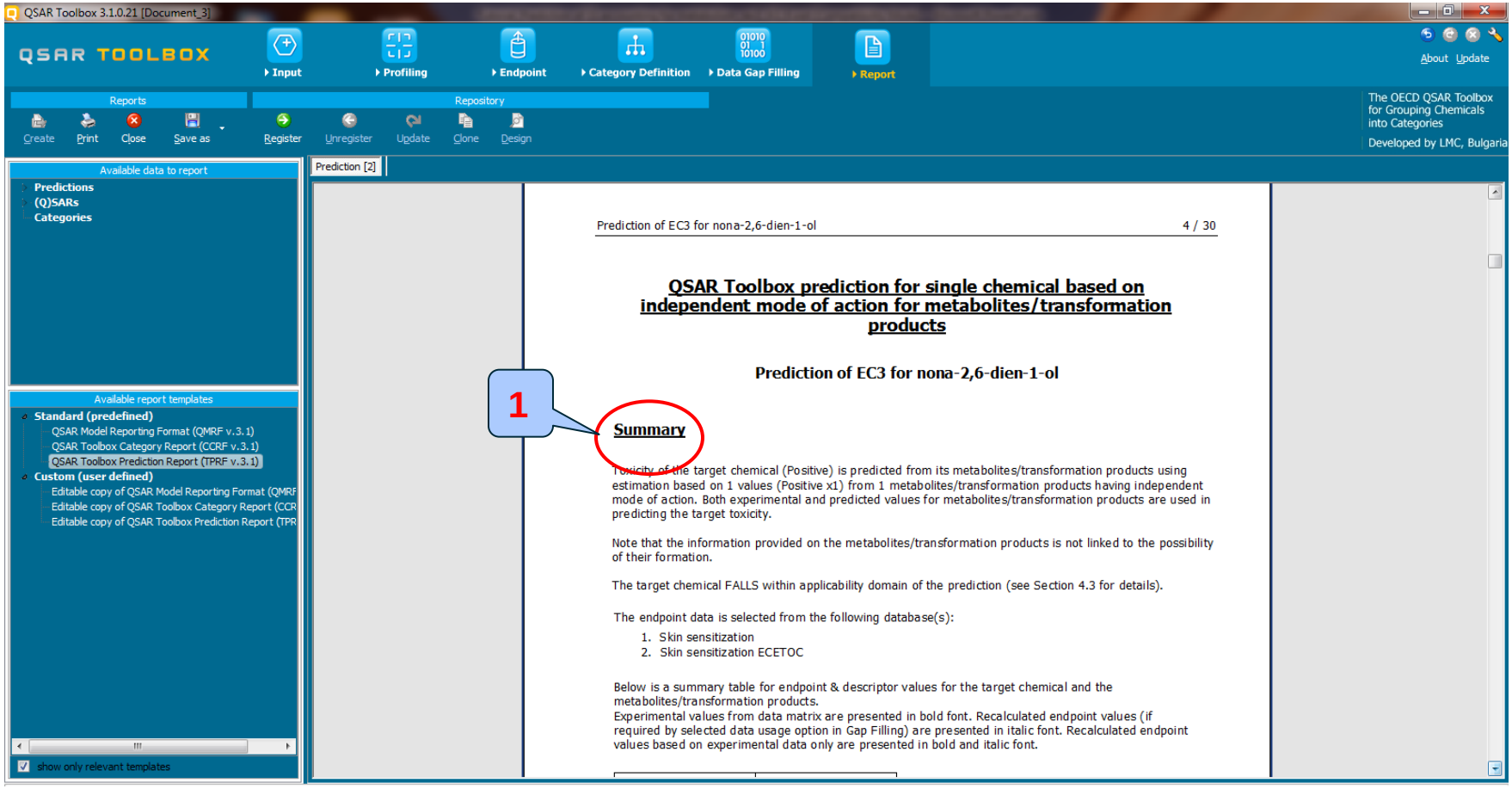


1. **Select** prediction
2. **Right Click** and **Select** Report

Report

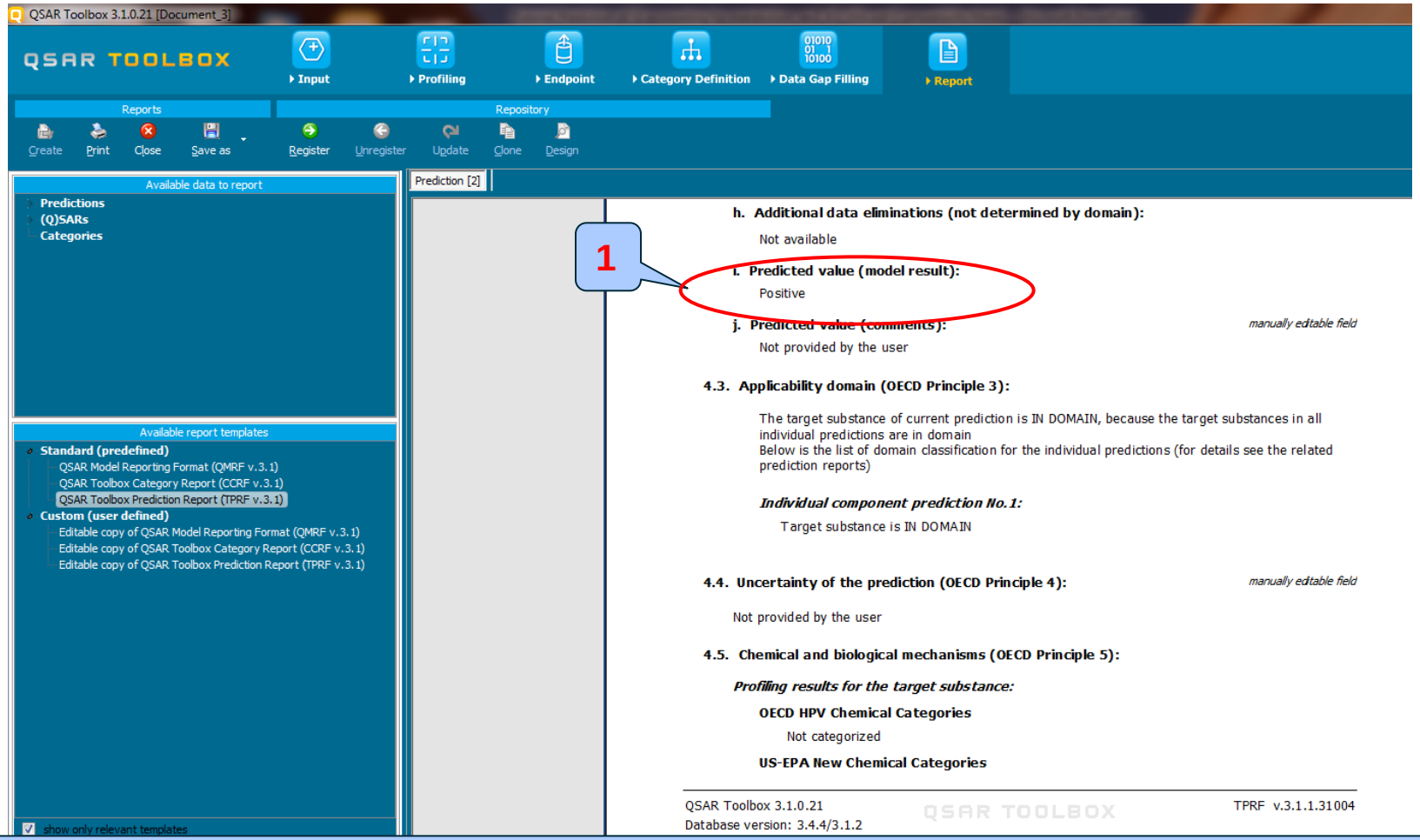

QSAR Toolbox 3.1.0.21 [Document_3]
QSAR TOOLBOX
Input
Profiling
Endpoint
Category Definition
Data Gap Filling
Report
Reports
Repository
Create
Print
Close
Save as
Register
Unregister
Update
Clone
Design
Available data to report
Predictions
(Q)SARs
Categories
Available report templates
Standard (predefined)
QSAR Model Reporting Format (QMRF v.3.1)
QSAR Toolbox Category Report (CCRF v.3.1)
QSAR Toolbox Prediction Report (TPRF v.3.1)
Custom (user defined)
Editable copy of QSAR Model Reporting Format (QMRF v.3.1)
Editable copy of QSAR Toolbox Category Report (CCRF v.3.1)
Editable copy of QSAR Toolbox Prediction Report (TPRF v.3.1)
show only relevant templates
Prediction [2]
Prediction of EC3 for nona-2,6-dien-1-ol
1 / 30
QSAR Toolbox prediction for single chemical using metabolism
(uses single component mode for handling of target chemical and its metabolites/transformation products)
The template of the current report is based on "GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS MODELS" published by OECD (September, 2007) and "GUIDANCE ON INFORMATION REQUIREMENTS AND CHEMICAL SAFETY ASSESSMENT / CHAPTER R.6: QSARS AND GROUPING OF CHEMICALS" published by ECHA (May, 2008).
The report provides information about the target substance, chemical characteristics

Report



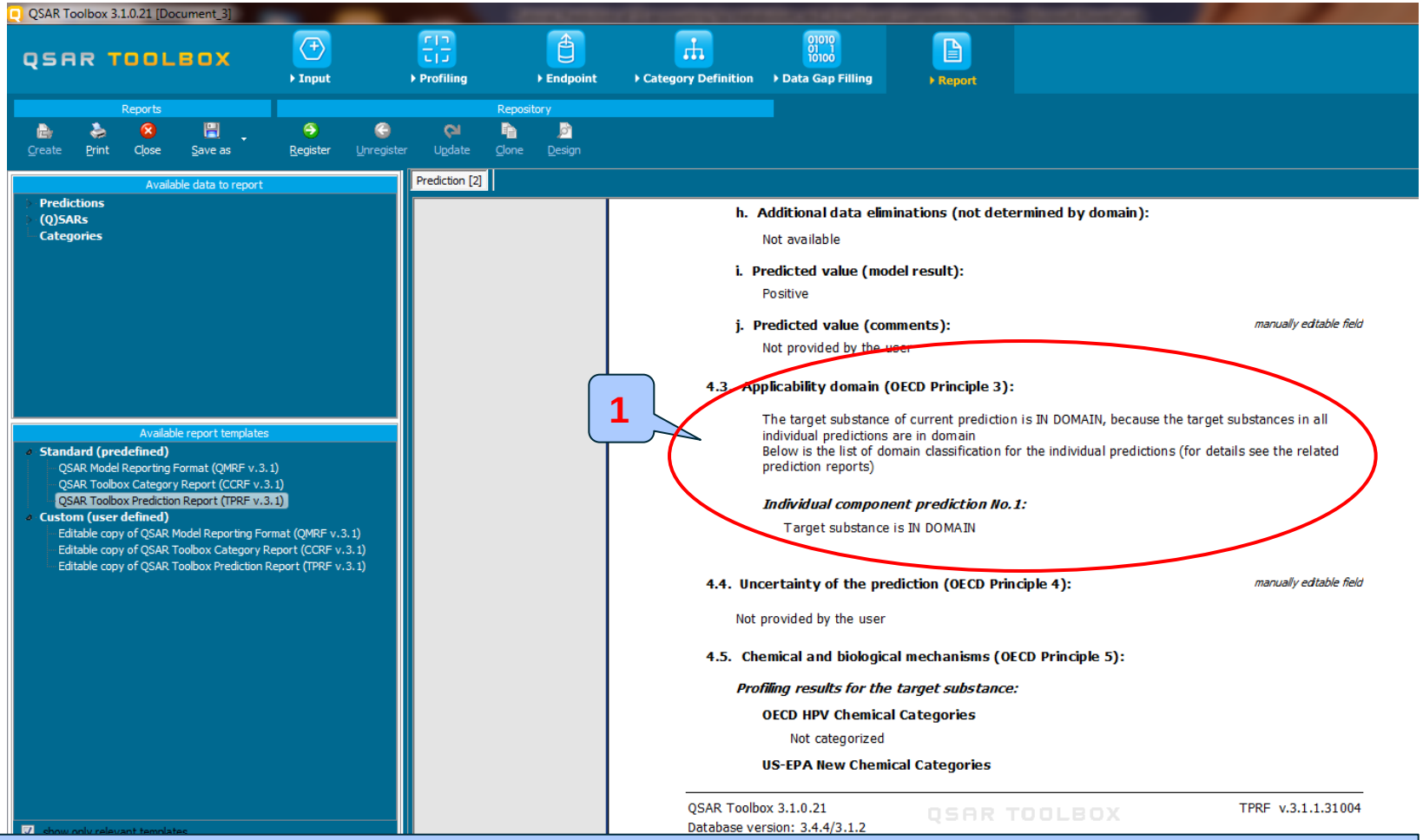
1. Summary information for prediction

Report



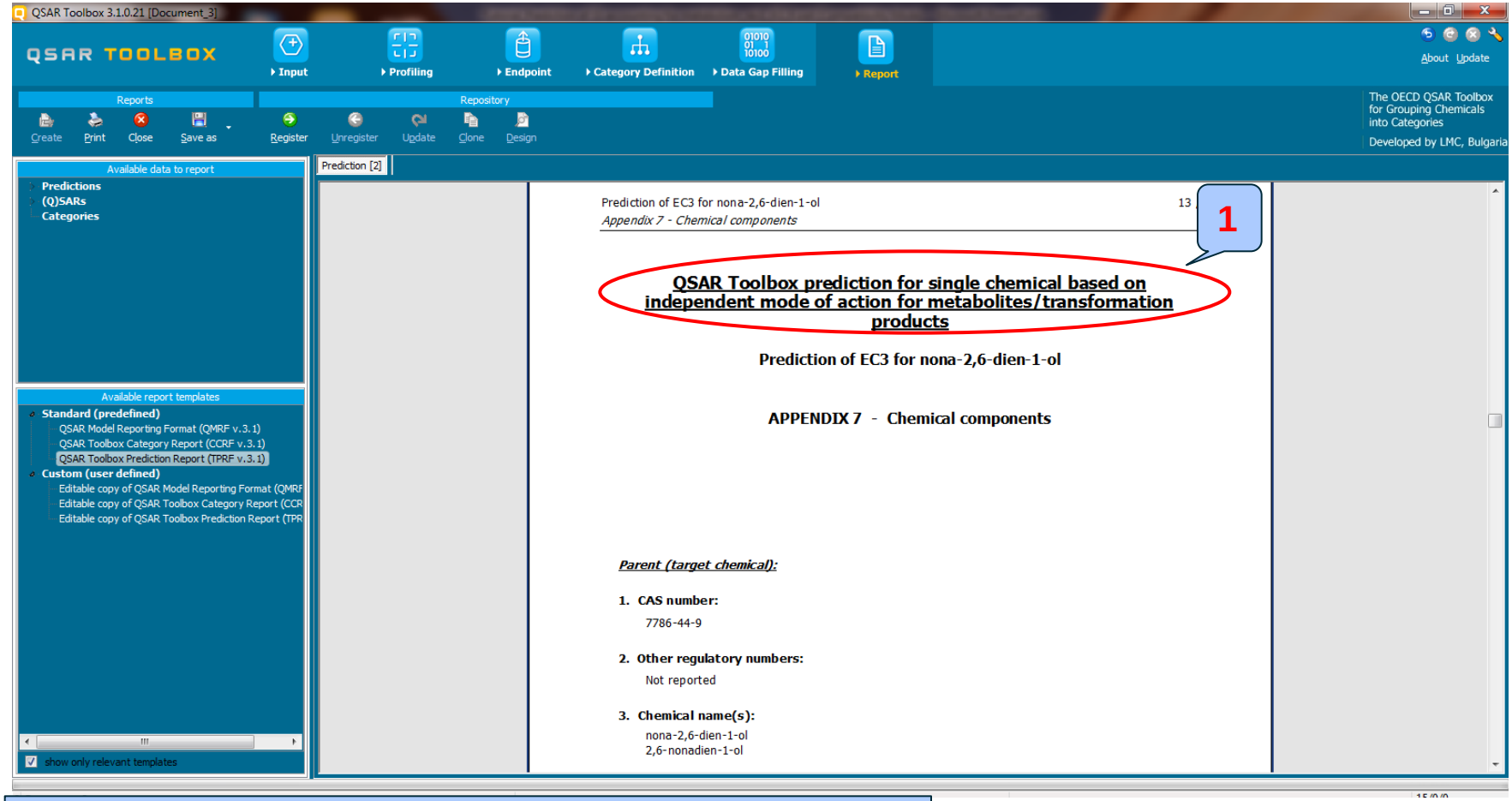
1. Predicted value

Report



1. Applicability domain
The target chemical is “In domain”, because the prediction of active metabolite is “In domain”.

Report



1. Individual component prediction