

OECD (Q)SAR Toolbox v.4.4.1

Tutorial on how to predict Skin sensitization potential
taking into account alert performance

Outlook

- **Background**
- Keywords
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow

Background

- This is a step-by-step presentation designed to take the Toolbox user through the workflow of a data filling exercise accounting for alert performance.

Outlook

- Background
- **Keywords**
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow

Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options (e.g. Profiling)

WORKFLOW – the use, in combination, of the different modules (e.g. prediction workflow: from input to report)

PROFILER - algorithm (rule set) for the identification of specific features of the chemicals. Several types of profilers are available, such as structural (e.g. Organic functional groups), mechanistic (e.g. Protein binding by OECD) and endpoint-specific (e.g. in vitro in vitro mutagenicity (Ames test) alerts by ISS) profilers.

ALERT - the profilers consist of sets of rules or alerts. Each of the rules consists of a set of queries. The queries could be related to the chemical structure, physicochemical properties, experimental data, comparison with the target or list with substances and external queries from other predefined profilers (reference queries).

CATEGORY – “group” of substances sharing same characteristics (e.g. the same functional groups or mode of action). In a typical Toolbox workflow, it consists of the target chemical and its analogues gathered according to the selected profilers

ENDPOINT TREE – Endpoints are structured in a branched scheme, from a broader level (Phys-Chem properties, Environmental Fate and transport, Ecotoxicology, Human health hazard) to a more detailed one (e.g. EC3 in LLNA test under Human health hazard-Skin sensitization)

ALERT PERFORMANCE – is used to define how much relevant to a target endpoint an alert is. It reflect usability of category formation

DATA MATRIX – Table reporting the chemical(s) and data (experimental results, profilers outcomes, predictions). Each chemical is in a different column and each data in a different row

Outlook

- Background
- Keywords
- **Objectives**
- Specific Aims
- Alert performance
- The exercise
- Workflow

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Define target endpoint;
- Relevancy of profiles and data availability;
- Calculation of alert performance (AP).

Outlook

- Background
- Keywords
- Objectives
- **Specific Aims**
- Alert performance
- The exercise
- Workflow

Specific Aims

- To introduce to the Toolbox user to the workflow for defining the target endpoint;
- To familiarize the user with the new interface of the Toolbox;
- To familiarize the user with the different highlighting of profiles and databases;
- To familiarize the user with the calculation of alert performance;
- To explain to the Toolbox user the rationale behind each step of the exercise.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- **Alert performance**
- The exercise
- Workflow

Alert performance

Overview

- Alert performance (AP) is used to define how relevant to a target endpoint an alert is;
- AP reflects the alerts usability for category formation;
- AP can be calculated for any endpoint and any profile; one should only have preliminary defined target endpoint;
- AP can be calculate for an alert with or without accounting for metabolism;

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Alert performance
- **The exercise**
- Workflow

The Exercise

- In this exercise we will illustrate how to calculate alert performance and to use this information for predicting skin sensitization potential (EC3 LLNA assay) for two untested compounds: *pyridaphenthion* [CAS# 119-12-0] and *2-methoxyphenol (guaiacol)* [CAS# 90-05-1] which will be the “target” chemicals.
- We will preliminary define the target endpoint.
- This prediction will be accomplished by calculation of alert performance.
- The category will be defined by the mechanism of protein binding common to all the chemicals in the category.
- The metabolic activation (skin sensitisation metabolism) will be taken into account in the prediction for the second target chemical [CAS# 90-05-1]
- The prediction itself will be made by “read-across”.

The Exercise

Theoretical considerations on Skin Sensitization

- Allergic contact dermatitis that results from skin sensitization is a significant health concern.
- Skin sensitization is a toxicological endpoint that is complex and conceptually difficult.
- Many organic chemicals have been shown to induce skin sensitization after covalent binding to skin proteins¹.
- Therefore, mechanisms by which organic chemicals bind with proteins are relevant to grouping chemicals that may be skin sensitizing agents.

¹ OECD (2014), *The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins*, OECD Series on Testing and Assessment, No. 168, OECD Publishing, Paris, <https://doi.org/10.1787/9789264221444-en>.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Alert performance
- The exercise
- **Workflow**

Workflow

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

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Alert performance
- The exercise
- **Workflow**
 - **Input**

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.

Input

Ways of Entering a Chemical

User Alternatives for Chemical ID:

A. Single target chemical

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

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Input Screen

Input target chemical by CAS#

Exercise 1: **CAS 119-12-0**

The screenshot shows the QSAR TOOLBOX software interface. A 'Search by CAS #' dialog box is open, displaying the search results for CAS 119-12-0. The results table shows the chemical name 'O-(1,6-Dihydro-6-oxo-1-phenyl-3', its SMILES, and sources. A chemical structure is shown next to the table. Numbered callouts (1-4) indicate the steps: 1. Click 'CAS#'; 2. Insert '119-12-0'; 3. Click 'Search'; 4. Click 'OK'.

1	CAS	119-12-0
	SMILES	CCOP(=S)(OCC)OC1C=CC(=...
	CS Relation	High
	Substance	Mono constituent
	Composition	
	Name	O-(1,6-Dihydro-6-oxo-1-phenyl-3
	Sources	DSSTOX ECHA PR

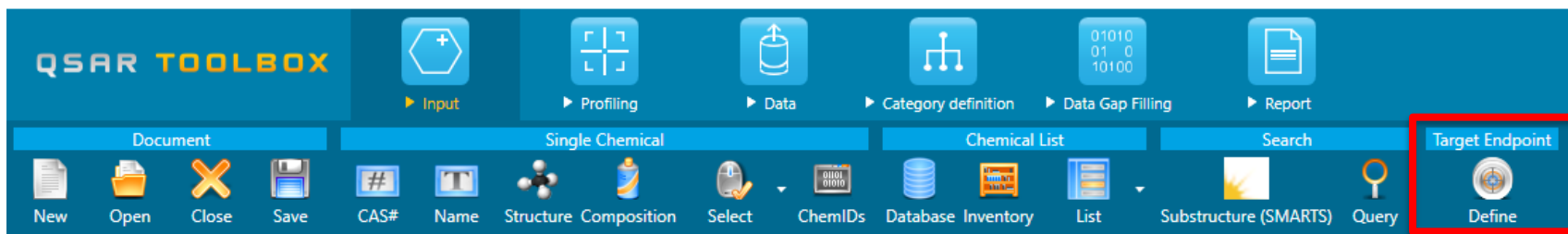
Click **CAS#** (1); Insert CAS **119-12-0** in the blank field (2) and click **Search** (3). When the structure appears, click **OK** (4).

Input

Define target endpoint

Calculation of alert performance (AP) is only possible if the target endpoint is preliminary defined.

Defining of the endpoint allows entering the endpoint of interest e.g. EC3, LC50, gene mutation etc., along with specific metadata information. Based on the metadata, different relevancy scores for profiles could be provided for the same endpoint.



Input

Define target endpoint

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Document, Single Chemical, Chemical List, Search, and Target Endpoint. The 'Target Endpoint' button is highlighted with a red box and labeled '1'. Below the menu bar, the 'Define' button is also highlighted with a red box and labeled '1'. The 'Select endpoint' dialog box is open, showing a list of endpoints under the 'Human Health Hazards' category. The 'Sensitization' endpoint is selected and labeled '3'. The 'Next' button at the bottom right of the dialog box is labeled '4'. The 'Structure' panel on the left shows a chemical structure and its properties.

1) Click **Define** icon; (2) "Select endpoint" dialogue appears, select **"Sensitization"** (3) and click on **Next** (4)

Input

Define target endpoint

On the next step you have to select the endpoint of interest and additional metadata if needed.

Select endpoint

Human Health Hazards
Sensitisation

Organ
Type of method
Assay
Endpoint

Skin
in Vivo
LLNA
EC3

Selection of additional metadata fields:

Add
Up
Clear

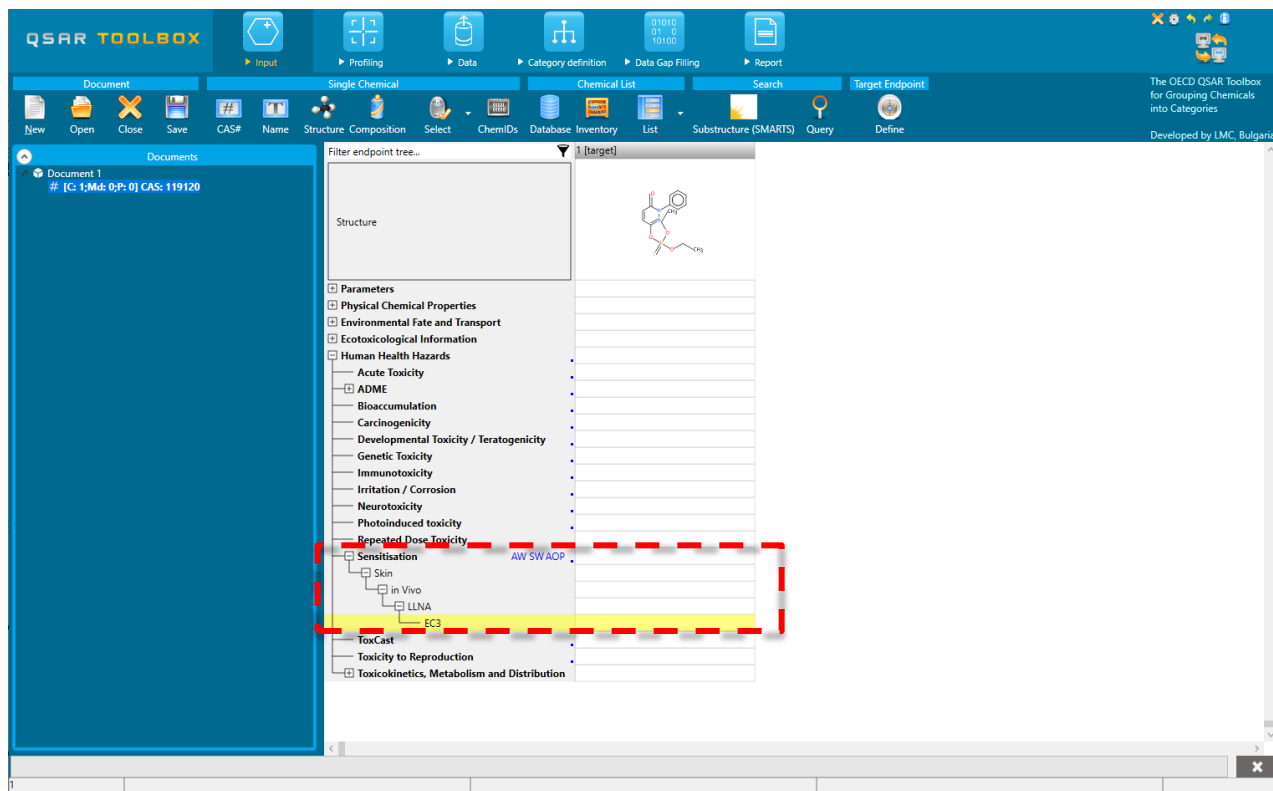
Undefine Back Finish

A new dialogue for defining additional details to the selected target endpoint appears;
(1) From the drop-down menus select the specific information for the metadata fields as follows:
Endpoint is **EC3**; Organ is **Skin**; Type of method is **in Vivo**; Assay is **LLNA** **(2)** Click **Finish**

Input

Define target endpoint

Once the endpoint is defined along with its metadata, they appear in the endpoint tree and the corresponding row of the data matrix is highlighted.



Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Alert performance
- 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

Sidebar to profilers' relevancy

Once the endpoint is selected, the relevant profilers and metabolic transformations are highlighted.

The screenshot displays the QSAR TOOLBOX software interface. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, the Profiling section is active, showing a sidebar with 'Profiling methods' and 'Metabolism/Transformations'. The 'Profiling methods' sidebar is divided into three categories: 'Suitable' (green), 'Plausible' (orange), and 'Unclassified' (no color). The 'Suitable' category includes 'Protein binding alerts for skin sensitization according to GHS', 'Protein binding alerts for skin sensitization by OASIS', and 'Protein binding by OASIS'. The 'Plausible' category includes 'Aquatic toxicity classification by ECOSAR', 'Chemical elements', 'Groups of elements', 'Keratinocyte gene expression', 'Lipinski Rule OASIS', 'OECD HPV Chemical Categories', 'Organic functional groups', 'Organic functional groups (nested)', 'Organic functional groups (US EPA)', 'Organic functional groups, Norbert Haider (checkmol)', 'Protein binding by OECD', 'Protein binding potency Cys (DPRA 13%)', 'Protein binding potency GSH', 'Protein Binding Potency h-CLAT', 'Protein binding potency Lys (DPRA 13%)', 'Respiratory sensitisation', 'Structure similarity', 'Substance type', and 'US-EPA New Chemical Categories'. The 'Unclassified' category includes 'Acute aquatic toxicity classification by Verhaar (Modified)', 'Acute aquatic toxicity MOA by OASIS', and 'Acute Oral Toxicity'. The 'Metabolism/Transformations' section is currently empty. The 'Filter endpoint tree...' panel on the right shows a tree structure with 'Structure' selected. The 'Structure' panel displays a chemical structure of a target molecule. The 'Filter endpoint tree...' panel also shows a list of endpoints, including 'Acute Toxicity', 'ADME', 'Bioaccumulation', 'Carcinogenicity', 'Developmental Toxicity / Teratogenicity', 'Genetic Toxicity', 'Immunotoxicity', 'Irritation / Corrosion', 'Neurotoxicity', 'Photoinduced toxicity', 'Repeated Dose Toxicity', 'Sensitisation', 'Skin', 'In Vivo', 'LLNA', and 'EC3'. The 'Sensitisation' endpoint is highlighted in blue, and the 'EC3' endpoint is highlighted in yellow.

- **Suitable** (in green) - developed using data/knowledge for the target endpoint;
- **Plausible** (in orange) - structure-based; form broader group of analogues;
- **Unclassified** (no color) – all profilers, which are not classified in any of the categories above.

Profiling

Profiling the target chemical

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

2. Click **Apply**

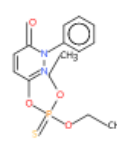
3. The profiling results of the target chemical appears in the data matrix

Profiling

Profiling the target chemical

Filter endpoint tree... 1 [target]

Structure



☒ Structure info
☒ Parameters
☒ Physical Chemical Properties
☒ Environmental Fate and Transport
☒ Ecotoxicological Information
☒ Human Health Hazards
☒ Profiling
 ☐ Endpoint Specific

Protein binding alerts for skin sensitization by OASIS

Alert 1 { Michael Addition
Michael Addition >> Michael addition on conjugated systems with electron withdrawing group
Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds
Schiff base formation

Alert 2 { Schiff base formation >> Pyrazolones and Pyrazolidinones derivatives
Schiff base formation >> Pyrazolones and Pyrazolidinones derivatives >> Pyrazolones and Pyrazolidinones

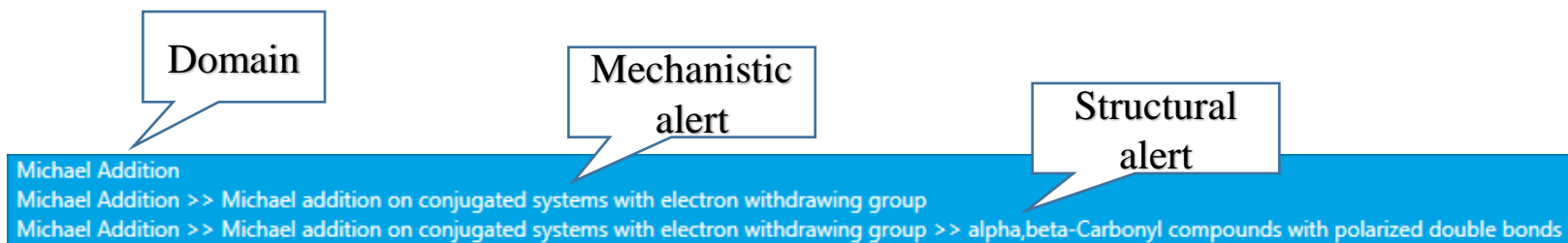
Alert 3 { SN2
SN2 >> Nucleophilic substitution at sp3 carbon atom
SN2 >> Nucleophilic substitution at sp3 carbon atom >> (Thio)Phosphates

Three protein binding alerts for skin sensitization are found in the target chemical.

Profiling

Sidebar on the hierarchical type profiles

Protein binding alerts for skin sensitization by OASIS is a hierarchical profile. The organization of the hierarchical profiles includes three levels of information for each category – domain, mechanistic alert and structural alert.



Right-click over the structural alert and select *Explain* to open the profiling scheme, where the user can see more details about the current alert.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read across and analogue approach
- 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

Sidebar on Data availability

Once the endpoint is selected, the databases, which contain such type of data, are highlighted in green.

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Data', 'Import', 'Export', and 'Delete'. The 'Data' sidebar on the left lists various databases and inventories. The 'Databases' section shows a list of databases, with 'ECHA REACH' and 'REACH Skin sensitisation database (normalise)' highlighted in green. The 'Inventories' section shows a list of inventories, with 'Canada DSL' and 'COSING' highlighted in green. The main window displays a chemical structure and a list of endpoints, with 'EC3' highlighted in yellow. The 'Filter endpoint tree...' panel on the right shows a tree structure of endpoints, with 'EC3' highlighted in yellow.

Data

Gather data

The screenshot shows the QSAR Toolbox interface. The 'Data' module is selected in the top toolbar (indicated by a callout '1'). The 'Gather' button is highlighted in the 'Data' dropdown menu (indicated by a callout '3'). In the 'Options' panel, the 'Skin Sensitization' database is selected (indicated by a callout '2'). A pop-up message box displays the text: 'No experimental data are available on the chemicals of interest.' The main window shows a chemical structure and a list of endpoints, with 'Skin Sensitization' expanded to show 'EC3' and 'ToxCast'.

1. Go to **Data** module;
 2. Select **Skin sensitization** database;
 3. Click **Gather**.
- A pop-up message informs that there is no experimental data for the target chemical.

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 (and therefore calculation of AP) is performed only among the chemicals which are listed in the selected databases. In this example only the Skin sensitization database is selected.
- In this example, an insert window appears stating there was “no data found” for the target chemical.

Recap

- In module one, you have entered the target chemical and defined the target endpoint.
- In the second module, you have profiled the target chemical with a profiler, which is suitable for the selected target endpoint.
- In the third module, you have seen the database corresponding to the defined target endpoint. You have found that no experimental data is currently available in the database for the structure.
- In other words, you have identified a data gap which you would like to fill in.
- Click “Category Definition” to move to the next module.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Data
 - **Category definition**

Category Definition

Overview

- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- This is the critical step in the workflow.
- Several options are available in the Toolbox to assist the user in refining the category definition.

Category Definition

Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity” so that within a category data gaps can be filled by read-across.
- For example, starting from a target chemical for which a specific protein binding mechanism is identified, analogues can be found which can bind by the same mechanism and for which experimental results are available.

Category Definition

Protein binding by OASIS grouping method

- This is one of the best grouping methods in the Toolbox. It is built on conventional organic chemical reactions and as such is qualitative in character.
- This method is particularly relevant for respiratory and skin sensitization and acute aquatic toxicity, but also for chromosomal aberration and acute inhalation toxicity.

Category Definition

Protein binding by OASIS grouping method

- This scheme includes 112 categories organized in three level of information:
 - ✓ Level I: Mechanistic Domains (11 categories)
 - ✓ Level II: Mechanistic alerts associated to each mechanistic domain are created on the basis of a common reactive centre being activated by a number of substituents (50 categories)
 - ✓ Level III: A number of structural alerts specifying the substituents to a common reactive center are made up each mechanistic alert (112 categories)

Category Definition

Protein binding by OASIS grouping method

- Each category from level III is presented by defined 2-dimensional structural alerts that is responsible for eliciting the toxic effects, such as skin sensitization which are a result of protein binding.
- The associated chemical reactions are in accordance with existing knowledge on electrophilic interaction mechanisms of various structural functionalities.

Category Definition

Protein binding by OASIS grouping method

- There is an agreement that most organic chemicals must react covalently with skin proteins in order to behave as skin sensitizers.
- Therefore, chemical reactions by which organic chemicals bind with proteins are relevant to grouping chemicals that may be skin sensitizing agents. So you have mechanistic plausibility for defining your category based on similar protein-binding mechanism.

Category Definition

- When more than one alert is found in the target structure before or after metabolic activation, Alert performance could be used to define which of them is the most suitable for primary categorization.

Alert performance

Overview

The performance of an alert represents the number of chemicals with data related to the predefined scale across all chemicals from the selected databases, which have the same alert. It provides and distribution of data according to a given effect (e.g. positive, negative) in percentages.

Category Definition

Calculation of Alert performance

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Category definition' menu is open, showing options: 'Define', 'Define with metabolism', 'Subcategorize', 'Combine', 'Clustering', and 'Category elements'. A red box labeled '2' highlights the 'Define' button. The 'Documents' panel on the left shows a list of documents, with 'Protein binding alerts for skin sensitization by OASIS' selected. A blue box labeled '1' highlights this selection. The 'Filter endpoint tree...' panel on the right shows a hierarchical list of endpoints, with 'Sensitisation' and 'EC3' highlighted. A chemical structure is displayed in the top right corner.

2

1

1. Select **Protein binding alerts for skin sensitization by OASIS**;
2. Click **Define**

Category Definition

Calculation of Alert performance

After clicking **Define**, the Categorization dialog appears. It consists of all protein binding alerts for SS found in the target structure.

The most suitable alert for category formation is determined by comparison of their alert performance.

Additional section for calculating "Alert performance" is designed in this dialogue, when the endpoint is preliminary defined (see on the next slide).

Alert performance can be calculated for only one alert, for combination of alerts or for all found alerts.

Category Definition

Calculation of Alert performance

alert 1

alert 2

alert 3

Grouping options (Protein binding alerts for skin sensitization by OASIS)

Target categories

- Michael Addition
 - Michael Addition >> Michael addition on conjugated systems with electron withdrawing group
 - Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds
- Schiff base formation
 - Schiff base formation >> Pyrazolones and Pyrazolidinones derivatives
 - Schiff base formation >> Pyrazolones and Pyrazolidinones derivatives >> Pyrazolones and Pyrazolidinones
- SN2
 - SN2 >> Nucleophilic substitution at sp3 carbon atom
 - SN2 >> Nucleophilic substitution at sp3 carbon atom >> (Thio)Phosphates

Options

Down Up Reset Options

All categories
(N/A)
Acylation
Acylation >> (Thio)carbamoylation of protein nucleophiles

Combine profiles

☐ Invert result ☐ Strict ☐ Sort results

☒ AND ☐ OR

Alert performance

Scales Calculate

OK Cancel

Category Definition

Calculation of Alert performance

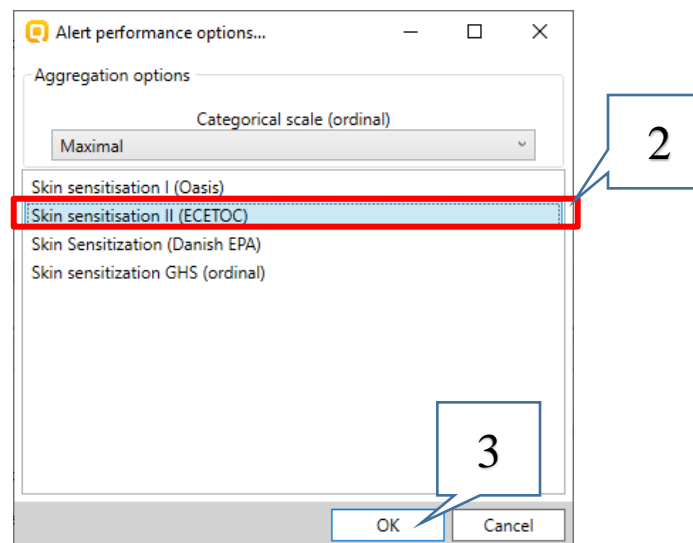
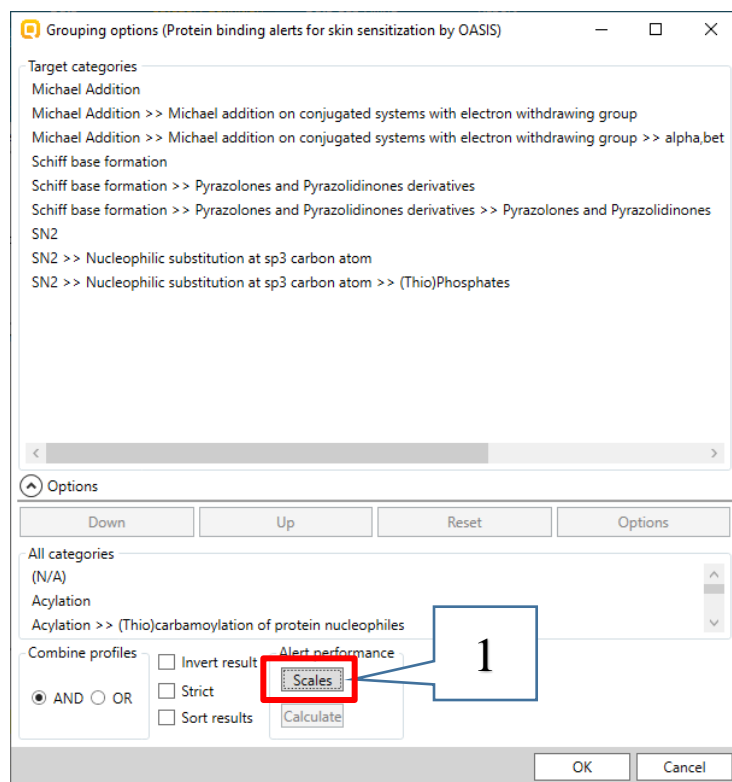
In order to calculate the performance of an alert, first of all you have to click on **Scales**.

The main purpose of the scales is to unify all data available in the Toolbox databases for a certain endpoint. Therefore, the most appropriate scale is "Skin Sensitisation II (ECETOC)". It is a dichotomous scale that converts all skin data into: Positive and Negative.

Additional option for applying different weight of the data that is available is also provided. Worst case scenario have been taken into account, i.e. "Maximal" data is set as default.

Category Definition

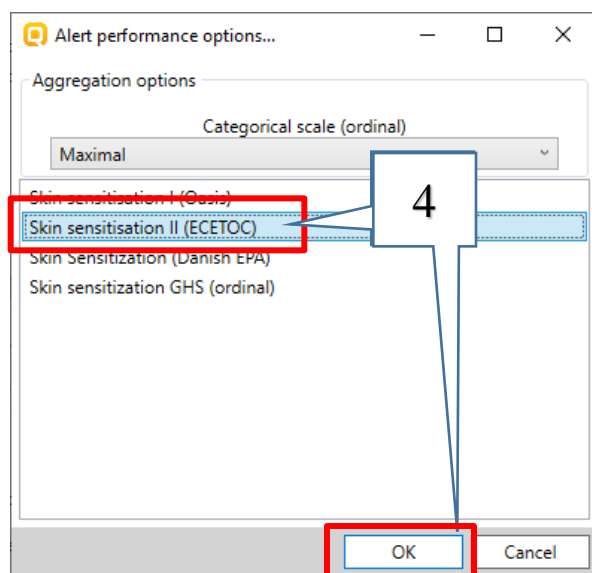
Calculation of Alert performance



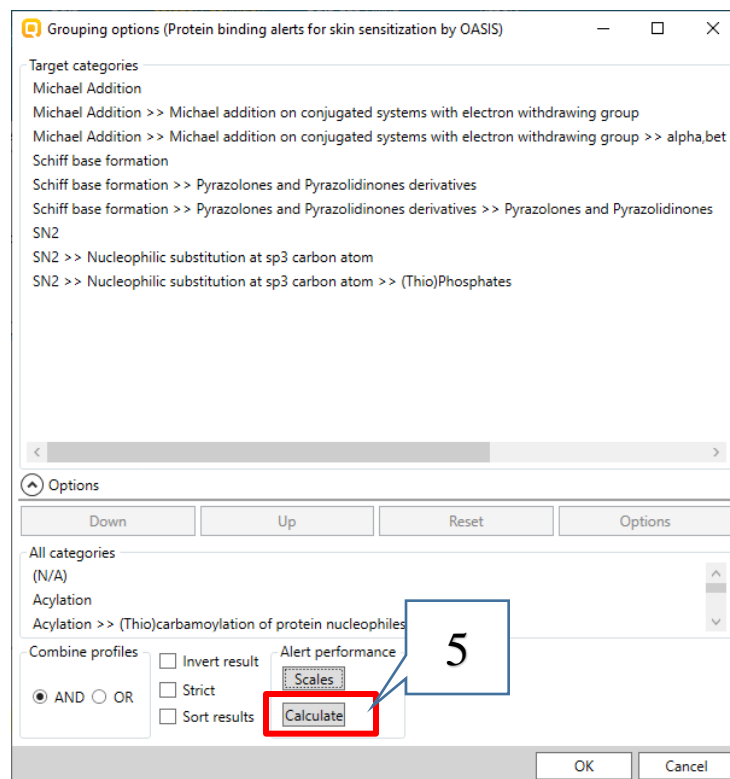
1. Click **Scales**;
2. Select **Skin sensitization II (ECETOC)** scale;
3. Confirm with "OK"

Category Definition

Calculation of Alert performance



4. After selection and scale and confirming with **OK**;
5. Click **Calculate**;



Category Definition

Calculation of Alert performance

Information for the calculated AP for each of the alerts appears in the following window:

1. Alert performance is calculated for the combination of **all alerts** (alerts are AND-ed);
2. And for **each alert individually**.

Alert performance results...

There were no chemicals with data for the alert/category.

SN2 >> Nucleophilic substitution at sp3 carbon atom >> (Thio)Phosphates<AND>Schiff base formation >> Pyrazolones and Pyrazolidinones derivatives >> Pyrazolones and Pyrazolidinones<AND>Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Positive	100.00%	Show chemicals... With data(1)...	Show all(1)...
SN2 >> Nucleophilic substitution at sp3 carbon atom >> (Thio)Phosphates	Negative	0.00%	Show chemicals... With data(0)...	
Schiff base formation >> Pyrazolones and Pyrazolidinones derivatives >> Pyrazolones and Pyrazolidinones	Positive	100.00%	Show chemicals... With data(3)...	Show all(3)...
	Negative	0.00%	Show chemicals... With data(0)...	
Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Positive	94.00%	Show chemicals... With data(47)...	Show all(50)...
	Negative	6.00%	Show chemicals... With data(3)...	

Close

Category Definition

Calculation of Alert performance for one alert

Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Positive	94.00%	Show chemicals... With data(47)...	Show all(50)...
	Negative	6.00%	Show chemicals... With data(3)...	

50 analogues having *alpha, beta-Carbonyl compounds with polarized double bonds* have been found in the database.

Of them:

- 47 out of 50 chemicals have positive data (94%)
- 3 out of 50 chemicals have negative data (6%).

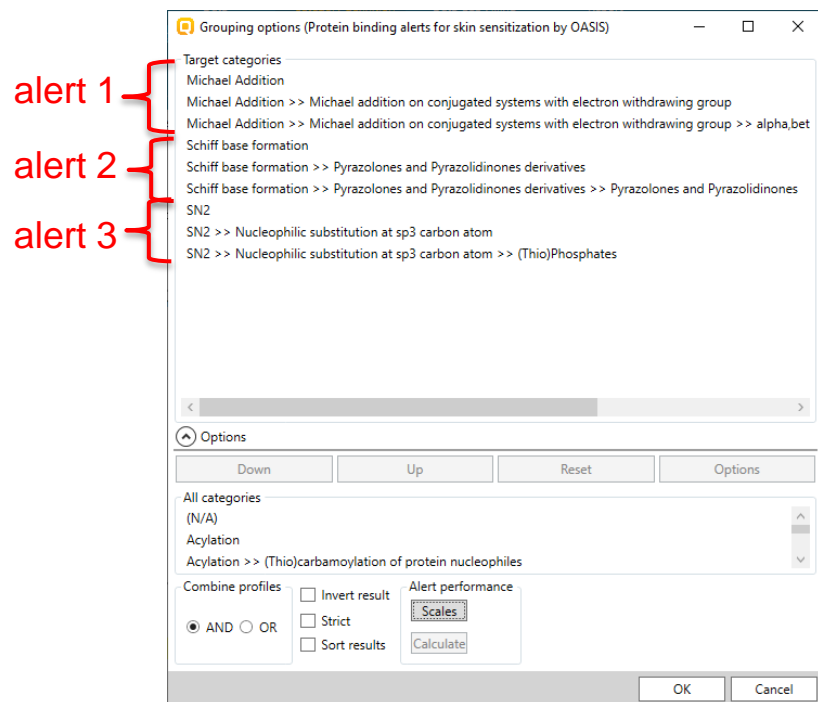


Keep in mind that the statistic is obtained from the chemicals and data, available in the selected databases

Category Definition

Calculation of Alert performance for one alert

In summary we see that the second alert (Schiff base formation >>....>> Pyrazolones and Pyrazolidinones) is the most suitable to define a category (100% positive alert). Read-across prediction is not illustrated in the forthcoming slides



Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds

Positive	94.00%
Negative	6.00%

Show chemicals...

With data(47)...

Show chemicals...

With data(3)...

Show all(50)...

alert 1

Schiff base formation >> Pyrazolones and Pyrazolidinones derivatives >> Pyrazolones and Pyrazolidinones

Positive	100.00%
Negative	0.00%

Show chemicals...

With data(3)...

Show chemicals...

With data(0)...

Show all(3)...

alert 2

SN2 >> Nucleophilic substitution at sp3 carbon atom >> (Thio)Phosphates

Positive	100.00%
Negative	0.00%

Show chemicals...

With data(1)...

Show chemicals...

With data(0)...

Show all(1)...

alert 3

Category Definition

Alert performance accounting for metabolism

Exercise 2: CAS 90-05-1

AP can be also calculated for alert(s) identified in the simulated metabolites after autooxidation simulation or metabolic simulation.

1. Click **Profiling** (1);
2. Check *Protein binding alerts for skin sensitization by OASIS* and *Skin metabolism simulator* (2);
3. Click **Apply** (3).

The screenshot displays the QSAR TOOLBOX software interface during the 'Profiling' step. The top navigation bar includes 'Input', 'Profiling' (highlighted with a red box and labeled '1'), 'Category definition', 'Data Gap Filling', and 'Report'. The 'Profiling' tab is active, showing 'Profiling methods' and 'Metabolism/Transformations' sections. In the 'Profiling methods' section, 'Suitable' is selected, and 'Protein binding alerts for skin sensitization by OASIS' is checked (highlighted with a red box and labeled '2'). In the 'Metabolism/Transformations' section, 'Suitable' is selected, and 'Skin metabolism simulator' is checked (highlighted with a red box and labeled '2'). The 'Apply' button is highlighted in the top left (labeled '3'). A chemical structure of CAS 90-05-1 is shown on the right. A filter endpoint tree on the right lists various endpoints, with 'Sensitisation' and 'Skin' highlighted.

Category Definition

Alert performance accounting for metabolism

Exercise 2: CAS 90-05-1

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Profiling Custom profile

Apply View New Delete

Documents

Profiling methods

Options Select All Unselect All Invert

☒ Suitable

- ☒ Protein binding alerts for skin sensitization according to GHS
- ☒ Protein binding alerts for skin sensitization by OASIS
- ☒ Protein binding by OASIS

☒ Plausible

- ☐ Aquatic toxicity classification by ECOSAR
- ☐ Chemical elements
- ☐ Groups of elements
- ☐ Keratinocyte gene expression
- ☐ Lipinski Rule Oase
- ☐ OECD HPV Chemical Categories
- ☐ Organic functional groups

Metabolism/Transformations

Options Select All Unselect All Invert

☒ Suitable

- ☒ Autooxidation simulator
- ☒ Skin metabolism simulator

☒ Plausible

- ☐ Autooxidation simulator (alkaline medium)
- ☐ Dissociation simulator
- ☐ Hydrolysis simulator (neutral)

☒ Unclassified

- ☐ Hydrolysis simulator (acidic)
- ☐ Hydrolysis simulator (basic)
- ☐ in vivo Rat metabolism simulator
- ☐ Microbial metabolism simulator

Filter endpoint tree... 1 [target]

Structure

Structure info

Parameters

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Human Health Hazards

Profiling

Endpoint Specific

Protein binding alerts for skin sensitiz... No alert found

Metabolism/Transformation

Skin metabolism simulator 4 metabolite(s)

Endpoint Specific

Protein binding alerts for skin sensitization by OASIS

1 x Michael Addition

1 x Michael Addition >> Michael addition on qin.

1 x Michael Addition >> Michael addition on qin.

1 x Schiff base formation

1 x Schiff base formation >> Schiff base formatio

1 x Schiff base formation >> Schiff base formatio

2 x No alert found

No protein binding alerts are found in the parent structure

Structural alerts are found in the generated metabolites after applying Skin metabolism simulator

Category Definition

Alert performance accounting for metabolism

The screenshot shows the QSAR TOOLBOX interface. At the top, there are icons for 'Input', 'Profiling', 'Data' (highlighted with a red box and a callout '1'), 'Category definition', and 'Data Gap Filling'. Below these are buttons for 'Data', 'Import', 'Export', and 'Delete'. The 'Databases' list on the left includes various toxicity databases, with 'Skin Sensitization' highlighted by a red box and a callout '2'. The 'Filter endpoint tree...' on the right shows a hierarchical structure of endpoints, with 'Skin' selected under 'Sensitisation'.

1. Go to **Data**;

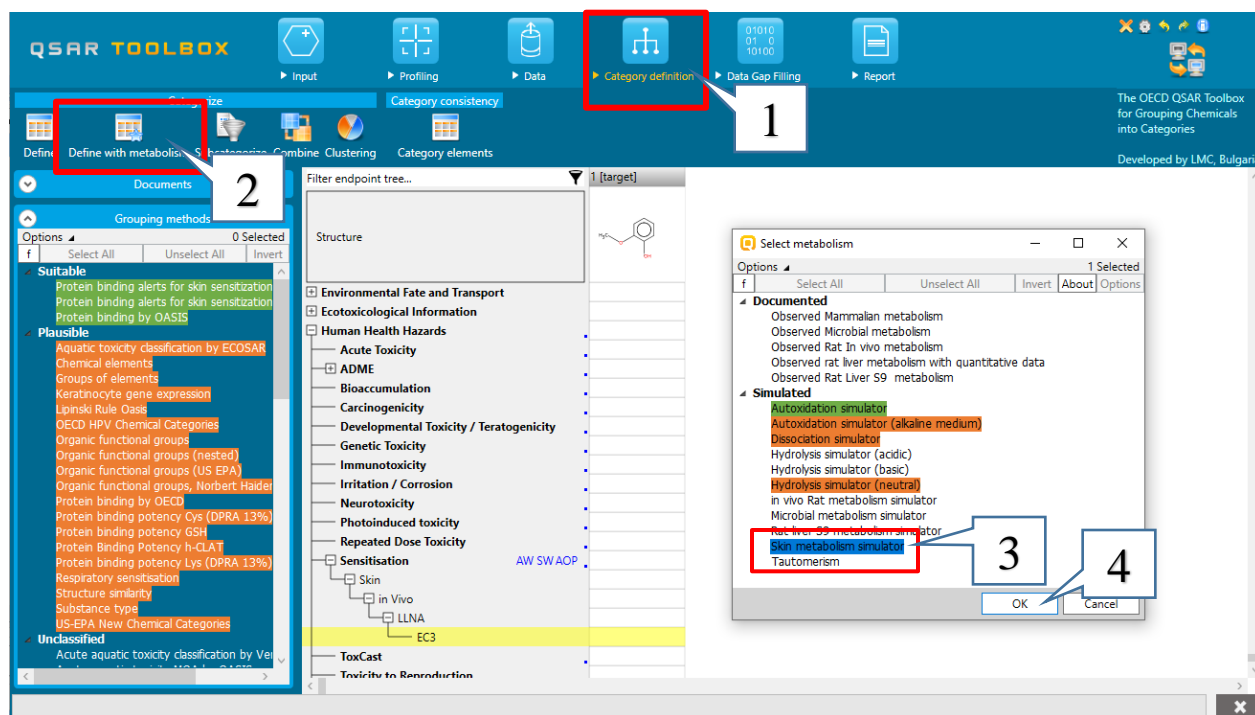
2. Select **Skin sensitization** database;

Category Definition

Alert performance accounting for metabolism

The target chemical has no alert for protein binding as parent but it is activated as a result of skin metabolism. In this respect, the primary category will be defined with accounting for the metabolic activation.

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







Category Definition

Alert performance accounting for metabolism

The system will search for chemicals which have similar distribution of the protein binding alerts as identified in the generated package parent and metabolites, accounting for the skin metabolism activation for the analogues.

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent 	none	No criteria.
Metabolite 1 	none	No criteria.
Metabolite 2 	none	No criteria.
All chemicals 	none	No criteria.

Alert performance

Scales

Calculate

OK Cancel

Category Definition

Alert performance accounting for metabolism

All chemicals

Parent & Metabolites	<div style="border: 1px solid red; padding: 2px; display: inline-block;">Profile ▾</div>	Profiler: <div style="border: 1px solid red; padding: 2px; display: inline-block;">Protein binding alerts for skin sensitization by OASIS ▾</div>	Options: <div style="border: 1px solid red; padding: 2px; display: inline-block;">Edit</div>
----------------------	--	---	--

1. Select **Profile** from the drop-down
2. Select **Protein binding alerts for skin sensitization by OASIS** from the drop-down menu ;.
3. Click **Edit**.

Target categories

Michael Addition

Michael Addition >> Michael addition on quinoid type compounds

Michael Addition >> Michael addition on quinoid type compounds >> Quinone methide(s)/imines; Quinoides

No alert found

Schiff base formation

Schiff base formation >> Schiff base formation with carbonyl compounds

Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes

Options

Down Up Reset Options

All categories (N/A)

Acylation

Acylation >> (Thio)carbamoylation of protein nucleophiles

Combine profiles

☒ AND ☐ OR

☐ Invert result

☐ Strict

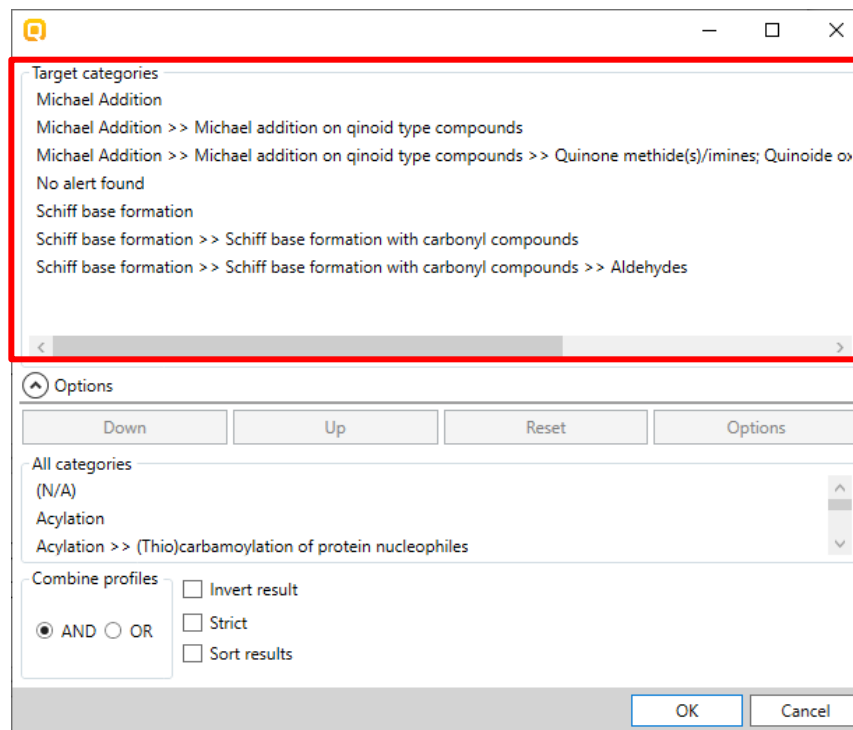
☐ Sort results

OK Cancel

Category Definition

Alert performance accounting for metabolism

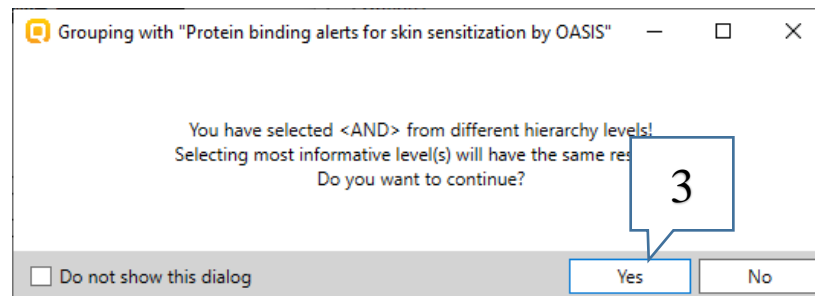
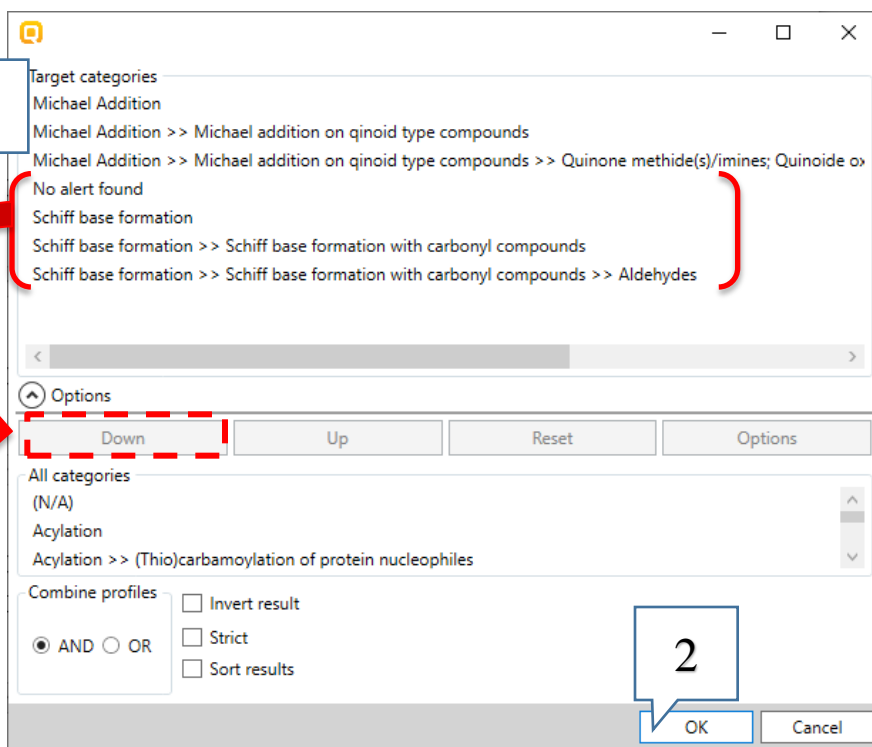
The found alerts could be seen by click on the **Edit** button. To calculate AP for one alert – remove all alerts except the alert for which AP will be calculated (three levels of mechanistic information are required – domain, mechanistic and structural alert), select a scale in the **Options** and click on **Calculate** (see next slide).



Category Definition

Calculation of Alert performance accounting metabolism for one alert

Now we will calculate AP for each of the alerts in order to see which of them is the most suitable for category formation.



1. Remove **No alert found** and the second alert by double click or using **"Down"** button;
2. Click **OK**.
3. Click **OK** in the Warning message.

Category Definition

Calculation of Alert performance accounting metabolism for one alert

Now we will calculate AP for each of the alerts in order to see which of them is the most suitable for category formation.

1. Click **Scales**;

2. Select Skin sensitization II (ECETOC)

3. Click **OK**.

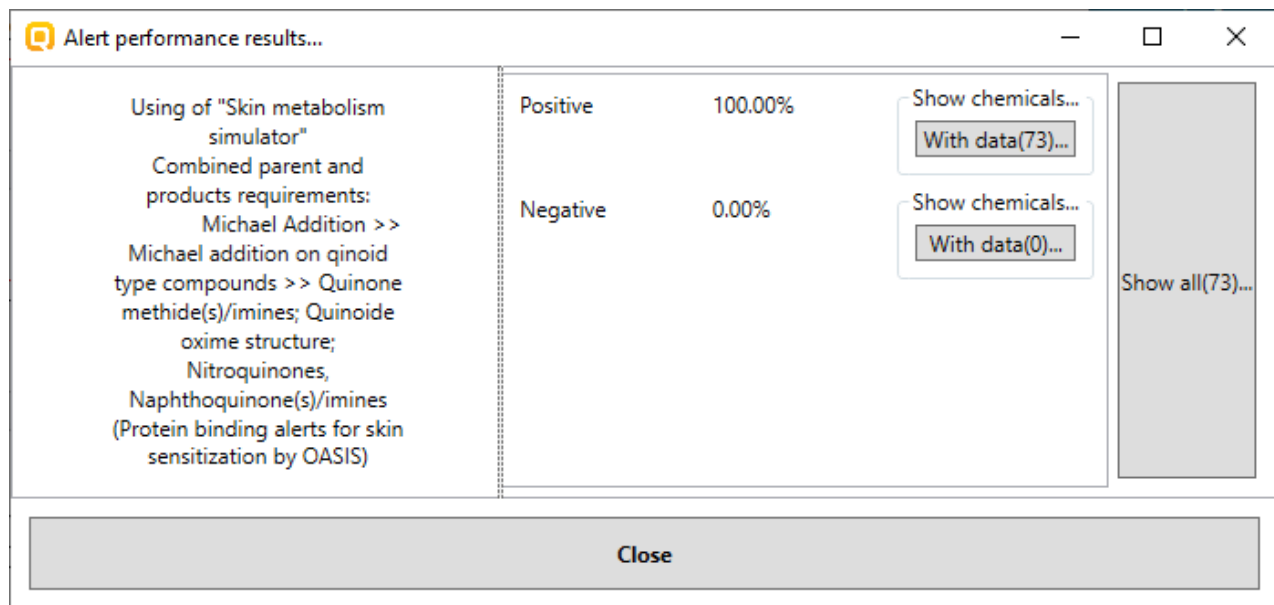
4. Then click **Calculate**

The screenshot shows the 'Alert performance options...' dialog box. The 'Aggregation options' tab is active, showing 'Maximal' selected for 'Categorical scale (ordinal)'. The 'Alert performance' tab is also visible, showing 'Scales' selected. A red 'X' is visible next to the 'Calculate' button. Numbered callouts 1 through 4 indicate the steps: 1. Click Scales; 2. Select Skin sensitization II (ECETOC); 3. Click OK; 4. Then click Calculate.

Category Definition

Calculation of Alert performance accounting metabolism for one alert

Performance of the first alert appears in the following window:



The screenshot shows a window titled "Alert performance results...". It displays the performance of a specific alert. The alert description is: "Using of 'Skin metabolism simulator' Combined parent and products requirements: Michael Addition >> Michael addition on quinoid type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines (Protein binding alerts for skin sensitization by OASIS)". The performance table shows 100.00% for Positive results and 0.00% for Negative results. There are buttons to "Show chemicals... With data(73)..." and "Show chemicals... With data(0)...". A "Show all(73)..." button is also present. A "Close" button is at the bottom.

Category	Positive	Negative
Using of "Skin metabolism simulator"	100.00%	0.00%

The system informs that 73 analogues with the searched alert accounting for skin metabolism have been found. All the chemicals have positive data (100%).

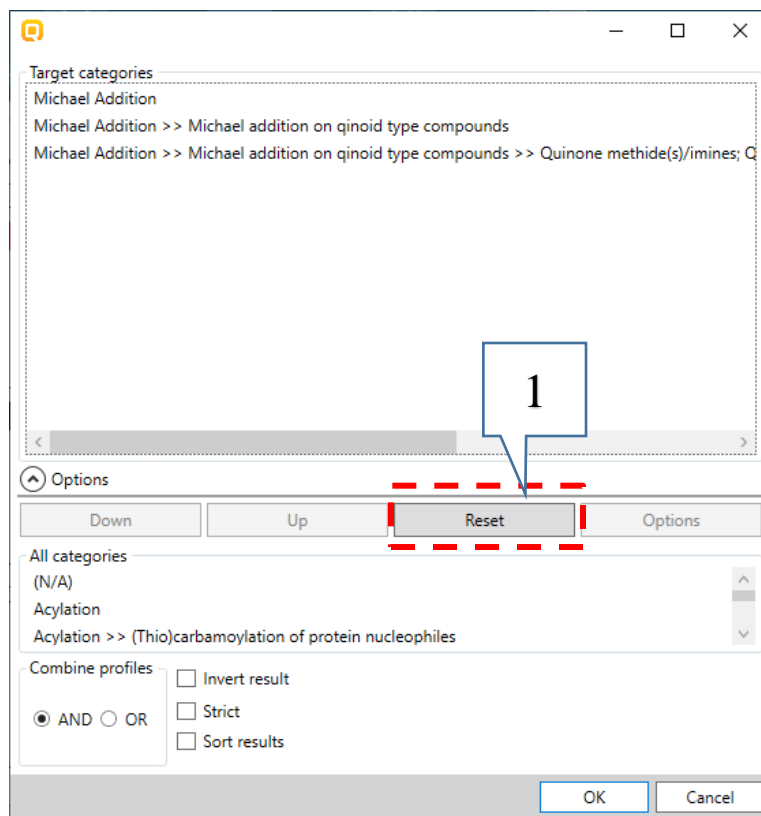


Keep in mind that the statistic is obtained from the chemicals and data, available in the selected databases

Category Definition

Calculation of Alert performance accounting metabolism for one alert

Click on the **Reset (1)** button and repeat the alert performance calculation steps for the second alert.



Category Definition

Calculation of Alert performance accounting metabolism for one alert

In summary we see that the first alert (**alert 1**) is the most suitable to define a category (because of its higher performance).

alert 1 {

- Target 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

alert 2 {

- No alert found
- Schiff base formation
- Schiff base formation >> Schiff base formation with carbonyl compounds
- Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes

Positive	100.00%	Show chemicals... With data(73)...	Show all(73)...
Negative	0.00%	Show chemicals... With data(0)...	

alert 1

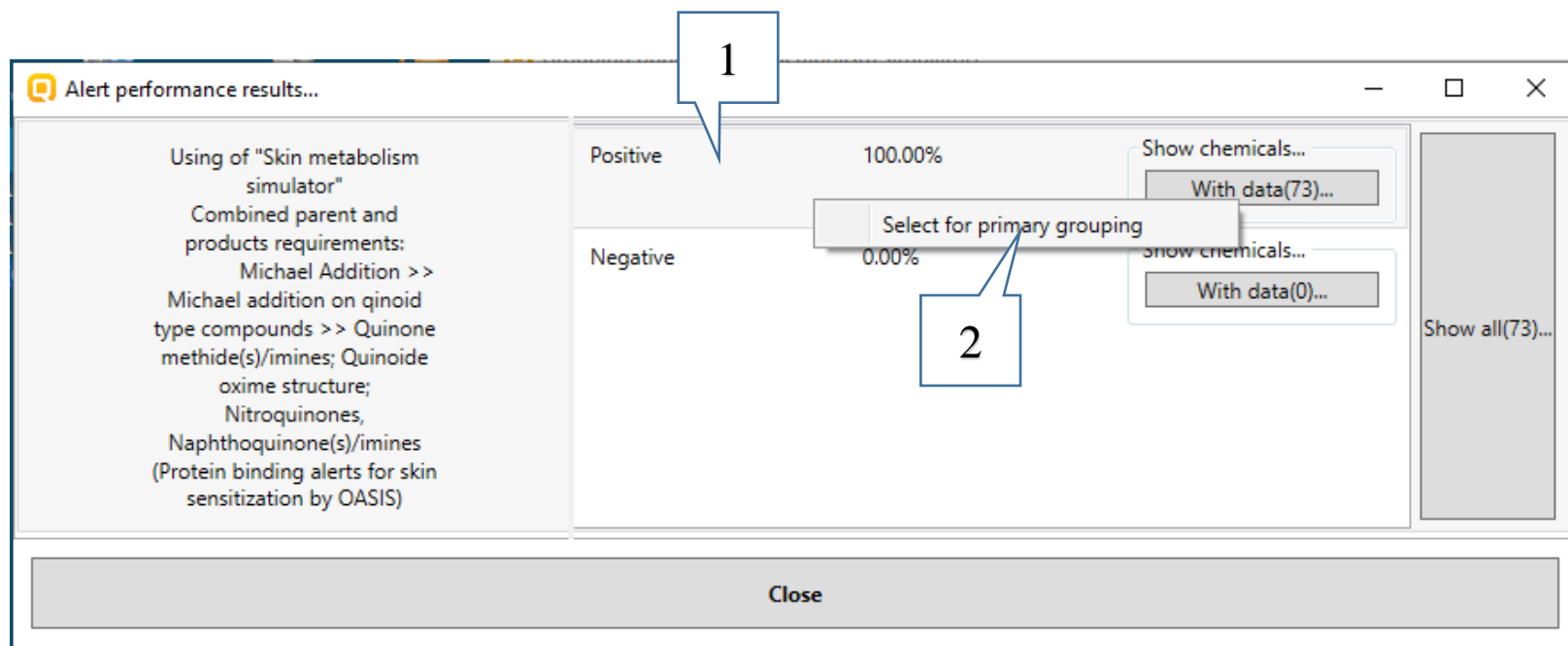
Positive	80.12%	Show chemicals... With data(137)...	Show all(171)...
Negative	19.88%	Show chemicals... With data(34)...	

alert 2

Category Definition

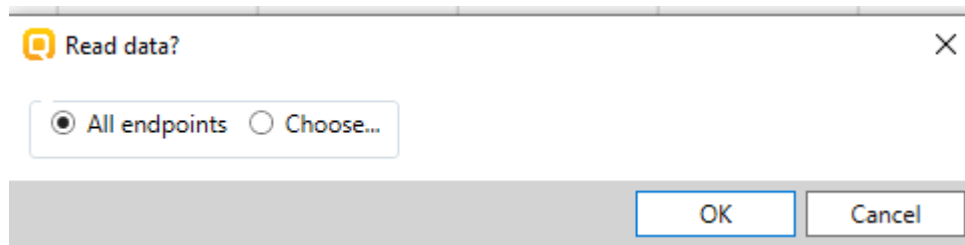
Calculation of Alert performance accounting metabolism for one alert

Right-click on the results (Positive/Negative) (1) to display **Select for primary grouping**, select **Select for primary grouping** (2)



Category Definition Analogues

- Based on the defined category (**Michael Addition > Michael addition on quinoid type compounds > Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphtoquinone(s)/imines**) 152 analogues have been identified (including the target chemical CAS: 90-05-1).
- The Toolbox automatically requests the user to select the endpoint data that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).



Only Skin sensitization database is selected in this example and we click **OK**.

Category Definition Analogues

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Category definition' tab is active, showing a tree view of categories. A 'Gather data' dialog box is open, displaying '347 points added across 152 chemicals.' and an 'OK' button. A callout box with the number '1' points to the 'OK' button. The background shows the 'Category definition' tab with a tree view of categories like Neurotoxicity, Sensitisation, and Toxicity to Reproduction.

1. The Toolbox automatically informs the user for the number of collected data points across the chemicals in the category. Click **OK** to confirm.

Category Definition

Summary information for Analogues

73 chemicals with 206 experimental results related to the defined target endpoint are found.

The screenshot shows the QSAR TOOLBOX interface with the 'Category definition' workflow selected. The 'Filter endpoint tree...' panel on the left lists various endpoints, with 'EC3' selected and highlighted in yellow. A red circle highlights the '73/206' count next to 'EC3'. The main table displays chemical structures and their associated experimental data for various endpoints. A blue callout box with a green exclamation mark icon points to the '73/206' count, stating: 'Chemical statistics presenting the number of chemicals and the available experimental data.'

Recap

- You have identified two protein binding alerts for the target chemical (2-methoxyphenol (guaiacol), CAS 90-05-1).
- You have calculated and compared the alert performance for each of the alerts.
- You have now retrieved in the available experimental results on skin sensitisation (EC3) values for 73 chemicals with the same mechanism of protein binding as the target compound, which were found in the “Skin Sensitisation” database.
- The user can now proceed to the next module; click on *“Data Gap Filling”*.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Data
 - Category definition
 - **Data Gap Filling**

Data Gap Filling Overview

- “Data Gap Filling” module give access to five different data gap filling tools:
 - Read-across
 - Trend analysis
 - (Q)SAR models
 - Standardized workflow
 - Automated workflow
- 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.
 - Standardized and Automated workflows are developed to facilitate the users work. Once started, they follow the implemented logic and finish with prediction. The general differences between the two type of workflows are represented on the next slide.

In this example we will use the manual read-across approach.

Data Gap Filling

Apply Read across

The screenshot displays the QSAR Toolbox software interface. The top menu bar contains 'Data Gap Filling' (highlighted with a red box and callout 2). The left sidebar shows 'Read across' (highlighted with a red box and callout 3). The central table lists chemical data with columns for 'Structure', 'Endpoint', and 'Data'. The 'EC3' endpoint is highlighted in yellow (callout 1). The 'Possible data inconsistency' dialog box is open on the right, showing metadata and options for scale/unit conversion (callout 4). The 'OK' button in the dialog is highlighted (callout 5).

1. Click on the row with the target endpoint and the cell corresponding to the target chemical (the yellow highlighted row);

2. Go to **Data gap Filling** module;

3. Click on **Read-across** button;

4. Select **Skin sensitisation II (ECETOC)**;

5. Click **OK**.

Data Gap Filling

Apply Read across

Filter endpoint tree...

Structure

Structure info

Parameters

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Human Health Hazards

Acute Toxicity

ADME

Bioaccumulation

Carcinogenicity

Developmental Toxicity / Teratogenicity

Genetic Toxicity

Immunotoxicity

Irritation / Corrosion

Neurotoxicity

Photoinduced toxicity

Repeated Dose Toxicity

Sensitisation

Skin

in Vivo

GPMT

HRIPT

LLNA

EC3

Miscellaneous

ToxCast

1 [target] 6 9 11 20 21 26 27 28 29

Information

7 observed values for 5 chemicals were excluded due to missing X descriptor value(s)

OK

19/19 M: Strong sensi...

6/12

M: 4.02E+03 µg/...

M: Strong sensi... M: Strong sensi...

68/199 M: 3.5 % M: 0.355 % M: 16.4 % M: 0.3 % M: Strongly posit... M: 5.79 % M: 0.32 % M: 0.6 % M: 7.49 %

11/20 M: Moderately s...

1. The Toolbox informs the user that 7 observed values for 5 chemicals were excluded due to missing X descriptor values. The reason for exclusion is that log Kow for these chemicals cannot be calculated (most probably these are mixtures, UVCB substances or some metal containing chemicals for which EPISUITE program cannot return result). The log Kow is the default X-descriptor for the read-across approach.
2. Click **OK**.

Data Gap Filling

Apply Read across

The screenshot displays the OECD QSAR Toolbox interface. On the left, the 'Subcategorization' window is open, showing the 'Predefined' section with 'US-EPA New Chemical Categories' selected (indicated by callout 2). In the 'Simulated' section, 'Hydrolysis simulator (neutral)' is selected (indicated by callout 3). The 'Analogues' list shows 'Phenols (Acute toxic)' as the selected target. The main window shows the 'Data Gap Filling' results table, which lists various chemical categories and their associated data points. A 'Read-across prediction for EC3, based on 14 values' plot is shown below the table, with a 'Predicted: Positive' result. On the right, the 'Select / filter data' panel is visible, with 'Subcategorize' selected (indicated by callout 1). The bottom of the interface shows a list of 'In nodes below' with counts for 'Automated workflows' and 'Standardized workflows'.

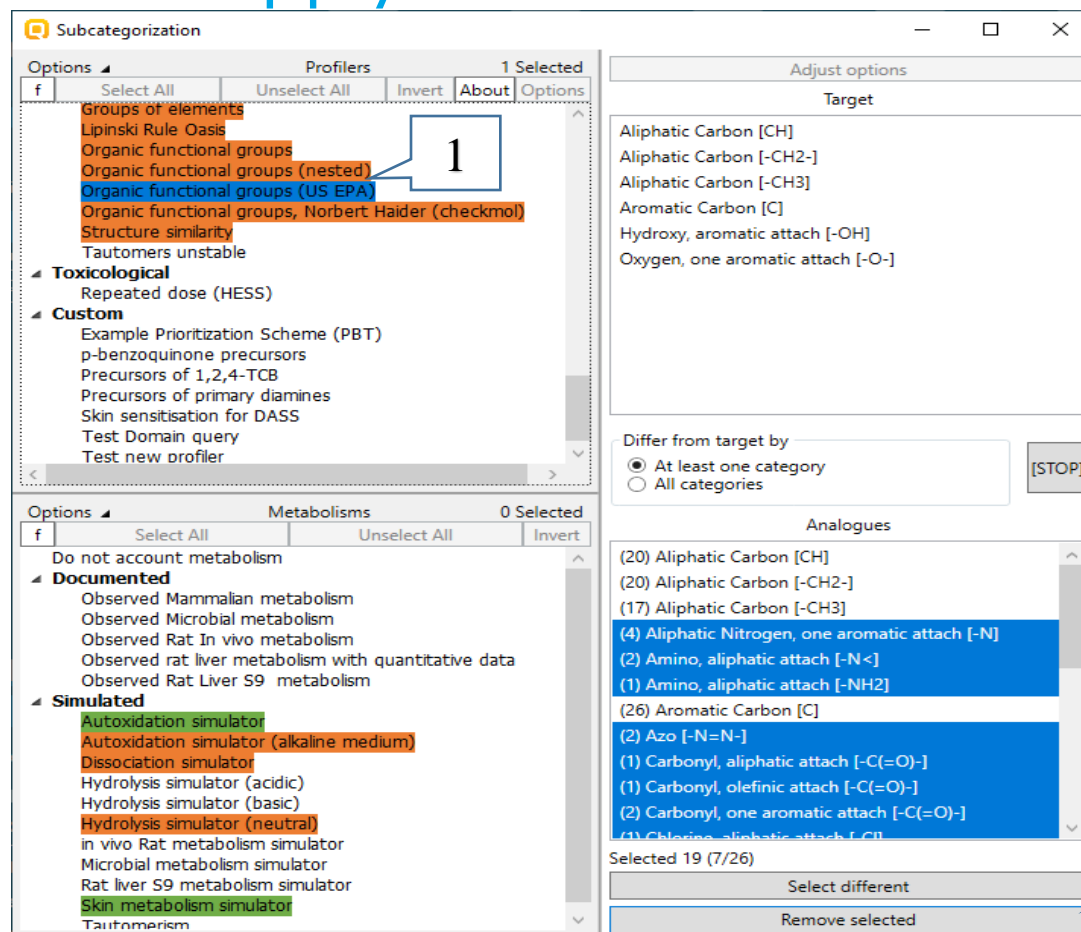
1. Go to **Select/filter data** and click **Subcategorize**;

2. Select the **US-EPA New Chemical Categories** profiler;

3. Click **Remove selected**

Data Gap Filling

Apply Read across



1. Select the **Organic functional groups (US EPA)** profiler;
2. Click **Remove selected**

Data Gap Filling

Accepting the predicted result

The screenshot displays the QSAR Toolbox interface. At the top, a 'Filter endpoint tree...' sidebar shows various endpoints, with 'EC3' highlighted. The main table lists chemical structures and their predicted EC3 values. A 'Confirm' dialog box is open, asking 'Are you sure you want to accept this prediction?'. Below the table, a 'Read-across prediction for EC3, based on 22 values' plot shows log Kow values on the x-axis and EC3 status (Positive/Negative) on the y-axis. A 'Select / filter data' panel on the right contains buttons for 'Accept prediction' and other actions. Numbered callouts 1 and 2 highlight the 'Accept prediction' button and the 'Confirm' dialog respectively.

1. Click **Accept prediction**;
2. Confirm with "Yes"

Data Gap Filling

Accepting the predicted result

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (Q)SAR Standardized Automated

Documents

Document 1

- # [C: 1;Md: 0;P: 1] CAS: 90051
- [C: 150;Md: 345;P: 1] Grouping with metabolism: Pre
- [C: 69;Md: 250;P: 1] Enter GF(RA)
- [C: 27;Md: 132;P: 1] Subcategorized: US-EPA Nev
- [C: 8;Md: 44;P: 1] Subcategorized: Organic fu

Data Gap Filling Settings

☒ Only endpoint relevant

At this position:

Select a cell with a rigid (bold) path

Automated workflows 0

Standardized workflows 0

Filter endpoint tree...

Structure

Structure info

Parameters

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Human Health Hazards

- Acute Toxicity
- ADME
- Bioaccumulation
- Carcinogenicity
- Developmental Toxicity / Teratogenicity
- Genetic Toxicity
- Immunotoxicity
- Irritation / Corrosion
- Neurotoxicity
- Photoinduced toxicity
- Repeated Dose Toxicity
- Sensitisation
- Skin
- In Vivo
- GPMT 57/57 M: Strong sensitizer
- HRIPT 6/12 M: Moderate sen...
- LLNA
- EC3 74/207 R: Positive
- Miscellaneous 52/69
- Undefined Assay 1/1
- ToxCast
- Toxicity to Reproduction
- Toxicokinetics, Metabolism and Distribution

1 [target]

2

3

4

5

6

7

8

57/57 M: Strong sensitizer

6/12 M: Moderate sen...

74/207 R: Positive

52/69

1/1

M: Category B

M: Category A

M: Category A

M: 3.5 %

M: Moderately s...

M: Category B

The OECD QSAR Toolbox for Grouping Chemicals into Categories

Developed by LMC, Bulgaria

Recap

- Read-across is the appropriate data-gap filling method for “qualitative” endpoints like skin sensitisation. Since the most of the analogues and all five neighbouring tested chemicals in the category were positive, it was easy to accepting the prediction of positive for the target chemical.
- You are now ready to complete the final module and to create the report.
- Click on “Report” to proceed to the last module.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
 - Chemical Input
 - Profiling
 - Endpoint
 - Category definition
 - Data Gap Filling
 - **Report**

Report Overview

- The report module can generate reports of predictions performed with the Toolbox.
- The report module contains a predefined report template which users can customize.

Report Generation report

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Reports' and 'Export'. The 'Reports' menu is open, showing options like 'Prediction', 'Data Matrix', 'Category', and 'QMRF'. The 'Prediction' option is highlighted with a red box and a callout '3'. The 'Customize report content and appearance' dialog box is open, showing the 'Wizard pages' on the left and the 'Customization' options on the right. The 'Customization' options include 'Add RAAF scenario', 'Prediction', 'Category', and 'Data matrix'. The 'Prediction' section is checked, and the 'Create report' button is highlighted with a red box and a callout '4'. The background shows a 'Filter endpoint tree...' window with a list of endpoints and a table of results.

1. Go to the **Report** module;

2. Click on the cell with the read-across prediction;

3. Click **Prediction**;

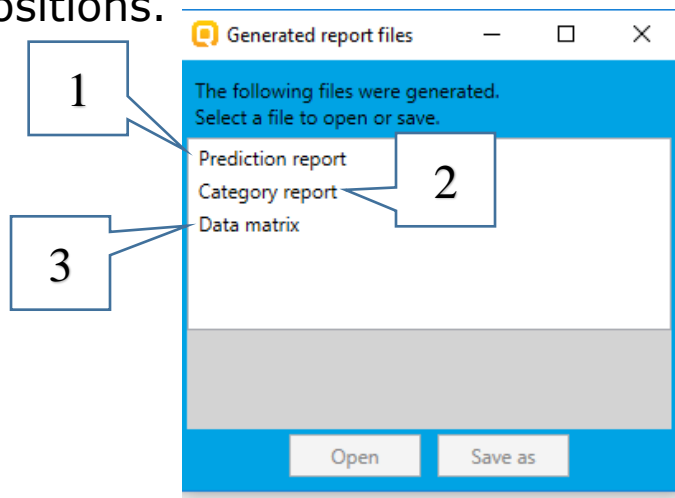
4. Click **Create report**.

Report

Generation report

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

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



Report

Generated report files

Prediction report

Prediction of EC3 for guaiacol

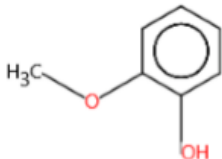
1 / 8

QSAR Toolbox prediction for single chemical

Date: 15 Apr 2020

Author(s):

Contact details:

Target information		
Structural information	Numerical identifiers	Chemical names
SMILES: COc1ccccc1O	CAS#: 90-05-1 Other: EC Number:2019647	"2-methoxyphenol (guaiacol);guaiacol;phe nol, 2-methoxy;2-me thoxy-phenol;2-metho xyphenol;o-methoxyph enol;o-methoxy pheno l;phenol, o-methoxy- ;guaiacol [jan]" 2-methoxy-phenol 2-Methoxyphenol
Structure 		

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

Category report

QSAR Toolbox report for category

1. Category definition

1.1. Category definition

Category name

manually editable field

Not provided by the user

Covered (target) endpoint(s)

- Human Health Hazards/Sensitisation: EC3, LLNA, in Vivo, Skin

Category hypothesis

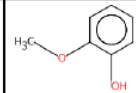
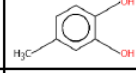
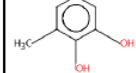
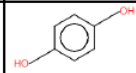
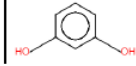
manually editable field

Not provided by the user

1.2. Category members

Information of category members

Table of category members

#	CAS	Name	SMILES	Structure
1	90-05-1	guaiacol	COc1ccccc1O	
2	452-86-8	4-Methylcatechol	Cc1ccc(O)c(O)c1	
3	488-17-5	3-methylcatechol	Cc1cccc(O)c1O	
4	123-31-9	p-Quinol	Oc1ccc(O)cc1	
5	108-46-3	resorcin	Oc1cccc(O)c1	

Congratulations!

- You have now been introduced to the definition of target endpoint;
- You have now been familiarized with the meaning of the different colouring of the profilers and databases.
- You have now been introduced to the consecutive steps of the calculation of alert performance.
- Note, proficiency comes with practice!