

OECD QSAR Toolbox

Step-by-step example of how to build an user-defined linear profiling scheme

Outlook

- **Background**
- Objectives
- Profiling
- The exercise

Background

- This is a step-by-step presentation designed to provide guidance to users of the Toolbox on how to create their own profiling scheme.

Outlook

- Background
- **Objectives**
- Profiling
- The exercise

Objectives

- **This presentation demonstrates how to build a new profiling scheme including the:**
 - building linear/hierarchical scheme
 - naming of the new scheme
 - building a category by defining different type queries
 - saving the new profiler

Outlook

- Background
- Objectives
- **Profiling**
- The exercise

Profiling Overview

- As you are aware “Profiling” refers to the electronic process of retrieving relevant information on a compound which is stored in the Toolbox, other than fate and toxicity data.
- The Toolbox has many predefined profilers but it also allows the user to development new profilers.

Outlook

- Background
- Objectives
- Profiling
- **The exercise**

The Exercise

In this example we build a profiler that identifies chemicals:

➤ Case 1

- that are aldehydes (rule 1)
- that can react with proteins by “Schiff base formation” (rule 2), but
- that do not react with proteins by “Michael-type nucleophilic addition” (rule 3)
- which also have the fragment C(=O)(O)c1ccccc1 in their structure (rule 4), and
- that have a value of Log KOW between 1 and 7 (rule 5).

➤ Case 2 – chemicals that belong to a predefined list of structures (rule 1)

➤ Case 3 – chemicals structurally similar to eugenol (COc1cc(CC=C)ccc1O) (rule 1)

The Exercise

Start building a new profiler

We are going to create a new profiling scheme:

- Open the Toolbox.
- Move to the Profiling module
(see next screen shot).

Building of a new profiler

Define the name of the new profiler

1. Select the **Profiling** section; 2. Click **New**; 3. Enter the name of the new scheme, for example **"Test Profiler"**; 4. Click **OK**.

Building of a new profiler

Define type of new profiler

The screenshot displays the QSAR Toolbox software interface. The main window has a top toolbar with icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this is a sub-toolbar for Profiling with buttons for Apply, View, New, and Delete. The left sidebar contains three main sections: Documents (with 'Document 1'), Profiling methods (with '3 Selected' and options like Predefined, General Mechanistic, and (AOT)Protein binding by OASIS v1), and Metabolism/Transformations (with '2 Selected' and options like Documented and Simulated). A dialog box titled 'Type of scheme to create' is open in the center. It has a close button (X) in the top right. The first option, 'Linear or hierarchical scheme', is highlighted with a red dashed box and labeled with a callout '1'. The second option is 'Dendroid scheme (Prioritization scheme)'. At the bottom right of the dialog are 'OK' and 'Cancel' buttons, with the 'OK' button labeled with a callout '2'.

1. Select **Linear or hierarchical scheme**; 2. Click **OK**.

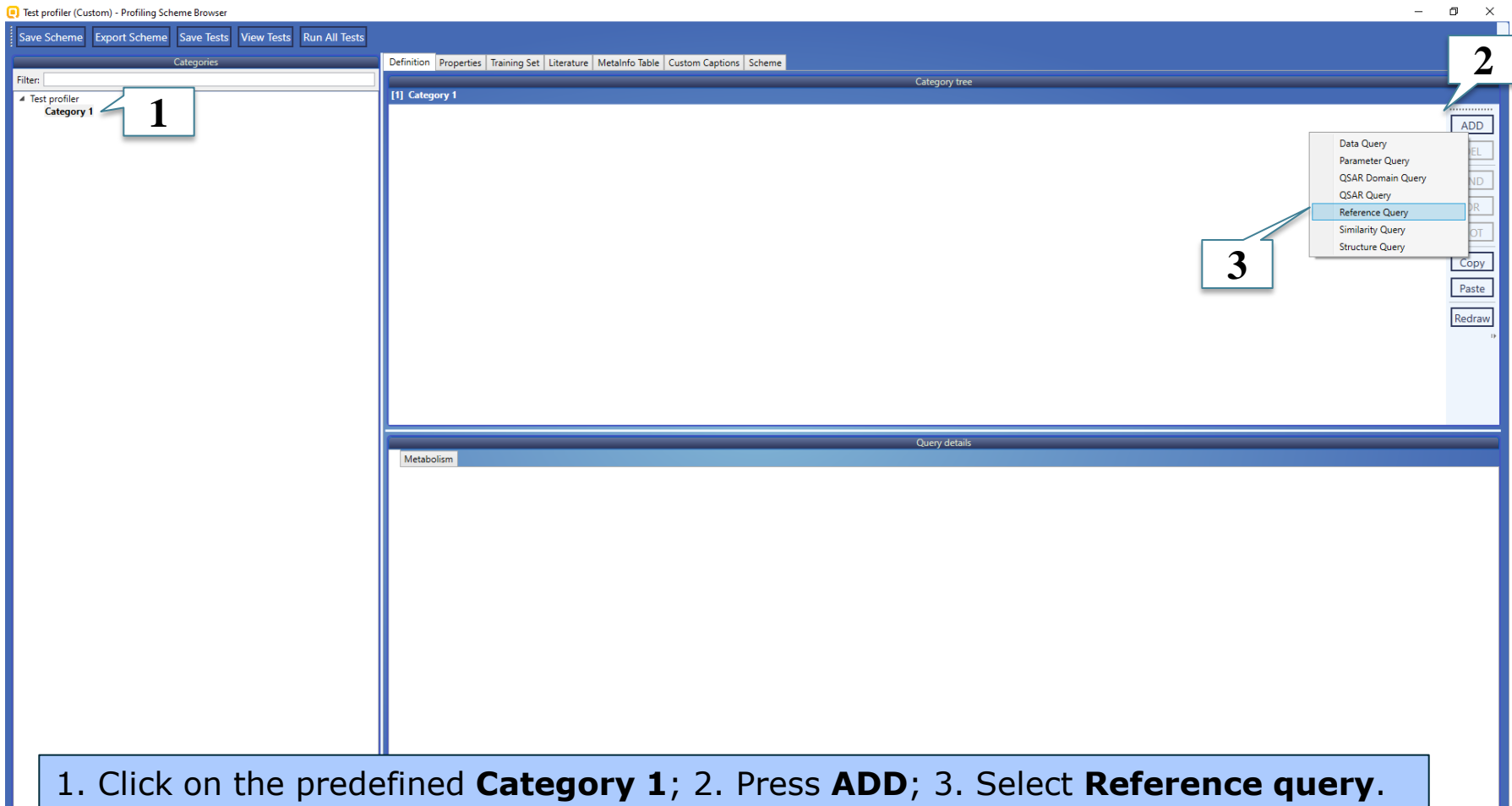
Building of a new profiler

Building the category definition – rule 1

To specify aldehydes, include a referential query making use of the predefined category definition "Aldehydes" within the Organic functional groups profiler.

Building of a new profiler

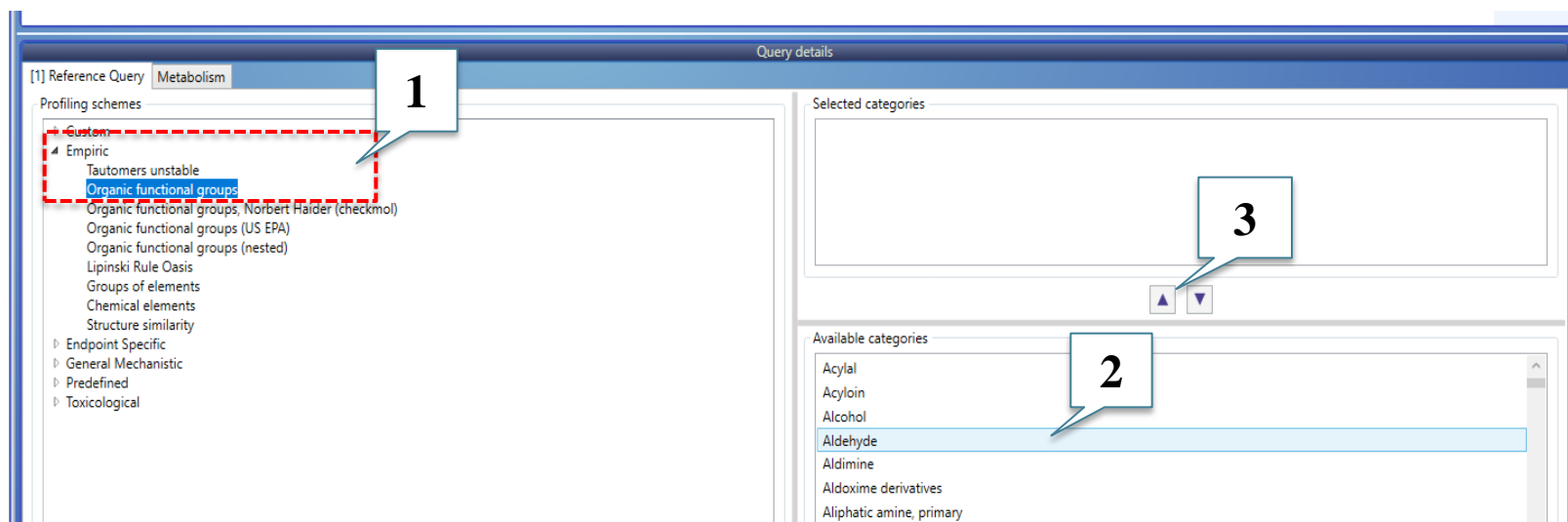
Building the category definition – rule 1



1. Click on the predefined **Category 1**; 2. Press **ADD**; 3. Select **Reference query**.

Building of a new profiler

Building the category definition – rule 1



1. Expand **Empiric** profilers and highlight **Organic functional groups**; 2. Select **Aldehydes** from the available categories; 3. Press up arrow button 

Building of a new profiler

Building the category definition – rule 1

The screenshot displays the QSAR Toolbox software interface. At the top, there are buttons for 'Save Scheme', 'Export Scheme', 'Save Tests', 'View Tests', and 'Run All Tests'. Below these, the 'Categories' tab is selected, showing a 'Filter:' field and a tree view with 'New Profiler' and 'Category 1'. The main workspace is titled 'Category tree' and contains a single icon labeled '1', which is circled in red. To the right of the workspace are buttons for 'ADD', 'DEL', 'AND', 'OR', 'NOT', 'Copy', 'Paste', and 'Redraw'. At the bottom, the 'Query details' panel is visible, with the 'Metabolism' tab selected. Under 'Profiling schemes', the 'Empiric' sub-tab is active, listing various schemes. The 'Selected categories' list on the right of this panel contains 'Aldehyde', which is also circled in red.

Building of a new profiler

Restriction of the category definition

To restrict the category definition by mechanism, add two additional referential queries:

- the first one specifies chemicals that can react with proteins by forming a Schiff base (rule 2) according to Protein binding by OECD.
- the other one specifies chemicals that do NOT react with proteins by Michael-type nucleophilic addition (rule 3) according to Protein binding by OECD.

Building of a new profiler

Building the category definition – rule 2

The screenshot shows the QSAR Toolbox software interface. The top section is titled 'Category tree' and contains two icons labeled '1' and '2'. The middle section is titled 'Query details' and contains a 'Reference Query' dropdown set to 'Metabolism'. Below this is a 'Profiling schemes' list with various chemical and biological parameters. The 'Protein binding by OECD' item is highlighted with a red box and labeled '2'. To the right of the 'Profiling schemes' list is a 'Selected categories' panel. Below this is an 'Available categories' panel. The 'Schiff Base Formers >> Direct Acting Schiff Base Formers' category is highlighted with a red box and labeled '3'. To the right of the 'Available categories' panel is a 'Category tree' panel with buttons for 'ADD', 'DEL', 'AND', 'OR', 'NOT', 'Copy', and 'Paste'. A callout box labeled '1' points to the 'ADD' button.

1. Create new **Referential query**; 2. From the **General mechanistic** profilers select **Protein binding by OECD**; 3. Select **Schiff base formers>>Direct Acting Schiff Base Formers** category from the panel with **Available categories** and move it to the panel **Selected categories**

Building of a new profiler

Building the category definition – rule 3

1 ADD

2 Protein binding by OECD

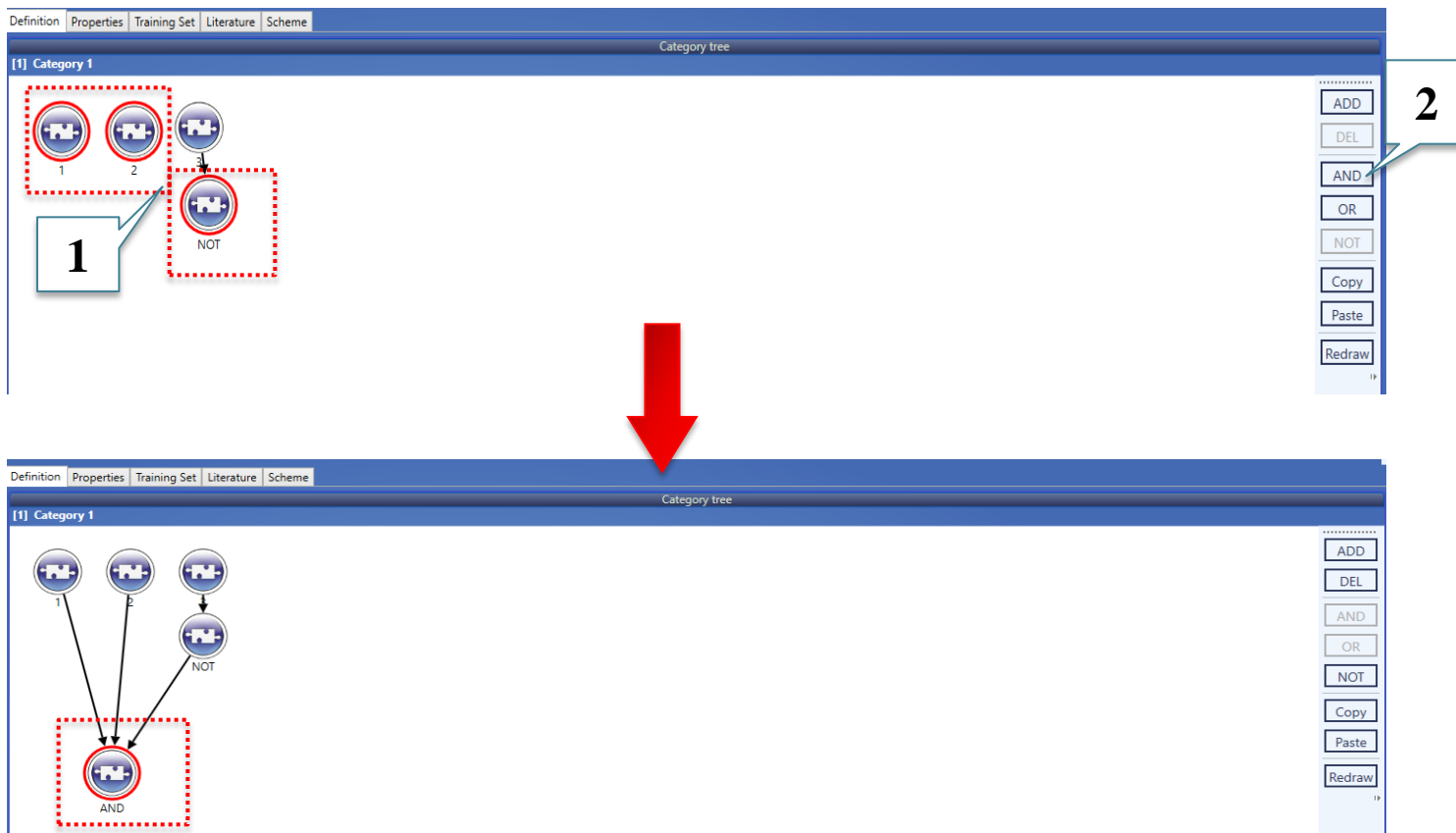
3 Michael addition

4 NOT

1. Add a new Referential query; 2. Select **Protein binding by OECD**; 3. Select **Michael addition category** and move it in the upper panel; 4. Click **NOT** to negate the query.

Building of a new profiler

Grouping the referential queries



1. To select the three queries keep **Crtl** button pressed and click on each **query** to get **RED** circle; 2. Click **AND**.

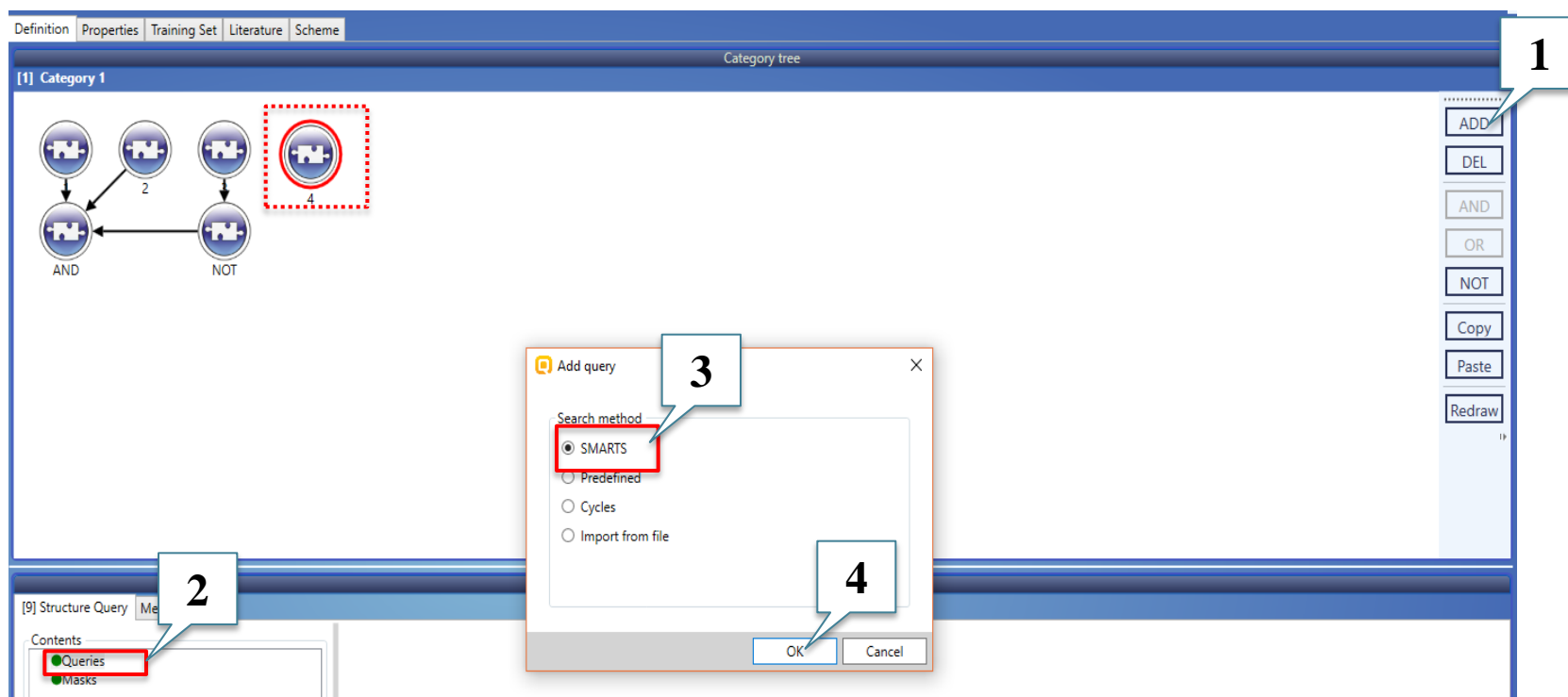
Building of a new profiler

Restriction of the category definition

To complement the category definition add an instruction for ignoring those structures which have the fragment C(=O)(O)c1ccccc1 (where the aldehyde group is deactivated - no binding with protein) - **rule 4**.

Building of a new profiler

Building the category definition – rule 4



1. **ADD** a new **Structure Query**; 2. Click on **Queries**; 3. **SMART** query is selected (by default); 4. Click **OK**;

Building of a new profiler

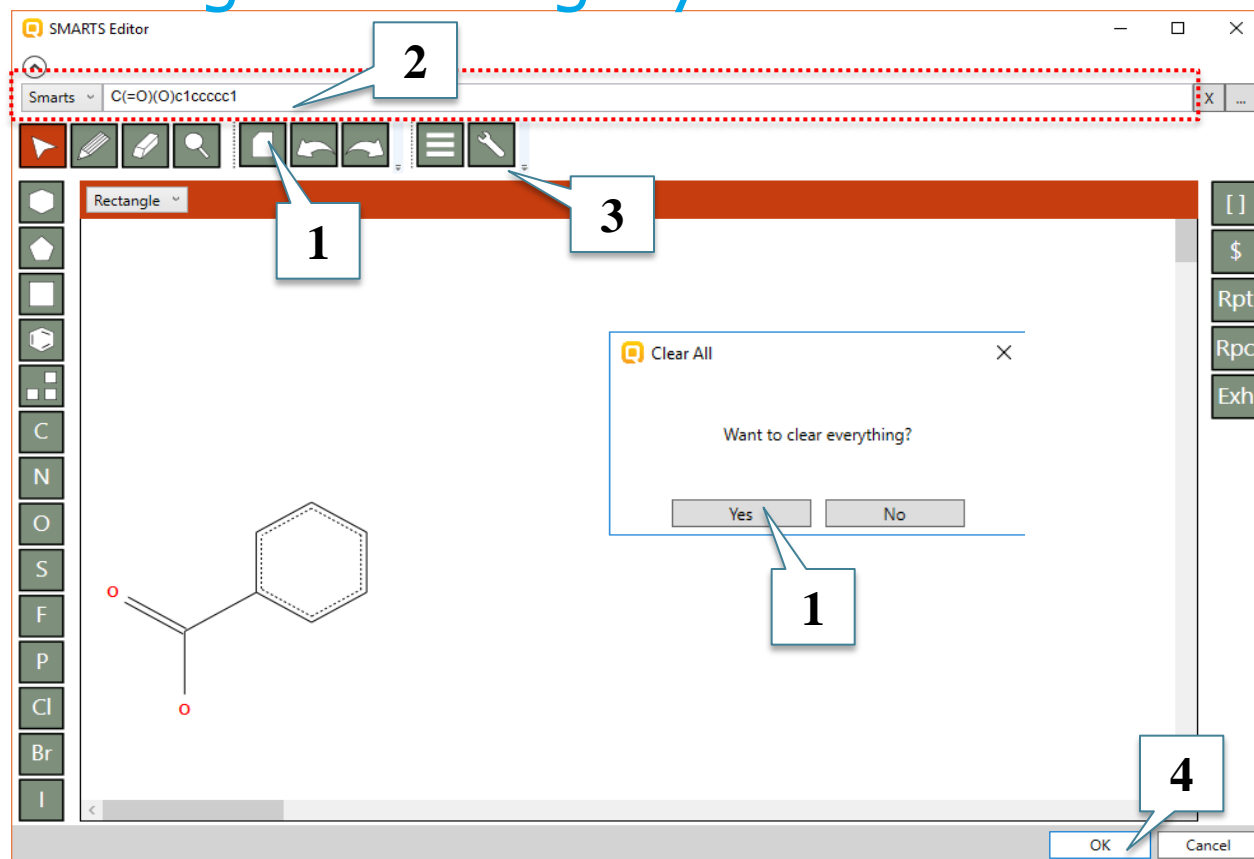
Building the category definition – rule 4


The screenshot displays the QSAR Toolbox interface. The top panel, 'Category tree', shows a logical diagram with nodes labeled 'AND' and 'NOT', and a node labeled '4' circled in red. The bottom panel, 'Query details', shows the 'SMARTS' search string 'c1ccccc1' and a chemical structure of benzene. Two callout boxes with numbers 1 and 2 point to the 'Search 1: SMARTS' entry in the 'Contents' list and the 'Edit' button, respectively.

1. Select **Search 1: SMART**; 2. Click on **Edit**.

Building of a new profiler

Building the category definition – rule 4



1. Clear default structure, appearing of message, select "Yes"; 2. Type the **SMART** of the fragment C(=O)(O)c1ccccc1 or draw it in the 2D editor window; 3. Click  **button**; 4. Click **OK**.

Building of a new profiler

Building the category definition – rule 4

The screenshot displays the QSAR Toolbox interface. The top panel, 'Category tree', shows a hierarchical structure of queries. A node is highlighted with a red dashed box and labeled '1'. The bottom panel, 'Query details', shows the 'Structure Query' tab. It includes a 'Contents' list on the left, a 'SMARTS' input field with the string 'C(=O)O)c1ccccc1', and a 'View mode' dropdown set to 'Facade'. The 'Navigation mode' is set to 'Cascade'. A chemical structure of benzaldehyde is shown in the center. On the right side of the 'Query details' panel, a toolbar contains buttons for 'ADD', 'DEL', 'AND', 'OR', 'NOT', 'Copy', 'Paste', and 'Redraw'. The 'NOT' button is highlighted with a red dashed box and labeled '2'.

1. Click over the structural query in order to select it (it is **RED**); then 2. Click **NOT** to negate the query.

Building of a new profiler

Parametric range

To specify the needed parametric range, add a parametric query and combine it together with the rest queries in an AND query. The parametric range is log Kow between 1 and 7 (rule 5).

Building of a new profiler

Building the category definition – rule 5

The screenshot displays the QSAR Toolbox interface. The top panel, 'Category tree', shows a logical structure with five nodes. Node 1 is connected to node 2 by an 'AND' operator. Node 3 is connected to node 4 by a 'NOT' operator. Node 5 is also connected to node 4 by a 'NOT' operator. The bottom panel, 'Query details', shows a list of parameters on the left. The 'log Kow' parameter is selected. The 'Expression' field shows 'between 1 7'. The 'Origin' and 'Destination' fields are set to 'None'.

1. **ADD** new **parametric query**; 2. From the list with 2D parameters select **log Kow**;
3. Select qualifier **between** and enter the requisite values (**1 to 7**).

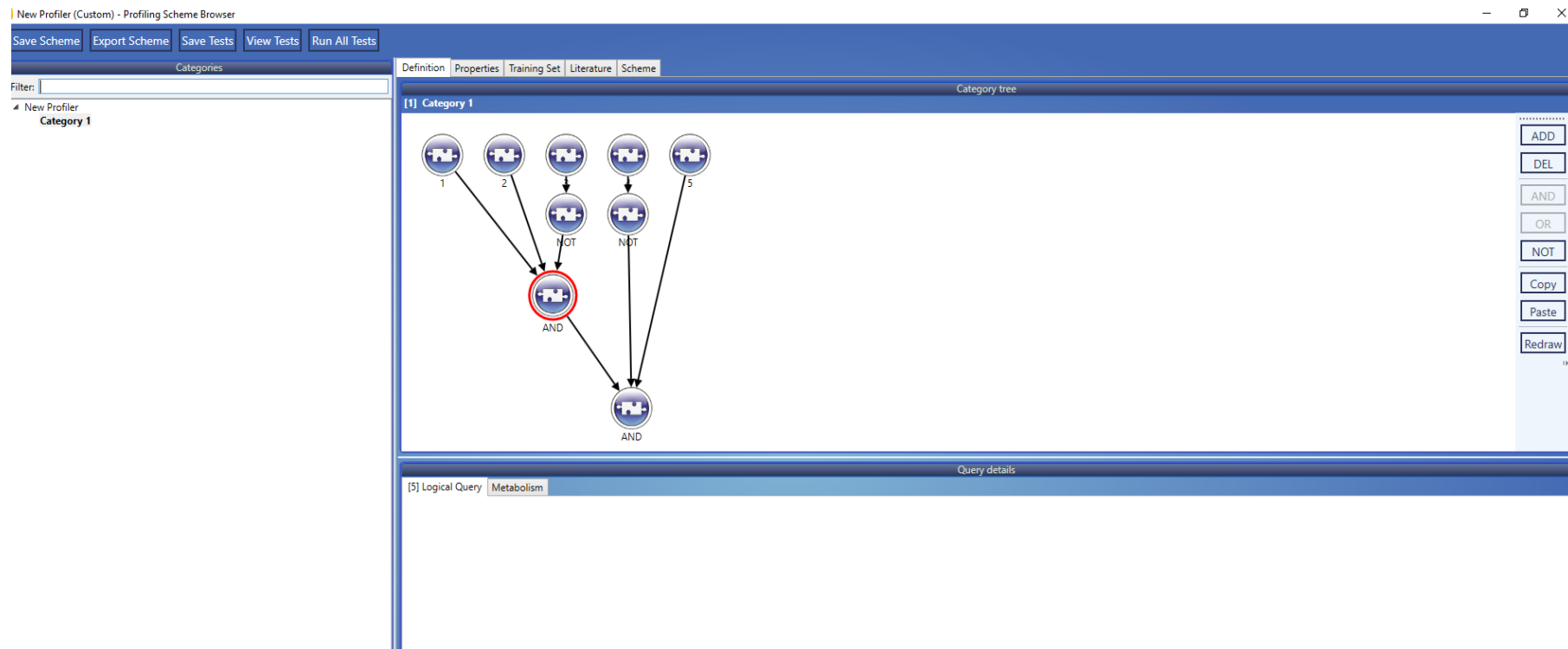
Building of a new profiler

Combine queries

Finally, group the combined referential queries, the structural restriction and the parametric range in one single **AND** query.

Building of a new profiler

Combine queries



Building of a new profiler

Rename a category

When create a new profile by default it contains a category named **Category 1**. The user has possibility to rename it. In this example it will be changed to **Case 1**.

Building of a new profiler

Rename a category

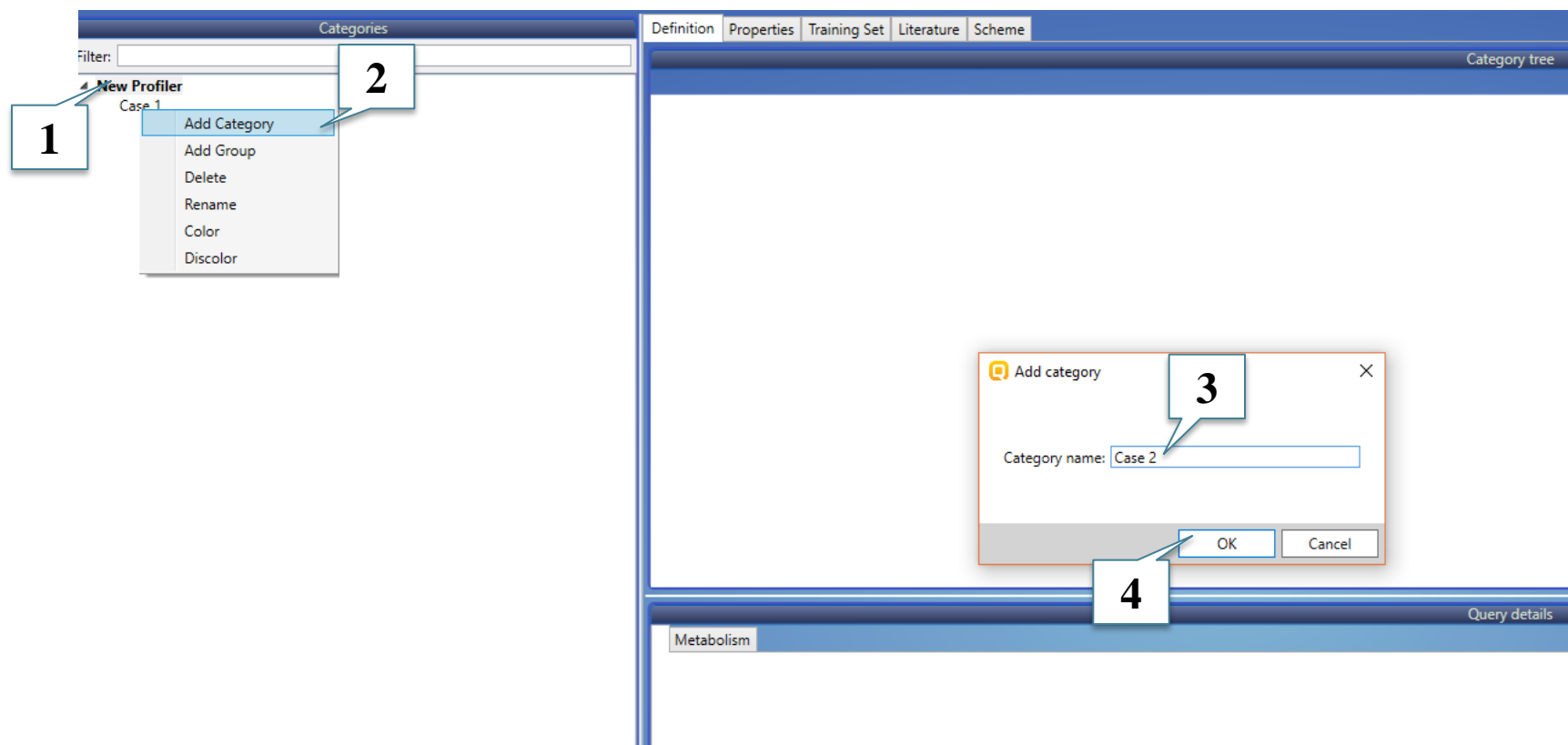
The screenshot shows the 'New Profiler (Custom) - Profiling Scheme Browser' window. The 'Categories' tab is active, displaying a tree structure with '[1] Case 1'. A context menu is open over 'Case 1', with 'Rename' selected. A 'Rename category' dialog box is open, showing 'Case 1' in the 'Category name' field. The 'Logical Query' tab is also visible at the bottom.

1. Right click on **Category 1**; 2. Select **Rename**; 3. Type the new name for example "**Case 1**"; 4. Click **OK**.

1. Right click on **Category 1**; 2. Select **Rename**; 3. Type the new name for example "**Case 1**"; 4. Click **OK**.

Building of a new profiler

Add a new category



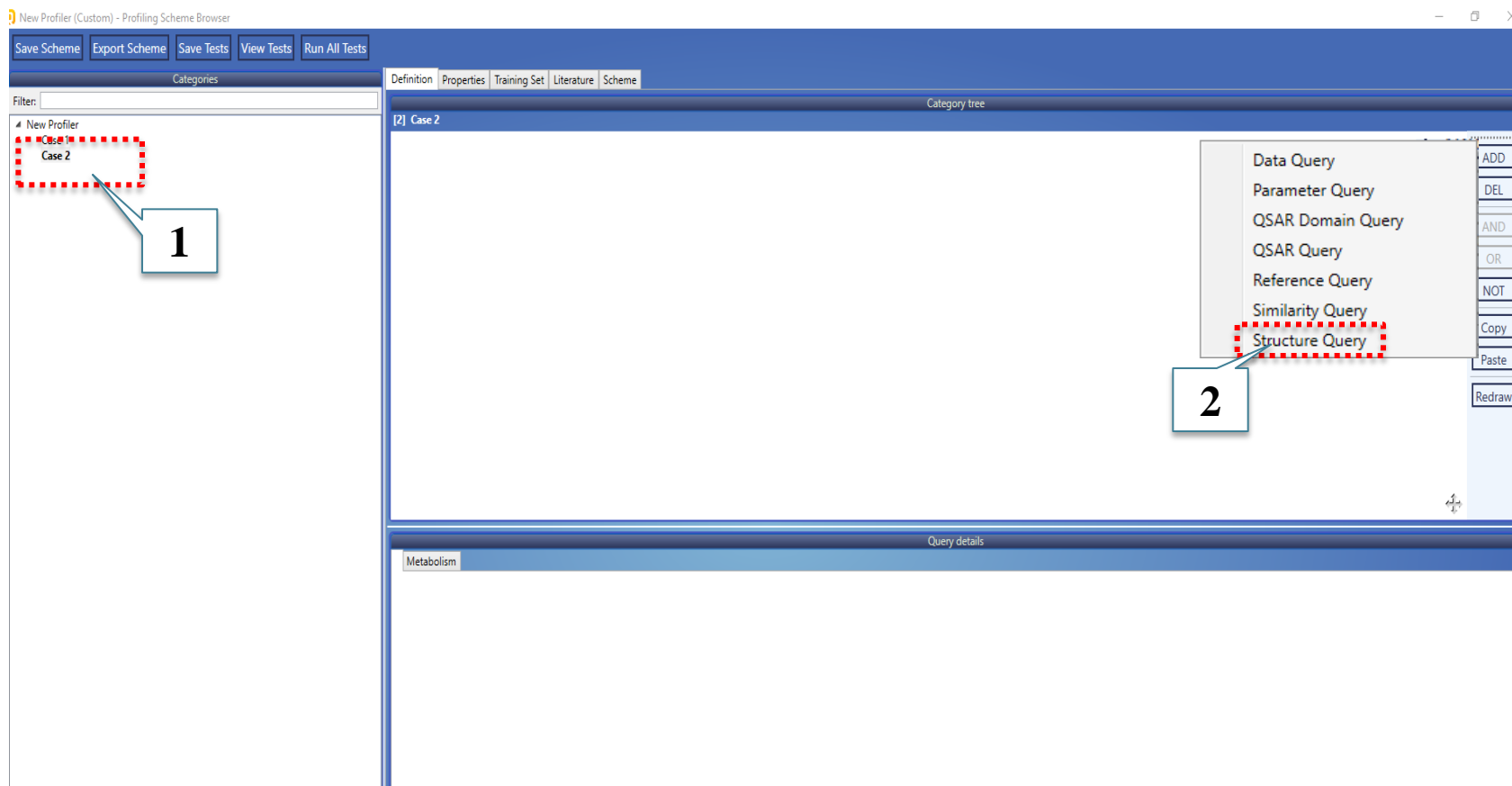
Building of a new profiler

Search by a list of structures

- Structural query provides possibility to search by list of structures (SMARTS). If the target chemical belongs to the loaded in query list, it fulfills the category definition.
- It is possible to logically combine the that type of query with rest query types.

Building of a new profiler

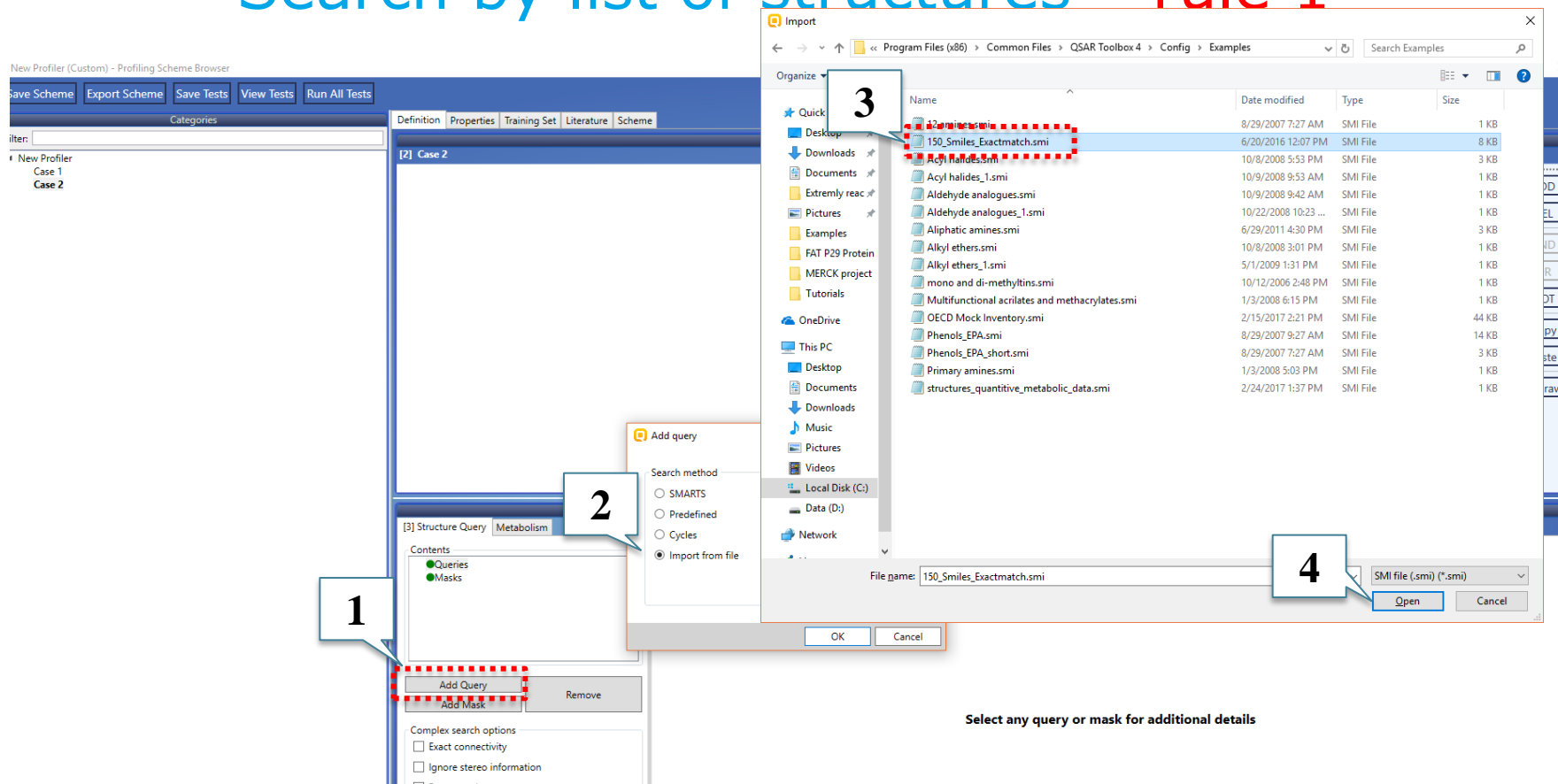
Search by list of structures – rule 1



1. Click on category; 2. **ADD** a new **Structure Query**.

Building of a new profiler

Search by list of structures – rule 1



1. Click **Add query**; 2. Select **Import from file**; 3. Load to add a pre-defined **smi** file which contains SMART*; 4. Click **Open**.

* The example file with 150 discrete SMART could be found in the example folder of TB installation:
C:\Program Files (x86)\Common Files\QSAR Toolbox 4.4\Config\Examples.

Building of a new profiler

Search by list of structures – Additional options

The screenshot displays the QSAR Toolbox software interface. At the top, the 'Definition' tab is active for 'Case 2'. A red circle highlights a puzzle piece icon labeled '1'. Below this, the 'Structure Query' section is visible, showing a list of queries (Search 1 to Search 7, all SMARTS) and a 'Queries execution mode' dropdown set to 'Any', which is also highlighted with a red box and a callout labeled '1'. The 'Query details' section shows the SMARTS string 'CC(O)=O' and a chemical structure of acetic acid.

1. Select **Any** from Query execution mode – it means that the SMART in the list are OR-ed and it is enough the target chemical to match at least one of them.

Building of a new profiler

Search by similarity query

- Similarity query provides possibility to search chemicals similar to a predefined target chemical. In this exercise will search chemical similar to eugenol more than 60%, using default Similarity options – **rule 1**.
- The Similarity option are explained in details in the following tutorial (https://qsartoolbox.org/wp-content/uploads/2020/02/Tutorial_26_New-options-of-the-structure-similarity.pdf)

Building of a new profiler

Search by similarity query

The screenshot displays the QSAR Toolbox software interface. At the top, there are tabs for 'Save Scheme', 'Export Scheme', 'Save Tests', 'View Tests', and 'Run All Tests'. Below these is a 'Categories' section with a filter input and a tree view showing 'New Profiler', 'Case 1', 'Case 2', and 'Case 3'. A red dashed box labeled '1' highlights the 'Case 3' icon. To the right, a 'Category tree' panel shows a list of query types: 'Data Query', 'Parameter Query', 'QSAR Domain Query', 'QSAR Query', 'Reference Query', 'Similarity Query', and 'Structure Query'. A red box labeled '2' highlights the 'Similarity Query' option. Below the main interface, the 'Query details' panel is visible, showing settings for a 'Similarity Query' under the 'Metabolism' tab. It includes fields for 'Similarity, %' (Expression: =, Value: 0), 'Options' (Mode: Hologram, CombineAllFeatures), 'Measure' (-Dice), 'Molecular features' (-AtomCenteredFragments), and 'Atom characteristics' (-AtomType, -CountHAttached, -Hybridization). A 'Target' section with a 'SMILES' input field is also present.

1. Add new category "**Case 3**"; 2. **ADD** a new **Similarity Query**;

Building of a new profiler

Search by similarity query

The screenshot displays the QSAR Toolbox software interface. On the left, a 'Categories' panel shows a tree structure with 'New Profiler' and 'Case 1', 'Case 2', 'Case 3'. The main window has a 'Definition' tab selected. Below the 'Definition' tab, a 'Category tree' section shows a single node '1' with a puzzle piece icon. The 'Query details' section at the bottom is titled '[2] Similarity Query Metabolism'. It contains a 'Similarity, %' section with an 'Expression' field set to '= 0'. Below this is an 'Options' button, which is highlighted with a red box and a callout labeled '1'. To the right of the 'Options' button is a 'Target' section with a 'SMILES' field and an 'Edit' button. The 'Options' button is located below the 'Similarity, %' section.

1. Click **Options** to define the target chemical

Building of a new profiler

Search by similarity query

1 Similarity options

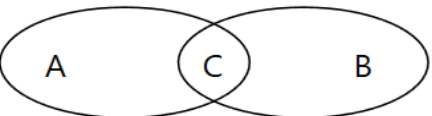
Measure

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

Formula

$$\frac{c}{0.5 [(a + b) + (b + c)]}$$

Description



Molecular features

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

Options

Calculation

- ☐ Fingerprint
- ☒ Hologram

Average by

- ☐ Average by
- ☒ Combine a

Atom character

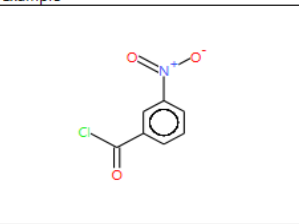
- ☒ Atom type
- ☒ Count H a
- ☐ Count he
- ☒ Hybridiza
- ☐ Incident p
- ☐ Valency
- ☐ Charge
- ☐ Cyclic

Structure

COc1cc(CC=C)ccc1O

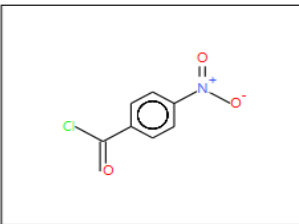
Define

Example



A	B	C
2	2	10

Similarity = 83.333% Details



2 Define

3 OK

4 OK

1. **Similarity options** set by default; 2. **Define** target structure by pasting the SMART for eugenol (COc1cc(CC=C)ccc1O) or drawing it using the 2D Editor (click Define); 3. Click **OK**; 4. Click **OK** to finalize.

Building of a new profiler

Search by similarity query

The screenshot shows the QSAR Toolbox interface. The 'Query details' panel is active, showing a similarity query configuration. A red box highlights the 'Expression' field set to '>' and the 'Similarity %' field set to '70'. A blue callout box with the number '1' points to this red box. The 'Query details' panel also shows 'Options' for the similarity query, including 'Model: Hologram, CombineAllFeatures', 'Measure: -Dice', 'Molecular features: -AtomCenteredFragments', and 'Atom characteristics: -AtomType, -CountHAttached, -Hybridization'. The 'Target' field contains the SMILES string 'COc1cc(C=C)ccc1O' and a chemical structure of Eugenol is displayed.

1. Define more than 70% structural similarity between Eugenol and profiled chemicals;

Building of a new profiler

Additional functionalities

- When the queries in the scheme are done the user can switch to the tabs: *Properties*, *Training sets*, *Literature*, and *Scheme* in order to enrich information supporting queries.

Building of a new profiler

Additional functionalities: Properties section

Definition | **Properties** | Training Set | Literature | Scheme

General Properties

ID: 3 (ProfilingNode)

Caption: Case 3

Literature Key: Eugenol structural similarity.htm

Color: #00000000

Description

Comments

Examples

Yes Example

No Example

Referential Node

Metabolism

Packet Options

- ☒ Any
- ☐ All
- ☐ Accumulative
- ☐ Individual
- ☐ Parent

Another way to change the category name is to put the text in the **Caption** field

Literature key is related to the justification of the category. Here is needed to put the name of *htm* (web page filtered) file containing information.

Description and **Comments** are fields for free text.

The category can appear on the data matrix (after profiling) with different colors. It could be selected here.

Here by **right click** could be pasted SMILES for YES and NO examples.

Building of a new profiler

Additional functionalities: Literature

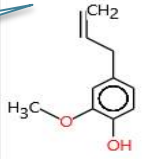
Categories

Filter:

- New Profiler
 - Case 1
 - Case 2
 - Case 3

Visualization of the Literature (.htm) file

Eugenol structural similarity



Similarity options

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

Calculation

- ☐ Fingerprint
- ☒ Hologram
- ☐ Average by features
- ☒ Combine all features

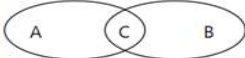
Atom characteristics

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

Formula

$$r' = \frac{0.5 [(a + b) + (b + c)]}{c}$$

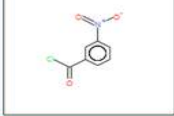
Description



Structure

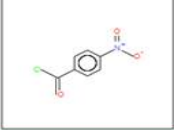
COc1cc(C=C)ccc1O

Example



A	B	C
2	2	10

Similarity = 83.333% Details



Default Help
OK Cancel

Building of a new profiler

Additional functionalities: Training set

The screenshot shows the QSAR Toolbox software interface. The 'Training Set' tab is selected in the top menu. Below the menu, there are 'Load' and 'Clear' buttons. A table lists training set data with columns: CAS#, Name, Smiles, and Representative h-CLAT DC56/DC84. An 'Open' file dialog is overlaid on the interface, showing the 'Downloads' folder. The file 'Vinyl pyridines.smi' is selected in the list. The 'File name' field at the bottom of the dialog contains 'Vinyl pyridines.smi'. To the right of the dialog, the chemical structure of 1-vinylpyridine is displayed, consisting of a pyridine ring with a vinyl group (-CH=CH₂) attached to the nitrogen atom.

1. Click **Load**; 2. Browse the training set file*; 3. Click **Open**

*The training set file should be a tab delimited file (.smi or .sdf) containing the following columns: CAS#, Name, SMILES, Parameters. It should have title row. Empty positions are acceptable.

Building of a new profiler

Additional functionalities: Scheme section

The **Scheme name** could be changed here

The **Counter profiler** appears on the data matrix for chemicals which do not correspond to any of the profiler categories

In the **Literature** is needed to put the specific art of the path* (Reference\Name of folder with .htm files) to the folder containing literature files for all categories.

Default Color related to appearance of all categories on the data matrix. It is also related with counter profiler.

Fields for free text

Labels order: Available labels: Case 1, Case 2, Case 3

Ordered labels:

Definition | Properties | Training Set | Literature | **Scheme**

Scheme Name: New Profiler

Nature: Custom

Version:

Counter Profile:

Literature: References\New profiler

GUID: 1eccfe90-343c-4c86-b2ce-9de27c587464

Type: Linear

Default color: #00000000

Author:

Donator:

Website:

Adopted:

Last Modified: 13.04.2017

Description:

Changelog:

Disclaimer:

Literature (.htm) file should be collected in a folder and placed in the following directory *C:\Program Files (x86)\Common Files\QSAR Toolbox 4.4\Config\Addins\LMC.Toolbox.Server.Profiling\References

Building of a new profiler

Save scheme

The screenshot shows the 'Profiling Scheme Browser' window. The 'Save Scheme' button is highlighted with a red box and labeled '1'. A confirmation dialog box with the text 'Scheme successfully saved on the server.' and an 'OK' button is shown, with the 'OK' button labeled '2'. The 'Close' button (X) in the top right corner is labeled '3'. The main window displays a category tree on the left, a definition area in the center, and query details on the right, including a chemical structure of 4-(3-methoxyphenyl)but-3-en-2-ol.

1. Press the **Save scheme** button; 2. Click **OK** to confirm the newly created profiler and 3. Close the window.

New profiler

The new profiler can now be applied to a target chemical or a list of chemicals complying with the newly constructed category definitions (see next screen shot).

New profiler Results

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Apply View New Delete

Documents

Document 1
[C: 14;Md: 0;P: 0] Chemical name: contai

Filter endpoint tree...

Structure

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

Profiling

Custom

New profiler

Options 1 Selected
f Select All Unselect All Invert

☒ Crowded anilines
☐ ECHA' test
☐ Example Prediction Scheme (PBT)
☒ New profiler
☐ Precursor of 120000
☐ Precursor of p-BQ
☐ Precursors of primary diamines
☐ Skin sensitisation for DASS
☐ Test dendro for matita
☐ Test dendro scheme

Metabolism/Transformations

Options 0 Selected
f Select All Unselect All Invert

☒ Documented
☐ Observed Mammalian metabolism
☐ Observed Microbial metabolism
☐ Observed Rat In vivo metabolism
☐ Observed rat liver metabolism with qu
☐ Observed Rat Liver S9 metabolism

	1	2	3	4	5	6	7	8	9	10
Structure										
Case 2	Case 2	Case 2 Case 3	Case 2	Case 2	Case 2	Case 2	Case 2	Case 2	Case 2	Case 2

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

Building of a new profiler

Create a new dendroid profiler

How to build a dendroid/prioritization profiler is explained in details in “*Manual for creating prioritization schemes*” (http://oasis-lmc.org/media/74346/Manual_for_creating_prioritization_schemes.pdf).

Congratulation

- You have used several new functions to create a new profiler for use with the Toolbox.
- Continual use of the Toolbox will increase your skills.