

## OECD QSAR Toolbox v.4.4.1

Evaluating alert performance accounting for the  
metabolic activation of chemicals

# Outlook

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

# Background

- This is a step-by-step presentation designed to take the Toolbox user through the workflow of a data gap filling exercise taking into account the performance of the alerts identified in the target structure or in its metabolites.

# Outlook

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

# Keywords

**TARGET CHEMICAL** - chemical of interest

**MODULE** – a Toolbox module is a section dedicated to specific actions and options

**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).

**ALERT PERFORMANCE** – Performance of an alert to predict the target endpoint based on the chemicals in the selected databases having the same alert and experimental data available for them. The alert performance (AP) is also of help for selection the most appropriate alert(s) for defining categories.

**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)

**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 a target endpoint;
- Illustration of relevancy of the profilers and data availability;
- Define the primary group (i.e. searching for analogues) by accounting for a metabolism;
- Calculation of an alert performance (AP) accounting for the metabolic activation of the chemicals;

# Outlook

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



## Specific Aims

- To familiarize the user with the Alert performance (AP) functionality;
- To introduce to the user the calculation of AP accounting for the metabolic activation of the chemicals;
- 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

- Alerts are the main category-building units of many profiling schemes (profilers) and their definition is based on a theoretical knowledge and empirical observations.
- The alert performance is estimated based on the distribution of the chemicals having (a) specific alert(s) across the available experimental data for a defined endpoint.
- AP can be applied for endpoints for which the experimental data exists as potency categories (e.g. Positive, Equivocal, Negative; Strong, Weak, Non sensitizer, etc.).
- The outcome of the estimation provides percent of the Positive and Negative performance and the number of chemicals used to calculate the performance.

# Outlook

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

## The Exercise

- In this exercise the skin sensitization potential of *1,3-Propanediamine, N-(3-aminopropyl)* [CAS# 56-18-8] (i.e. the target chemical) will be predicted.
- The target endpoint (i.e. Skin sensitization) will be preliminary defined.
- Category will be defined accounting for the skin metabolism of the chemicals.
- Alert performance (AP) will be evaluated for the alerts found in the package a *parent & metabolites*.
- The prediction itself will be made by “read-across”.
- The alert performance item generated for the report will be shown.

# Outlook

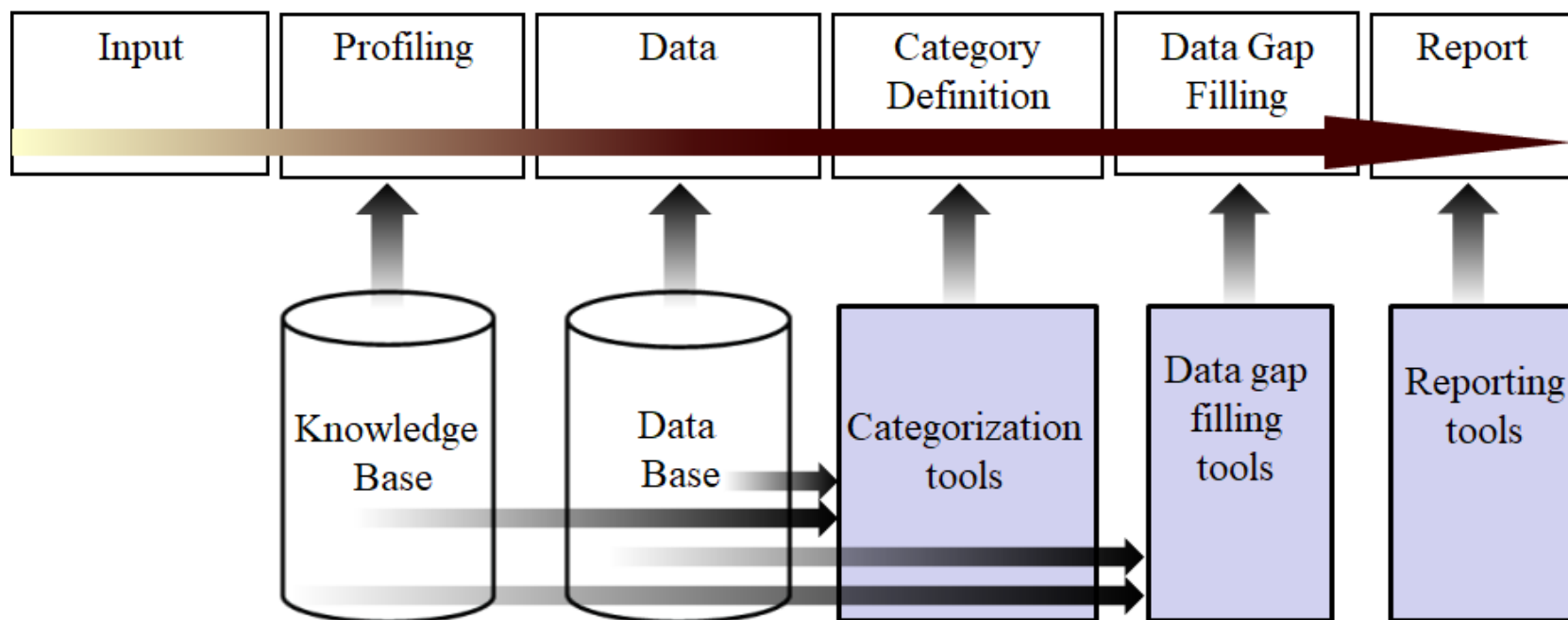
- Background
- Keywords
- Objectives
- Specific Aims
- Alert performance
- The exercise
- **Workflow**

# Workflow

- **The Toolbox has six modules which are used in a sequential workflow:**
  - Input
  - Profiling
  - Data
  - Category Definition
  - Data Gap Filling
  - Report
- **We will go through all of them with the exercise**

# Workflow

**Scheme illustrating the Toolbox workflow**





# 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 the chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.
- Other key functionalities such as “Define a target endpoint” are also placed in the Input module.
- Possibility to enter the chemical via additional chemical ID (e.g. EC number) is also available since version 4.3 of the Toolbox.

# Input

## Entering a Chemical by CAS #

The screenshot shows the QSAR Toolbox 4.4.1 interface. The 'Input' module is selected in the top toolbar. The 'CAS#' icon is highlighted with a red box and labeled '1'. The 'Search by CAS #' dialog box is open, showing the search field with '56188' and the 'Search' button labeled '3'. The 'OK' button is labeled '4'. The dialog box also displays the chemical name '1,3-Propanediamine, N-(3-aminopropyl)', its SMILES string 'NCCCNCCCN', and its chemical structure.

1	CAS	56-18-8
	SMILES	NCCCNCCCN
	CS Relation	High
	Substance	Mono constituent
	Composition	
	Name	1,3-Propanediamine, N-(3-aminopropyl)
	Sources	NICNAS Canada DSI

Chemical structure: NCCCNCCCN

Click on **CAS#** icon (1) in the *Input* module. "Search by CAS#" dialogue appears. Type CAS number **56-18-8** in the field (2) and click the **Search** button (3). Confirm by **OK** (4).

# Input

## Define the target endpoint

QSAR Toolbox 4.4.1 [Document 1]

The screenshot shows the QSAR Toolbox 4.4.1 interface. The top toolbar has a red box around the 'Input' icon (a hexagon with a plus sign). Below the toolbar, the 'Document' menu is open, showing options like 'New', 'Open', 'Close', 'Save', 'CAS#', 'Name', 'Structure', 'Composition', 'Select', 'ChemIDs', 'Database', 'Inventory', 'List', 'Substructure (SMARTS)', 'Query', and 'Define'. The 'Target Endpoint' menu is also visible. The main window displays a 'Filter endpoint tree...' dialog with a list of endpoint categories: Structure, Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards. The 'Structure' category is selected, and a chemical structure is shown in the preview area.

Defining of the target endpoint is needed for calculation of AP later on. This is a two-step way process:

- First, the main endpoint position has to be specified, e.g. *Human Health Hazard / Sensitization*
- Second, the specific meta data fields such as "type of method", "assay", etc. related to the main endpoint tree position has to be defined (see next slides).

# Input

## Define the target endpoint

QSAR Toolbox 4.4.1 [Document 1]

Document: Document 1  
# [C: 1;Md: 0;P: 0] CAS: 56188

Filter endpoint tree... 1 [target]

Structure

- Structure info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards

Target Endpoint

Define

1

2

3

Step 1: Define the general target endpoint

Select endpoint

Filter:

- 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**
  - ToxCast
  - Toxicity to Reproduction
  - Toxicokinetics, Metabolism and Distribution

Undefine

Next

Click on the **Define** icon (1). The "Select endpoint" dialogue appears. First select the general target endpoint - "**Sensitisation**" (2) located under the Human Health Hazard and then click on **Next** (3)

# Input

## Define the target endpoint

Step 2: Define the specific target endpoint by addition of metadata

(1) From the drop-down menus define the following specific information:  
Endpoint - **EC3**; Organ - **Skin**; Type of method - **in Vivo**; Assay - **LLNA**

(2) Click on **Finish**.

# 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' tab is active, showing a 'Filter endpoint tree...' dialog. The tree lists various endpoints, with 'Sensitisation' expanded to show 'Skin', 'in Vivo', and 'LLNA'. The 'EC3' endpoint is highlighted in yellow, and a red dashed box encloses the 'Sensitisation' section. The chemical structure of the target endpoint is shown as NCCNCCN (1,3-bis(aminomethyl)propane).

Once the target endpoint is defined, the related row of the endpoint tree is highlighted.

# Profiling Overview

- As you know, the “Profiling” refers to the electronic process of retrieving relevant information on a compound which is stored in the Toolbox, other than fate and (eco)toxicity data.
- The available information includes likely mechanism(s) of action, as well as observed and/or simulated metabolites.
- Based on the already defined target endpoint (slides #20-22) the profilers and metabolism simulators are automatically grouped by their relevancy to the endpoint\*.

\*More details regarding the grouping of profilers by relevancy could be seen in tutorial “*Example for predicting skin sensitisation taking into account alert performance*”.

# Profiling

## Profiling the target

The screenshot shows the QSAR Toolbox Profiling module. The interface includes a top toolbar with icons for Input, Profiling (highlighted with a red dashed box), Data, Category definition, Data Gap Filling, and Report. Below the toolbar is a sub-menu with 'Profiling' and 'Custom profile'. The 'Profiling' sub-menu has 'Apply' (highlighted with a red box and callout 4), 'View', 'New', and 'Delete'. The main window displays a 'Documents' list on the left, a 'Filter endpoint tree...' on the right, and a 'Profiling' dialog box on the right. The 'Profiling' dialog box contains a message: 'Selected profiles will be applied on all metabolites! Do you want to continue?' with 'Yes' and 'No' buttons (callout 5). The 'Profiling' dialog box also has a checkbox for 'Do not show this dialog'.

Numbered callouts indicate the steps:

- Go to **Profiling** module and unselect all previously ticked profilers (click on **Unselect All** button)
- Select the profilers defined as suitable to the target endpoint by clicking on the box in front of "**Suitable**" level.
- Similarly to (2) select the "Suitable" simulators.
- Click on **Apply** button.
- Information message appears to notify that profilers will be applied on all generated metabolites. Click on **Yes**.



# Profiling

## Profiling the target

The screenshot displays the QSAR Toolbox interface during a profiling task. On the left, the 'Documents' panel lists a chemical structure with CAS: 56188. Below it, the 'Profiling methods' panel shows three selected methods: 'Suitable', 'Plausible', and 'Unclassified'. The 'Metabolism/Transformations' panel shows two selected methods: 'Autoxidation simulator' and 'Skin metabolism simulator'. The main 'Filter endpoint tree...' panel shows a tree structure with 'EC3' selected. The 'Profiling' panel shows results for 'Protein binding by OASIS' and 'Metabolism/Transformation'.

**As a result of the applying of the profiles and metabolic simulators suitable to the target endpoint:**

- **No alerts are found** in the target structure;
- The target chemical does not autoxidized.
- Four metabolites are simulated by the Skin metabolism simulator
- **Protein binding alerts are identified** in some of the generated skin metabolites.
- More details for the mechanism of interaction and some additional information is also provided (*see next page*)

Profiling result of the target structure

Profiling result of the generated metabolites

# Profiling

## Profiling the target

4

Q SAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Profiling Custom profile

Apply View New Delete

Documents

Document 1  
# [C: 1;Md: 0;P: 0] CAS: 56188

Profiling methods

Options 3 Selected  
f Select All Unselect All Invert

☒ Suitable

- ☒ Protein binding alerts for skin sensitization
- ☒ Protein binding alerts for skin sensitization
- ☒ Protein binding by OASIS

☒ Plausible

- ☐ Aquatic toxicity classification by ECOS
- ☐ Chemical elements
- ☐ Groups of elements
- ☐ Keratinocyte gene expression
- ☐ I ninski Rule Oas

Metabolism/Transformations

Options 2 Selected  
f 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)

Filter endpoint tree... 1 [target]

Structure

Profiling results

- 4 metabolite(s)
  - 1 x No alert found  
NCCCCC(O)=O
  - 3 x Schiff base formation
    - 3 x Schiff base formation with carbonyl compounds
      - 2 x Aldehydes  
NCCCCC=O  
OC(=O)CCNCCC=O
      - 1 x Bis aldehydes  
O=CCNCCC=O

2

Details

3

Close

Explanation for Protein binding alerts for skin sensitization by OASIS -> Schiff base formation -> Schiff base formation with carbonyl...

Categories

Filter:

- Schiff base formation
  - Benzoyl Schiff base formation
  - Benzoyl phosphine oxides
  - Direct acting Schiff base formers
  - 1,2-Dicarbonyls and 1,3-Dicarbonyls
  - Di-substituted alpha,beta-unsaturated aldehydes
  - Pyrazolones and Pyrazolidinones derivatives
  - Pyrazolones and Pyrazolidinones
  - Schiff base formation with carbonyl compound: Activated Carbonyl compounds
  - Aldehydes
  - alpha-Ketoesters
  - Aromatic carbonyl compounds
  - Bis aldehydes
- SN1
  - Carbenium ion formation (enzymatic)
  - Carbenium ion
- Nucleophilic substitution (SN1) on alkyl (aryl) 1
  - Mercury compounds
- SN2
  - Activated allyl and nonaromatic type fragment

Definition Properties Training Set Literature MetaInfo Table Custom Captions Scheme

**Mechanistic Domain:** Schiff base formation

**Mechanistic Alert:** Schiff base formation with carbonyl compounds

**Structural Alert:** Aldehydes

All aliphatic aldehydes can potentially undergo **Schiff base formation** with a primary amine, which is a reversible reaction (optimal at pH 3-4) and proceeds in two stages via a tetrahedral intermediate.

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

R = any carbon atom

According to Roberts *et al.*, 2015 simple mono-aldehydes are not highly reactive. However, the nature of the substituents will affect the reactivity of the carbonyl groups. For example, in TIMES SS model chemicals having the following structural boundaries will be predicted as strong skin sensitizers:

$$R-C(=O)-H$$

R = H, -Csp3Ar, -Csp2C=C, -C(=O)OH, -OC

1

2

3

4

Double click on the cell with profiling a result (1) (or right click and select "Explain");  
A new dialogue appears where the SMILES of the generated metabolites are provided along with the respective profiling result. Select a metabolite (2) and click on **Details** (3). The explanation of the mechanism associated with the identified alert will be shown. Close the explanation window (4).

# Data Overview

- “Data” module refers to the electronic process of retrieving environmental fate, ecotoxicity and toxicity data stored in the Toolbox databases.
- The 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 single or limited number of endpoints).
- Once the target endpoint is defined, the system highlights the databases where the data for the defined endpoint could be found

# Data

## Collecting the experimental data

The screenshot displays the QSAR Toolbox software interface. The top toolbar includes buttons for Input, Profiling, Data (highlighted with a red dashed box and labeled '1'), Category definition, Data Gap Filling, and Report. Below the toolbar, the 'Data' module is active, showing a 'Gather' button (labeled '3') and a 'Documents' panel (labeled '2') containing a list of databases. The 'Databases' panel shows 'REACH Skin sensitisation database (no)' and 'Skin Sensitization' selected. A 'Filter endpoint tree...' panel on the right shows a chemical structure and a list of endpoints. A 'Read data?' dialog box (labeled '4') is open, asking to 'Read data?' with options 'All endpoints' (selected) and 'Choose...'. The dialog has 'OK' and 'Cancel' buttons.

- (1) Go to the **Data** module
- (2) Select all highlighted databases except ECHA REACH.
- (3) Click on the **Gather** button
- (4) A message appears asking to Read data/All endpoints. Click on **OK**

# Data

## Collecting the experimental data

The screenshot displays the QSAR Toolbox interface. The 'Data' menu item in the top toolbar is highlighted with a red dashed box. Below the toolbar, the 'Documents' panel shows 'Document 1' with chemical formula # [C: 1;Md: 5;P: 0] CAS: 56188. The 'Databases' panel shows a list of databases with 'REACH Skin sensitisation database' and 'Skin Sensitization' selected. The 'Inventories' panel shows a list of inventories. The main window displays the chemical structure of 1,3-bis(2-aminoethyl)urea, NCCNCNCCNCN, and a list of endpoints. A dialog box in the foreground states '5 points added across 1 chemicals.' and has an 'OK' button.

- The experimental data for the target chemical appear on the Data matrix
- Additionally, the system shows how many data points have been collected. In the current example five experimental data points for the target chemical have been collected from the selected databases.

# Category Definition

## Overview

- The next step of the TB workflow is to collect analogues of the target chemical. This is the critical step in the workflow.
- The analogues search happens in the Category definition module.
- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- 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 a read-across.
- The knowledge implemented in the system in the form of profilers appear here as grouping methods.
- For example, if a specific protein binding mechanism is identified in the target chemical, it could be used for searching analogues acting by the same mechanism of interaction with biomacromolecules.
- The profilers, relevant to the defined target endpoint, are highlighted and could be used to define the category.
- The group could be defined accounting for the metabolic activation of the chemicals in the cases where an alert is identified in some of the target` metabolites.

# Category Definition

## Define the category accounting for a metabolism

The screenshot displays the QSAR Toolbox software interface. The top toolbar shows the 'Category definition' step selected. The 'Filter endpoint tree...' window is open, showing a hierarchical tree of endpoints. The 'Suitable' and 'Plausible' methods are highlighted in green and orange respectively. The chemical structure of the target molecule is shown in the top right.

**Documents:** Document 1  
# [C: 1;Md: 5;P: 0] CAS: 56188

**Grouping methods:** 0 Selected  
f Select All Unselect All Invert

**Suitable (Green):**  
Protein binding alerts for skin sensitization  
Protein binding alerts for skin sensitization  
Protein binding by OASIS

**Plausible (Orange):**  
Aquatic toxicity classification by ECOSAR  
Chemical elements  
Groups of elements  
Keratinocyte gene expression  
Lipinski Rule Oass  
OECD HPV Chemical Categories  
Organic functional groups  
Organic functional groups (nested)  
Organic functional groups (US EPA)  
Organic functional groups, Norbert Haider  
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  
US-EPA New Chemical Categories

**Unclassified:**  
(AOT)Protein binding by OASIS v1  
Acute aquatic toxicity classification by Me

**Filter endpoint tree...** 1 [target]

Structure: NCCCCNCCN

**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  
LLNA  
EC3  
Miscellaneous  
ToxCast  
Toxicity to Reproduction  
Toxicokinetics, Metabolism and Distribution  
Profiling

- The grouping methods that are relevant to the defined target endpoint are highlighted (the "green" are suitable and the orange are "plausible").
- However, for the current example we saw that it has "No protein binding alert" as a parent but is getting activated as a result of a skin metabolism (see slides 24-26).
- In this respect, the primary group will be defined with accounting for the metabolism activation of the target.



# Category Definition

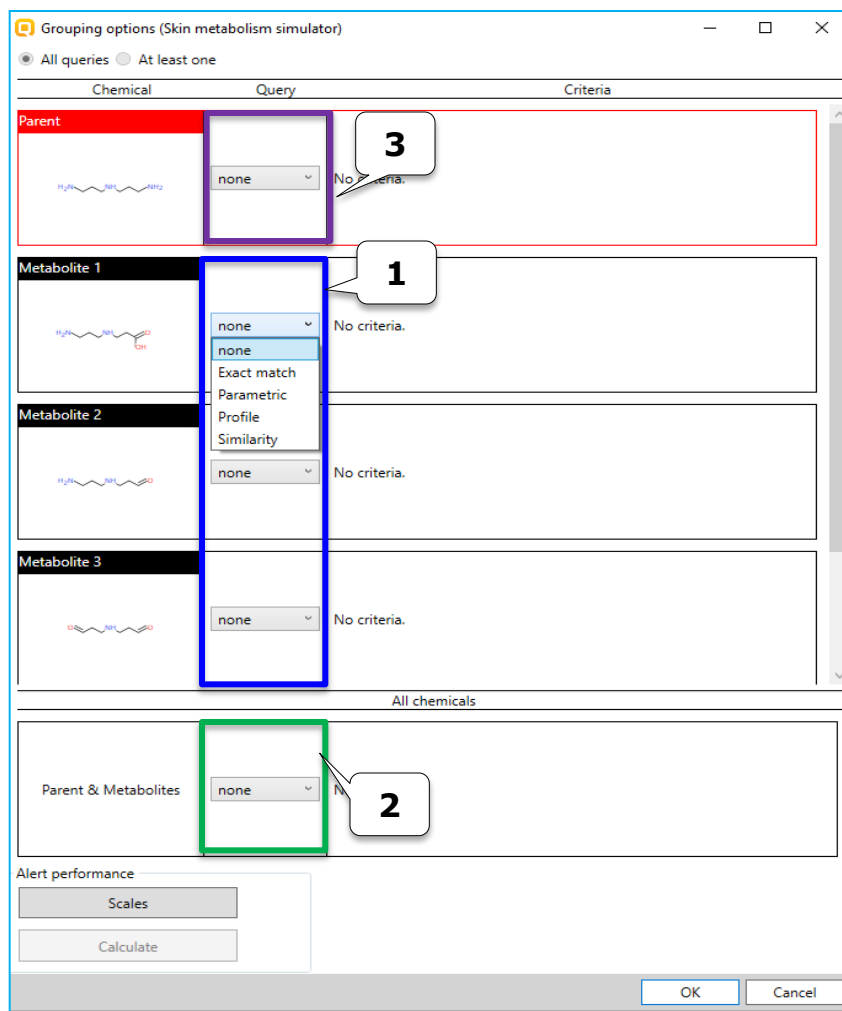
## Define the category accounting for a metabolism

The screenshot displays the QSAR TOOLBOX software interface. The top toolbar contains icons for 'Input', 'Profiling', 'Data', 'Category definition' (labeled 1), 'Data Gap Filling', and 'Report'. Below the toolbar, the 'Category consistency' section features buttons for 'Define', 'Define with metabolism' (labeled 2), 'Subcategorize', 'Combine', 'Clustering', and 'Category elements'. The main workspace is divided into three panes. The left pane, 'Documents', shows 'Document 1' with chemical formula '[C: 1; Md: 5; P: 0] CAS: 56188'. The center pane, 'Filter endpoint tree...', shows a tree structure with 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', and 'Human Health Hazards'. The right pane, 'Structure', shows the chemical structure of a diamine. A 'Select metabolism' dialog box is open on the right, showing a list of metabolic simulators. The 'Simulated' section is expanded, and 'Skin metabolism simulator' is highlighted with a red dashed box and labeled 3. The 'OK' button is labeled 4.

- (1) Move to the **Category definition** module;
- (2) Click on the **Define with metabolism** button;  
A new dialogue appears with all available documented and simulated metabolic simulators highlighted in the respective color (suitable – green, plausible–orange).
- (3) Select the **Skin metabolism simulator**;
- (4) Click **OK** to confirm the selection.

## Category Definition

## Define the group with accounting for a metabolism



- The newly appeared window shows the parent and all generated metabolites produced by the selected metabolic simulator (*Skin metabolism simulator* in the current example).
- The user is able to set a searching criteria for each of the metabolites (1) or for the whole package "*Parent & Metabolites*" (2).
- Query for the parent (3) could be also defined as an addition. However, it is not possible to define searching criteria for the parent, only.
- The following queries could be set:
  - **None** – default options; no criteria is set;
  - **Exact** – provides opportunity to search for chemicals having exactly the same metabolite;
  - **Parametric** – the searched chemical (parent or metabolite) to have specific value or a range of variation of a defined parameter (a list with all parameters currently available in the Toolbox is provided);
  - **Profile** – the searched chemical (parent or metabolite) to have a specific category by the selected profiler (a list with all profilers is provided);
  - **Structural** – the searched chemical (parent or metabolite) to be structurally similar to the current chemical above a defined threshold.
- Calculation of the AP will take into account all defined criteria.

# Category Definition

## Define the group with accounting for a metabolism

QSAR Toolbox 4.4.1 [Document 1]

Grouping options (Skin metabolism simulator)

All queries At least one

Chemical	Query	Criteria
Parent <chem>CCCCCCCC</chem>	none	No criteria.
Metabolite 1 <chem>CCCCCCCC(=O)O</chem>	none	No criteria.
Metabolite 2 <chem>CCCCCCCC(=O)O</chem>	none	No criteria.
Metabolite 3 <chem>CCCCCCCC(=O)O</chem>	none	No criteria.

Parent & Metabolites

Profile Profiler: Protein binding alerts for skin sensitization by OASIS Options: Edit

Alert performance

Scales Calculate

Target categories

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

Schiff base formation >> Schiff base formation with carbonyl compounds >> Bis aldehydes

Grouping with "Protein binding alerts for skin sensitization by OASIS"

You have selected <AND> from different hierarchy levels!  
Selecting most informative level(s) will have the same results!  
Do you want to continue?

Options

Down Up Do not show this dialog

All categories (N/A)

Acylation

Acylation >> (Thio)carbamoylation of protein nucleophiles

Combine profiles

☐ Invert result

☒ AND ☐ OR

☐ Strict

☐ Sort results

OK Cancel

1 2 3 4 5

(1) Select the **profile** query for the package "Parent & Metabolites";

(2) Select the **Protein binding alerts for skin sensitization by OASIS** profiler;

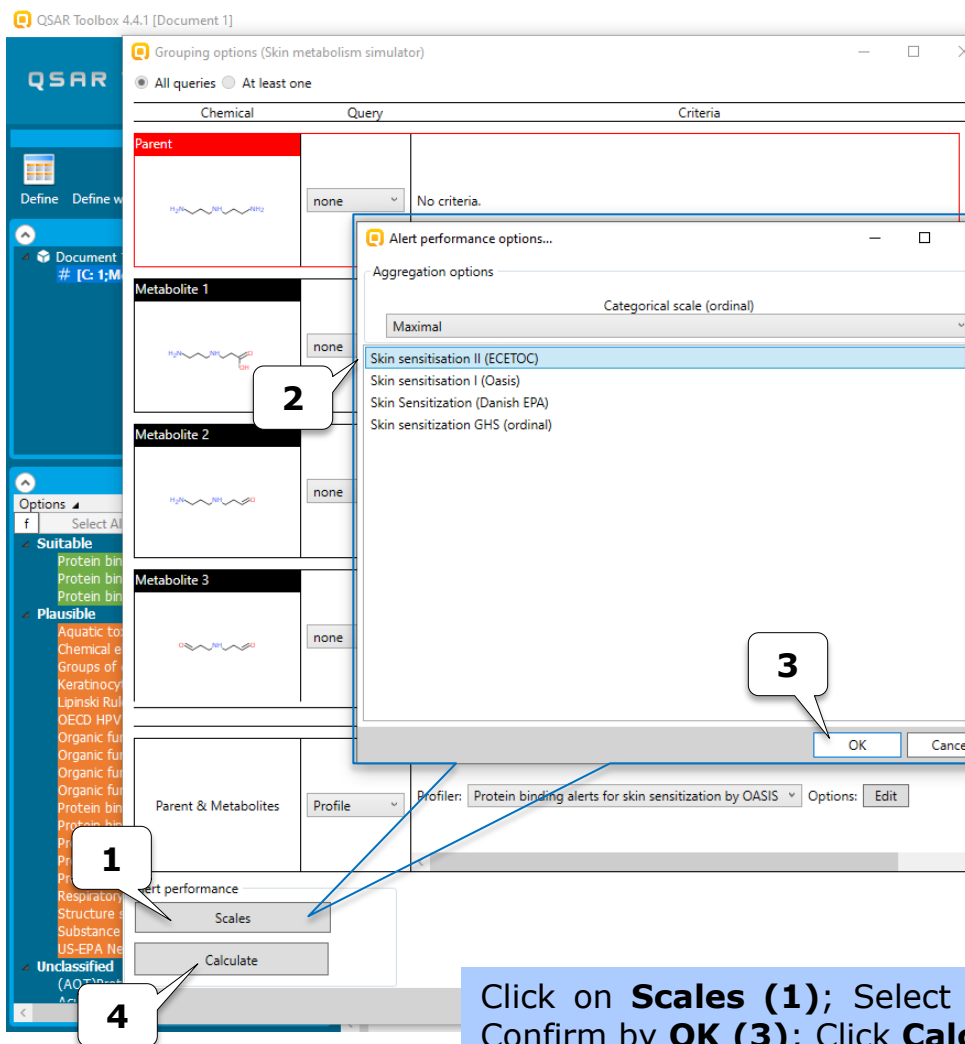
(3) Click **Edit** in order to see all identified alerts in the parent and its metabolites

(4) Click **OK**.

(5) Click OK on the appeared message

## Category Definition

### Calculation of the alert performance



The alert performance results depend on:

- **The defined target endpoint** – The AP is endpoint-dependent. SS, EC3 is defined in the current example.
- **Selected databases** – The AP results will be based on the chemicals presenting in the selected databases. Skin sensitization and REACH Skin sensitization database (normalized) are selected in our case.
- **Selected scale** – the available scales vary based on the defined target endpoint. For Skin sensitization the most appropriate scale is *Skin sensitization II (ECETOC)*. This is a dichotomous scale that converts the data into positive/negative. In this way the experimental data in different scales could be combined in order to provide the full AP statistic.
- **Aggregation options** – this options takes a role when a chemical from the selected databases has more than one experimental data that could be converted to the selected scale. The *Maximal* mode (the worst case scenario) is set by default (e.g. if a chemical has simultaneously positive and negative data, only the positive data will be taken when calculate AP).

Click on **Scales (1)**; Select the ***Skin sensitization II (ECETOC)*** scale **(2)**;  
Confirm by **OK (3)**; Click **Calculate (4)**

# Category Definition

## Calculation of the alert performance

**Alert performance results...**

Search Criteria	Positive	Negative	Chemicals
Using of "Skin metabolism simulator" Combined parent and products requirements: Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes<AND>Schiff base formation >> Schiff base formation with carbonyl compounds >> Bis aldehydes<AND>No alert found (Protein binding alerts for skin sensitization by OASIS)	80.00%	20.00%	Show chemicals... With data(12)...
Using of "Skin metabolism simulator" Combined parent and products requirements: Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes (Protein binding alerts for skin sensitization by OASIS)	51.09%	48.91%	Show chemicals... With data(187)...
Using of "Skin metabolism simulator" Combined parent and products requirements: Schiff base formation >> Schiff base formation with carbonyl compounds >> Bis aldehydes (Protein binding alerts for skin sensitization by OASIS)	82.35%	17.65%	Show chemicals... With data(14)...
Using of "Skin metabolism simulator" Combined parent and products requirements: No alert found (Protein binding alerts for skin sensitization by OASIS)	46.22%	53.78%	Show chemicals... With data(611)...

Close

Once the calculation of the AP is finished, a new window appears providing the following information:

- 1) The AP statistic accounting for all set criteria and all identified alerts in case of a selected *profile* query.

- 2) The AP statistic for each of the searching criteria (i.e. for each of the alerts)

- 3) The percentages of different data (positive/negative) and number of chemicals used for the statistic. The user is also able to see the corresponding chemicals by single click (the parent chemicals are shown, only).

By analyzing of the provided information the user can take a decision whether to use all identified alerts for searching for analogues or just one of them.

You can see the chemicals used for the AP calculation by single click

File

71 18516-18-2	16 931419-77-1	557 8063-07-8
------------------	-------------------	------------------

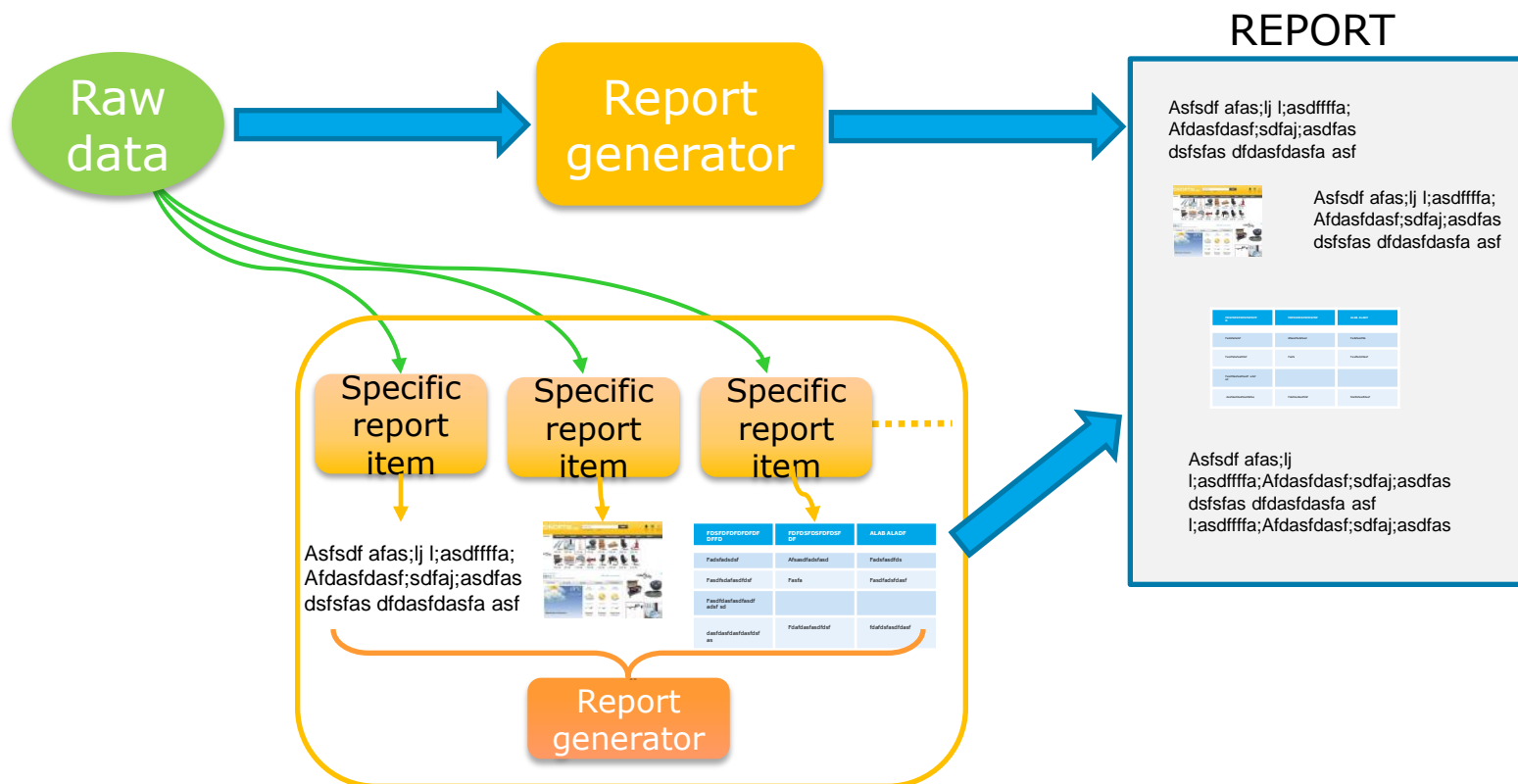
Save to smi Search OK



**Calculation of the alert performance creates a specific report item stored in the so-called *Report basket* (see the next slide).**

# Sidebar on the Report basket

- Specific report items are collected during the workflow (e.g. calculation of AP). All items are stored in the "Report basket" and can be used in the report to support or justify the consistency of a category.
- Items with external content - text and/or image (e.g. schemes of mechanisms of interactions, metabolic maps, snapshots from external modeling sources, etc.) could be also created within the Report basket and further included to the report.



# Category Definition

## Selection of an alert for category definition

Alert performance results...

Alert Description	Positive	Negative	Show chemicals...	Show all...
Using of "Skin metabolism simulator" Combined parent and products requirements: No alert found found<AND>Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes<AND>Schiff base formation >> Schiff base formation with carbonyl compounds >> Bis aldehydes (Protein binding alerts for skin sensitization by OASIS)	80.00%	20.00%	Show chemicals... With data(12)...	Show all(15)...
Using of "Skin metabolism simulator" Combined parent and products requirements: No alert found (Protein binding alerts for skin sensitization by OASIS)	46.22%	53.78%	Show chemicals... With data(611)...	Show all(1322)...
Using of "Skin metabolism simulator" Combined parent and products requirements: Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes (Protein binding alerts for skin sensitization by OASIS)	51.09%	48.91%	Show chemicals... With data(187)...	Show all(366)...
Using of "Skin metabolism simulator" Combined parent and products requirements: Schiff base formation >> Schiff base formation with carbonyl compounds >> Bis aldehyde (Protein binding alerts for skin sensitization by OASIS)	82.35%	17.65%	Show chemicals... With data(14)...	Show all(17)...

Select for primary grouping

Close

By analyzing the results, we can see that the **Bis aldehydes** alert shows the best predictability (82%) with respect to the defined endpoint and selected databases (i.e. 14 out of 17 chemicals having the same alert are positive by experimental data).

So, we can use only this alert for category definition. The category will consists of chemicals having Bis aldehyde alert as parents or as a result of skin metabolism.

(1) Apply right click on the row for Bis aldehydes alert and click **Select for primary grouping**.

# 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:
  - The 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 the 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.

Additionally, two workflows - Standardized and Automated have been developed in order to facilitate the users work. Once started, they follow the implemented logic and finish with the prediction.

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



# Data Gap Filling

## Apply Read across

1) Go to the **Data Gap Filling** module;

2) Click on the cell corresponding to the target chemical and defined endpoint;

3) Click the **Read across** button;

4) *Skin sensitisation II (ECETOC)* scale is selected by default. Confirm by **OK**.

# Data Gap Filling Subcategorization

The screenshot displays the QSAR Toolbox Subcategorization window. The interface includes a top navigation bar with buttons for Data, Category definition, Data Gap Filling, and Report. The main window is divided into several sections:

- Options:** A list of profilers with checkboxes. The 'Protein binding alerts for skin sensitization by OASIS' profiler is highlighted (3). The 'Autoxidation simulator' is also highlighted (4).
- Target:** A section for defining the target chemical.
- Chemicals:** A table listing chemicals with their molecular weights (M), categories, and various alerts. The table has columns for different chemical classes and their corresponding M values.
- Read-across prediction for EC3:** A scatter plot showing the relationship between log Kow and EC3 values. The plot includes a legend for 'Observed: Positive (x3); Predicted: Positive'.
- Buttons:** A vertical sidebar on the right contains buttons for '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', and 'Clear existing marks'. The 'Subcategorize' button is highlighted (1).
- Bottom Bar:** A bar at the bottom contains buttons for 'Remove selected' (5) and 'Remove selected'.

**Subcategorization 1:** Because our target chemical does not autoxidized (see slide 25), first we will remove the chemicals that have some alerts as a result of abiotic activation.

In order to subcategorize your category (i.e. to keep only the most similar analogues to the target chemical) Go to **Select / filter data** >> **Subcategorize** (1). The Subcategorization window with all available profilers will appear (2). Select the **Protein binding alerts for skin sensitization by OASIS** profiler (3) and then the **Autoxidation simulator** (4). The chemicals different to the target will be highlighted. Remove them by click on **Remove selected** button (5).

# Data Gap Filling Subcategorization

The screenshot shows the 'Subcategorization' window of the QSAR Toolbox. The interface includes a left sidebar with various profiler options, a central table of chemical data, and a bottom plot area.

**1** Points to the 'Structure similarity' profiler in the left sidebar.

**2** Points to the 'Adjust options' section, specifically the 'Target' dropdown set to '[90%,100%]'.

**3** Points to the 'Analogues' section, where the range '(3) [0%,10%]' is selected.

**4** Points to the 'Remove selected' button at the bottom of the 'Analogues' section.

**5** Points to the 'Accept prediction' button in the bottom right corner.

The central table displays chemical data for various categories. The 'EC3' column shows values like '9/21', '3/23', etc. The 'Similarity' column shows ranges like '[90%,100%]', '[30%,40%]', etc. The 'log Kow' column shows values like '-9', '-8', etc.

The bottom plot area shows a 'Read-across prediction for EC3, based on 15 values' with 'Observed: Positive (x3); Predicted: Positive'. The plot shows a scatter of points on a grid with 'log Kow' on the x-axis and 'EC3' on the y-axis.

**Subcategorization 2:** Because our category is already mechanistically similar (i.e. the category consists only of chemicals that are not abiotically activated, but have an alert as a result of skin metabolism), the second subcategorization will be based on structural similarity.

Select the **Structural similarity** profiler (1), click on the range with the lowest similarity (0% to 10%) (2), press the **Ctrl** button of the keyboard and click on the second similarity range (20% to 30%) (3). Remove the selected dissimilar chemicals by click on **Remove selected** button (4). Finally, click on **Accept prediction** button (5).

# Report Overview

- The report module can generate a report on predictions performed within the Toolbox.
- The report module contains a predefined report template which the users can customize.
- Three type of report files are generated:
  - *A Prediction report* – containing information for the target
  - *A Category report* – containing information for the final category (target plus used analogues)
  - *A Data matrix* – containing information for the analogues used for the prediction.
- Additionally a specific RAAF scenario could be selected. Selection of one of the scenarios will append automatically the related assessment elements (AE) related to the corresponding report sections.
- The *Report basket* (and Alert performance item, respectively) could be used for supporting information to the appropriate category elements or RAAF AE.

# Report

## Generating a prediction report

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Report' module is selected, indicated by a red dashed box and a callout '1'. The 'Export' menu is open, showing options like 'Prediction Data Matrix', 'Category', 'QMRF', 'SMI File', 'SDF File', 'CAS List', and 'Data Matrix'. The 'Prediction Data Matrix' option is highlighted with a red box and a callout '3'. The 'Documents' panel on the left shows a tree structure of chemical groups. The 'Filter endpoint tree...' panel in the center shows a list of endpoints, with 'EC3' selected and a callout '2' pointing to its prediction result 'R: Positive'. The 'Wizard pages' dialog box is open on the right, showing the 'Customize report' tab. It includes sections for 'Prediction', 'Category', and 'Data matrix', each with checkboxes for various report sections. The 'Add RAAF scenario' section is also visible. The 'Create report' button is at the bottom right of the dialog.

- 1) Go to the **Report** module; 2) Click on the cell with the prediction; 3) Click on the **Prediction** button.
- 3) The **Wizard pages** editor appears.

# Report

## Generating a prediction report

Customize report content and appearance

**Wizard pages**

**Customization**

Customize report

**Prediction**

Target and prediction summary

Prediction details (II)

Target profiles

Analogues selection details

**Category**

Category definition and members

**Consistency check**

Options

**Data matrix**

Options

2.1. Physicochemical similarity

2.2. Structural similarity

2.3. Mechanistic similarity

Mechanistic similarity

Mechanistic similarity profilers

Options 3 Selected

Select All Unselect All Invert

☒ Suitable

☐ Plausible

☐ Unclassified

Simulators

Options 2 Selected

Select All Unselect All Invert

☒ Suitable

☐ Plausible

☐ Unclassified

Add / Remove

Table with profiling results for "Protein binding alerts for skin sensi" Edit Preview

Table with profiling results for "Protein binding alerts for skin sensi" Edit Preview

Table with profiling results for "Protein binding by OASIS" Edit Preview

Table with profiling similarity accounting for metabolism ("Skin me" Edit Preview

Table with profiling similarity accounting for metabolism Edit Preview

Alert performance Preview

Comments on mechanistic similarity

2.4. Additional endpoints

Back Next Cancel Create report

Alert performance  
Scale=Skin sensitisation II (ECETOC); Endpoint=EC3; Metabolism=Skin metabolism simulator; Data aggregation=Maximal

#	Alert name	Alert performance, %		Number of chemicals	
		Positive	Negative	Positive	Negative
1	Using of "Skin metabolism simulator" Combined parent and products requirements: No alert found<AND>Aldehydes<AND>Bis aldehydes (Protein binding alerts for skin sensitization by OASIS)	80.00	20.00	12	3
2	Using of "Skin metabolism simulator" Combined parent and products requirements: No alert found (Protein binding alerts for skin sensitization by OASIS)	46.22	53.78	611	711
3	Using of "Skin metabolism simulator" Combined parent and products requirements: Aldehydes (Protein binding alerts for skin sensitization by OASIS)	51.09	48.91	187	179
4	Using of "Skin metabolism simulator" Combined parent and products requirements: Bis aldehydes (Protein binding alerts for skin sensitization by OASIS)	82.35	17.65	14	3

1) Go to the **Consistency check** section of the report; 2) Open section **2.3. Mechanistic similarity**; 3) The **Alert performance** item appears below the other automatically included items; 4) Click **Preview** button to see the information of this item will look like in the generated report; 5) Finally, click the **Create report** button to generate the report files. The AP item will be included in the Category report file.

**!** If the Alert performance is calculated more than once by setting different searching criteria, information for the latest calculation will be stored in the Report basket.

## Congratulation

- You now know how to define a target endpoint;
- You now know how to calculate the alert performance accounting for the metabolic activation of the chemicals;
- You now know where the Alert performance item appears in the report;
- Note, proficiency comes with practice!