

OECD QSAR Toolbox v.3.4

Predicting skin sensitisation potential of eugenol (CAS 97-53-0) using a new categorization tool taking into account its abiotic activation

Outlook

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

Background

- This is a step-by-step presentation designed to take the user through the Toolbox workflow for predicting skin sensitization potential of eugenol using a newly implemented categorization tool taking into account its abiotic activation.

Outlook

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

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Profiling the target chemical.
- Identifying analogues of the target chemical.
- Filling data gaps for target chemical by read-across.
- Profiling target chemical taking into account its (a)biotic activation.
- Identifying analogues of the target using a new categorization functionality allowing (a)biotic activation to be taken into account.
- Filling data gaps by read across when (a)biotic activation is taken into account.

Outlook

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

The Exercise

- In this exercise we will predict the skin sensitization potential of target chemical **Eugenol** [**CAS# 97-53-0**].
- Profile and gather data for the target chemical.
- Two types categorizations are applied:
 - Identifying analogues using well-known categorization group
 - Identifying analogues based on autoxidation activation of the target illustrating new categorization functionality
- Filling data gap by read-across.

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
 - Categorization
 - 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

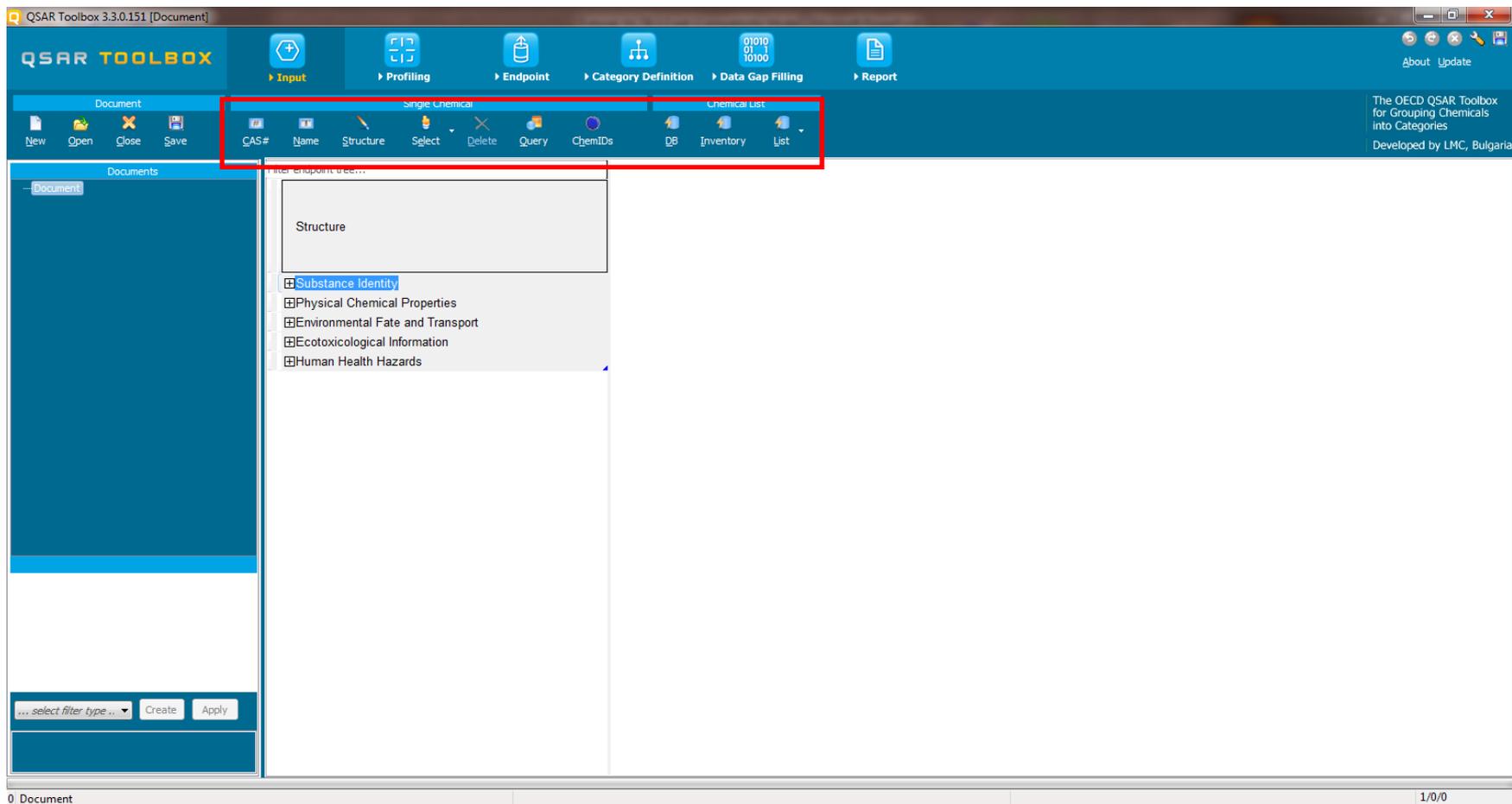
B. Group of chemicals

- 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#

The screenshot shows the QSAR Toolbox software interface. The 'Input' menu is highlighted with a red box and a callout bubble containing the number '1'. A 'Search by CAS #' dialog box is open in the foreground, showing a search input field, a 'Tautomeric sets' checkbox, and a search button. Below the search field are buttons for 'Select All', 'Clear All', and 'Invert Selection', with 'Selected 0 of 0' displayed. The dialog box also contains a table with columns: Selected, CAS, Smiles, Depiction, Names, CAS/Name, ZD/Name, and CAS/ZD.

1. Click on "CAS#"

Chemical Input

Enter CAS# of 2-methoxy-4-(2-propenyl)phenol (Eugenol)

Search by CAS #

97-53-0 Tautomeric sets

Select All Clear **1** Invert Selection Selected 1 of 1 **2** **3**

Selected	CAS	SMILES	Depiction	CAS/Name	2D/Name	CAS/2D
1. Yes	97-53-0	COc1cc(C=C)cc1O		1:: Low C 1:: Al 2:: High 1: 1:: Ba 2: 2:: C 3: 3:: C 4: 4:: C 5: 5:: Cl 6: 6:: D 7: 7:: D 8: 8:: E 9: 9:: E 10: 10:: E 11: 11:: C 12: 12:: F	1:: Low C 1:: Al 2:: High 1: 1:: E 2: 2:: RI 3: 3:: C 4: 4:: C 5: 5:: C 6: 6:: Ke 7: 7:: M 8: 8:: M 9: 9:: M 10: 10:: S 11: 11:: F 12: 12:: U	: High

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-dimensional depiction.

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
1. Yes	97-53-0	COc1cc(C		1: 2: 3: 4: 5: 6: 7: 8: 9: 10:	1:: Low C 1:: Ar 2:: High 1:: Ba 2:: C 3:: C 4:: C 5:: Cl 6:: D 7:: D 8:: E 9:: E 10:: F	1:: Low C 1:: Ar 2:: High 1:: E 2:: R 3:: C 4:: C 5:: C 6:: Ke 7:: M 8:: M 9:: M 10:: S	: High : High : High : High : High : High : High : High : High : High

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-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 screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. Below this is a toolbar with icons for Document, Single Chemical, and Chemical List operations. The main workspace is divided into several panels:

- Documents Panel:** Shows a document named 'Document_1' with a CAS number of 97-53-0. Below this, the chemical structure is visualized with its SMILES string: COc1cc(CC=C)ccc1O.
- Filter endpoint tree...:** A search filter is set to '1 [target]'. The 'Structure' field contains the chemical structure of the target.
- Substance Identity Panel:** This panel is expanded to show various identification fields:
 - CAS Number:** 97-53-0
 - Chemical IDs:** EINECS:2025891
 - Chemical Name:** eugenol (4-allyl-2-methoxyphenol), 4-allyl-2-methoxyphenol, phenol, 2-methoxy-4-allyl-, 2-methoxy-4-(prop-2-enyl)phenol, phenol, 4-allyl-2-methoxy-, 1-allyl-3-methoxy-4-methoxy-, 2-methoxy-4-(2-propenyl)phenol, 4-allyl-2-methoxyphenol, p-allylguaiacol, C10H12O2, COc1cc(CC=C)ccc1O
 - Physical Chemical Properties**
 - Environmental Fate and Transport**
 - Ecotoxicological Information**
 - Human Health Hazards**
 - Profile**

A red circle highlights the 'Substance Identity' section, specifically the list of chemical names and IDs, indicating the target chemical's 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 alerts for skin sensitization by OASIS v1.4 and clicking on “View” (see next screen shot)).

Profiling Side-Bar to Profiling

The screenshot shows the QSAR Toolbox Profiling Scheme Browser. On the left, the 'Profiling methods' list has 'Protein binding alerts for skin sensitization' checked (1). The 'Apply' button is circled in red (2). The 'Advanced' button is circled in blue (4). The main window displays a tree view of categories (3) with 'Aldehydes' selected. The right pane shows a 'Textual description' for Aldehydes, including a chemical reaction scheme for Schiff base formation and a 'Mechanistic alert' (5) and 'Structural alert' (6).

1 Highlight the profiler

2 Click "View"

3 Click over "Aldehydes" to see textual description associated with the category. In order to see more details about structural boundaries coding the rule you should click "Advanced" button (4) (see next slide)

4 Advanced

5 Mechanistic alert

6 Structural alert

Textual description:

Mechanistic Domain: Schiff base formation

Mechanistic Alert: Schiff base formation with carbonyl compounds

Structural Alert: Aldehydes

The chemical causes skin sensitization effect as a result of Schiff base formation with aldehydes:

$$R-C(=O)H + Pr-NH_2 \rightleftharpoons R-C(=N-Pr)H$$

R = H, alkyl

Aldehydes are highly reactive molecules, many of which are strong sensitizers and their direct conjugation to protein nucleophiles is thought to be responsible. Simple aldehydes react readily with the amino groups of lysine residues on proteins to form imines or Schiff bases. All aliphatic aldehydes can potentially undergo Schiff base formation with a primary amine, which is a reversible reaction (optimal at pH 3-4) and proceeds in two stages via a tetrahedral intermediate. Aryl aldehydes possess a deactivated aldehyde group and such compounds are expected to be non-sensitizers if there is no alternative functionality. In the TIMES SS model the reaction of Schiff base formation is also incorporated with COREPA model by using the parameters E_{HOMO} and Molecular weight (MW). (Form)aldehydes could be presented in parent structure or could be generated after abiotic or biotical transformations. According to De Groot et al for chemicals synthesized from formaldehyde that may still contain residues of free formaldehyde

Profiling Side-Bar to Profiling

The screenshot shows the QSAR Toolbox Profiling Scheme Browser. The left sidebar lists various profiling methods, with 'Protein binding alerts for skin sensitization' selected. The main window displays a tree view of rules, with 'Michael addition on polarised Alkenes' expanded. A callout box labeled '1' points to a 'Structural boundaries' icon in the rule definition area. Another callout box labeled '2' points to a chemical structure of a carbonyl group (C=O) with an R group and an H atom, labeled 'Structural fragment'.

1. Illustrates structural boundary coding the rule
2. Illustrates structural fragment used for defining the rule

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 OECD – general mechanistic
 - Protein Binding Potency – general mechanistic
 - Protein binding alerts for skin sensitization by OASIS v1.4 – endpoint specific

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, go through the general and endpoint specific profiling mechanisms and highlight those that are relevant to skin sensitization effect (see next screen shot).

Profiling

Profiling the target chemical

The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is open, showing options: 'Apply', 'New', 'View', and 'Delete'. The 'Apply' button is circled in red with a callout '2'. Below the menu, the 'Profiling methods' list is shown with several items checked, including 'Protein binding by OASIS v1.', 'Protein binding by OECD', 'Protein binding potency', and 'Protein binding alerts for skin sensitization', all circled in red with a callout '1'. The 'Metabolism/Transformations' section is also visible at the bottom of the list.

1. **Select** protein binding profiles from “General Mechanistic” and “Endpoint specific” group mentioned on slide 26
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 result obtained by the specific protein-binding profilers.
- No protein binding alert has been found for the target compound (eugenol) based on three protein binding profilers. Therefore no skin sensitization effect is expected.

Profiling

Profiling the target chemical

The screenshot displays the QSAR Toolbox software interface. The main window is titled "QSAR Toolbox 3.3.0.151 [Document_1]". The top menu bar includes "Input", "Profiling", "Endpoint", "Category Definition", "Data Gap Filling", and "Report". The "Profiling" menu is active, showing options for "Apply", "New", "View", and "Delete".

The "Profilng methods" panel on the left lists various methods, with the following checked:

- Protein binding by OASIS v1.3
- Protein binding by OECD
- Protein binding potency
- Protein binding alerts for skin sensitization by OASIS

The "Filter endpoint tree..." panel in the center shows a tree structure with "Protein binding by OASIS v1.3" selected. The "1 [target]" panel on the right shows the chemical structure of the target and a table of results:

Endpoint	Alert
Protein binding by OASIS v1.3	No alert found
Protein binding by OECD	No alert found
Protein binding potency	Not possible to cla...
Protein binding alerts for skin sensitization by OASIS	No alert found

A red box highlights the text: "The target chemical has no protein binding alert. In this respect no skin sensitization effect is expected". A red circle highlights the "No alert found" result for "Protein binding by OASIS v1.3" in the table.

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

The screenshot displays the QSAR Toolbox software interface. The top menu bar has 'Endpoint' selected, indicated by callout 1. The left sidebar shows a tree view of endpoints, with 'Human Health Hazards' expanded, indicated by callout 2. Within this section, 'Skin sensitization' and 'Skin sensitization ECETOC' are selected, indicated by callout 3. The main workspace shows a chemical structure and a list of selected databases, including 'Skin sensitization' and 'Skin sensitization ECETOC'. The 'Gather' button in the top menu bar is highlighted by callout 4.

1. Click on "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.

The screenshot shows the QSAR Toolbox software interface. The 'Endpoint' menu is active, and a 'Read data?' dialog box is open. The dialog box has three radio buttons: 'All endpoints' (selected), 'Choose...', and 'from Tautomers'. There are 'OK' and 'Cancel' buttons. A blue callout bubble with the number '1' points to the 'OK' button. Below the dialog box, a blue box contains the text: '1. Click "OK" to read all available data'.

Endpoint Gather data

The screenshot displays the QSAR Toolbox interface. The 'Endpoint' tab is active. In the 'Databases' panel on the left, 'Skin sensitization' is selected. The 'Filter endpoint tree...' window shows a hierarchical tree of endpoints. Under 'Skin' > 'In Vivo', the 'LLNA' endpoint is highlighted with a blue callout box containing the number '1'. The result for 'LLNA' is '(1/1) M: Positive'. A chemical structure of the target molecule is shown in the top right corner of the main window.

1. Positive experimental data for skin sensitization is found for the target chemical.

Endpoint Gather data

The screenshot shows the QSAR Toolbox software interface. The main window is titled 'Endpoint' and features a menu bar with options like 'Data', 'Import', 'Export', 'Delete', and 'Tautomerize'. Below the menu bar, there are several toolbars and a main workspace. On the left, there is a 'Databases' panel with a tree view of various categories, including 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', and 'Human Health Hazards'. The 'Human Health Hazards' category is expanded, showing a list of sub-categories like 'Acute Oral Toxicity database (ChemID)', 'Bacterial mutagenicity ISSSTY', etc. In the center, there is a 'Filter endpoint tree...' panel with a tree view of the selected endpoint, 'EC3'. To the right of the tree, there is a 'Structure' panel showing a chemical structure of a target molecule. A 'Data points' popup window is open, displaying a table of data points. The table has columns for '#', 'Endpoint', 'Value', 'Original value', 'Organ', 'Reference source', 'Phylum (common name)', 'Phylum', 'Test method / Data source', 'Type of method', 'Year', and 'Test organisms'. The first row is highlighted, showing data for 'EC3' with a value of 'Positive (Skin sensitisation II (ECETOC))' and a reference source of 'Dermatitis, 16 (4): 1-46'. A red speech bubble with the number '1' points to the 'Value' cell of this row.

#	Endpoint	Value	Original value	Organ	Reference source	Phylum (common name)	Phylum	Test method / Data source	Type of method	Year	Test organisms
1	EC3	Positive (Skin sensitisation II (ECETOC))	13 % (Skin sensitization EC3 (ratio))	Skin	Dermatitis, 16 (4): 1-46	Vertebrates	Chordata	LLNA	in Vivo	2005	mouse

1. **Double-click** on the cell displays metadata information for the observed data

Recap

- The first module, introduces the target chemical, ensure for correctness of the structure.
- The second module shows that there is no protein binding alert for the target chemical.
- In the third module, you have found that the target chemical has positive skin sensitization data.
- In the further read-across analysis we will try to reproduce positive skin sensitization data.
- The study continues with identifying analogues and applying read-across.

Outlook

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

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>

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.

Basic guidance for category formation and assessment

Suitable categorization phases:

1. Structure-related profilers (for primary categorization).
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**

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 the identified analogue is provided below. In order to preserve the basic functional groups available within the molecule: Allyl, Ether and Phenol, OFG is used for categorization purposes. US-EPA and ECOSAR are not used because they omit the other two important functionalities: Allyl and Ether. Str. similarity identifies small set of analogues and apparently could not be used for categorization.

Phase I categorization in Toolbox

128 analogues are identified.

102 analogues are identified

7 analogues are identified (in case all categories are preserved)
11 analogues are identified (in case Allyl, Ether and Phenol are preserved)

21 analogues are identified

Structural similarity, Dice ACF, 50%

Category Definition

Define category by OFG

1. Select "OFG"

2. Click "Define"

3. Combination of all seven organic functional groups identified seven analogues only (3). In order to expand the initial group the categories "Allyl", "Ether" and "Phenol" are used only.

4. Click "Cancel". See next slide

Category Definition

Define category by OFG

1. **Select** "OFG";
2. **Click** "Define" button;
3. **Select** "Alkenes", "Alkoxy" "Aryl" and "Precursors quinoid compounds" (highlighted in blue) and **click** arrow down  to remove them. They are moved in the panel down called "All profiles";
4. Arrow down
5. "Allyl", "Ether" and "Phenol" should remain in the upper panel only
6. **Click** "OK" button

Category Definition

Define category by OFG

The screenshot displays the QSAR Toolbox interface with the 'Category Definition' workflow active. The 'Define category name' dialog box is open, showing the category name '>Ether<AND>Phenol (Organic Functional groups)'. A 'Warning' dialog box is also open, asking 'You have selected different from target categories! Do you want to continue?'. A 'Read data?' dialog box is also open, with 'All endpoints' selected and 'from Tautomers' checked. Red callout boxes with numbers 1, 2, and 3 point to the 'Yes' button in the warning dialog, the 'OK' button in the 'Define category name' dialog, and the 'OK' button in the 'Read data?' dialog, respectively.

1. A message informs for different categories from the target have been selected. **Click "Yes"**
2. **Click "OK"** to confirm the name of the category
3. **Click "OK"** to read all available data

Category Definition

Gather data for analogues chemicals

The screenshot shows the QSAR Toolbox interface with the 'Category Definition' workflow active. The 'LLNA' method is selected, and a data matrix is displayed. The first row shows chemical structures for target and analogues. The second row shows experimental data for the LLNA method, with a red oval highlighting the data row. A red box contains the text: 'The analogues along with their experimental data appears on data matrix'.

LLNA	1 (target)	2	3	4	5	6	7	8
Structure								
Substance Identity								
Human Health Hazards								
Sensitisation								
Skin								
In Vivo								
LLNA	(10/10)	M: Positive						
EC3								

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Categorization
 - **Data gap filling without taken into account metabolism**

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.

Data gap filling

Apply Read-across

The screenshot displays the QSAR Toolbox interface during a data gap filling operation. The 'Data Gap Filling Method' is set to 'Read-across'. A dialog box titled 'Possible data inconsistency' is open, showing the following details:

- Scale/Unit: Skin sensitization EC3(ratio) and Skin sensitization I (Oasis)
- Gap filling scale/unit:
 - Skin sensitization EC3(ratio)
 - Skin sensitization I (Oasis)
 - Skin sensitization II (ECETOC)
 - Skin Sensitization (Danish EPA)
- converted data:
 - 9 from scale Skin sensitization EC3(ratio)
 - 1 from scale Skin sensitization I (Oasis)
- Selected [10/10] points

The main interface shows a tree view of 'Human Health Hazards' with 'LLNA EC3' selected under 'In Vivo'. A red circle highlights the 'M: Positive' cell in the table, and another red circle highlights the 'Apply' button in the top left corner.

1. **Click** on the cell corresponding to "Sensitization>>Skin>>In Vivo>>LLNA>>EC3"
 2. **Select** "Read-across"
 3. **Click** "Apply"
- Additional window informing for more than one data/scale has been used appears.
More details about scale definitions is provided on next slide.

Data gap filling

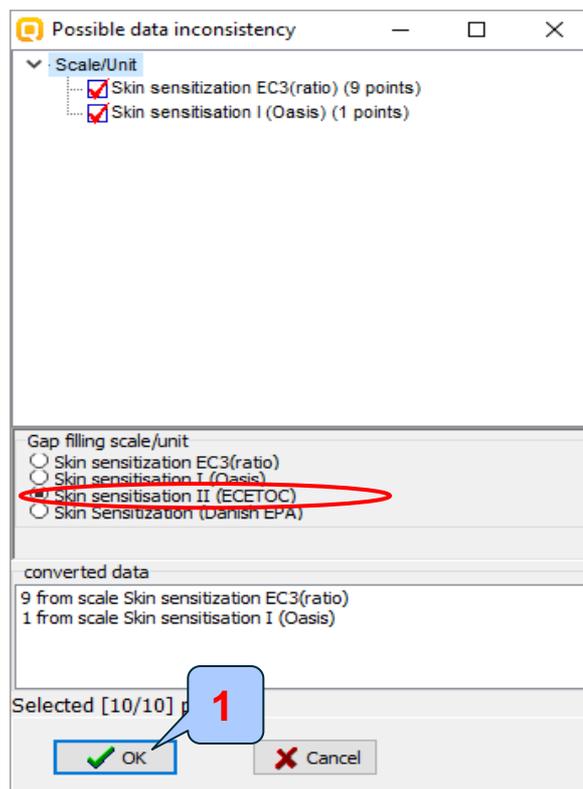
Scale definition

- Skin sensitisation is a “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

Scale definition

Back to our example



Verify that the default scale "Skin sensitisation II (ECETOC)" is selected
1. Click "OK"

Data gap filling

Read-across

- The resulting plot place the experimental results of all analogues (Y axis) according to a descriptor (X axis) which by default is 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

Read-across

The screenshot shows the QSAR Toolbox interface with the following details:

- Navigation Bar:** Input, Profiling, Endpoint, Category Definition, **Data Gap Filling**, Report.
- Left Panel:**
 - Data Gap Filling Method:** Read-across (selected), Trend analysis, (Q)SAR models.
 - Target Endpoint:** Human Health Hazards Sensitisation Skin In Vivo LNA EC3.
- Main Area:**
 - Structure:** LNA (10/10) M. Positive
 - Grid:** 8 columns of chemical structures with predicted values: M. Positive, M. Positive.
 - Descriptors:** Prediction
 - Graph:** Scatter plot of EC3 (obs.) vs log Kow. Text: "Read across prediction of EC3, taking the highest mode from the nearest 5 neighbours, based on 7 values from 7 neighbour chemicals, Observed target value: 'Positive', Predicted target value: 'Positive'".
 - Descriptor X:** log Kow
 - Right Panel:** Accept prediction, Return to matrix, Select/filter data, Selection navigation, Gap filling approach, Descriptors/data, Model/(Q)SAR, Calculation options, Visual options, Information, Miscellaneous.

Initial graph without any subcategorizations

Data Gap Filling Subcategorizations

In this example, the following subcategorizations are applied in order to eliminate dissimilar analogues (see slide 44):

- Organic functional group (US-EPA) – phase I is repeated in order to eliminate multifunctional analogues (subcategorization 1)
- Protein binding alerts for skin sensitization by OASIS v1.4 (subcategorization 2)

See next screen shots.

Data gap filling

Subcategorization 1: Organic functional groups (US EPA)

The screenshot displays the 'Subcategorization' window in the QSAR Toolbox. On the left, the 'Grouping methods' list has 'Organic functional groups (US EPA)' selected, indicated by a red circle and a '2'. The 'Adjust options' dialog shows 'Target' as 'Alcohol, olefinic attach [-OH]' and 'Differ from target by' set to 'At least one category', with 'Analogues' listed below. A table of predicted analogues is shown with columns for log Kow and predicted target values. A scatter plot at the bottom shows the log Kow distribution of the target and analogues. The 'Accept prediction' panel on the right has 'Return to matrix' selected, and 'Select/filter data' is highlighted with a red circle and a '1' callout.

log Kow	Target	Analogue	Predicted Target
2.80	Alcohol, olefinic attach [-OH]	Alcohol, olefinic attach [-OH]	Positive
2.90	Alcohol, olefinic attach [-OH]	Aliphatic Carbon [-CH2-]	Positive
3.00	Alcohol, olefinic attach [-OH]	Aliphatic Carbon [-CH2-]	Positive
3.10	Alcohol, olefinic attach [-OH]	Aliphatic Carbon [-CH2-]	Positive
3.20	Alcohol, olefinic attach [-OH]	Aliphatic Carbon [-CH2-]	Positive
3.30	Alcohol, olefinic attach [-OH]	Aliphatic Carbon [-CH2-]	Positive
3.40	Alcohol, olefinic attach [-OH]	Aliphatic Carbon [-CH2-]	Positive
3.50	Alcohol, olefinic attach [-OH]	Aliphatic Carbon [-CH2-]	Positive
3.60	Alcohol, olefinic attach [-OH]	Aliphatic Carbon [-CH2-]	Positive

The identified analogues are similar to target chemical with respect to Organic functional groups (US EPA)

1. Open "Select/filter data/Subcategorize" 2. Select "Organic functional groups (US EPA)"

Data gap filling

Subcategorization 2: Protein binding alerts for skin sensitization by OASIS v.1.4

The screenshot displays the QSAR Toolbox interface during a subcategorization task. On the left, the 'Subcategorization' window is active, showing a list of grouping methods. The method 'Protein binding alerts for skin sensitization' is highlighted and circled in red. The 'Adjust options' panel for this method shows 'No alert found' for both 'Target' and 'Analogues', with the latter also circled in red. The main window shows a table of chemical structures with their predicted values (M. Positive) and a scatter plot of log Kow values. A red text overlay states: 'No protein binding alerts are identified for target and analogues, which can not be explained by positive experimental data found. In this respect metabolism should be taken into account (see next slide)'.

1. Select "Protein binding alerts for skin sensitization by OASIS v1.4"

Data gap filling

Subcategorization when metabolism is taken into account

1

2

Read across prediction of EC3, taking the highest mode from the nearest 5 neighbours, based on 7 values from 7 neighbour chemicals, Observed target value: 'Positive', Predicted target value: 'Positive'

The metabolites of target chemical and its analogues possess same distribution of protein binding alerts. This could explain positive experimental data and respectively positive read-across prediction

Chemical	log Kow	Target
1 [target]	~2.75	M: Positive
2	~2.85	M: Positive
3	~2.95	M: Positive
4	~3.05	M: Positive
5	~3.15	M: Positive
6	~3.25	M: Positive
7	~3.35	M: Positive

Now subcategorization will be applied accounting for autoxidation simulation in combination with "Protein binding alerts" on the target and its analogues. Follow the steps:

1. The "Protein binding alerts for skin sensitization by OASIS v1.4" has been already selected;
2. **Click over** "Autoxidation simulator"

Data gap filling

Interpreting Read-across

- In this example the target and all analogues have no protein binding alerts.
- All analogues along with the target possess same distribution of positive protein binding alerts when autoxidation is taken into account.
- The latter could explain the positive experimental data of the target compound.
- Once ready go back to data matrix, when click on "Return to matrix" button (see next slide).

Data gap filling

Return to data matrix

The screenshot displays the QSAR Toolbox software interface during a data gap filling process. The main window is titled "Data Gap Filling" and shows a workflow with steps: Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The "Data Gap Filling" step is active, showing a table of 8 chemical structures. Below the table is a plot of log Kow values, with a red dot indicating the target chemical and a blue dot indicating the predicted chemical. A red callout box with the number "1" points to the "Return to matrix" button in the right-hand panel.

Return to matrix

Select/filter data

- Subcategorize
- Mark chemicals by descriptor value
- Filter points by test conditions
- Mark focused chemical
- Mark focused points
- Remove marked chemicals/points
- Clear existing marks
- Selection navigation
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
- Information
- Miscellaneous

1. Click "Return to matrix"

10 Allyl<AND>Ether<AND>Phenol (Organic Functional groups)

Create prediction by gap filling

1/1/0

Data gap filling

Next actions

- The study continues with second data gap filling where a category of analogues is defined by using new categorization functionality allowing to define category accounting for (a)biotic activation of the target.
- Before proceeding with Data gap filling the following two items will be illustrated intended to explain and support the analysis. Following the steps is not necessary.
 - Multiplication of the target chemical
 - Profiling the parent and metabolites based on (a)biotic activation

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Categorization
 - Data gap filling without taken into account metabolism
 - **Multiplication of the target chemical**

Multiplication of the target chemical

- Multiplication of the target chemical could be accomplished by two ways:
 - In the **Input** section outside data gap filling module (scenario 1)
 - slide 66
 - In the **Profiling** section (scenario 2) – slide 69
- Both scenarios will be demonstrated on next few slides

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Categorization
 - Data gap filling without taken into account metabolism
 - **Multiplication of the target chemical**
 - **In the Input section (scenario 1)**

Multiplication of target chemical in the Input section (scenario 1)

The screenshot illustrates the steps to multiply a target chemical in the QSAR Toolbox. The interface shows the 'Input' section with a target chemical selected. A context menu is open, and the 'Autoxidation simulator' option is chosen. The resulting metabolites are displayed in a tree-like structure, including the original chemical and various oxidation products.

1. Go to "Input"
2. Click over the SMILES of the target chemical and perform **right click** on it, then
3. **Select** "Multiplication-Metabolism/Transformations"
4. **Select** "Autoxidation simulator"
5. Generated metabolites appeared in a tree-like form. They could be visualized in two modes. See next slide

Multiplication of target chemical in the Input section (scenario 1)

Visualization the set of parent and metabolites

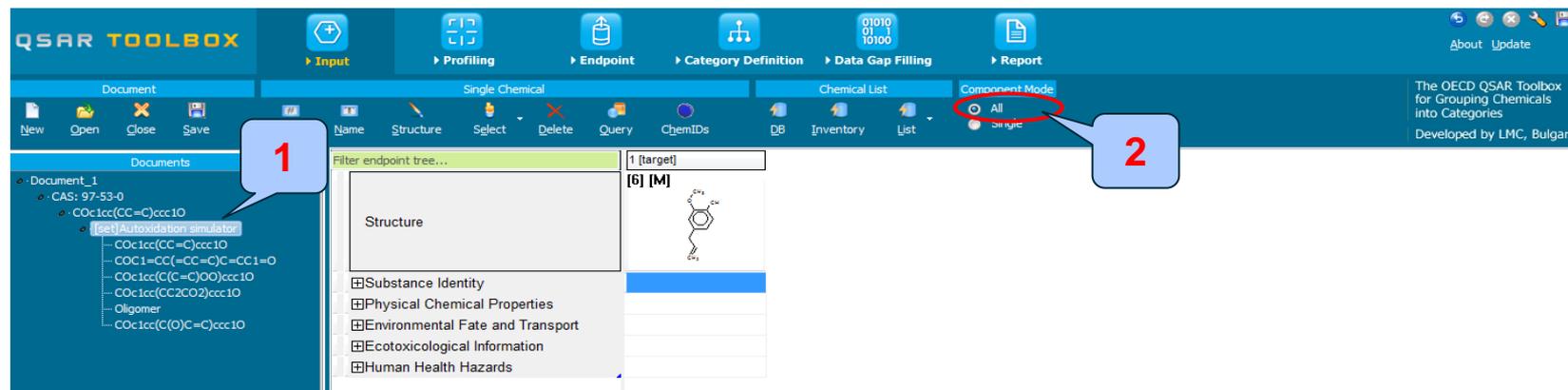
- Two component modes are implemented:
 - **Set Mode** - all metabolites are analysed as a package
 - **Individual Component Mode** - each metabolite is analysed individually

(graphical illustration of both modes is provided next screenshot)

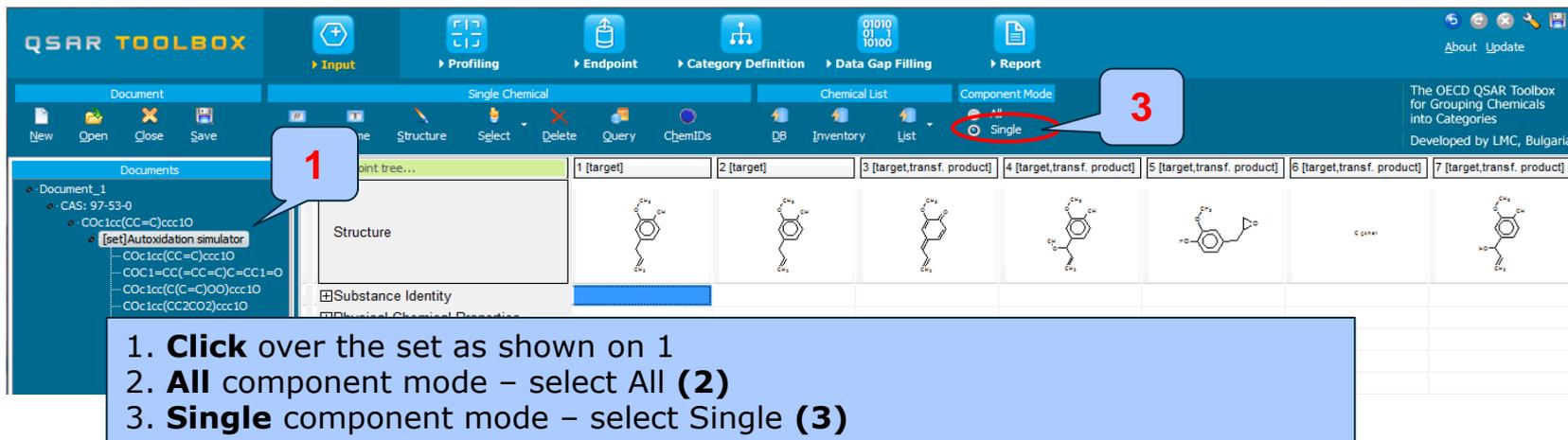
Multiplication of target chemical in the Input section (scenario 1)

Visualization the set of parent and metabolites

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



- Single Mode – each metabolite is analyzed individually



Protein binding result for parent and metabolites multiplied in the Input section

The profiling result indicates no protein binding alerts for target chemical. However, three of simulated AO metabolites exhibit interaction with proteins via three different protein binding mechanisms (Michael Addition, Radical reactions, and SN2).

Autoxidation simulator

The screenshot shows the QSAR Toolbox Profiling interface. The 'Protein binding alerts for skin sensitization by OASIS v1.4' checkbox is checked. The table below displays the results for the parent chemical and seven metabolites.

Endpoint	1 [target]	2 [target]	3 [target,transf. product]	4 [target,transf. product]	5 [target,transf. product]	6 [target,transf. product]	7 [target,transf. product]
Protein binding alerts for skin sensitization by OASIS v1.4	No alert found	No alert found	Michael Addition Michael Addition >... Michael Addition >...	Radical reactions Radical reactions ... Radical reactions ...	SN2 SN2 >> Ring openi... SN2 >> Ring openi...	(N/A)	No alert found

Once the chemical is multiplied in the Input section and metabolites are visualized (distributed on data matrix) via "Single mode" 1. **Go** to "Profiling"; 2. **Check** "Protein binding alerts for skin sensitization by OASIS v1.4"; 3. **Click** "Apply"

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Categorization
 - Data gap filling without taken into account metabolism
 - **Multiplication of the target chemical**
 - In the Input section (scenario 1)
 - **In the Profiling section (scenario 2)**

Multiplication of target chemical in the Profiling section (scenario 2)

1 Profiling

4 Apply

2 Protein binding alerts for skin sensitization by OASIS v1.3

3 Autoxidation simulator

5 5 metabolites

Generated metabolites

Reference: Autoxidation simulator

1 CAS# N/A	2 CAS# N/A	3 CAS# N/A	4 CAS# N/A	5 CAS# N/A
---------------	---------------	---------------	---------------	---------------

Save to smi Search OK

Filter endpoint tree... 1 [target]

Structure

Substance Identity

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Human Health Hazards

Profile

- Endpoint Specific
- Protein binding alerts for skin sensitization by OA...
- Metabolism/Transformations
 - Autoxidation simulator
 - Endpoint Specific

No alert found

5 metabolites

Profiling methods

Select All Unselect All Invert About

- Eye irritation/corrosion Inclusion rules by BFR
- In vitro mutagenicity (Ames test) alerts by ISS
- In vivo mutagenicity (Micronucleus) alerts by ISS
- Keratinocyte gene expression
- Oncologic Primary Classification
- Protein binding alerts for skin sensitization by OASIS v1.3
- Respiratory sensitization
- Retinoic Acid Receptor Binding
- rER, Expert System ver. 1 - USEPA
- Skin irritation/corrosion Exclusion rules by BFR
- Skin irritation/corrosion Inclusion rules by BFR

Empiric

- Chemical elements
- Groups of elements
- Lipinski Rule Oasis
- Organic Functional groups
- Organic Functional groups (nested)
- Organic functional groups (US EPA)
- Organic functional groups, Norbert Haider (checkmol)
- Tautomers unstable

Toxicological

- Repeated dose (HESS)

Metabolism/Transformations

Select All Unselect All Invert About

Documented

- Observed Mammalian metab...
- Observed Microbial metabi...
- Observed Rat In vivo metab...
- Observed Rat Liver SP...

Simulated

- Autoxidation simulator
- Autoxidation simulator (alkaline medium)
- Dissociation simulator
- Hydrolysis simulator (acidic)
- Hydrolysis simulator (basic)

1. Once the chemical is entered into the system into a new document, **go** to "Profiling"
2. **Select** "Protein binding for skin sensitization by OASIS v1.4" profiler from **Endpoint specific** group
3. **Select** "Autoxidation simulator" from **Metabolism/Transformations** menu
4. **Click** "Apply"
5. **Double click** over the cell corresponding to **5 metabolites** to see the generated metabolites

Protein binding result for parent and metabolites multiplied in the Profiling section

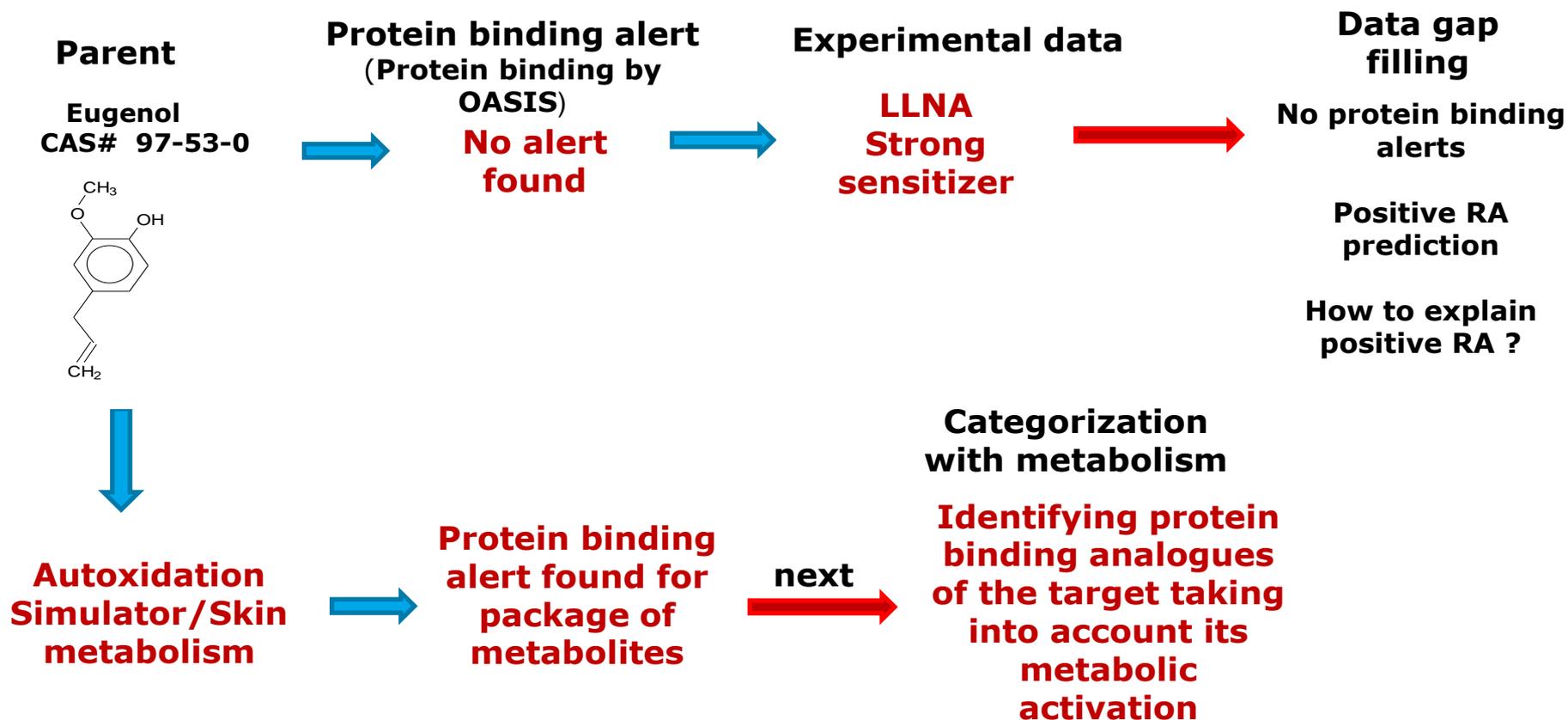
The screenshot shows the QSAR Toolbox interface with the Profiling section active. The 'Autoxidation simulator' node is selected in the endpoint tree, and its results are displayed in a table. Two callouts, labeled 1 and 2, highlight the selection of the node and the double-clicking of a cell in the results table, respectively.

Structure	1 [target]
<chem>Cc1ccc(C)cc1</chem>	
Substance Identity	
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	
Human Health Hazards (1/4)	M: Negative, Positive, Positive, Positive
Profile	
Endpoint Specific	
Protein binding alerts for skin ...	No alert found
Metabolism/Transformations	
Autoxidation simulator	5 metabolites
Endpoint Specific	
Protein binding alerts for ...	1 x (N/A) 1 x Michael Addition 1 x Michael Addition >> Quinoide type compo... 1 x Michael Addition >> Quinoide type compo... 1 x No alert found 1 x Radical reactions 1 x Radical reactions >> Free radical formation 1 x Radical reactions >> Free radical formatio... 1 x SN2 1 x SN2 >> Ring opening SN2 reaction 1 x SN2 >> Ring opening SN2 reaction >> Ep...

1. Open node "Autoxidation simulator"

2. Double click over the cell to investigate the profiling results obtained for the metabolites

Recap



Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Categorization
 - Data gap filling without taken into account metabolism
 - Multiplication of the target chemical
 - **Categorization applying metabolism**

Categorization applying metabolism

The advantages of the new functionality are:

- Application of metabolism for analogues identification during process of categorization. Metabolism could be used for primary categorization.
- Possibility to expand the chemical domain of the category and to identify analogues based on metabolism approach.
- Before proceeding with categorization accounting for (a)biotic activation of the target input the target in a new document (see next slide).

Categorization applying metabolism

The screenshot shows the QSAR Toolbox software interface. The 'Input' section is active, and a search dialog box is open. The search dialog box has a search bar with '97-53-0' entered and an 'OK' button highlighted. Below the search bar is a table with the following columns: Selected, CAS, Smiles, Depiction, Names, CAS/Name, 2D/Name, CAS/2D. The table contains one entry for CAS 97-53-0 with a chemical structure and a list of names.

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
1. Yes	97-53-0	COc1cc(C=C)ccc1O		1: 2: 3: 4: 5: 6: 7: 8: 9: 10: 11:	1:: Low 1:: Ar 2:: High 1:: Br 2:: Cl 3:: Cl 4:: Cl 5:: Cl 6:: D 7:: Di 8:: Et 9:: Et 10:: F	1:: Low 1:: Ar 2:: High 1:: Et 2:: R 3:: Cl 4:: Cl 5:: Cl 6:: K 7:: M 8:: M 9:: M 10:: S	High Ar High Et R Cl Cl Cl K M M S

1. Go to "Input" section; number of the target
2. Click on "New" button;
3. Click on "CAS#" button;
4. Enter the CAS
5. Click "OK"

Categorization applying metabolism

The screenshot shows the QSAR Toolbox interface with the following elements highlighted:

- 1:** The "Define with metabolism" button in the top toolbar.
- 2:** The "Protein binding alerts for skin sensitization by OASIS v1.4" option in the left sidebar.
- 3:** The "Define with metabolism" button in the top toolbar (repeated).
- 4:** The "Autoxidation simulator" option in the "select metabolism" dialog box.
- 5:** The "OK" button in the "select metabolism" dialog box.

1. Go to "Category Definition" section;
2. Click on "Protein binding alerts for skin sensitization by OASIS v1.4";
3. Click on "Define with metabolism" button;
4. Select "Autoxidation simulator"
5. Click "OK" (additional window appears, see next slide)

Note: In some cases this process may take longer time, due to not indexed results for the rest of the simulators.

Categorization applying metabolism

Grouping methods

- Bioaccumulation - metabolism half-lives
- Biodegradation fragments (BioWIN MITI)
- Carcinogenicity (genotox and nongenotox)
- DART scheme v.1.0
- DNA alerts for AMES, MN and CA by OASIS
- Eye irritation/corrosion Exclusion rules by OASIS
- Eye irritation/corrosion Inclusion rules by OASIS
- in vitro mutagenicity (Ames test) alerts by OASIS
- in vivo mutagenicity (Micronucleus) alerts by OASIS
- Keratinocyte gene expression
- Oncologic Primary Classification
- Protein binding alerts for Chromosomal aberrations
- Protein binding alerts for skin sensitization by OASIS
- Respiratory sensitisation
- Retinoic Acid Receptor Binding
- rER Expert System ver. 1 - USEPA
- Skin irritation/corrosion Exclusion rules by OASIS
- Skin irritation/corrosion Inclusion rules by OASIS

Filter endpoint tree... 1 [target]

Structure

Substance Identity

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Package of Protein profiling result for parent and its autoxidation products

Autoxidation simulator<WTH>Protein binding alerts for skin sensitization by OASIS v1.3

Target(s) profiles

(N/A)

Michael Addition

Michael Addition >> Quinoid type compounds

Michael Addition >> Quinoid type compounds >> Quinone methide(s)/mines; Quinoid oxime structure; Nitroquinones, Naphthoquinones

No alert found

Radical reactions

Radical reactions >> Free radical formation

Radical reactions >> Free radical formation >> Hydroperoxides

SN2

SN2 >> Ring opening SN2 reaction

SN2 >> Ring opening SN2 reaction >> Epoxides, Aziridines and Sulfuranes

All profiles

Acylation

Acylation >> Acyl transfer via nucleophilic addition reaction

Acylation >> Acyl transfer via nucleophilic addition reaction >> Carbodimides

Acylation >> Acyl transfer via nucleophilic addition reaction >> Isocyanates, Isothiocyanates

Acylation >> Direct acylation involving a leaving group

Acylation >> Direct acylation involving a leaving group >> (Thio)Acetates

Acylation >> Direct acylation involving a leaving group >> (Thio)Acyl and (thio)carbamoyl halides and cyanides

Acylation >> Direct acylation involving a leaving group >> Anhydrides (sulphur analogues of anhydrides)

Acylation >> Direct acylation involving a leaving group >> Azlactones and unsaturated lactone derivatives

Acylation >> Direct acylation involving a leaving group >> Carbamates

Acylation >> Direct acylation involving a leaving group >> Diacyl peroxides

Acylation >> Direct acylation involving a leaving group >> Sulphur analogues of diacyl peroxides

Acylation >> Direct acylation involving a leaving group >> N-Acylated heterocycles

Acylation >> Direct acylation involving a leaving group >> N-Acyloxysuccinimides

Acylation >> Direct acylation involving a leaving group >> N-Acylsulfonylureas

Combine profiles logically

AND OR

Invert result Strict

1

OK

Cancel

1. Select "OK"

Categorization applying metabolism

The screenshot shows the QSAR Toolbox interface with the 'Category Definition' step active. The 'Filter endpoint tree...' panel is open, showing a tree structure of toxicity endpoints. The 'EC3' endpoint is selected, and a table below it shows data for five chemical structures. A red box highlights the text 'Category of 3 analogues has been defined with EC3 data'.

Endpoint	1 [target]	2	3	4	5
ACB					
(1/1) M: Negative					
(1/1) M: Positive					
(2/4) M: Positive, Positiv...					M: 8E3 µg/cm2
(4/4) M: Positive		M: Positive	M: Positive	M: Positive	
(1/1) M: Positive					

The forthcoming two slides illustrates how consistent is the identified category with respect to protein binding alerts when metabolism is taken into account

Categorization applying metabolism

Profiling results for parent and metabolites

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Profiling methods' and 'Metabolism/Transformations' sections. The main window shows a tree view of endpoints and a table of results for five metabolites. A red box highlights the following result:

Protein binding alerts for skin sensitization by OASIS v1.4	No alert found				
---	----------------	----------------	----------------	----------------	----------------

The text inside the red box states: "The profiling results indicates no protein binding alerts for target chemical. There are protein binding alerts identified in the autoxidation products."

1. Go to "Profiling";
2. Check "Protein binding alerts for skin sensitization by OASIS v1.4"
3. Check "Autoxidation simulator";
4. Click "Apply"

Categorization applying metabolism Profile statistic

The screenshot displays the QSAR Toolbox interface with the 'Profile statistic' window open. The window shows a table of categories and their counts, a bar chart for the selected category, and a list of chemicals. A red circle highlights the 'Protein binding alerts for skin sensitization by OASIS' category in the table below the graph, indicating that all metabolites share this distribution.

#	Category	Count
1	(N/A)+Michael Addition+Michael Addition >>	5

Chemical	Protein binding alerts for skin sensitization by OASIS
1 CAS# 97-53-0	1 x (N/A)
2 CAS# 186743-26-0	1 x Michael Addition
3 CAS# 186743-25-9	1 x Michael Addition
4 CAS# 186743-24-8	1 x Michael Addition
5 CAS# 97-53-1	1 x Michael Addition

All metabolites have same distribution of the protein binding alerts after metabolic transformation

Categorization applying metabolism

The screenshot displays the QSAR Toolbox software interface during the 'Categorize' step. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The main workspace is divided into several sections:

- Grouping methods:** A list of various toxicological endpoints such as 'Bioaccumulation - metabolism half-lives', 'Carcinogenicity', and 'Repeated Dose Toxicity'.
- Filter endpoint tree...:** A hierarchical tree view showing the selected endpoint 'EC3' under 'Sensitization'.
- Structure:** A view showing five chemical structures corresponding to the data rows in the table.
- Data Table:** A table with 5 columns (labeled 1 to 5) and multiple rows. The row (4/4) is highlighted with a red box, indicating the next action: 'Apply read-across for EC3 LLNA data for 3 analogue chemicals'. The table contains data such as '(1/1) M: Positive', '(2/4) M: Positive, Positiv...', and 'M: 8E3 µg/cm2'.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Categorization
 - Data gap filling without taken into account metabolism
 - Categorization applying metabolism
 - **Data gap filling handling metabolism of the target chemical**

Data gap filling

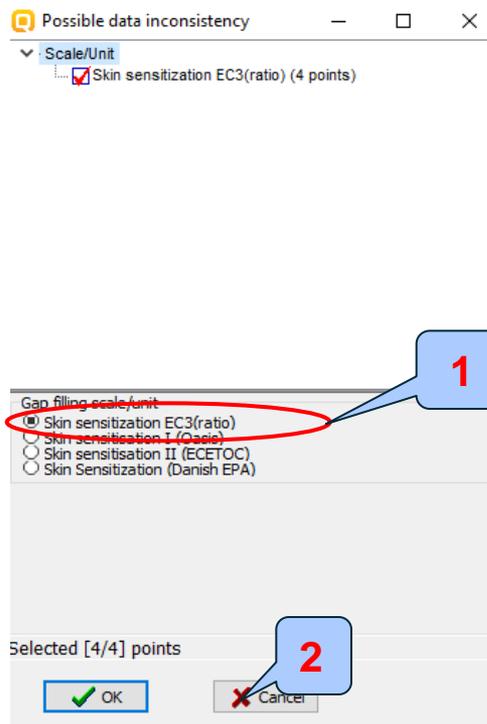
Apply Read across

The screenshot displays the QSAR Toolbox interface during a data gap filling session. The top navigation bar includes buttons for 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows the 'Data Gap Filling Method' menu, with 'Read-across' selected. The main workspace is divided into a 'Filter endpoint tree...' on the left and a data table on the right. The table has columns for target chemicals and rows for various endpoints. A red circle highlights the 'Read-across' button in the sidebar (labeled '3'), and another red circle highlights the 'Apply' button in the top left (labeled '2'). A red arrow points from the 'Apply' button to a cell in the table corresponding to 'Sensitization > Skin > In Vivo > LLNA > EC3' (labeled '1').

1. **Click** on the cell corresponding to "Sensitization>>Skin>>In Vivo>>LLNA>>EC3" for the target chemical 2. **Select** "Read-across" 3. **Click** "Apply"

Data gap filling

Scale definition



1. **Select** scale "Skin sensitisation EC3 (ratio)"
2. **Click** "OK"

Data gap filling

Read-across

QSAR TOOLBOX

Input Profiling Endpoint Category Definition **Data Gap Filling** Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Read-across

Target Endpoint
Human Health Hazards Sensitisation Skin In Vivo LLNA EC3

Structure

1 [target] 2 3 4

EC3 (4/4) M. Positive M. Positive M. Positive M. Positive

Descriptors Prediction

Read across prediction of EC3, taking the average from the nearest 5 neighbours, based on 3 values from 3 neighbour chemicals, Observed target value: 13.0 %, Predicted target value: 20.7 %

Descriptor X: log Kow

Accept prediction
Return to matrix
Select/filter data
Selection navigation
Gap filling approach
Descriptors/data
Model/(Q)SAR
Calculation options
Visual options
Information
Miscellaneous

Initial graph without any subcategorizations

Data Gap Filling Subcategorizations

In this second data gap filling, the following subcategorizations are applied (see slide 44):

- Protein binding alerts for skin sensitization by OASIS v1.4 (subcategorization 1)
- Protein binding alerts skin sensitization by OASIS v1.4 taking into account autoxidation metabolism (subcategorization 2)

See next screen shots.

Data gap filling

Subcategorization 1: Protein binding alerts for skin sensitization by OASIS v1.4

The screenshot shows the 'Data Gap Filling' workflow in the OECD QSAR Toolbox. The left sidebar lists various alert categories, with 'Protein binding alerts for skin sensitization by OASIS v1.4' selected (callout 2). The central workspace shows the target chemical and its analogues, all with 'No alert found' (callout 2). The right sidebar shows the 'Select/filter data' menu with 'Subcategorize' selected (callout 1). The central workspace also displays a read-across prediction for EC3 based on 3 neighbors, with an observed target value of 13.0% and a predicted target value of 20.7%.

There are no protein binding alerts found for target chemical and its analogues

1. **Open** "Select filter data/subcategorize";
2. **Select** "Protein binding alerts for skin sensitization by OASIS v1.4"

Data gap filling

Subcategorization 2: Protein binding alerts for SS when AO is taken into account

The screenshot shows the 'Data Gap Filling' workflow in the OECD QSAR Toolbox. The interface is divided into several sections:

- Left Sidebar:** Contains 'Grouping methods' and 'Endpoint Specific' categories. The 'Protein binding alerts for skin sensitization by OASIS v1.4' is highlighted with a red circle and callout '2'. The 'Autoxidation simulator' is highlighted with a red circle and callout '3'.
- Top Bar:** Shows navigation buttons for 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'.
- Main Area:** Displays a 'Target' chemical structure and its 'Analogues'. Below this, a 'Read across prediction of EC3' section shows a graph of log Kow vs. predicted target value. The text reads: 'Read across prediction of EC3, taking the average from the nearest 5 neighbours, based on 3 values from 3 neighbour chemicals, Observed target value: 13.0 %, Predicted target value: 20.7 %'. A red text overlay states: 'The same autoxidation products of target chemical and its analogues explain the positive experimental data'.
- Right Sidebar:** Contains 'Accept prediction' and 'Return to matrix' options. The 'Subcategorize' button is highlighted with a red circle and callout '1'.

1. **Open** "Select filter data", **click** "Subcategorize";
2. **Select** "Protein binding alerts for skin sensitization by OASIS v1.4";
3. **Select** "Autoxidation simulator"

Data gap filling results

QSAR TOOLBOX | Input | Profiling | Endpoint | Category Definition | **Data Gap Filling** | Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Data Gap Filling Method

- Read-across
- Trend analysis
- (Q)SAR models

Target Endpoint

Human Health Hazards Sensitisation Skin In Vivo LLNA EC3

EC3 data scale is used in RA

Gap filling scale/unit

- Skin sensitization EC3(ratio)
- Skin sensitisation I (Oasis)
- Skin sensitisation II (ECETOC)
- Skin Sensitization (Danish EPA)

Selected [4/4] points

Structure

1 [target] 2 3 4

(4/4) M. Positive M. Positive M. Positive M. Positive

Target chemical is predicted as positive skin sensitizer

Read across prediction of EC3, taking the average from the nearest 5 neighbours, based on 3 values from 3 neighbour chemicals, Observed target value: 13.0 %, Predicted target value: 20.7 %

EC3 (obs.), %

log Kow

Observed data for the target chemical is 13%

Predicted value for the target chemical is 20.7%

Accept prediction

Return to matrix

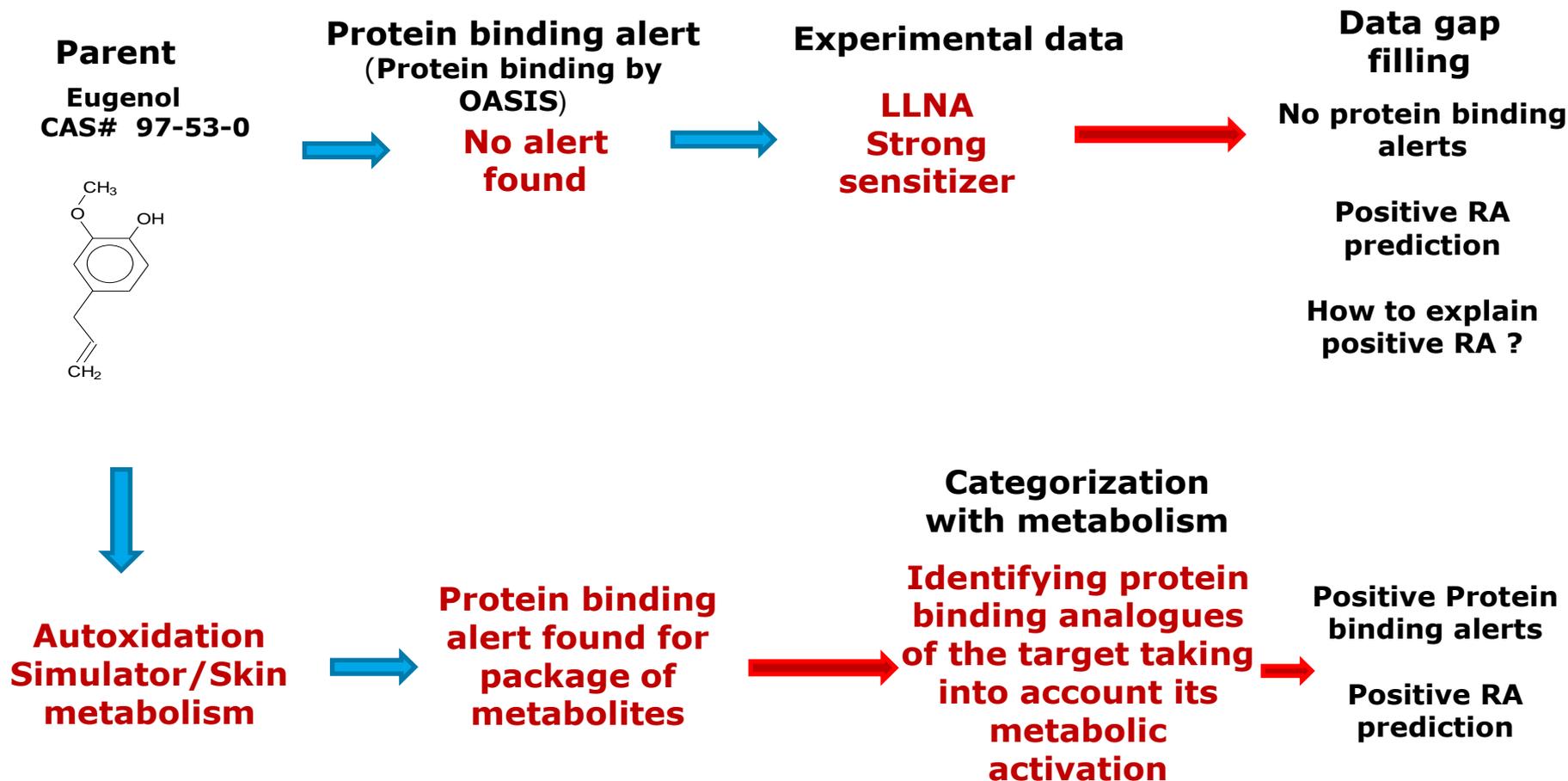
- Select/filter data
 - Subcategorize
 - Mark chemicals by descriptor value
 - Filter points by test conditions
 - Mark focused chemical
 - Mark focused points
- Selection navigation
 - Gap filling approach
 - Descriptors/data
 - Model/(Q)SAR
 - Calculation options
 - Visual options
 - Information
 - Miscellaneous

Data gap filling results

The screenshot displays the 'Data Gap Filling' module of the QSAR Toolbox. The interface includes a top navigation bar with icons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The main workspace shows a scatter plot of EC3 (obs.) % versus log Kow. A warning dialog box is open, stating: 'Warning: The predicted value is not reliable due to following reason(s): - The value of the target chemical for log Kow (2.73) is out of the range of the values for neighbours (3.28 - 3.28). Would you like to accept the current prediction anyway?'. An information dialog box below it says: 'Information: The current prediction was accepted'. On the right, a 'Select/filter data' panel is visible with options like 'Subcategorize', 'Mark chemicals by descriptor value', etc. Red callouts 1-4 point to the 'Accept prediction' button, the 'Yes' button, the 'OK' button, and the 'Return to matrix' button respectively.

1. **Click** "Accept prediction";
2. A message informing the user that the target is out of parametric domain
3. **Click** "OK" on the appeared message
4. **Click** "Return to matrix"

Recap



Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Categorization
 - Data gap filling without taken into account metabolism
 - Categorization applying metabolism
 - Data gap filling handling metabolism of the target chemical
- **Report**

Report

- The report module allows you to generate a 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

The screenshot shows the QSAR Toolbox software interface during the 'Report' phase. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows the 'Data Gap Filling Method' with options: 'Read-across', 'Trend analysis', and '(Q)SAR models'. The main area displays a 'Filter endpoint tree...' on the left and a table of predictions on the right. A context menu is open over a prediction, with 'Report' highlighted. Red callouts '1' and '2' point to the selected prediction and the 'Report' menu item, respectively.

1. **Select** prediction
2. **Right click** and **Select "Report"**

Report

The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Report' section is active, displaying a report titled 'QSAR Toolbox prediction based on read-across' for 'Prediction of EC3 for eugenol (4-allyl-2-methoxyphenol)'. A callout box with the number '1' points to the 'Summary' section of the report.

Summary

Toxicity of the target chemical (20.7 %) is predicted from category members using read-across based on 3 values within the range 13.0 - 32.0 % from 3 nearest neighbours compared by prediction descriptors. Category members are single chemicals or mixtures and are selected based on the profile of the target chemical. Only chemicals having experimental data are listed in the category.

The target chemical DOES NOT FALL within applicability domain of the prediction (see Section 4.3 for details).

The data used for calculating the current prediction is taken from 3 experimental values selected from the following database(s):

1. Skin sensitization

Below is a summary table for endpoint & descriptor values for the target chemical and the category members. Experimental values from data matrix are presented in bold font. Recalculated endpoint values (if required by selected data usage option in Gap Filling) are presented in italic font. Recalculated endpoint values based on experimental data only are presented in bold and italic font.

	<i>Endpoint(s)</i>	<i>Descriptor(s)</i>
	Human Health Hazards; Sensitisation	log Kow
	%	-
<i>Target chemical</i>	13.0	2.73
<i>Cat. member No. 1</i>	32.0	3.28
<i>Cat. member No. 2</i>	13.0	3.28
<i>Cat. member No. 3</i>	17.0	3.28

1. Summary information for prediction

Report

Available data to report

- Predictions
- (Q)SARs
- Categories

Available report templates

- Standard (predefined)
 - QSAR Model Reporting Format (QMRF v. 3.4)
 - QSAR Toolbox Category Report (CCRF v.3.0)
 - QSAR Toolbox Prediction Report (TPRF v.3.0)
- Custom (user defined)
 - Editable copy of QSAR Model Reporting Form
 - Editable copy of QSAR Toolbox Category Report
 - Editable copy of QSAR Toolbox Prediction Report

show only relevant templates

Prediction [1]

1 Prediction of EC3 for eugenol (4-allyl-2-methoxyphenol) 6 / 18

1) Referential boundary:
 The target chemical should be classified as (N/A) AND Michael Addition AND Michael Addition >> Michael addition on quinoid type compounds AND Michael Addition >> Michael addition on quinoid type compounds >> Quinone methide(s)/imines; Quinoid oxime structure; Nitroquinones, Naphthoquinone(s)/imines AND No alert found AND Radical reactions AND Radical reactions >> Free radical formation AND Radical reactions >> Free radical formation >> Hydroperoxides AND SN2 AND SN2 >> Ring opening SN2 reaction AND SN2 >> Ring opening SN2 reaction >> Epoxides, Aziridines and Sulfuranes. by Protein binding alerts for skin sensitization by OASIS v1.4

1M) Metabolism:
 In boundary [1] profiling scheme "Protein binding alerts for skin sensitization by OASIS v1.4" was combined with "Autoxidation simulator"

Additional info:
 Process :All, Match : Accumulative. ~~Separate query criteria may be met by different metabolites (or the parent)~~

2) Parametric boundary:
 The target chemical should have a value of log Kow which is >= 3.28

3) Parametric boundary:

1. Information that metabolism was taken into account when predicting skin sensitization is available

Report

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Reports: Create Print Close Save as

Repository: Register Unregister Update Clone Design

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Available data to report

- Predictions
- (Q)SARs
- Categories

Available report templates

- Standard (predefined)
 - QSAR Model Reporting Format (QMRF v.3.4)
 - QSAR Toolbox Category Report (CCRF v.3.)
 - QSAR Toolbox Prediction Report (TPRF v.3.)
- Custom (user defined)
 - Editable copy of QSAR Model Reporting Form
 - Editable copy of QSAR Toolbox Category Report
 - Editable copy of QSAR Toolbox Prediction Report

show only relevant templates

Prediction [1]

f. Input for prediction (target chemical):
SMILES

g. Descriptor and endpoint values for target chemical (if applicable):

Descriptor(s)	
log Kow	2.73
Endpoint (dep. variable)	
Human Health Hazards; Sensitisation	13.0 %

h. Additional chemicals eliminations (not determined by domain):
Not available

i. Additional data eliminations (not determined by domain):
Not available

j. Predicted value (model result):
20.7 %

k. Predicted value (comments):
Not provided by the user *manually editable field*

4.3. Applicability domain (OECD Principle 3):
The target chemical DOES NOT FALL within applicability domain (see Section 3.1.b for detailed description of the domain)

1. Predicted value

Report

Available data to report

Predictions
[1] 12.07.2016 14:22 [R]: 20.7(-4.22;45.5) %; Estimation for EC
(QSARs)
Categories

Available report templates

Standard (predefined)
- QSAR Toolbox Prediction Report (TPRF v.3.4)

Custom (user defined)
- Editable copy of QSAR Toolbox Prediction Report (TPRF v.3.4)

show only relevant templates

Prediction [1]

Sensitisation

h. Additional chemicals eliminations (not determined by domain):
Not available

i. Additional data eliminations (not determined by domain):
Not available

j. Predicted value (model result):
20.7 %

k. Predicted value (comments): *manually editable field*
Not provided by the user

4.3. Applicability domain (OECD Principle 3):
The target chemical DOES NOT FALL within applicability domain
(see Section 3.1.b for detailed description of the domain)

1. M
2. AND
3. AND
4. AND
5. AND

4.4. Uncertainty of the prediction (OECD Principle 4): *manually editable field*

QSAR Toolbox 3.4.0.17
Database version: 3.8.8/2.1.2

QSAR TOOLBOX
TPRF v.3.4.1.34101

1. Applicability domain

The target chemical is "Out of domain", because does not fall within parametric range of Log Kow[3.28-4.75]

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Categorization
 - Data gap filling without taken into account metabolism
 - Categorization applying metabolism
 - Data gap filling handling metabolism of the target chemical
 - Report
 - **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

The screenshot displays the QSAR Toolbox application with a 'Save As' dialog box open. The dialog shows the file path: < Back > TOOLBOX 4.0 > Toolbox 3.4 > 5.07.2016. The file list includes 'Tutorial 15.tbw' (68 KB) and 'Tutorial 16.tbw' (598 KB). The 'File name' field is set to 'Tutorial 23.tbw' and the 'Save as type' is 'Toolbox work file (*.tbw)'. Three numbered callouts indicate the steps: 1. Click on 'Save' button; 2. Browse and put name of the file; 3. Click Save button. The background shows the software's main window with a sidebar for 'Predictions' and 'Available report templates'.

1. **Click** on "Save" button; 2. **Browse** and **put** name of the file; 3. **Click** Save button

Open saved file

The screenshot illustrates the steps to open a saved file in the QSAR Toolbox. The interface is divided into several sections: a top menu bar, a left sidebar for document management, a central file selection dialog, and a main workspace for chemical analysis. The file selection dialog is the primary focus, showing a list of files in the 'TOOLBOX 4.0 > Toolbox 3.4 > 5.07.2016' directory. The file 'Totorial 23.tbw' is selected, and the 'Open' button is highlighted. The background workspace shows a chemical structure and a table of results.

Assay	Result	Value	Unit
In Vivo			
GPMT	(1/1) M: Positive		
HRIPT	(2/4) M: Positive, Positiv...		M: 8E3 µg/cm2
LLNA	(4/5) M: Positive	M: Positive	M: Positive
EC3	(1/1) M: Positive		
Undefined Assay			
Toxicity to Reproduction			
Toxicokinetics, Metabolism and Distribution			
Basic Toxicokinetics			
Dermal Absorption			
Profile			

Once the file has been saved 1. **Go** to "Input"; 2. **Create** new document 3. **Click** "Open"; 4. **Browse** and **select the file**; 5. **Click** "Open" button