

## OECD QSAR Toolbox v.4.1

Step-by-step example for predicting  
skin sensitization accounting for abiotic  
activation of chemicals

# 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 Toolbox 4.1:**

- Profiling the target chemical.
- Identifying analogues of the target chemical.
- Filling in data gaps of target chemical by means of read-across.
- Profiling target chemical by taking into account its (a)biotic activation.
- Identifying analogues of the target by using a new categorization functionality allowing (a)biotic activation to be taken into account.
- Filling in 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 by using **well-known categorization group**
  - Identifying analogues based on autoxidation activation of the target illustrating new categorization functionality
- **Filling in data gaps 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
  - Data
  - 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

### **Alternative ways for input of Chemical:**

#### 1. Single target chemical

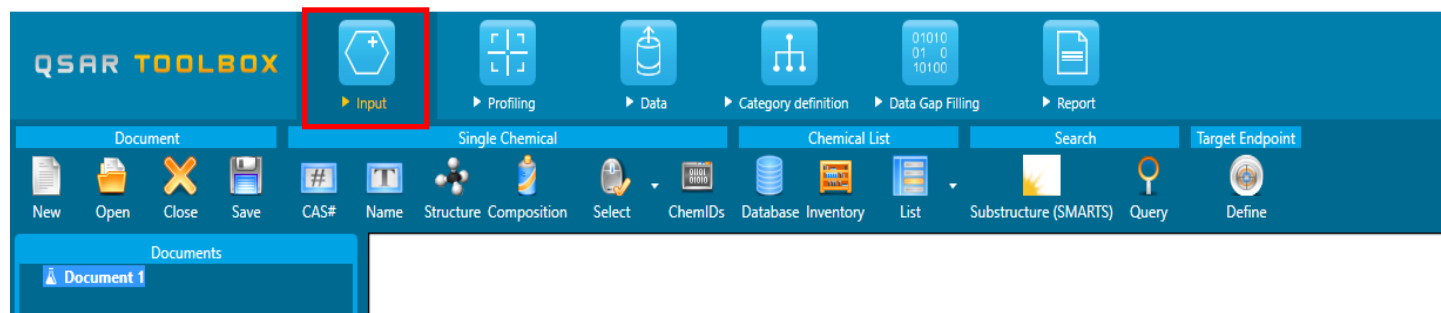
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi- SMART
- Drawing chemical structure
- Select from User List/Inventory/Databases

#### 2. Group of chemicals

- User's List/Inventory
- Customized Databases

# Chemical Input Input Screen

- Open the Toolbox.
- The six modules in the workflow are seen listed next to “QSAR TOOLBOX” title.
- Click “Input”



# Chemical Input

## Input target chemical by CAS#

Search by CAS #

97530 Search OK Cancel

Select All Unselect All Invert Selection Selected 1 of 1

1	CAS	97-53-0	
	SMILES	COc1cc(CC=C)ccc1O	
	CS Relation	High	
<input checked="" type="checkbox"/>	Substance	Mono constituent	
	Composition		
	Name	1-ALLYL-3-METHOXY-4-...	
		1-allyl-3-methoxy-4-hydrox...	
		2-methoxy-4-(2-propenylp...	

Chemical structure diagram of Eugenol: COc1cc(CC=C)ccc1O

1. Click **CAS#**; 2. Enter the CAS#97-53-0 in the blank field; 3. Click **Search** : Eugenol chemical is found in TB databases; 4. Click **OK**.

# Chemical Input

## Target chemical identity

Expanding **Structure Info** (1) displays the chemical identification information.

The screenshot shows the QSAR Toolbox interface. The 'Filter endpoint tree...' panel on the left has 'Structure info' expanded, indicated by a red circle and a callout box with the number '1'. The '1 [target]' panel on the right displays the chemical structure and its identification data:

97-53-0
High
[eugenol]eugenol (4-
C10H12O2
Mono constituent
COc1cc(CC=C)ccc1O

# Chemical Input

## Target chemical identity

The labels indicated different reliability of the CAS-SMILES relationship are as follows:

### High:

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

### Moderate:

This reliability corresponds to moderate reliability of CAS-SMILES relation. The moderate label is assigned if the chemical belongs to three “Distribute to QA” data sources.

### Low:

This reliability corresponds to poor reliability of CAS-SMILES relation. This label is assigned if the chemical belongs to less than three, but at least one “Distribute to QA” data sources.



# Outlook

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

# Profiling

## Overview

- “Profiling” refers to the electronic process of retrieving relevant information for the target compound, other than its environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- Available information includes the probable mechanism(s) of action, as well as observed or simulated metabolites.

# Profiling

- For most of the profiles, background information can be retrieved by highlighting one of the profiles (for example, Protein binding alerts for skin sensitization by OASIS and clicking "View" (see next screen shot)).

# Profiling

## Profiling the target chemical

1. Highlight the profiler (1); 2. Click **View** (2); 3. Click "Isocyanates, Isothiocyanates" 4. Click **Literature** to see textual description associated with the category.

1. Highlight the profiler (1); 2. Click **View** (2); 3. Click "Isocyanates, Isothiocyanates" 4. Click **Literature** to see textual description associated with the category.

# Profiling

## Profiling the target chemical

1. Click **Definition** in order to see more details about the structural boundaries of *Isocyanates, Isothiocyanates*;
2. The structural boundaries implemented in the category are shown in (2);
3. The structural fragment in the first structural boundary (circled in red) is shown in (3).

The screenshot displays the 'Protein binding alerts for skin sensitization by OASIS (Endpoint Specific) - Profiling Scheme Browser' window. The interface includes a menu bar with 'Save Scheme', 'Export Scheme', 'Tests', 'View Tests', and 'Run All Tests'. Below the menu is a 'Definition' tab, which is highlighted with a red box and a callout '1'. The main content area shows a tree view of categories, with 'Isocyanates, Isothiocyanates' selected. A red circle highlights a sub-category, with a callout '2' pointing to it. Below this, a 'Query details' panel shows a 'Structure Query' with a SMARTS string: [\*6][\*7]=[\*6]=[\*16]. A red circle highlights a specific structural fragment within the query, with a callout '3' pointing to it. The interface also includes a 'Contents' panel on the left and a 'View mode' dropdown set to 'Facade'.

# Profiling

## Profiling the target chemical

- The outcome of the profiling determines the most appropriate way to search for analogues.
- 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 – 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 (the green check mark disappears) profilers.
- An option color by ***Endpoint selected in the data matrix*** is implemented, which highlights all relevant profiles (see next slide).

# Profiling

## Profiles' relevancy- new functionality

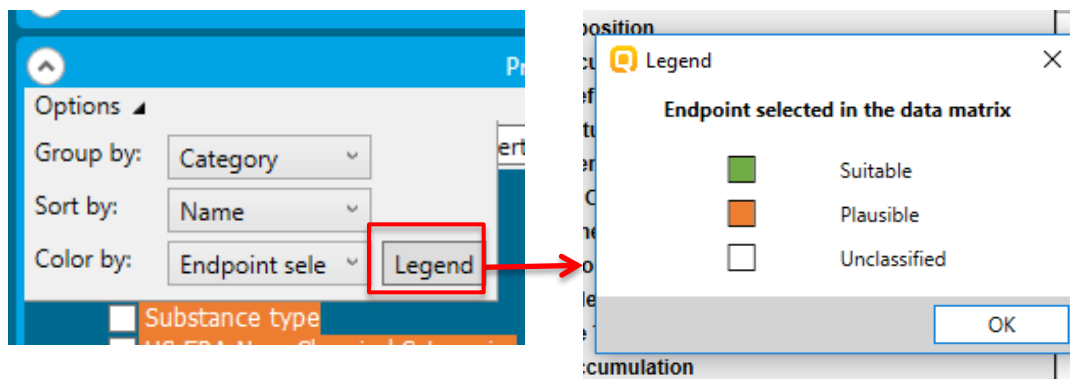
The screenshot displays the QSAR Toolbox Profiling interface. On the left, the 'Documents' panel shows 'Profiling methods' with an 'Options' dropdown menu (callout 2) and a list of endpoints. The 'Endpoint selected in the data matrix' (callout 3) is highlighted. Below this, the 'Metabolism/Transformations' panel shows a list of simulators. On the right, the 'Filter endpoint tree...' panel shows a tree structure with 'Sensitisation' selected (callout 1). The '1 [target]' panel shows a chemical structure and associated data.

1. Open the endpoint tree and select the cell at the **Sensitization** level (1)
2. Expand **Options** (2);
3. Select **Color by Endpoint selected in the data matrix** (3);
4. All profiles and metabolic simulators relevant to the selected endpoint will be colored (see next slide for color legend).



# Profiling

## Profiles' relevancy- new functionality



- **Suitable** - developed using data/knowledge for the target endpoint;
- **Plausible** – not endpoint specific; structure-based; form broader group of analogues;
- **Unclassified** – all profiles, which are not classified in any of the categories above.

# Profiling

## Profiling the target chemical

The screenshot shows the QSAR Toolbox Profiling interface. Key elements include:

- Profiling Methods (Left Panel):** A list of endpoints with checkboxes. Two endpoints are checked and highlighted with a red box and a '1' callout:
  - Protein binding by OASIS
  - Protein binding by OECD
- Profiling Methods (Bottom):** A list of endpoints with checkboxes. Two endpoints are checked and highlighted with a red box and a '1' callout:
  - Keratinocyte gene expression
  - Protein binding alerts for skin sensitization by OASIS
- Filter endpoint tree... (Middle Panel):** A tree view of endpoints. The 'Human Health Hazards' section is expanded, and 'Sensitisation' is highlighted with a blue bar and a '1' callout.
- Structure (Right Panel):** Displays chemical information for 'eugenol' (CAS 97-53-0, SMILES: COc1cc(C=C)ccc1O).
- Buttons:** 'Apply', 'View', and 'Delete' buttons are visible at the top left.

1. Tick the checkboxes of Protein binding alert by OASIS, Protein binding by OECD and protein binding alerts for skin sensitization by OASIS.
2. Click **Apply**;

# Profiling

## Profiling the target chemical

The screenshot shows the 'Filter endpoint tree...' window in the QSAR Toolbox. The 'Structure' tab is active, displaying the chemical structure of eugenol. The 'Profile' section is expanded and highlighted with a red box, showing the following results:

Endpoint	Result
General Mechanistic	
Protein binding by OASIS	No alert found
Protein binding by OECD	No alert found
Endpoint Specific	
Protein binding alerts for skin sensitization ...	No alert found

- The actual profiling takes up to several seconds depending on the number and type of profiles selected.
- The results of profiling automatically appear at the bottom of the endpoint tree (1).
- No protein binding alert has been found for the target compound (eugenol) based on three protein binding profilers.

# Outlook

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

# Data Overview

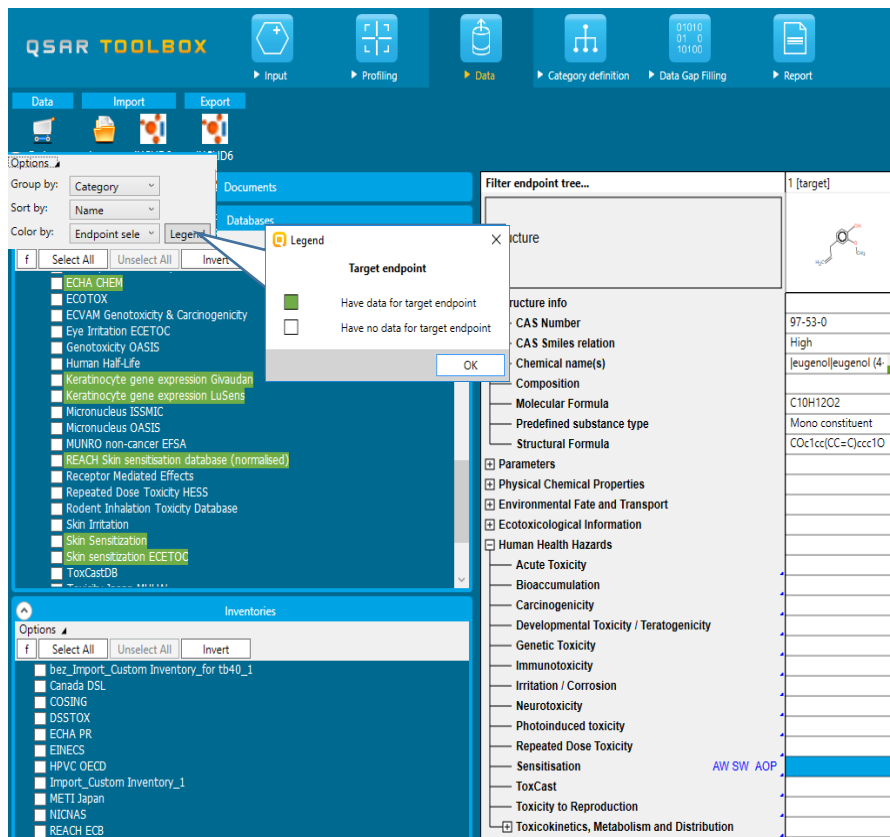
- “Data” refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.
- Data gathering can be executed in a global fashion (i.e., collecting all data for all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).

# Data

## 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).

# Data Database` relevancy-new functionality



1. Click the cell at the **Sensitization** level (1);
2. Expand **Options** (2);
3. Select **Colour by Endpoint selected in the data matrix** (3);
4. All databases that have data for the target endpoint are coloured in green.
5. Click **Legend** to open the colour legend (5).

# Data Gather data

The screenshot displays the QSAR Toolbox interface. At the top, there are navigation buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below these are tabs for Data, Import, and Export. The 'Databases' panel on the left contains a list of databases with checkboxes. Two items, 'Skin Sensitization' and 'Skin sensitization ECETOC', are checked and circled in red, with a callout '1' pointing to them. The 'Gather' button is highlighted with a blue callout '2'. The 'Filter endpoint tree...' panel on the right shows a hierarchical list of endpoints, with 'Sensitisation' selected and highlighted in blue. The rightmost panel shows a table of properties for the selected endpoint, including CAS Number (97-53-0), Chemical name(s) (|eugenol|eugenol (4)), Molecular Formula (C10H12O2), and Structural Formula (COc1cc(CC=C)ccc1O).

1. Tick *Skin Sensitization* and *Skin Sensitization ECETOC* (1);
2. Click **Gather** (2);



# Data

## 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** .

# Data Gather data

The screenshot displays the QSAR Toolbox software interface. At the top, there are three main menu items: 'Data', 'Import', and 'Export'. Below these are icons for 'Gather', 'Import', 'IUCLID6', and 'IUCLID6'. The left sidebar shows a 'Documents' section and a 'Databases' section with a search bar and 'Options' (Select All, Unselect All, Invert). A list of databases is shown, with 'Human Health Hazards' selected. The main area is divided into 'Filter endpoint tree...' and '1 [target]'. The 'Filter endpoint tree...' shows a tree structure with categories like 'Structure info', 'Parameters', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', and 'Human Health Hazards'. The '1 [target]' section shows a chemical structure and its properties: CAS Number (97-53-0), CAS Smiles relation (High), Chemical name(s) (eugenol|eugenol (4-)), Composition (C10H12O2), Predefined substance type (Mono constituent), and Structural Formula (COc1cc(CC=C)ccc1O). A 'Read data?' dialog box is open in the foreground, with 'All endpoints' selected and 'from Tautomers' unchecked. The 'OK' button is highlighted with a red box.

1. A "Read data?" window appears.
2. You could choose to gather "all" or "endpoint specific" data.
3. Click **OK** to read all available data;

# Data Gather data

Filter endpoint tree... 1 [target]

Structure

Structure info

- CAS Number: 97-53-0
- CAS Smiles relation: High
- Chemical name(s): |eugenol|eugenol (4-)
- Composition: C10H12O2
- Molecular Formula: Mono constituent
- Predefined substance type: COc1cc(CC=C)ccc1O
- Structural Formula:

Parameters

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

**Human Health Hazards (1/7) M: 8E+03 µg/cm2**

Profile

- General Mechanistic
- Protein binding by QASIS: No alert found

Filter endpoint tree... 1 [target]

Structure

CAS SMILES relation

- Chemical name(s): |eugenol|eugenol (4-)
- Composition: C10H12O2
- Molecular Formula: Mono constituent
- Predefined substance type: COc1cc(CC=C)ccc1O
- Structural Formula:

Parameters

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Human Health Hazards

- Acute Toxicity
- Bioaccumulation
- Carcinogenicity
- Developmental Toxicity / Teratogenicity
- Genetic Toxicity
- Immunotoxicity
- Irritation / Corrosion
- Neurotoxicity
- Photoinduced toxicity
- Repeated Dose Toxicity
- Sensitisation
- Respiratory Tract (1/1) M: Negative
- Skin**
  - in Vivo
    - GPMT (1/1) M: Positive
    - HRIPT (1/3) M: 8E+03 µg/cm2
    - LLNA (1/1) M: Positive
    - Undefined Assay (1/1) M: Positive

1. The extracted data is displayed on the data matrix;
2. Expand the Human Health node to display all skin experimental data;
3. Positive experimental data is found.

# Endpoint Gather data

The screenshot displays the 'Data points' window and a portion of the assay tree. The 'Data points' window contains the following table:

Datapoints	#	Value	Original value	Assay
Human Health Hazards;Sensitisation	1	M: Positive (Skin sensitisation II (ECETOC))	Moderate sensitizer (Skin sensitisation IV (GPMT))	GPMT

Below the table, the assay tree is visible. The 'M: Positive' cell is highlighted in blue and enclosed in a red box. A red arrow points from this cell to the 'Value' column of the 'Data points' table. A blue callout bubble with the number '1' points to the highlighted cell.

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
  - Data
  - **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.

# 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-categorization).
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

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

Metabolism accounted for

### Phase III. Eliminating dissimilar chemicals

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

**Broad grouping  
Endpoint Non-specific**

**Subcategorization  
Endpoint Specific**

**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

**Broad grouping**  
**Endpoint Non-specific**

Each of the above grouping method is applied to the target chemical and number of the identified analogues are provided below. In order to preserve the basic functional groups available within the molecule: Allyl, Ether and Phenol, OFG are used for categorization purposes. US-EPA and OECD 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

The screenshot shows the 'Filter endpoint tree...' panel with the following sub-endpoints under 'Structure':

- Structure info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards
  - Acute Toxicity
  - Bioaccumulation
  - Carcinogenicity
  - Developmental Toxicity / Teratogenicity
  - Genetic Toxicity
  - Immunotoxicity
  - Irritation / Corrosion
  - Neurotoxicity
  - Photoinduced toxicity
  - Repeated Dose Toxicity
  - Sensitisation
  - ToxCast
  - Toxicity to Reproduction
  - Toxicokinetics, Metabolism and Distribution
- Profile
  - Predefined
    - OECD HPV Chemical Categories
    - US-EPA New Chemical Categories
  - Empiric
    - Organic functional groups (107 analogues)
    - Organic functional groups (nested) (127 analogues)
    - Structure similarity (2 analogues)

1007 analogues are identified.

127 analogues are identified

2 analogues are identified (in case all categories are preserved)  
10 analogues are identified (in case Allyl, Ether and Phenol are preserved)

38 analogues are identified

# Category Definition

## Define category by OFG

The image shows the QSAR Toolbox software interface. On the left, the 'Documents' panel lists various categories, with 'Organic functional groups' selected (callout 1). The 'Define' button is highlighted (callout 2). The 'Grouping options (Organic functional groups)' dialog box is open, showing a list of groups. Most groups are highlighted in blue (callout 3), while 'Allyl', 'Ether', and 'Phenol' are not. The 'Down' button is highlighted (callout 4). The 'Grouping options' dialog box is shown again, with the 'Target' list containing only 'Allyl', 'Ether', and 'Phenol' (callout 5). The 'OK' button is highlighted (callout 6).

1. Select **Organic functional groups**; 2. Click **Define** 3. Click all groups (highlighted in blue) but Allyl, Ether and Phenol by also holding the Ctrl button. 4. Click **Down** to remove them. They are moved in the panel down; 5. Only Allyl, Ether and Phenol should remain in the upper panel; 6. Click **OK**.

# Category Definition

## Define category by OFG

1. A warning message is displayed informing that the selected category differ from the target ones. Click Yes.
2. Click OK to confirm the grouping results;
3. Click OK to read all available data;

# Category Definition

## Gather data for analogues chemicals

The screenshot shows the QSAR Toolbox software interface during the 'Category definition' step. The main window displays a data matrix for the category 'EC3'. The matrix has 8 columns representing different data points and 1 row for 'EC3'. The data points are: (10/10), M: Positive, M: Positive, M: Positive, M: Positive, M: Positive, M: Negative, and M: Positive. A red oval highlights this row. A dialog box titled 'Gather data' is open, showing '22 points added across 10 chemicals.' and an 'OK' button. Three callout boxes with numbers 1, 2, and 3 point to the dialog, the filter box, and the data matrix respectively.

1. An information message about the number of data points pops up. Click **OK**. All data is displayed on the data matrix.

2. Type **EC3** in the filter box, then click enter from the keyboard to select only the **EC3** data (2, 3).

# Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - **Data gap filling without accounting metabolism**

# Data Gap Filling Overview

- “Data Gap Filling” module provides three different data gap filling tools:
  - Read-across
  - Trend analysis
  - (Q)SAR models
- The most relevant data gap mechanism should be used by 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' session. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' sub-menu is active, showing 'Gap Filling' and 'Read across' (highlighted with a red box and callout '2'). The main workspace shows a hierarchical tree of 'Human Health Hazards' with 'Sensitization' expanded to 'Skin' > 'In Vivo' > 'LLNA' > 'EC3'. A table of chemical structures is displayed, with the cell containing 'M: Positive' highlighted (callout '1'). A 'Possible data inconsistency' dialog box is open, showing 'Native scale/unit' options (checked for 'Skin sensitization I (Oasis)' and 'Skin sensitization EC3(ratio)') and 'converted data' (9 from scale 'Skin sensitization EC3(ratio)', 1 from scale 'Skin sensitization I (Oasis)'). The dialog also shows 'Data 10/10; Chemicals 10/10' and 'OK'/'Cancel' buttons (callout '3').

1. Click the cell corresponding to *Sensitization* > *Skin* > *In Vivo* > *LLNA* > *EC3*;
2. Select **Read-across**;
3. A pop-up window informing about possible data inconsistency 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).
- The skin sensitisation potential of chemicals is coded in different scales depending on their source (for example: data from John Moores University of Liverpool is classified as: *Strongly sensitizing, Moderately sensitizing etc.*; data from European centre for Ecotoxicology and Toxicology of chemicals is classified as: *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.

# Data gap filling

## Scale definition

Back to our example:

Possible data inconsistency

**Native scale/unit**

- Skin sensitisation I (Oasis) (1 data; 1 chemicals)
- Skin sensitization EC3(ratio) (9 data; 9 chemicals)

Gap filling scale/unit

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

converted data

9 from scale %(Skin sensitization EC3(ratio))  
1 from scale Negative(Skin sensitisation I (Oasis))

Data 10/10; Chemicals 10/10

OK Cancel

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

# Data gap filling

## Read-across

- The resulting plot places the experimental results of all analogues (Y axis) against the descriptor (X axis) which by default is  $\log K_{ow}$  (see next screen shot).
- The **RED** point represents predicted results for the target chemical.
- The **BROWN** points represent the experimental data of the analogues that are used for the read-across (by default are the five nearest neighbours to the target)
- The **BLUE** points represent the experimental data of the analogues not used in the read-across.

# Data gap filling Read-across

The screenshot displays the QSAR Toolbox interface for a data gap filling task. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The main workspace is divided into several sections:

- Documents:** Shows a list of documents, including 'ent 1' and '97530', with a note 'Enter GF(RA) with 10 chemicals, 10 data points'.
- Filter endpoint tree:** A hierarchical tree showing 'Human Health Hazards' > 'Sensitisation' > 'Skin' > 'In Vivo' > 'LLNA' > 'EC3'. The predicted outcome for EC3 is 'M: Positive'.
- Table:** A table with 10 columns representing different chemical structures. The first column is labeled '1 [target]'. The table shows predicted outcomes for each structure, such as 'M: Positive' or 'M: Negative'.
- Read-across prediction for EC3, based on 7 values:** A graph showing the relationship between log Kow (x-axis, ranging from 2.6 to 3.7) and EC3 (y-axis, with 'Positive' and 'Negative' categories). The graph displays 7 data points, with 5 positive and 2 negative outcomes. The observed prediction is 'Positive' and the predicted prediction is 'Positive'.
- Data Gap Filling Settings:** A sidebar panel with checkboxes for 'Only endpoint relevant' and 'Only chemical relevant'. It also includes options for 'Automated workflows' and 'Standardized workflows'.

Initial graph without any subcategorization.

# Data Gap Filling Subcategorizations

In this example two subcategorizations are applied in order to eliminate dissimilar analogues:

- 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 (subcategorization 2)

See next slides.

# Data gap filling

## Subcategorization 1: Organic functional groups (US EPA)

The screenshot shows the QSAR Toolbox interface during a subcategorization task. On the left, the 'Organic functional groups (US EPA)' category is selected in the sidebar, indicated by a blue callout '2'. The main window displays a list of functional groups, with a red circle highlighting the list and a red arrow pointing to the text 'Profiling results of target'. Below this, a graph shows 'Read-across prediction for EC3, based on 7 values' with 'Observed: Positive; Predicted: Positive'. The graph plots log Kow (x-axis, 2.6 to 3.7) against predicted values. A red circle highlights the list of functional groups in the graph area, with a red arrow pointing to the text 'Profiling results of analogues'. A red box contains the text: 'The profiling results of the analogues are similar to target chemical with respect to Organic functional groups (US EPA)'. In the bottom right, a '1' callout points to the 'Select / filter data' button in the 'Subcategorize' menu.

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

The screenshot shows the QSAR Toolbox interface during a subcategorization step. The 'Subcategorization' dialog box is open, with the 'Protein binding alerts for skin sensitization by OASIS' option selected and highlighted in red. A callout bubble with the number '1' points to this option. In the 'Adjust options' section, the text 'No alert found' is circled in red. Below this, the 'Differ from target by' section has 'At least one categ.' selected. The main window displays a table of chemical structures and their predicted outcomes (e.g., M: Positive, M: Negative). A red box highlights a text overlay: 'No protein binding alerts are identified for target and analogues, which can not explained the positive experimental data found. In this respect metabolism should be taken into account (see next slide).'. The bottom right corner shows a 'Accept prediction' button with a green checkmark.

1. Select *Protein binding alerts for skin sensitization by OASIS*.



# Data gap filling

## Subcategorization when metabolism is taken into account

**1** Protein binding alerts for skin sensitization by OASIS

**2** Autoxidation simulator

**3** Adjust options

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

Subcategorization is applied by taking into account autoxidation simulation in combination with *Protein binding alerts for skin sensitization by OASIS*:

1. Select *Protein binding alerts for skin sensitization by OASIS*;
2. Select *Autoxidation simulator*; 3. Close the *Subcategorization* window;

# 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 the 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.

# Data gap filling

## Return to data matrix

The screenshot shows the QSAR Toolbox interface. The top navigation bar includes icons for Input, Profiling, Data, Category definition, and Data Gap Filling. Below this are tabs for Gap Filling and Workflow. The main workspace is divided into three sections:

- Documents:** A tree view showing 'Document 1' with a sub-entry for '# CAS: 97530'. A red callout box with the number '1' points to this entry. Below it, there are options for 'Organic functional groups' and a prompt to 'Enter GF(RA) with 10 chemicals, 10 data points'.
- Data Gap Filling Settings:** A section with checkboxes for 'Only endpoint relevant' and 'Only chemical relevant'. Below this, a table shows the current status:
 

At this position:	
QSARs	1
Automated workflows	1
Standardized workflows	1
- Filter endpoint tree...:** A hierarchical tree of endpoints. The 'Sensitisation' node is expanded, showing sub-nodes for 'Respiratory Tract' and 'Skin'. The 'Respiratory Tract' node is further expanded, showing 'AW SW AOP' with values '(1/1)' and '(1/6)'. The 'Skin' node is also expanded, showing 'M: Negative' and 'M: 8E+03 µg/cm2'.

To return to data matrix go to the document tree and click the node CAS: 97530 (1).

# Data gap filling

## Next actions

- The study continues with a second data gap filling where a category of analogues is defined by using a new categorization functionality allowing to define category accounting for (a)biotic activation of the target.
- Before proceeding with Data gap filling the following two procedures 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
  - Data
  - 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 in two ways:

- Scenario 1: In the **Input** section outside data gap filling module (slide 65)
- Scenario 2: In the **Profiling** section (slide 73)

# Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - 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 displays the QSAR Toolbox software interface. The 'Input' menu is highlighted with a red '1'. A context menu is open over a chemical entry in the 'Documents' panel, with 'Multiplication' selected and a red '2' next to it. A sub-menu is open for 'Metabolism/Transformations', with 'Autoxidation simulator' selected and a red '3' next to it. The 'Autoxidation simulator' sub-menu lists various simulation options, including 'Autoxidation simulator (alkaline medium)', 'Dissociation simulator', 'Hydrolysis simulator (acidic)', 'Hydrolysis simulator (basic)', 'Hydrolysis simulator (neutral)', 'in vivo Rat metabolism simulator', 'Microbial metabolism simulator', 'Observed Mammalian metabolism', 'Observed Microbial metabolism', 'Observed Rat In vivo metabolism', 'Observed rat liver metabolism with quantitative data', 'Observed Rat Liver S9 metabolism', 'Rat liver S9 metabolism simulator', and 'Skin metabolism simulator'. The chemical entry in the background is identified as '# CAS: 9753-0' and 'leugenol|eugenol (4-...'.

1. Go to *Input*
2. Select the CAS of the target chemical and right-click on it .
3. Select *Multiplication-Metabolism/Transformations/Autoxidation simulator*



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

The screenshot shows the QSAR Toolbox interface. On the left, the 'Documents' tree is expanded to show a folder named 'Autoxidation simulator' containing five entries: 'metabolite #1', 'metabolite #2', 'metabolite #3', 'metabolite #4', and 'metabolite #5'. This folder is highlighted with a red box and a callout bubble containing the number '2'. The main window displays a data matrix with columns for 'Parent chemical [target]' and five metabolites. The 'Parent chemical' column shows the chemical structure of Eugenol and its properties. The metabolite columns show their respective structures and properties. An 'Information' dialog box is open over the 'Parent chemical' row, displaying the message 'A parent list with 5 child lists were created' and an 'OK' button, which is highlighted with a callout bubble containing the number '1'.






Filter endpoint tree...	Parent chemical [target]	metabolite #1	metabolite #2	metabolite #3	metabolite #4	metabolite #5
Structure						Oligomer
Structure info	97-53-0			Invalid CAS number: 0-		
— CAS Number	High			Not applicable		
— CAS Smiles relation	Chemical name(s)					
— Chemical name(s)	eugenol eugenol (4-					
— Composition	Information					
— Molecular Formula	C10H12O2					Oligomer
— Predefined substance type	Mono con					Mono constituent
— Structural Formula	COc1cc(CC					Oligomer
Parameters						
Physical Chemical Properties						
Environmental Fate and Transport						
Ecotoxicological Information						
Human Health Hazards						
— Acute Toxicity						
— Bioaccumulation						

1. Select **OK** and then parent and product(s) are visualized on data matrix.
2. The generated metabolites are listed in the Documents tree ;

# 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 autoxidation metabolites exhibit interaction with proteins via three different protein binding mechanisms (Michael Addition, Radical reactions, and SN2).

## Autoxidation simulator

Parent chemical	metabolite #1	metabolite #2	metabolite #3	metabolite #4	metabolite #5
					Oligomer
97-53-0	Invalid CAS number: 0-0	Invalid CAS number: 0-0	Invalid CAS number: 0-0	Invalid CAS number: 0-0	Invalid CAS number: 0-0
High	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
eugenoleugenol (4-					
C10H12O2	C10H10O2	C10H12O4	C10H12O3	C10H12O3	Oligomer
Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent
COc1cc(CC=C)ccc1O	COC1=CC(C=CC1=O)=	COc1cc(ccc1O)C(O)C=	COc1cc(ccc1O)C(O)C=	COc1cc(CC2CO2)ccc1O	Oligomer
Parameters					
Physical Chemical Properties					
Environmental Fate and Transport					
Ecotoxicological Information					
Human Health Hazards	(1/7)				
Profile					
General Mechanistic					
Protein binding by OASIS					
Protein binding by OECD					
Protein binding potency					
Endpoint Specific					
Protein binding alerts for skin sensitization a ...	No alert found	Michael Addition >> M	Radical reactions >> Fr	No alert found	SN2
Protein binding alerts for skin sensitization ...	No alert found	Michael Addition >> M	Radical reactions >> Fr	SN2 >> Ring opening S	SN2 >> Ring opening S

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

# Outlook

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

The screenshot displays the QSAR Toolbox software interface. The top toolbar shows the 'Profiling' icon (1). The 'Documents' panel on the left shows 'Document 2' selected (4). The 'Filter endpoint tree' on the left lists various endpoints, with 'Protein binding alerts for skin sensitization by OASIS' checked (2). The 'Metabolism/Transformations' panel shows 'Autoxidation simulator' checked (3). A 'Profiling' dialog box (5) is open, asking 'Selected profiles will be applied on all metabolites! Do you want to continue?'. The 'Generated metabolites' window (6) shows 5 metabolites, all with 'Invalid CAS number: 0'. A red bracket underlines the 'Generated metabolites' window, and a red arrow points from the '5 metabolites' cell in the 'Autoxidation simulator' table to the 'Generated metabolites' window.

**Generated metabolites**

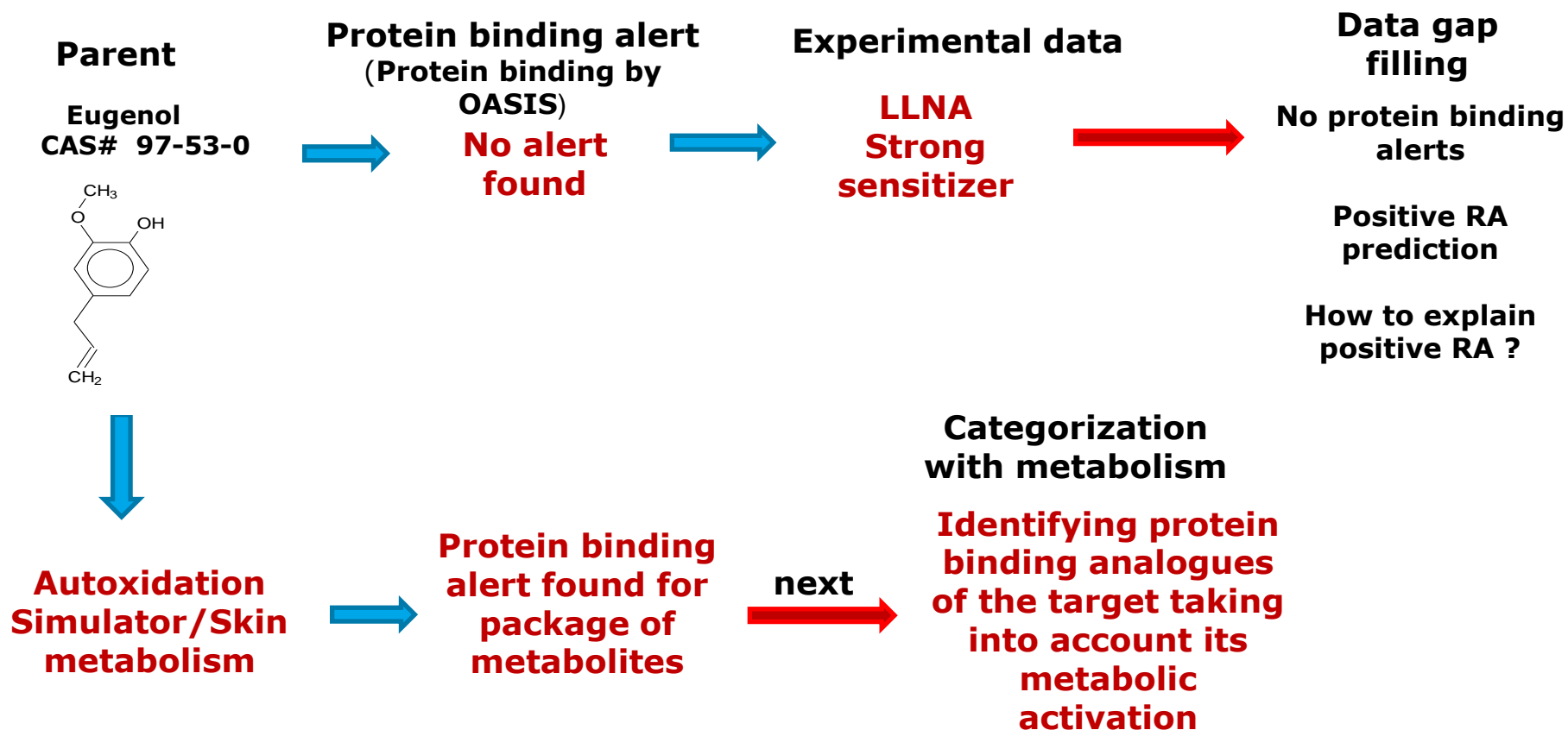
1. Once the chemical is entered in Toolbox in a new document, go to *Profiling* ;
2. Tick *Protein binding for skin sensitization by OASIS* profiler from **Endpoint specific** group;
3. Tick *Autoxidation simulator* from *Metabolism/Transformations*;
4. Click **Apply**; 5. Click **Yes**;
6. Double click the cell containing the 5 metabolites in order to visualize them.

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

**Package of Protein profiling result for parent and its autoxidation products**

1. Open node "Autoxidation simulator";
2. Double click the cell with profiling results obtained for the metabolites;
3. Appearing the window to be investigate profiling result for each metabolite;

# Recap



# Outlook

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

## Category definition with applying metabolism

The advantages of the new functionality are:

- Application of metabolism for analogues identification during process of categorization.
- A category can be defined with and without metabolism.
- This is a critical step in the workflow.
- 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).



# Category definition with applying metabolism

The screenshot shows the QSAR Toolbox software interface. The main window is titled 'Document 1'. The menu bar includes 'Input', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Input' menu is expanded, showing options like 'New', 'Open', 'Close', 'Save', 'CAS#', 'Name', 'Structure', and 'Composition'. A search window is open, showing a search for CAS number 97530. The search results table lists chemical properties for CAS 97-53-0, including SMILES, CS Relation, Substance, and Name. A chemical structure is also displayed next to the results.




1	CAS	97-53-0
	SMILES	COc1cc(CC=C)ccc1O
	CS Relation	High
<input checked="" type="checkbox"/>	Substance	Mono constituent
	Composition	
	Name	1-ALLYL-3-METHOXY-4-... 1-allyl-3-methoxy-4-hydrox... 2-methoxy-4-(2-propenyl)p...

1. Go to **Input**; 2. Click **New**; 3. Click **CAS#**; 4. Enter the CAS number of the target; 5. Click **Search** 5. Click **OK**.

# Category definition with applying metabolism

1

2

Parent Chemical	metabolite #1	metabolite #2	metab
			
97-53-0	Invalid CAS number: 0-0	Invalid CAS number: 0-0	Invalid
High	Not applicable	Not applicable	Not a
eugenol eugenol (4-			
C10H12O2	C10H10O2	C10H12O4	C10H
Mono constituent	Mono constituent	Mono constituent	Monc
COc1cc(C=C)ccc1O	COc1=CC(C=C)O=	COc1cc(ccc1O)C(O)C=	COc1.

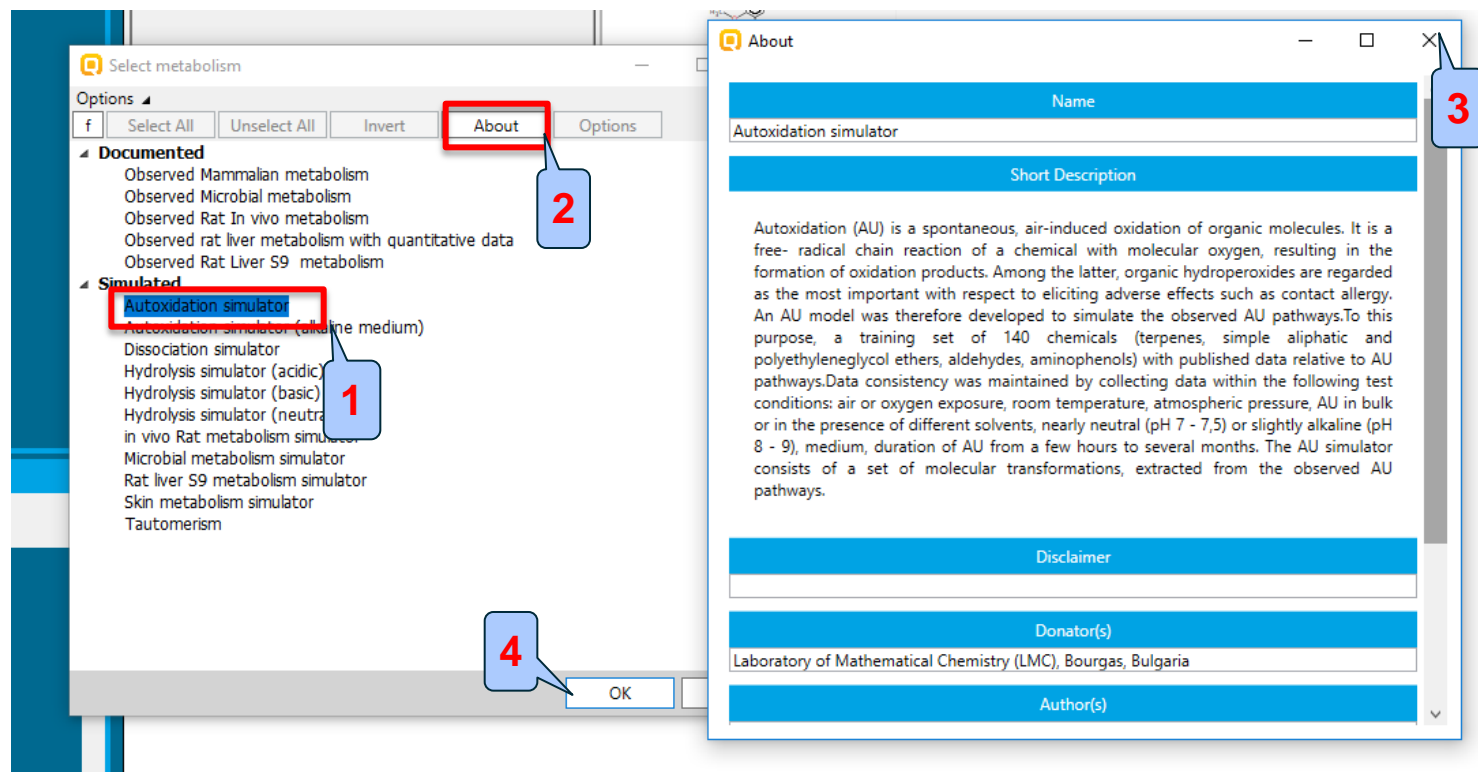
1. Go to **Data**; 2. Tick *Skin Sensitization* and *Skin Sensitization ECETOC* databases;

# Category definition with applied metabolism

The screenshot displays the QSAR Toolbox software interface. The top toolbar contains several icons, with the 'Category definition' icon (a tree structure) highlighted by a red box and labeled with a blue callout '1'. Below the toolbar, the 'Define with metabolism' button is highlighted by a red box and labeled with a blue callout '2'. The main workspace shows a document titled 'Document 1' with CAS number 97531. A 'Structure' panel is visible, showing a chemical structure and a list of properties: Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards. A 'Select metabolism' dialog box is open, showing a list of metabolism options. The 'Autoxidation simulator' option is highlighted by a red box and labeled with a blue callout '3'. The 'OK' and 'Cancel' buttons at the bottom of the dialog box are labeled with a blue callout '4'.

1. Go to **Category definition**; 2. Click **Define with metabolism**; 3. Select **Autoxidation simulator**; 4. Click **OK**.

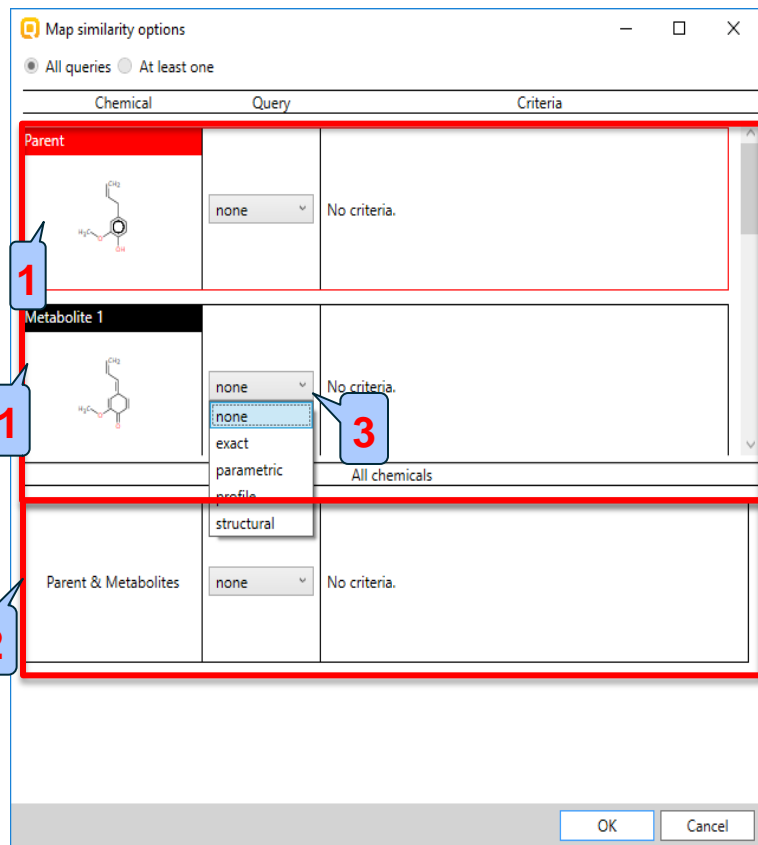
# Category definition with applied metabolism



All available transformation maps – documented and simulated in Toolbox can be used in the primary grouping.

1. Click *Autoxidation simulator*;
2. Click *About* to read the short description of the simulator;
3. Close the *About* window;
4. Click *OK*.

# Category definition with applied metabolism



*Map similarity options* dialogue appears. It shows all the generated metabolites of the target chemical (use the scroll bar to see them). It has two subsections:

- (1) shows the parent and each of the generated metabolites. This allows defining different criteria for each structure when looking for analogues.
- (2) treats the parent and its metabolites as a whole. i.e. the criteria is provided for the whole package but not for separate metabolites.

A drop down menu (3) is available for each of the structures (in the column "Query") which allow setting the type of criteria for looking for analogues.

# Category definition with applying metabolism

Explanation of different options from the drop down menu:

- **None** – default options; no criteria is set;
- **Exact** – search for analogues which metabolites have the exact structure of the target metabolite; only available for the metabolites and the package “parent + metabolites” but not for the parent chemical;
- **Parametric** – to have specific value or range of variation of defined parameter (a list with all parameters currently available in the Toolbox is provided);
- **Profile** – to have specific category by selected profiler (a list with all profilers is provided);
- **Structural** –allows structural similarity assessment based on the atom centered fragments.
- The user can select a profiling, parametric or structural query for both target and its metabolites.

# Category definition with applied metabolism

The image shows the 'Map similarity options' window in the QSAR Toolbox. It contains a table with columns for 'Chemical', 'Query', and 'Criteria'. The table has three rows: 'Parent', 'Metabolite 1', and 'Parent & Metabolites'. The 'Parent & Metabolites' row has a 'profile' dropdown menu (1), a 'Profiler' dropdown menu set to 'Protein binding alerts for skin sensitization by OASIS' (2), and an 'Edit' button (3). To the right is a 'Target' dialog box with a list of reaction types. The 'Acylation >> Direct acylation involving a leaving group >> (Thio)Acetyl and (thio)car...' option is selected (4). Below the list are 'Combine profiles' options: 'AND' (selected), 'OR', 'Invert result', and 'Strict'. The 'AND' option is selected (5). Both windows have 'OK' and 'Cancel' buttons at the bottom.

1. Select a profile option for the package "parent & metabolites";
2. Select "Protein binding alerts for SS by OASIS" profile;
3. Click **Edit** . The profiling results of the parent structure and its metabolites are shown.
4. Click **OK** to confirm the defined search criteria.
5. Click **OK** in Map similarity options window to execute the search.

# Category definition with applied metabolism

Category of 3 analogues has been defined with EC3 data

10 points added across 4 chemicals.

Read data?  
 All endpoints  Choose...  from Tautomers

	1 [target]	2	3	4
(1/1)	M: Negative			
(1/1)	M: Positive			
(1/3)	M: 8E+03 µg/cm2			
(4/4)	M: Positive	M: Positive	M: Positive	M: Positive
(1/1)	M: Positive			

1. Click **OK** to read all data; 2. An information window appears informing about the number of experimental data collected and the number of chemicals in the category (1 target and 3 analogues), click **OK**. 3. The analogues and their experimental data displayed on the matrix. Target and analogues have one experimental EC3 data each. The forthcoming two slides illustrates how consistent is the identified category with respect to protein binding alerts when metabolism is taken into account



# Categorization with applied metabolism

## Profiling results for parent and metabolites

The screenshot displays the QSAR Toolbox interface with the following components:

- Top Bar:** Navigation icons for Profiling, Data, Category definition, Data Gap Filling, and Report.
- Documents Panel:** Shows 'Document 1' with CAS: 97530 and a custom profile named 'Autoxidation simulator'.
- Profiling methods Panel:** Lists various methods. 'Protein binding alerts for skin sensitization according to GHS' and 'Protein binding alerts for skin sensitization by OASIS' are checked and circled in red.
- Metabolism/Transformations Panel:** Lists simulation methods. 'Autoxidation simulator' is checked and circled in red.
- Filter endpoint tree:** A hierarchical tree of endpoints. 'Protein binding alerts for skin sensitization according to GHS' and 'Protein binding alerts for skin sensitization by OASIS' are highlighted.
- Results Table:** A table with 4 columns representing target chemical structures. The table shows 'No alert found' for the parent compounds and lists various metabolic pathways (e.g., Michael Addition, Radical reactions, Schiff base forms, SN2 reactions) for the 5 metabolites of each target.

The profiling results indicate no protein binding alerts for the target chemical and the analogues, but protein binding alerts are identified in their autoxidation product

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

# Categorization with applied metabolism

## Profile statistic

The screenshot displays the QSAR Toolbox interface during a categorization task. The 'Filter endpoint tree...' window shows a hierarchical tree of endpoints. The 'Protein binding by OASIS' endpoint is selected, and a context menu is open over it. The 'Profile Statistics' dialog box is open, showing a table with one category and a bar chart with a single green bar. Callout boxes 1, 2, and 3 point to the selected node, the 'Profile Statistics' menu item, and the bar chart respectively.

**1** Right-click next to Metabolism/Transformations/Autoxidation simulator/General mechanistic/Protein binding by OASIS Protein binding by OASIS in the grey field;

**2** Click **Profile statistics**.

**3** All autoxidation products have the same distribution of the protein binding alerts.

1. Right-click next to Metabolism/Transformations/Autoxidation simulator/General mechanistic/Protein binding by OASIS Protein binding by OASIS in the grey field;
2. Click **Profile statistics**.
3. All autoxidation products have the same distribution of the protein binding alerts.

# Categorization with applied metabolism

Documents

Document 1  
# CAS: 97530  
Grouping with metabolism: 'Autoxidation simulator'

Profiling methods

Options

- Oncologic Primary Classification
- Protein binding alerts for Chromosomal aberration by
- Protein binding alerts for skin sensitization according
- Protein binding alerts for skin sensitization by OASIS
- Protein Binding Potency h-CLAT
- Respiratory sensitisation
- Retinoic Acid Receptor Binding
- rtER Expert System - USEPA
- Skin irritation/corrosion Exclusion rules by BFR
- Skin irritation/corrosion Inclusion rules by BFR
- Empiric

Metabolism/Transformations

Options

- Simulated
  - Autoxidation simulator
  - Autoxidation simulator (alkaline medium)
  - Dissociation simulator
  - Hydrolysis simulator (acidic)
  - Hydrolysis simulator (basic)
  - Hydrolysis simulator (neutral)
  - in vivo Rat metabolism simulator
  - Microbial metabolism simulator
  - Rat liver S9 metabolism simulator
  - Skin metabolism simulator
  - Tautomerism

Filter endpoint tree...

- Structure
- Structure info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards
  - Acute Toxicity
  - Bioaccumulation
  - Carcinogenicity
  - Developmental Toxicity / Teratogenicity
  - Genetic Toxicity
  - Immunotoxicity
  - Irritation / Corrosion
  - Neurotoxicity
  - Photoinduced toxicity
  - Repeated Dose Toxicity
  - Sensitisation
    - Respiratory Tract
    - Skin
      - in Vivo
        - GPMT
        - HRIPT
        - LLNA
        - EC3 (4/4) M: Positive
        - Undefined Assay (1/1) M: Positive
- ToxCast
- Toxicity to Reproduction
- Toxicokinetics, Metabolism and Distribution
- Profile
  - Endpoint Specific
    - Protein binding alerts for skin sensitization ...
  - Metabolism/Transformations
    - Autoxidation simulator
      - Endpoint Specific

	1 [target]	2	3	4
Structure				
Structure info				
Parameters				
Physical Chemical Properties				
Environmental Fate and Transport				
Ecotoxicological Information				
Human Health Hazards				
Acute Toxicity				
Bioaccumulation				
Carcinogenicity				
Developmental Toxicity / Teratogenicity				
Genetic Toxicity				
Immunotoxicity				
Irritation / Corrosion				
Neurotoxicity				
Photoinduced toxicity				
Repeated Dose Toxicity				
Sensitisation				
Respiratory Tract				
Skin				
in Vivo				
GPMT				
HRIPT				
LLNA				
EC3 (4/4) M: Positive	M: Positive	M: Positive	M: Positive	M: Positive
Undefined Assay (1/1) M: Positive				
ToxCast				
Toxicity to Reproduction				
Toxicokinetics, Metabolism and Distribution				
Profile				
Endpoint Specific				
Protein binding alerts for skin sensitization ...	No alert found	No alert found	No alert found	No alert found
Metabolism/Transformations				
Autoxidation simulator				
Endpoint Specific				
	5 metabolites	5 metabolites	5 metabolites	5 metabolites

Next action: Apply read-across for EC3 LLNA data based on the experimental data of the 3 analogues.

# Outlook

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

# Data gap filling

## Apply Read across

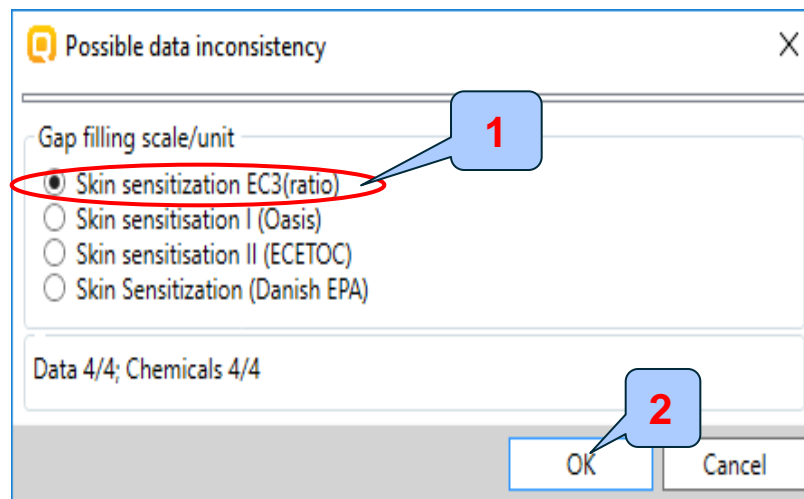
The screenshot shows the QSAR Toolbox interface with the 'Data Gap Filling' workflow selected. The 'Filter endpoint tree...' table is visible, showing a hierarchy of endpoints. The cell for 'Sensitization > Skin > In Vivo > LLNA > EC3' for target 1 is highlighted with a red circle and a blue callout box labeled '1'. The 'Read across' button in the top toolbar is also highlighted with a red circle and a blue callout box labeled '2'.

Filter endpoint tree...	1 [target]	2	3	4
Structure				
Parameters				
Physical Chemical Properties				
Environmental Fate and Transport				
Ecotoxicological Information				
Human Health Hazards				
Acute Toxicity				
Bioaccumulation				
Carcinogenicity				
Developmental Toxicity / Teratogenicity				
Genetic Toxicity				
Immunotoxicity				
Irritation / Corrosion				
Neurotoxicity				
Photoinduced toxicity				
Repeated Dose Toxicity				
Sensitisation				
Respiratory Tract				
Skin				
In Vivo				
GPMT	(1/1) M: Negative			
HRIPT	(1/1) M: Positive			
LLNA	(1/3) M: 8E+03 µg/cm2			
EC3	(4/4) M: Positive	M: Positive	M: Positive	M: Positive
Undefined Assay	(1/1) M: Positive			
ToxCast				
Toxicity to Reproduction				
Toxicokinetics, Metabolism and Distribution				
Profile				
Endpoint Specific				
Protein binding alerts for skin sensitization ...	No alert found	No alert found	No alert found	No alert found
Metabolism/Transformations				
Autoxidation simulator	5 metabolites	5 metabolites	5 metabolites	5 metabolites
Endpoint Specific				

1. Click the cell corresponding to *Sensitization >> Skin >> In Vivo >> LLNA >> EC3* for the target chemical; 2. Click *Read-across* .

# Data gap filling

## Scale definition



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

# Data gap filling Read-across

The screenshot displays the QSAR Toolbox software interface during a data gap filling workflow. The top navigation bar includes options for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows a document list and 'Data Gap Filling Settings' with checkboxes for 'Only endpoint relevant' and 'Only chemical relevant'. The main workspace is divided into several panels:

- Filter endpoint tree...:** A hierarchical tree view showing various endpoints such as 'respiratory tract', 'Skin', 'In Vivo', 'GPMT', 'HRIPT', 'LLNA', 'EC3', 'Undefined Assay', 'ToxCast', 'Toxicity to Reproduction', 'Toxicokinetics, Metabolism and Distribution', 'Profile', 'Endpoint Specific', 'Protein binding alerts for skin sensitization ...', 'Metabolism/Transformations', 'Autoxidation simulator', and 'Endpoint Specific'.
- Data Gap Filling Table:** A table with 4 columns representing different targets and rows for various endpoints. The row for 'EC3' is highlighted in yellow, showing 'M: Positive' for all four targets. Other rows include 'respiratory tract (1/1) M: negative', 'Skin (1/1) M: Positive', 'In Vivo (1/1) M: Positive', 'HRIPT (1/3) M: 8E+03 µg/cm2', 'LLNA (4/4) M: Positive', and 'Undefined Assay (1/1) M: Positive'.
- Read-across prediction for EC3, based on 3 values:** A scatter plot showing 'Observed: 13 %' and 'Predicted: 20.7 %'. The x-axis is 'log Kow' (ranging from 2.75 to 3.25) and the y-axis is 'EC3 [%]' (ranging from 20 to 30). The plot shows four data points: one at approximately (2.75, 20), one at (2.75, 25), one at (3.25, 15), and one at (3.25, 25).

Additional elements include a 'Documents' panel on the left, a 'Data Gap Filling Settings' panel, and a 'Read-across prediction' panel at the bottom right with a 'Accept prediction' button.

Initial graph without any subcategorizations.

# Data gap filling

## Subcategorization 1: Protein binding alerts for skin sensitization by OASIS

The screenshot displays the 'Subcategorization' window in the QSAR Toolbox. On the left, the 'Options' panel is open, showing a list of alert categories. The category 'Protein binding alerts for skin sensitization by OASIS' is selected and highlighted with a red circle and a blue callout box labeled '3'. Below this, the 'At this position:' section shows 'Automated workflows' and 'Standardized workflows' with a count of 1. In the center, the 'Adjust options' panel shows 'No alert found' circled in red. Below that, a table shows 'Differ from target by' with 'At least one category' selected. At the bottom, a scatter plot shows 'Read-across prediction for EC3, based on 3 values' with 'Observed: 13 %; Predicted: 20.7 %'. On the right, the 'Data Gap Filling' window shows a table with four columns representing target and three analogues. Each cell contains a chemical structure and the text 'No alert found'. A red box highlights this text with the message: 'There are no protein binding alerts found for the target chemical and its analogues'. At the bottom right, a 'Select / filter data' menu is open, with 'Subcategorize' selected and circled in red, and a blue callout box labeled '2'. Another blue callout box labeled '1' points to the 'Select / filter data' menu.

1. Open *Select filter data/subcategorize*; 2. Click **Subcategorize**; 3. Select Protein binding alerts for skin sensitization by OASIS



# Data gap filling

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

The screenshot displays the QSAR Toolbox interface during a subcategorization process. On the left, the 'Subcategorization' window is open, showing a list of options. Three callouts are present: Callout 1 points to the 'Subcategorize' button in the bottom right; Callout 2 points to 'Protein binding alerts for skin sensitization by OASIS' in the options list; Callout 3 points to 'Autoxidation simulator' in the 'Simulators' section. The 'Adjust options' window is also open, showing a list of alerts with 'Michael Addition' and 'Radical reactions' selected. The main window shows a 'Data Gap Filling' report with a table of chemical structures and their predicted values. A red box highlights the text: 'The same autoxidation products of target chemical and its analogues explain the positive experimental data'. At the bottom, a scatter plot shows 'EC3 [%]' vs 'log Kow' with a 'Read-across prediction for EC3, based on 3 values' showing 'Observed: 13%; Predicted: 20.7%'. A 'Descriptors' panel on the left shows 'log Kow' as the active descriptor.

1. Click **Select/filter data**;
2. Click **Subcategorize**;
3. Select **Protein binding alerts for skin sensitization by OASIS**;
4. Select **Autoxidation simulator**.

# Data gap filling prediction

QSAR TOOLBOX

Input | Profiling | Data | Category definition | Data Gap Filling | Report

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

Documents

- Document 1
  - # CAS: 97530
    - Grouping with metabolism: 'Autoxidation simulator'
      - Enter GF(RA) with 4 chemicals, 4 data points

EC3 data scale is used in RA

Predicted value for the target chemical is 20.7%

Target chemical is predicted as positive skin sensitizer

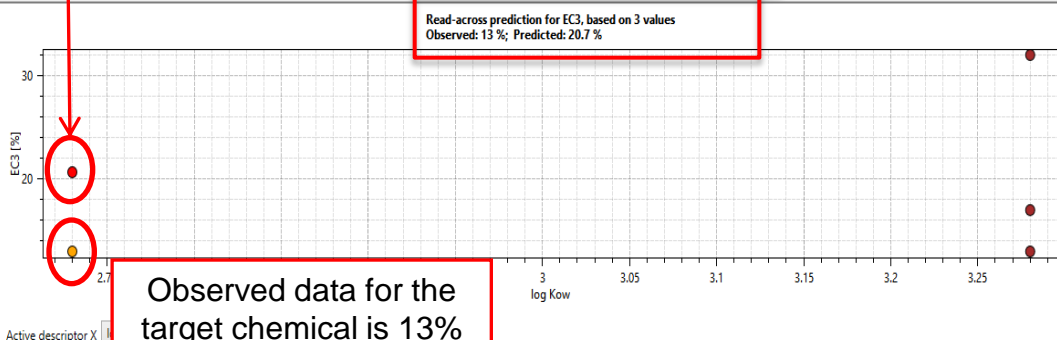
Read-across prediction for EC3, based on 3 values  
Observed: 13% ; Predicted: 20.7%

Observed data for the target chemical is 13%

Filter endpoint tree...

Structure

1 [target]	2	3	4
respiratory tract (1/1) M: Negative			
Skin			
In Vivo			
GPMT (1/1) M: Positive			
GPRT (1/3) M: 8E+03 µg/cm2			
M: Positive	M: Positive	M: Positive	M: Positive
M: Positive			
ToxC			
Toxic			
Toxic			
Profile			
Protein binding alerts for skin sensitization ...	No alert found	No alert found	No alert found
Metabolism/Transformations	5 metabolites	5 metabolites	5 metabolites
Autoxidation simulator			
Endpoint Specific			



Select / filter data

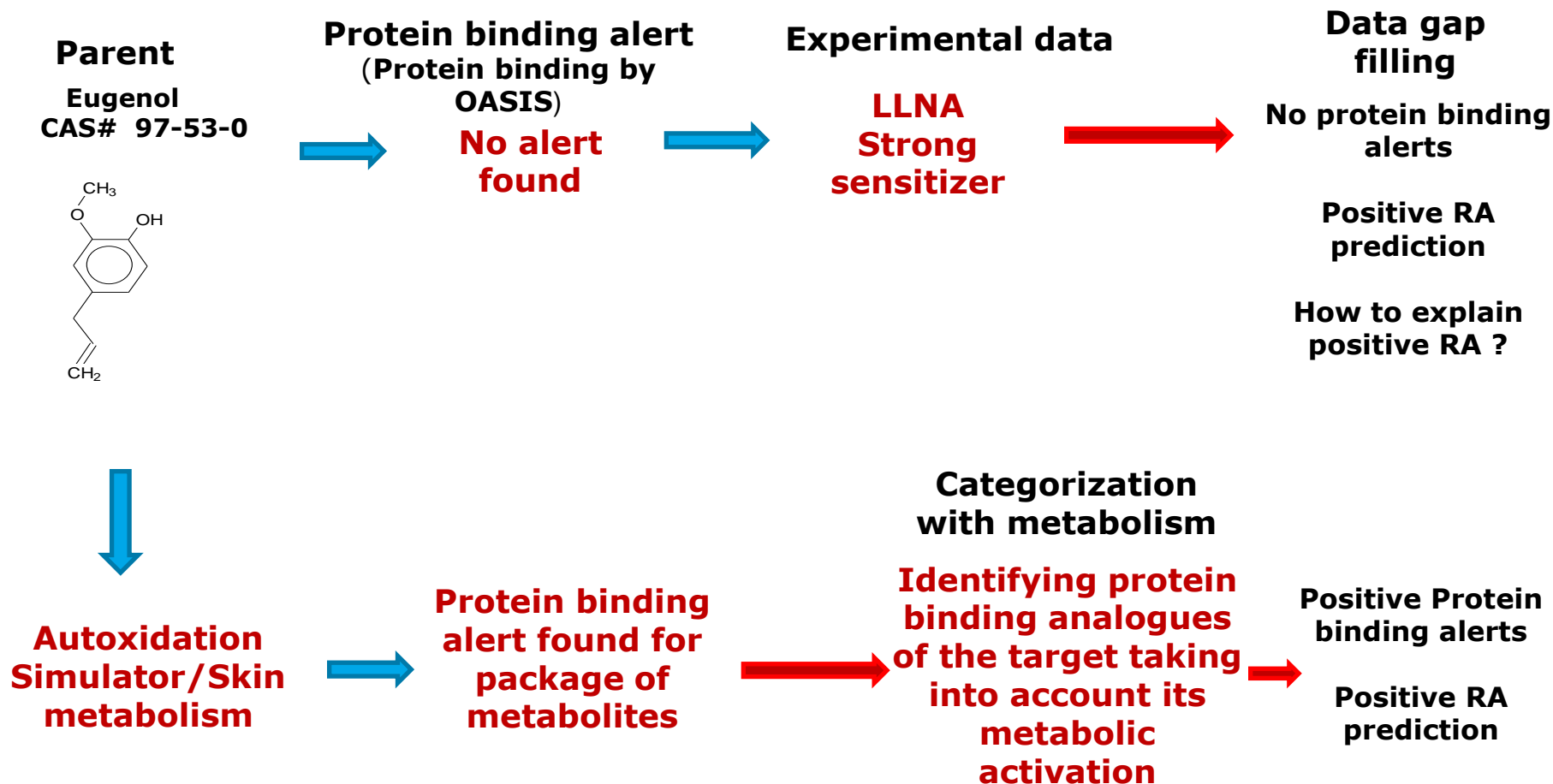
- Subcategorize
- Mark chemicals by WS
- Mark chemicals by descriptor value
- Filter points by test conditions
- Mark focused chemical
- Mark focused points
- Remove marked data
- Clear existing marks
- Accept prediction

# Data gap filling prediction

The screenshot displays the QSAR Toolbox interface during a data gap filling prediction. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Documents' and 'Data Gap Filling Settings' with options for endpoint and chemical relevance. The central area features a 'Filter endpoint tree...' on the left and a 'Structure' table on the right. The 'Structure' table contains four chemical structures under the heading '1 [target]'. A 'Confirm' dialog box is overlaid on the table, asking 'Are you sure you want to accept this prediction?' with 'Yes' and 'No' buttons. A red circle highlights the 'Accept prediction' button in the bottom right corner, with a blue callout containing the number '1'. Another blue callout with the number '2' points to the 'Yes' button in the 'Confirm' dialog.

1. Click Accept prediction; 2. Click **Yes** to confirm the prediction.

# Recap



# Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - Categorization applying metabolism
  - Data gap filling by taking into account metabolism
- **Report**

# Report

- The report module allows you to generate a report on the predictions obtained with the Toolbox.
- This module contains predefined report templates .
- The report can then be printed or saved in different formats.

*The Generating of a report is shown on next slides.*

# Report

1

2

3

Documents

- Document 1
  - # CAS: 97530
    - Autoxidation simulator
      - Grouping with metabolism: 'Autoxidation simulator'
        - Enter GF(RA) with 4 chemicals, 4 data points

Filter endpoint tree...

Structure

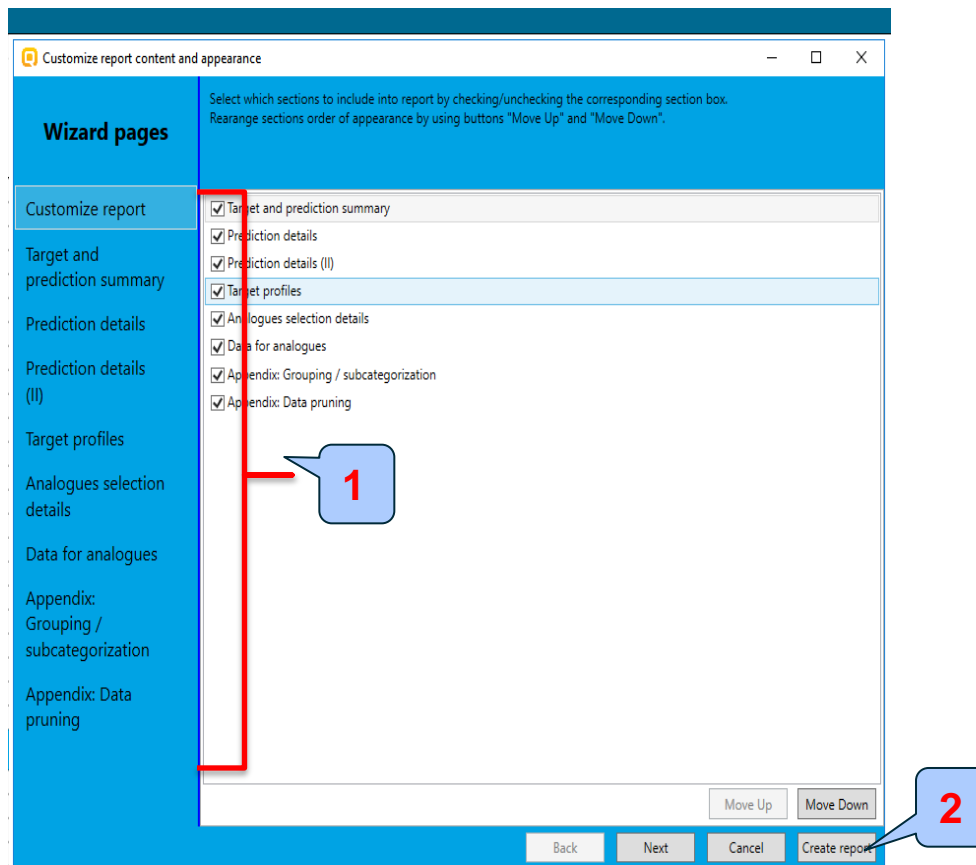
- Structure info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards
  - Acute Toxicity
  - Bioaccumulation
  - Carcinogenicity
  - Developmental Toxicity / Teratogenicity
  - Genetic Toxicity
  - Immunotoxicity
  - Irritation / Corrosion
  - Neurotoxicity
  - Photoinduced toxicity
  - Repeated Dose Toxicity
  - Sensitisation
    - Respiratory Tract
    - Skin
      - in Vivo
        - GPMT
        - HRIPT
        - LLNA
        - EC3
      - Undefined Assay
- ToxCast
  - Toxicity to Reproduction

	1 [target]	2	3	4
Structure				
Structure info				
Parameters				
Physical Chemical Properties				
Environmental Fate and Transport				
Ecotoxicological Information				
Human Health Hazards				
Acute Toxicity				
Bioaccumulation				
Carcinogenicity				
Developmental Toxicity / Teratogenicity				
Genetic Toxicity				
Immunotoxicity				
Irritation / Corrosion				
Neurotoxicity				
Photoinduced toxicity				
Repeated Dose Toxicity				
Sensitisation				
Respiratory Tract				
Skin				
in Vivo				
GPMT				
HRIPT				
LLNA				
EC3				
Undefined Assay				
ToxCast				
Toxicity to Reproduction				
AW SW AOP				
(1/1)	M: Negative			
(1/1)	M: Positive			
(1/3)	M: 8E+03 µg/cm2			
(4/5)	M: Positive R: Positive	M: Positive	M: Positive	M: Positive
(1/1)	M: Positive			

1. Go to **Report** section; 2. Highlight the prediction result – “Positive”; 3. Click **Prediction**;

# Report

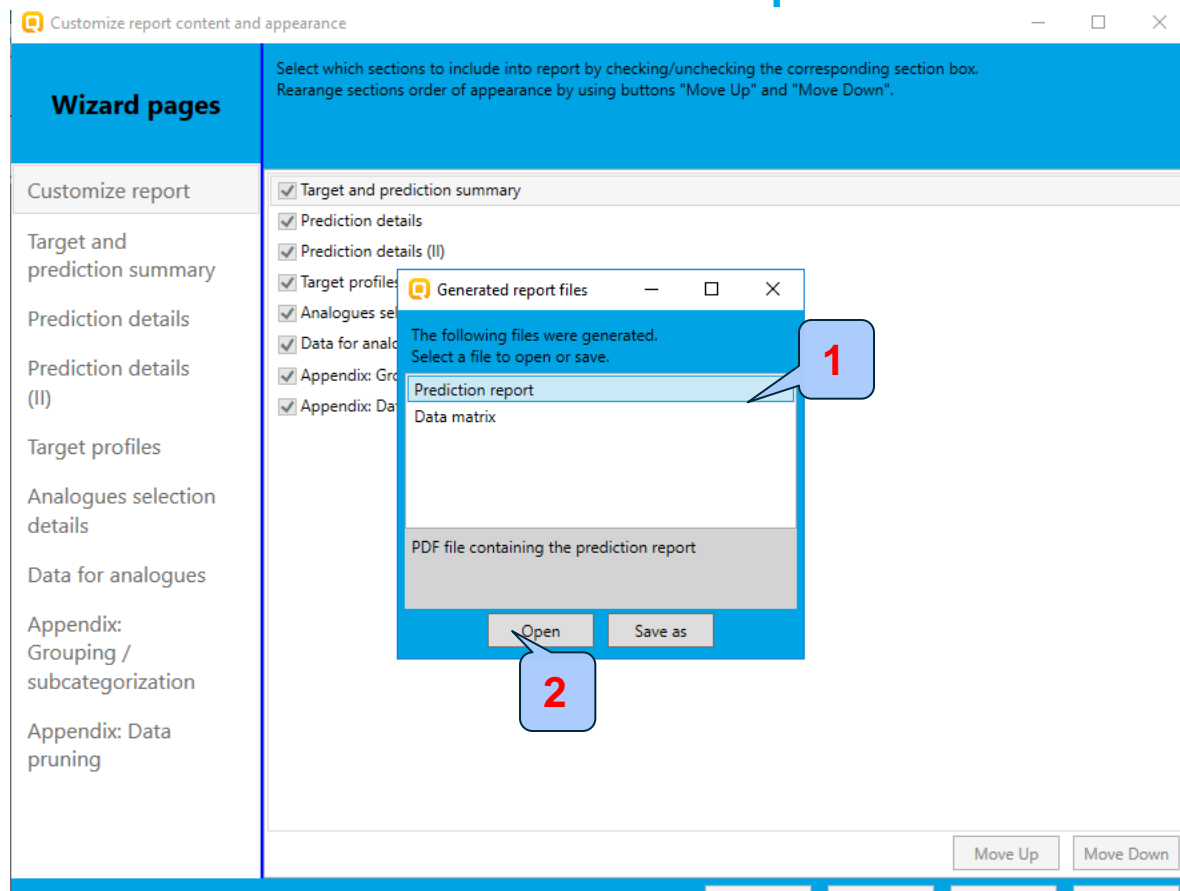
## Inserting additional information



1. The user could fill in additional information in some some of the fields;
2. Select **Create report**;



# Report Generation report



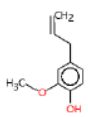
1. Two files are generated, which can be selected from the Generated report files window;
2. Select the file and then click **Open**.

# Report Overview

Prediction of EC3 for Eugenol 1 / 7

QSAR Toolbox prediction for single chemical

Date: 2 Aug 2017  
 Author(s):  
 Contact details:

Target information		
<b>Structural information</b>  SMILES: <chem>COc1cc(CC=C)ccc1O</chem>  Structure  	<b>Numerical identifiers</b>  EC#: N/A CAS#: 97-53-0 Other: N/A	<b>Chemical names</b>   eugenol eugenol (4-allyl-2-methoxypheno l) 4-allyl-2-methoxy -phenol 4-allyl-2-me thoxy-  phenol, 2-methoxy-4 -(2-propenyl)- eugen ol 1-ALLYL-3-METHOXY-4- HYDROXYBENZENE

Prediction summary
Predicted endpoint: EC3; No effect specified; No species specified; No duration specified; No guideline specified
<b>Predicted value: Positive</b>
Unit/scale: Skin sensitisation II (ECETOC)
Data gap filling method: Read-across analysis
Summary: <i>manually editable field</i> Not provided by the user

1

2

1. The prediction report is a file in PDF format
2. Predicted value is included in the Prediction summary

# Report

Prediction of EC3 for Eugenol 4 / 7

Target profiles (OECD principle 5 - Chemical and biological mechanisms)	
Profiles used for grouping/subcategorization	
Grouping with metabolism: 'Autoxidation simulator' (primary grouping)	Parent and 5 metabolites; Has all of the required categories: Michael Addition, Michael Addition >> Michael Addition on quinoid type compounds, Michael Addition >> Michael addition on quinoid type compounds >> Quinone methide(s)/imines; Quinoid oxime structure; Nitroquinones, Naphthoquinone(s)/imines, 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, Undefined
log Kow (calculated): 2.73	

1

Information that primary grouping with metabolism (Autoxidation simulator) was taken into account when predicting skin sensitization is included (1).

# Report

1

Prediction of EC3 for Eugenol

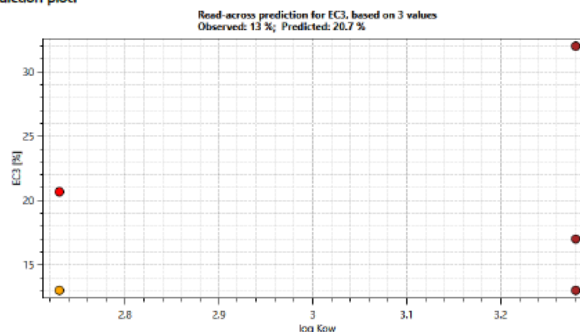
2/7

### Prediction details (1)

Predicted value: 20.7 %, conf.range: (-29.1 ; 70.4) at 95.0%

Predicted endpoint (OECD Principle 1 - Defined endpoint): Human Health Hazards -> Sensitisation -> EC3 -> LLNA -> in Vivo -> Skin

Prediction plot:



Calculation approach (OECD principle 2 - Unambiguous algorithm): takes the arithmetic mean (average) value from the nearest 3 neighbours

Active descriptor: log Kow (calculated)

Data usage: Arithmetic mean (average) value\*

\*When multiple values are available for the same chemical, their arithmetic mean (average) value is taken in prediction calculations

## 1. Predicted value

# Data matrix Overview

1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			Target chemical			Neighbour #1			Neighbour #2			Neighbour #3		
2	Substance identity													
3	Structure													
4	CAS number		97-53-0			186743-26-0			186743-25-9			186743-24-8		
5	Chemical name		Eugenol			3-METHYL_EUGENOL			5-METHYL_EUGENOL			6-METHYL_EUGENOL		
6	Other identifier													
7	SMILES		COc1cc(CC=C)ccc1O			COc1c(O)ccc(CC=C)c1C			COc1cc(CC=C)c(C)cc1O			COc1cc(CC=C)cc(C)c1O		
15	Measured and predicted data													
16	Data used for prediction													
17	environment	endpoint	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen	value	unit	species, duration, test type, type of method, assay, strain, test guideline, year, referen
18	Sensitisation	EC3			mouse in Vivo LLNA 2005 Dermatitis, 16 (4): 1-46 Gerberick, G.F.,Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. Compilation of historical Local Lvmph	32	%	mouse in Vivo LLNA 2005 Dermatitis, 16 (4): 1-46 Gerberick, G.F.,Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. Compilation of historical Local Lvmph	13	%	mouse in Vivo LLNA 2005 Dermatitis, 16 (4): 1-46 Gerberick, G.F.,Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. Compilation of historical Local Lvmph	17	%	mouse in Vivo LLNA 2005 Dermatitis, 16 (4): 1-46 Gerberick, G.F.,Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. Compilation of historical Local Lvmph

1. The data matrix is an *Excel* file, which contains information about the analogues

# Outlook

- Background
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Categorization
  - Data gap filling without taken into account metabolism
  - Categorization applying metabolism
  - Data gap filling by taking into account metabolism
  - Report
  - **Save the prediction**

## Saving the prediction result

- This functionality allows 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

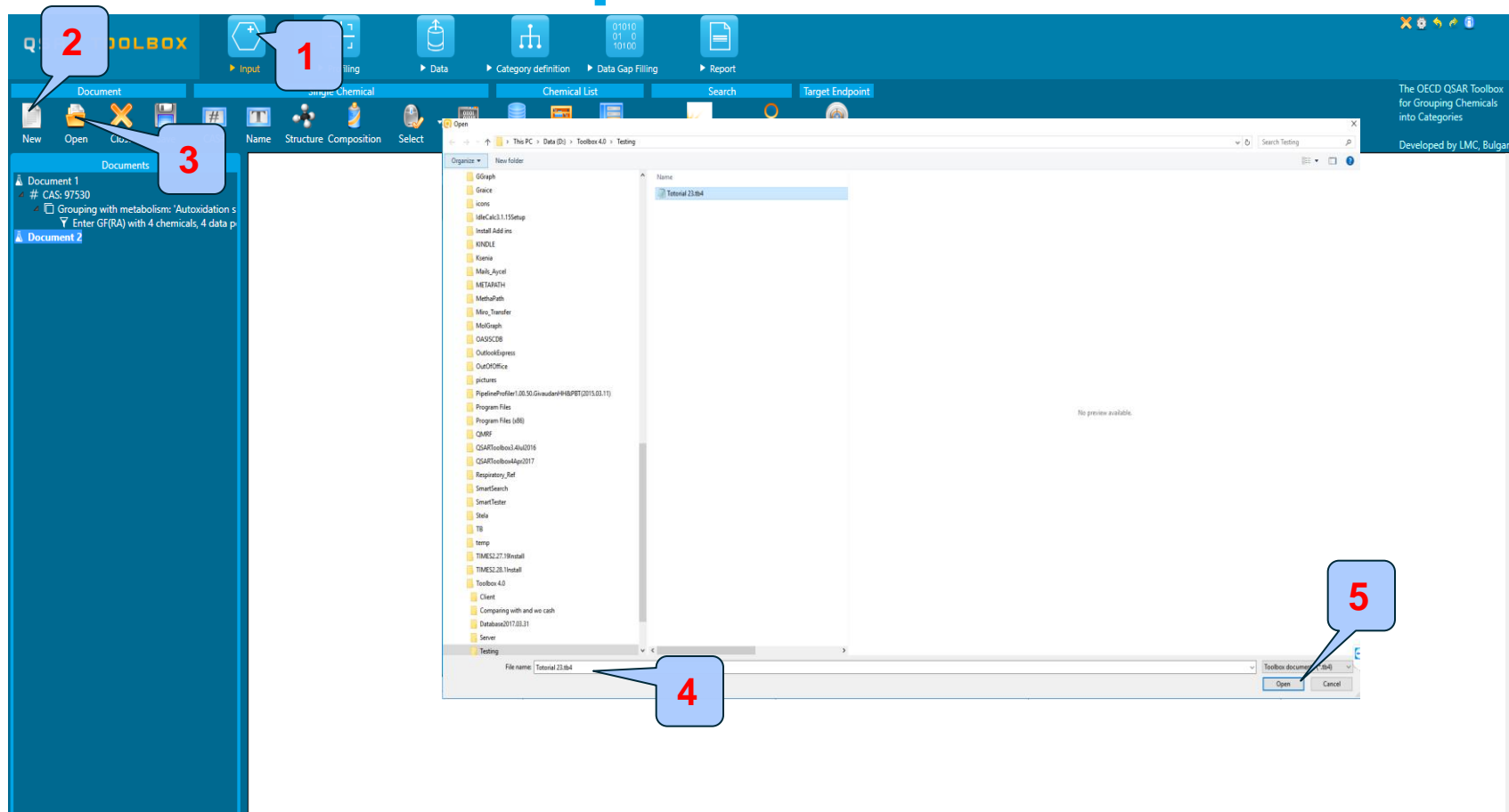
The screenshot displays the QSAR Toolbox software interface. A 'Save document' dialog box is open, showing a file explorer view of the 'Testing' folder. The dialog box has a 'File name' field containing 'Testing' and 'Save' and 'Cancel' buttons. Three red callouts with numbers 1, 2, and 3 indicate the steps: 1. Click 'Save' button; 2. Browse and type in the name of the file; 3. Click 'Save'. The background shows the 'Define' window with a chemical structure and a table of results.

(4/5)	M: Positive	M: Positive	M: Positive	M: Positive
(1/1)	M: Positive			

1. Click **Save** button; 2. Browse and type in the name of the file; 3. Click **Save**.



# Open saved file



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**;