

OECD QSAR Toolbox v.4.4.1

Tutorial illustrating new options for grouping with metabolism

Outlook

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

Background

- Grouping with metabolism is a procedure for finding analogues accounting for metabolism activation of the chemicals;
- This is a step-by-step presentation designed to take the user through the options for grouping with metabolism.

Outlook

- Background
- **Keywords**
- Objectives
- Specific Aims
- 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).

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
- The exercise
- Workflow

Objectives

This presentation demonstrates a number of functionalities for searching of analogues accounting for metabolism:

- Identify analogues based on the metabolites with:
 - common specific structure features;
 - common specific profiling results;
 - common specific parameter results.
- Identify analogues based on the parent and metabolites package.

Outlook

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

Specific Aims

- To familiarize the user with the metabolic transformation map similarity options when defining a category with metabolism.

Outlook

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

The Exercise

- In this exercise we will search for suitable analogues of 1,2-Ethanediamine (CAS# 107-15-3) for predicting of skin sensitization potential.
- Even though the chemical has experimental EC3 data in Skin sensitization database its structural features are suitable for the concept of this exercise.
- The target chemical has no protein binding alert for skin sensitization.
- Skin metabolism of target chemical will be accounted for.
- Different metabolic transformation map similarity options will be applied for defining a category.

Outlook

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

Workflow

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

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- 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

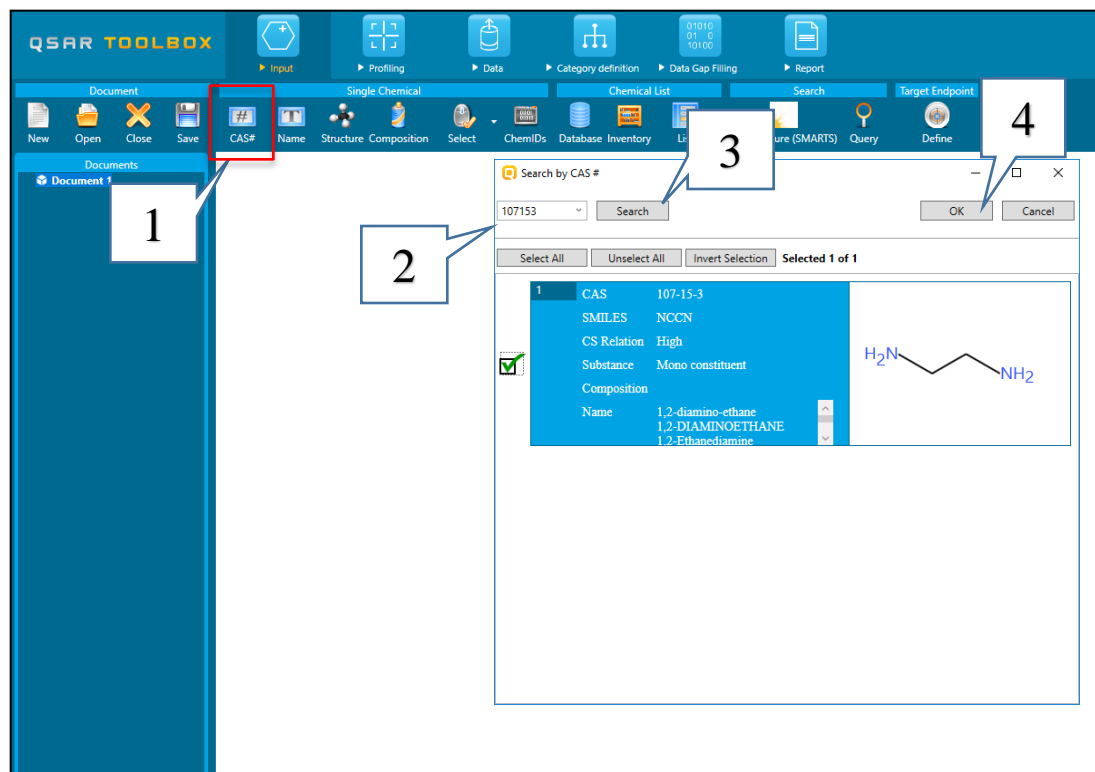
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation
- Drawing chemical structure
- Select from User List/Inventory/Databases

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Input Screen

Input target chemical by CAS#



1. Click **CAS#**;
2. Type **107153** in the blank field;
3. Click **Search**;
4. Confirm by clicking **"OK"**.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- The exercise
- **Workflow**
 - Chemical 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 metabolisms.

Profiling

Profiling the target chemical

- The actual profiling will take up to several seconds depending on the number and type of profilers selected.
- Including the metabolic simulator is more time consuming.
- The results of profiling automatically appear under the target chemical
- This result will be used to search for suitable analogues in the next steps of the exercise.

Profiling

Profiling the target chemical

The screenshot displays the QSAR Toolbox Profiling interface. The top navigation bar includes icons for Input, Profiling (highlighted with a red box and labeled '1'), Data, Category definition, Data Gap Filling, and Report. Below this, the 'Profiling' tab is active, showing 'Apply', 'View', 'New', and 'Delete' buttons. The 'Documents' section shows 'Profiling methods' with a list of endpoints. The 'Endpoint Specific' section is expanded, and the checkbox for 'Protein binding alerts for skin sensitization by OASIS' is checked (highlighted with a red box and labeled '2'). The 'Options' section shows '1 Selected'. The 'Filter endpoint tree...' panel on the right shows the chemical structure of the target (H₂N-CH₂-CH₂-NH₂) and a list of endpoints. The 'Profile' section shows 'Endpoint Specific' with 'Protein binding alerts for skin sensitization by OASIS' selected. The 'No alert found' message is displayed (labeled '4').

1. Click **Profiling**;
2. Check the box next to *Protein binding alerts for skin sensitization by OASIS* profiler;
3. Click **Apply**;
4. *No alert is found* in the target structure.

Profiling

Profiling the target chemical

Profiling results for the target metabolites can be also retrieved.

1. Check the *Skin metabolism simulator* in addition to the selected profile.

2. Click **Apply**;

3. Five metabolites are generated;

4. Structural alerts for SS are found in two of the metabolites (see next slide).

Profiling

Explain of profiling results

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...

Metabolism/Transformation

Skin metabolism simulator

Endpoint Specific

Protein binding alerts for skin sensitization by OASIS

No alert found

5 metabolite(s)

2 x Schiff base formation

2 x Schiff base formation >> Schiff base formation with carbonyl compounds

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

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

3 x No alert found

Find

Protein binding alerts for skin sens... OASIS

Skin metabolism simulator

Explain

Delete prediction

Explain prediction

Transfer to target

Set AOP target

Use for AOP

Copy

1. Right-click over the cell containing the results.
2. Select **Explain** for more details.

Profiling

Explain of profiling results

The image shows two windows from the QSAR Toolbox. The main window, titled "Profiling results", displays a hierarchical list of metabolites. A right-click context menu is open over the entry "2 x Schiff base formation with carbonyl c", with the "Display chemicals" option selected. A second window, titled "File", shows the chemical structures for "metabolite #3" and "metabolite #4", both of which have "No CAS number". Metabolite #3 is shown with a chemical structure of an aldehyde with an amino group (H₂N-CH₂-CHO). Metabolite #4 is shown with a chemical structure of a dialdehyde (O=CH-CH₂-CHO). The bottom of the "File" window has buttons for "Save to smi", "Search", and "OK".

1. Right-click on the alert;

2. Select **Display chemicals** to see for which structures (3) have alerts;

3. The structures can be saved as .smi file.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- 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

Selecting databases

The screenshot displays the QSAR Toolbox interface. At the top, there is a navigation bar with icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this is a toolbar with buttons for Gather, Import, IUCLID6, IUCID6, and Database Inventory. The main window is divided into several panes. On the left, the 'Documents' pane shows 'Document 1' with CAS: 107153. The 'Databases' pane is open, showing a list of databases. Two databases are selected: 'Skin Sensitization' and 'Skin sensitization ECETOC'. A red box highlights these two selections, and a callout bubble with the number '1' points to them. The right pane shows a 'Filter endpoint tree...' with a chemical structure and a list of endpoints. The 'Skin Sensitization' endpoint is selected, and the results show 'No alert found'.

1. Select **Skin sensitization** and **Skin sensitization ECETOC** databases

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- 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.
- A category can be defined with and without metabolism.
- Grouping with accounting for metabolic transformation is a procedure for finding analogues accounting for metabolism activation of the chemicals.

Category definition

Define with metabolism

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 highlighted with a red box and labeled '1'. The left sidebar shows 'Documents' and 'Grouping methods'. The 'Grouping methods' panel is open, showing a list of methods. The 'Sensitization' method is selected, highlighted with a red box and labeled '2'. The 'Define with metabolism' option is highlighted with a red box and labeled '3'. The 'Select metabolism' dialog box is open, showing a list of metabolism simulators. The 'Skin metabolism simulator' is selected, highlighted with a red box and labeled '4'. The 'OK' button is highlighted with a red box and labeled '5'.

1. Go to *Category definition*;

2. Click on **Sensitization** endpoint;

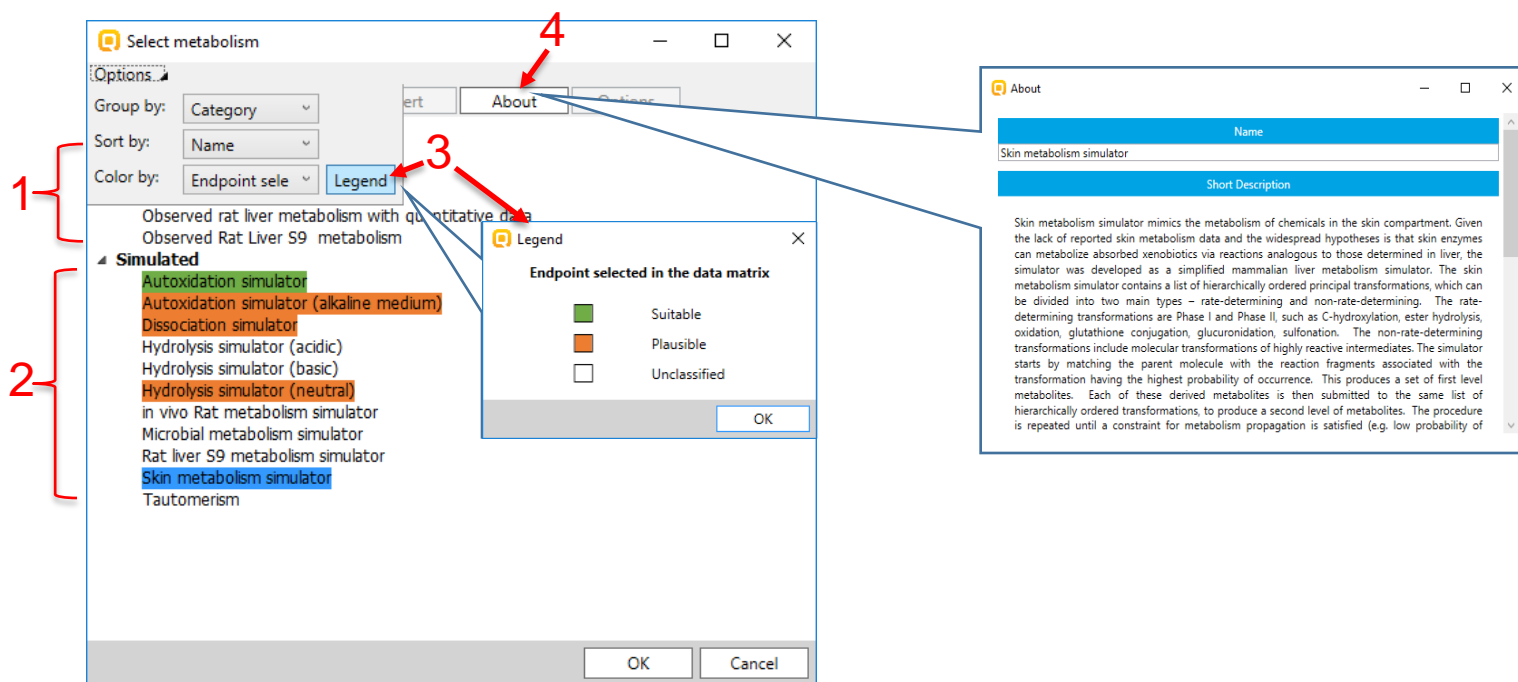
3. Click **Define with metabolism**;

4. Select **Skin metabolism simulator**;

5. Click **OK**.

Category definition

On defining with metabolism



All available transformation maps – documented (1) and simulated (2) in Toolbox can be used in the primary grouping. The simulators are coloured if a target endpoint is selected in the data matrix. The meaning of the colours is explained in the Legend (3). Short description for each of the metabolic transformations can be seen if you click **About**.

Category definition

Define with metabolism

When the transformation map is selected, the “Map similarity options” (Grouping options) dialogue appears. It shows all the generated metabolites of the target chemical by the simulator that was preliminary selected. The dialogue has two subsections:

- First subsection (1) shows parent and each of the generated metabolites (by the preliminary selected metabolism simulator) in separate rows. This allows defining of different criteria for each of structures for finding analogues.
- Second subsection (2) is working with whole package “parent + metabolites”, i.e. the criteria is provided for the whole package but not for separate metabolite.

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



See on the next slide.

Category definition

Define with metabolism

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent 	none	No criteria.
Metabolite 1 	none none Exact match Parametric Profile Similarity	No criteria.
Parent & Metabolites	none	No criteria.

OK Cancel

1

2

3

Category definition

Define with metabolism

Explanation of different options from the drop down menu:

- **None** – default options; no criterion is set;
- **Exact** – provides opportunity to search for metabolites in the analogues having exact to the specified metabolite structure; only available for the metabolites and the package “parent + metabolites” but not for the parent chemical;
- **Parametric** – to have specific value or range of variation of defined parameter (a list with all parameters currently available in the Toolbox is provided);
- **Profile** – to have specific category by selected profiler (a list with all profilers is provided);
- **Structural** – to have specific similarity based on the atom centered fragments.

Category definition

Define with metabolism – Case 1

Different map similarity options will be examined.

Case 1: Searching of analogues based on a metabolite with defined profile

As you remember, structural alerts have been found for only two of the generated metabolites (*slide 22*).

The alerts are identical and therefore we can choose only one of the structures.

Category definition

Define with metabolism – Case 1

In order to find analogues based on metabolites with common profiling result, you have to follow the described steps below:

1. On the row with the target metabolite select a query which you will use for searching. Select “**profile**” as a criteria for the first metabolite. New drop-down menu with all available profiles appears;
2. Select *Protein binding alerts for skin sensitization by OASIS* from the drop-down menu;
3. You can see the found alerts in the metabolite by click on the **Edit** button. If more than one alert is available, the user can select whether they will search analogues with all or only one of the alerts.
4. In this case there is only one structural alert and you click **OK**.
5. When you are ready click on the **OK** button.
6. Then a system alerts you that the categories from different hierarchy levels are combined. This is just a warning message, which will appear every time when a hierarchical type profiling scheme is used.
7. You can manage whether this message to appear or not by selecting the box “Do not show this dialog”.
8. Finally click **OK** in the main form.

See the illustrated steps on the next slide.

Category definition

Define with metabolism – Case 1

1 Metabolite 2

2 Profile

3 Options: Edit

4 Target categories

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

5 Combine profiles

☒ AND ☐ OR

☐ Invert result

☐ Strict

☐ Sort results

6 Yes

7 Do not show this dialog

8 OK

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent <chem>NCCN</chem>	none	No criteria.
Metabolite 1 <chem>NCC(=O)O</chem>	none	No criteria.
Metabolite 2 <chem>NCC=O</chem>	Profile	Protein binding alerts for skin sensitization by OASIS
Metabolite 3		
All chemicals		
Parent & Metabolites	none	No criteria.

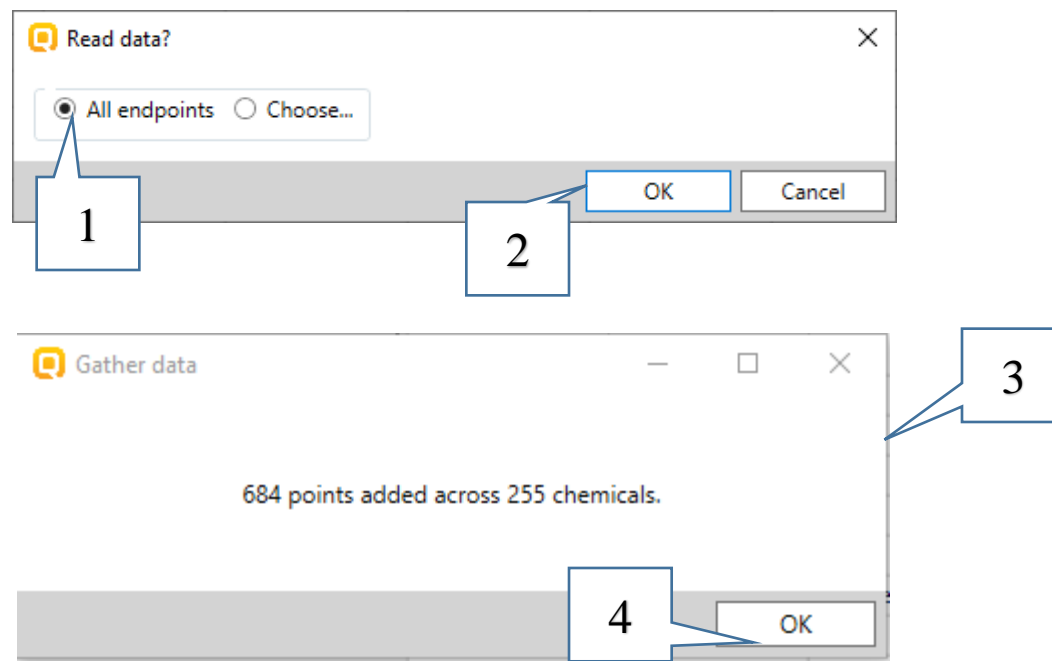
Grouping with "Protein binding alerts for skin sensitization by OASIS"

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

☐ Do not show this dialog

Category definition

Define with metabolism – Case 1



1. Keep **All endpoints** data checked;
 2. Click **OK**;
 3. A pop-up message is displayed where the number of chemicals (target chemical + analogues) and the total number of data points is shown;
- Note:** The numbers of chemicals and data points depend on the database version you are working with.
4. Click **OK**.

Category definition

Define with metabolism – Case 1

246 analogues are found. You can return to the profiling section and to check whether the found structures correspond to the defined query.

1. Go to *Profiling* section;

2. Check the used profile and metabolism (*Protein binding alerts for SS by OASIS* plus *Skin metabolism simulator*);

3. Click **Apply**;

4. Confirm by clicking Yes;

5. You can see that all of the found analogues produce one or more metabolites with the searched alert.

Category definition

Define with metabolism – Case 2

Case 2: Searching for analogues with a common metabolite

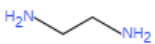
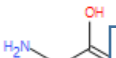
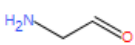
With the “**Exact**” option the user can search for analogues, which have a metabolite exactly the same as the selected metabolite generated for the target chemical.

Category definition

Define with metabolism – Case 2

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent 	none	No criteria.
Metabolite 1 	none	No criteria.
Metabolite 2 	Exact match	Matches exact structure.
Metabolite 3 		
All chemicals		
Parent & Metabolites	none	No criteria.

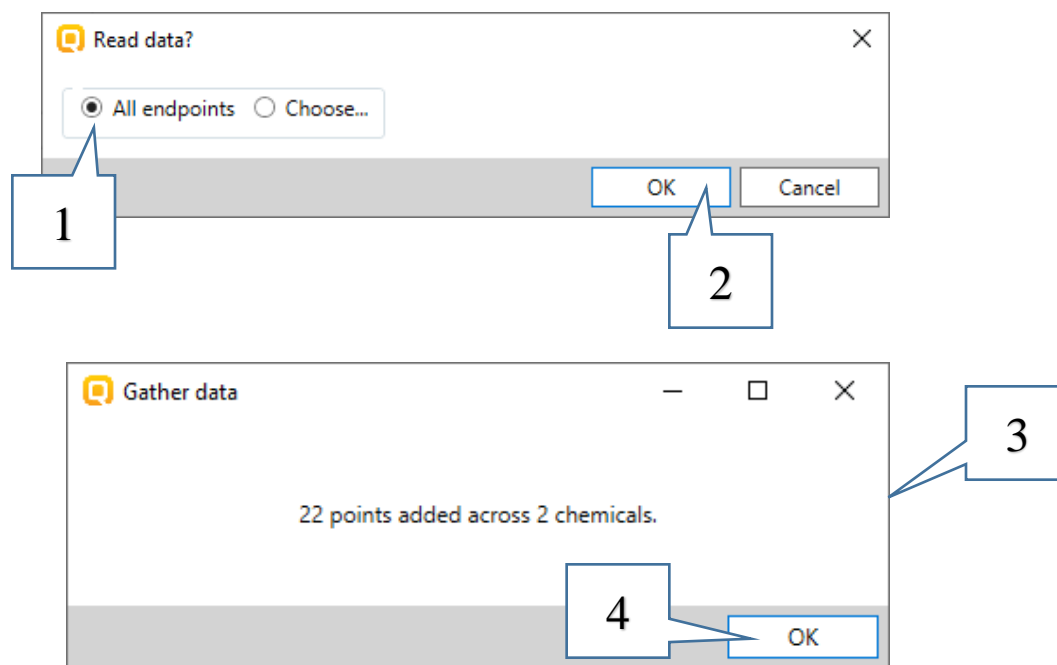
OK Cancel

Before repeating the steps illustrated below you need to follow once again the steps showed on slides # 30-31.

1. Go to the metabolite of interest
2. Select **Exact match** option from the drop-down menu (1);
3. "**Matches exact structure**" is displayed in the criteria;
4. Click **OK**.

Category definition

Define with metabolism – Case 2



1. Keep **All endpoints** data checked;
2. Click **OK**;
3. A pop-up message is displayed where the number of chemicals (target chemical + analogues) and the total number of data points are shown;
Note: The numbers of chemicals and data points depend on the database version you are working with;
4. Click **OK** .

Category definition

Define with metabolism – Case 2

One analogue is found in the selected databases.

To check whether the found structure fulfills the defined criteria, i.e. whether it produces the exact metabolite, you have to follow the steps:

1. Right-click over the found structure; 2 . Select "**Set as new target**".

The screenshot displays the QSAR Toolbox software interface. On the left, a sidebar contains a 'Filter endpoint tree...' section with expandable categories: Structure, Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards. The main workspace is divided into two columns, labeled '1 [target]' and '2'. Column 1 shows the chemical structure of ethylenediamine (NCCN). Column 2 shows the chemical structure of piperazine (C1CCNCC1). A right-click context menu is open over the piperazine structure in column 2. The menu items are: Set as new target (highlighted with a red dashed box and labeled '2'), Edit and set as new target, Chemical information, Edit substance, Add in category, Add target, Delete, Focus, Query tool matrix (with a keyboard shortcut Ctrl+F3), Set AOP target, and Copy. A callout box labeled '1' points to the piperazine structure, indicating the first step of the process.

Category definition

Define with metabolism – Case 2

1. Right-click over the newly appeared branch in the document tree:
Multiplication >> Metabolism/Transformations >> Skin metabolism simulator

The screenshot shows the QSAR Toolbox interface. The top toolbar includes buttons for Input, Profiling, Data, Category definition (highlighted), Data Gap Filling, and Report. Below the toolbar, the 'Categorize' section has buttons for Define, Define with metabolism, Subcategorize, Combine, and Clustering. The 'Category consistency' section has a button for Category elements. The 'Documents' tree on the left shows a document with a newly added branch 'Multiplication' under 'Metabolism/Transformations'. A right-click context menu is open over this branch, and the 'Skin metabolism simulator' option is highlighted at the bottom of the menu. The main window displays the chemical structure of a target molecule and a list of simulation options.

Documents

Document 1

[C: 1;Md: 21;P: 0] CAS: 107153

[C: 255;Md: 684;P: 0] Grouping with metabolism: Skin metabolism

[C: 2;Md: 22;P: 0] Grouping with metabolism: Skin metabolism

[C: 1;Md: 0;P: 0] Select from: Grouping with metabolism

1

Aquatic toxicity classification by ECOSAR

Options

f Select All Unselect All

Predefined

- Database Affiliation
- Inventory Affiliation
- OECD HPV Chemical Categories
- Substance type
- US-EPA New Chemical Categories

General Mechanistic

- Biodeg BioHC half-life (Biowin)
- Biodegradation primary (Biowin 4)
- Biodegradation probability (Biowin 1)

Filter endpoint tree...

1 [target]

Structure

Properties

and Transport

formation

ards

Multiplication

- Metabolism/Transformations
- Automerism
- Decomposition

Autoxidation simulator

Autoxidation simulator (alkaline medium)

Dissociation simulator

Hydrolysis simulator (acidic)

Hydrolysis simulator (basic)

Hydrolysis simulator (neutral)

in vivo Rat metabolism simulator

Microbial metabolism simulator

Observed Mammalian metabolism

Observed Microbial metabolism

Observed Rat In vivo metabolism

Observed rat liver metabolism with quantitative data

Observed Rat Liver S9 metabolism

Rat liver S9 metabolism simulator

Skin metabolism simulator

Category definition

Define with metabolism – Case 2

Information

A parent list with 5 child lists were created

OK

1

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Categorize

Define Define with metabolism Subcategorize Combine Clustering

Category consistency

Category elements

Documents

Document 1

[C: 1;Md: 21;P: 0] CA

[C: 1;Md: 0;P: 0] Skin metabolism simulator

[C: 6;Md: 0;P: 0] Skin metabolism simulator

[C: 1;Md: 0;P: 0] metabolite #1

[C: 1;Md: 0;P: 0] metabolite #2

[C: 1;Md: 0;P: 0] metabolite #3

[C: 1;Md: 0;P: 0] metabolite #4

[C: 1;Md: 0;P: 0] metabolite #5

2

3

Filter endpoint tree...

Structure

Structure info

Parameters

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Human Health Hazards

Parent chemical... metabolite #1 metabolite #2 metabolite #3 metabolite #4 metabolite #5

4

5

1. Click **OK** in the pop-up window; A parent list (2) (consists of the target structure and all generated metabolites) with six child lists (3) (for each of the metabolites) are created, which are also displayed in the data matrix (4). The chemical has the same metabolite (5) as the target chemical (CAS#107-15-3)

Category definition

Define with metabolism – Case 3

Case 3: Searching of analogues based on a metabolite with defined parameter value

With this option you can search analogues of the target chemical, which have metabolite(s) with defined parameter value.

When the parameter of interest is selected, it is automatically calculated for the current metabolite.

Category definition

Define with metabolism – Case 3

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent <chem>NCCN</chem>	none	No criteria.
Metabolite 1 <chem>NCC(=O)O</chem>	none	No criteria.
Metabolite 2 <chem>NCC=O</chem>	Parametric	Calculator: log Kow -1.64
Parent & Metabolites	none	No criteria.

1: Metabolite 2 (highlighted with a red box)

2: Query dropdown menu (set to Parametric)

3: log Kow selected from the calculator dropdown menu

As explained already before repeating the steps illustrated below you need to follow once again the steps showed on slides # 30-31.

Go to the metabolite of interest (1);
Select the **parametric** option from the drop-down menu (2). New calculator drop-down menu appears. Select **log Kow** (3). You can use and the filter option to find your parameter more quickly.

Category definition

Define with metabolism – Case 3

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent <chem>NCCN</chem>	none	No criteria.
Metabolite 1 <chem>N</chem>	none	No criteria.
Metabolite 2 <chem>NC(=O)O</chem>	none	No criteria.
Metabolite 3 <chem>NC=O</chem>	Parametric	Calculator: log Kow = -1.64 = ≠ < > ≤ ≥ range
Metabolite 4		
All chemicals		
Parent & Metabolites	none	No criteria.

OK Cancel

The parameter value is calculated automatically for the target metabolite (1). Then the user can decide to search with the exact value, to use any of the mathematical symbols or to search parameter values in a range (2).

Category definition

Define with metabolism – Case 3

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent <chem>NCCN</chem>	none	No criteria.
Metabolite 1 <chem>NCC(=O)O</chem>	none	No criteria.
Metabolite 2 <chem>NCC=O</chem>	Parametric	Calculator: log Kow range From: -1.6 To: -1.68
All chemicals		
Parent & Metabolites	none	No criteria.

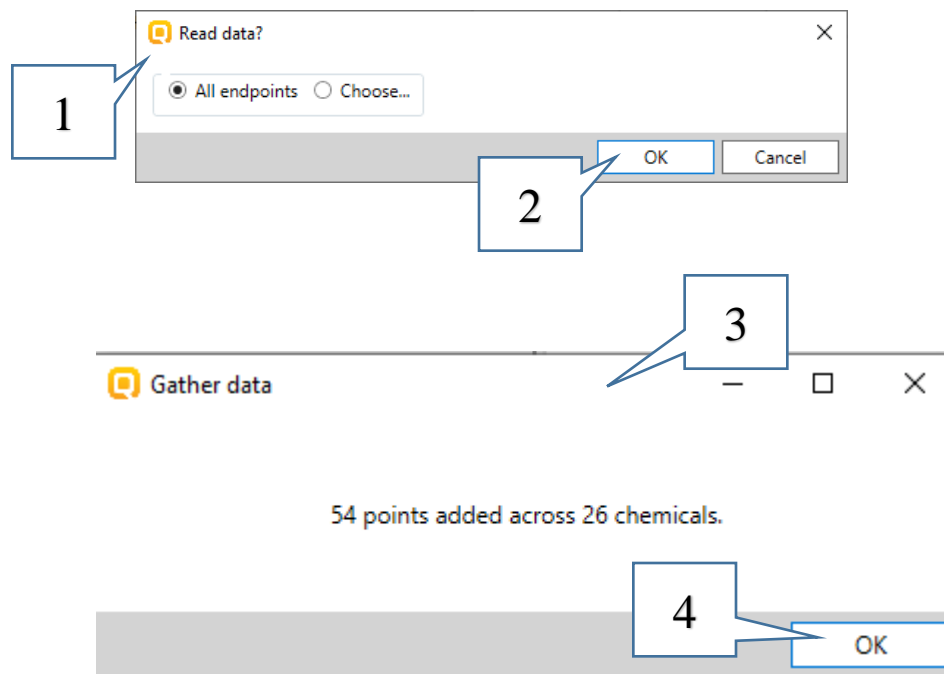
OK Cancel

We will select to search in a range (1).
 As you remember the calculated *logKow* value for the target metabolite is **-1.64**.
 Therefore, we will search analogues, which have metabolite(s) with similar *logKow* values.
 In this example we define a range from **-1.6** to **-1.68** (2).
 Click **OK** to execute the search (3).

Note: It will take up to several minutes if you make this example for first time.

Category definition

Define with metabolism – Case 3



1. Keep **All endpoints** data checked;
2. Click **OK**;
3. A pop-up message is displayed where the number of chemicals (target chemical + analogues) and the total number of data points is shown;
Note: The numbers of chemicals and data points depend on the database version you are working with
4. Click **OK**.

Category definition

Define with metabolism – Case 3

25 analogues are found in the selected database. Each of these structures posses metabolite(s) with logKow value in the previously defined range.

The screenshot displays the QSAR Toolbox software interface. The top toolbar contains icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below the toolbar, there are tabs for Profiling and Custom profile, with sub-tabs for Apply, View, New, and Delete. The main workspace is divided into a 'Documents' panel on the left and a grid of chemical structures and their corresponding metabolite categories on the right. The 'Documents' panel shows a list of documents, including 'Document 1' and 'Document 2'. The grid shows chemical structures for 9 different cases, with their metabolite categories listed below them.

Structure	1 [target]	2	3	4	5	6	7	8	9
	M: Positive	M: Category B	M: Moderately s...	M: Weak sensitizer	M: Moderate sen...	M: Strong sensiti...	M: Strong sensiti...	M: 2 %	M: N

Category definition

Define with metabolism – Case 4

Case 4: Searching of analogues based on similar metabolites

We can search for analogues of our target chemical, which have metabolites structurally similar to the target chemical metabolite(s).

1. Select **structural** option from the drop-down menu for the target metabolite;
2. The default settings could be seen by click on the **Options** button.
3. Close the window by **X** button;
4. Define the similarity threshold.

See the illustrated steps on the next slide.

Category definition

Define with metabolism – Case 4

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent <chem>NCCN</chem>	none	No criteria.
Metabolite 1 <chem>NCC(=O)O</chem>	none	No criteria.
Metabolite 2 <chem>NCC=O</chem>	Similarity	Similarity ≥ 50 %
All chemicals		
Parent & Metabolites	none	No criteria.

OK Cancel

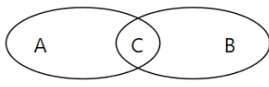
Similarity options

Measure
☐ Tanimoto (Jaccard)
☒ Dice
☐ Kulczynski-2
☐ Ochiai(Cosine)
☐ Yule

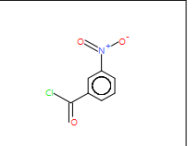
Molecular features
☐ Atom pairs
☐ Topologic torsions
☒ Atom centered fragments
☐ Path
☐ Cycles
☐ PubChem features

Options

Formula
 $c/0.5[(a+c)+(b+c)]$

Description


Structure
 NCC=O

Example


Similarity = 0.000% Details

Calculation
☐ Fingerprint
☒ Hologram

Average by features
☐ Average by features
☒ Combine all features

Atom characteristics
☒ Atom type
☒ Count H attached
☐ Count heavy atoms attached
☒ Hybridization
☐ Incident pi-bonds
☐ Valency
☐ Charge
☐ Cyclic

Define

Default Help
 OK Cancel

1. Select **Similarity** option from the drop-down menu next to the target metabolite;
2. The default settings could be seen by clicking **Options** button;
3. The settings can be modified by the user;
4. Click **OK**;
5. Define the similarity threshold.
6. Click **OK**.

Category definition

Define with metabolism – Case 4

7 analogues are found in the selected databases. Each of them possesses metabolite(s) with a similarity of 50% or more to the structure of the previously selected metabolite.

The screenshot displays the QSAR Toolbox interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Category definition' tab is active, showing a 'Filter endpoint tree...' on the left and a table of chemical structures on the right. The table has 8 columns, each containing a chemical structure. A red box highlights the first 7 structures. The bottom panel shows a list of toxicity endpoints and their associated categories.

Endpoint	Category	Category	Category	Category	Category	Category	Category
Acute Toxicity	M: Positive	M: Category B	M: 5.79 %	M: Negative	M: 2.2 %	M: Moderately s...	M: Not sensitising
ADME							
Bioaccumulation							
Carcinogenicity							
Developmental Toxicity / Terat...							
Genetic Toxicity							
Immunotoxicity							
Irritation / Corrosion							
Neurotoxicity							
Photoinduced toxicity							
Repeated Dose Toxicity							
Sensitisation	AW SW AOP 8/45	M: Positive	M: Category B	M: 5.79 %	M: Negative	M: 2.2 %	M: Moderately s...

Category definition

Define with metabolism – Case 5

Case 5: Searching for analogues based on defined criteria for the package “target and metabolites”

The user can select a profiling, parametric or structural query for both the target and its metabolites.

In this example we will use a defined profile to search for analogues.

Category definition

Define with metabolism – Case 5

1. Select a profile option for the package “parent & metabolites”;
2. Select *Protein binding alerts for SS by OASIS* profile;
3. Click on the **Edit** button. All found alerts in the parent structure and its metabolites are shown.
4. Check “**strict**” option to search only analogues with exact match these alerts.
5. Click on “**OK**” button to confirm the defined searching criteria.
6. Click on “**OK**” button in the general *Grouping options* window to execute the search.

See the illustrated steps on the next slide.

Category definition

Define with metabolism – Case 5

Grouping options (Skin metabolism simulator)

☒ All queries ☐ At least one

Chemical	Query	Criteria
Parent <chem>NCCN</chem>	none	No criteria.
Metabolite 1 <chem>NCC(=O)O</chem>	none	No criteria.
Metabolite 2 <chem>NCC=O</chem>	none	No criteria.
All chemicals		
Parent & Metabolites	Profile	Profiler: Protein binding alerts for skin sensitization by OASIS Options: Edit

OK Cancel

Options

Down Up Reset Options

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

All categories (N/A)

Acylation

Acylation >> (Thio)carbamylation of protein nucleophiles

Combine profiles

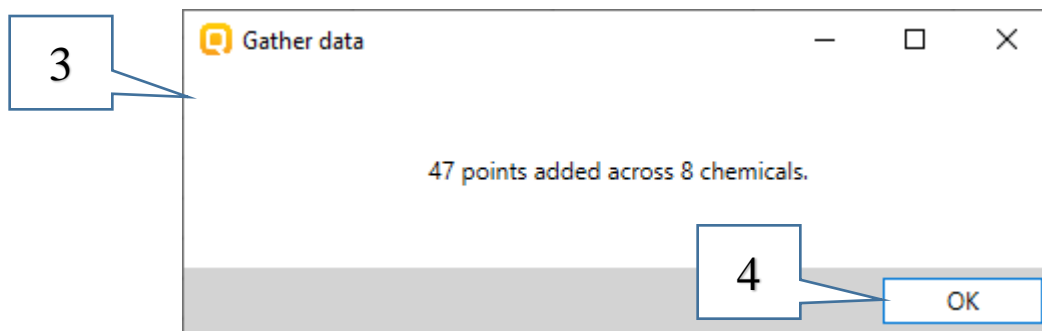
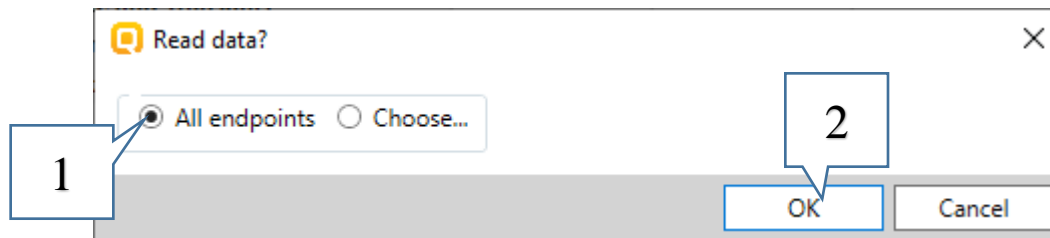
☐ Invert result

☒ AND ☐ OR ☒ Strict ☐ Sort results

OK Cancel

Category definition

Define with metabolism – Case 5



1. Keep **All endpoints** data checked;
2. Click **OK**;
3. A pop-up message is displayed where the number of chemicals (target chemical + analogues) and the total number of data points is shown;
Note: The numbers of chemicals and data points depend on the database version you are working with;
4. Click **OK**.

Category definition

Define with metabolism – Case 5

8 structures (along with the target) are retrieved. Only the searched structural alerts are found either in the parent structure or in the structures of its metabolites.

The screenshot displays the QSAR Toolbox interface during the 'Category definition' process. The 'Define with metabolism' option is selected under the 'Category' tab. The 'Documents' panel on the left shows a list of precursors for primary diamines, with 1 structure selected. The 'Filter endpoint tree' on the right shows a list of structures (1-8) with their chemical structures and associated hazard data. A red box highlights the 8 structures retrieved.

Structure	1 [target]	2	3	4	5	6	7	8
Structure								
Structure info								
Parameters								
Physical Chemical Properties								
Environmental Fate and Transport								
Ecotoxicological Information								
Human Health Hazards	8/47 M: Positive M: Ambiguous	M: Category B	M: 2.2 % M: 2.2 %	M: 27 %	M: Category C M: 0.882 %	M: Moderately s... M: Moderately s...	M: Moderately s... M: Moderately s...	M: Moderately s... M: Moderately s... M: Category B

Category definition

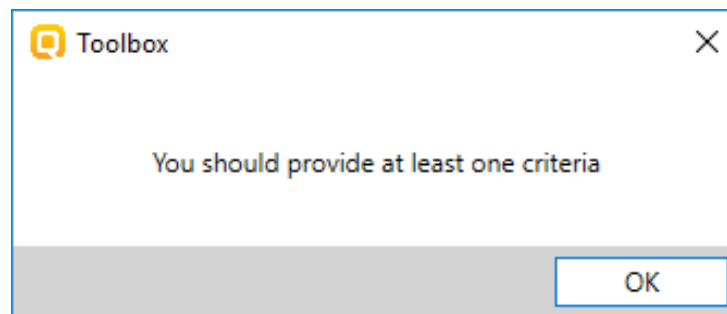
Define with metabolism – Case 6

Case 6: Searching of analogues based on combination of queries

The user can search for analogues with combination of different criteria for each of the metabolites as well as the parent structure.

Note: The user can search for analogues based on any of the target characteristics (profiling result, parameter value or structural similarity) and at least one other criteria for a metabolite.

If a criteria is set for the parent structure only, information message will appear:

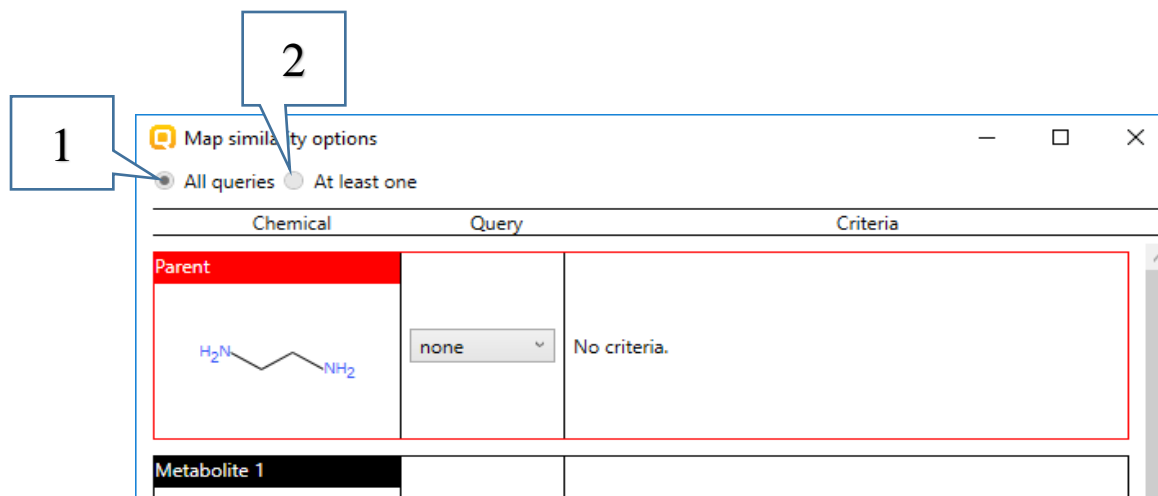


Category definition

Define with metabolism – Case 6

Case 6: Searching of analogues based on combination of queries

In addition the Toolbox user can choose whether the found analogues must meet all defined queries criteria (1) or at least one of them (2).



Category definition

Define with metabolism – Case 6

Now we will search for structures which have simultaneously:

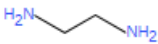
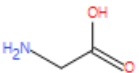

- Parent structure similar to the target structure;
- Exact metabolite structure;
- Metabolite with defined profile.

Category definition

Define with metabolism – Case 6

Grouping options (Skin metabolism simulator) — □ ×

☒ All queries ☐ At least one

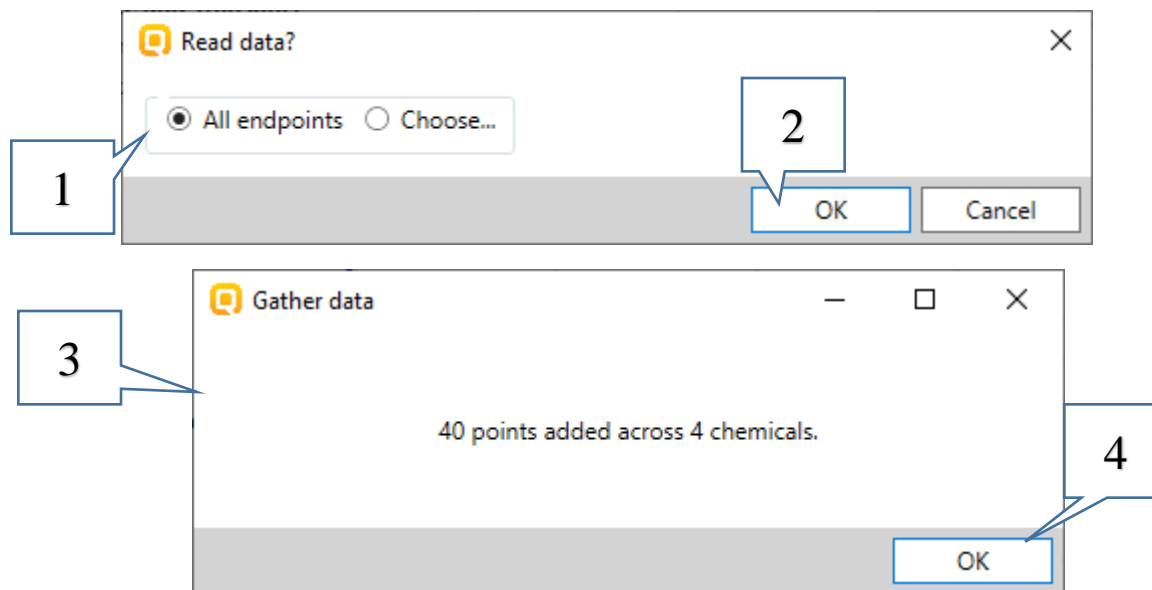
Chemical	Query	Criteria
Parent 	Similarity ▾ Similarity ≥ 50 % Options	<div>1</div>
Metabolite 1 	none ▾ No criteria.	
Metabolite 2 	Profile ▾ profiler: Protein binding alerts for skin sensitization by OASIS ▾ Options: Edit	<div>2</div>
All chemicals		
Parent & Metabolites	none ▾ No criteria.	<div>3</div>

OK Cancel

1. Select **structural** option from the drop-down menu for the parent; Set **50%** threshold with default options;
2. Select **profile** option and *Protein binding alerts for skin sensitization by OASIS profile* from the drop-down menus for the first metabolite.
3. Click **OK** to execute the search.

Category definition

Define with metabolism – Case 6



1. Keep **All endpoints** data checked;
 2. Click **OK**;
 3. A pop-up message is displayed where the number of chemicals (target chemical + analogues) and the total number of data points is shown;
- Note:** The numbers of chemicals and data points depend on the database version you are working with
4. Click **OK** .

Category definition

Define with metabolism – Case 6

Three analogues which fulfill all requirements are found (1).

The screenshot displays the QSAR Toolbox interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Category definition' menu is open, showing options like 'Define', 'Define with metabolism', 'Subcategorize', 'Combine', 'Clustering', and 'Category elements'. The 'Documents' pane on the left shows a list of documents, with 'Precursors of primary diamines' selected. The 'Filter endpoint tree...' pane on the right shows a list of endpoints, with 'Human Health Hazards' selected. The 'Structure' pane on the right shows four chemical structures, with the first one highlighted by a red box and labeled '1'.

Recap

In short, grouping with metabolism in Toolbox 4.4 allows finding analogues that have:

- metabolite categorized with a defined profiler
- the same metabolite as the target chemical
- metabolite with defined parameter value
- metabolite similar to a defined metabolite of the target chemical
- parent and its metabolites with defined profile, parameter value or structural similar
- combination of the above.

Congratulations!

- You have now been familiarized with different map similarity options for grouping with metabolism.
- Note, proficiency comes with practice!