

OECD QSAR Toolbox v.4.1

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

The screenshot displays the QSAR Toolbox software interface. The top menu bar shows 'Profiling' as the active mode. In the 'Documents' panel on the left, the 'New' button is highlighted with a red box and labeled '2'. A callout '1' points to the 'Profiling' icon in the top bar. A 'New profiler' dialog box is open in the center, with the 'Profiler name' field containing 'New Profiler' and labeled '3'. The 'OK' button in the dialog is labeled '4'. The background shows various toolbars and panels, including 'Profiling methods' and 'Metabolism/Transformations'.

1. Select the **Profiling** mode; 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 shows the QSAR Toolbox software interface. On the left, there are panels for 'Profiling methods' and 'Metabolism/Transformations'. The main window displays a dialog box titled 'Type of scheme to create'. The dialog box contains two radio button options: 'Linear or hierarchical scheme' and 'Dendroid scheme (Prioritization scheme)'. The 'Linear or hierarchical scheme' option is selected and highlighted with a red dashed box and a callout '1'. The 'OK' button is highlighted with a callout '2'.

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

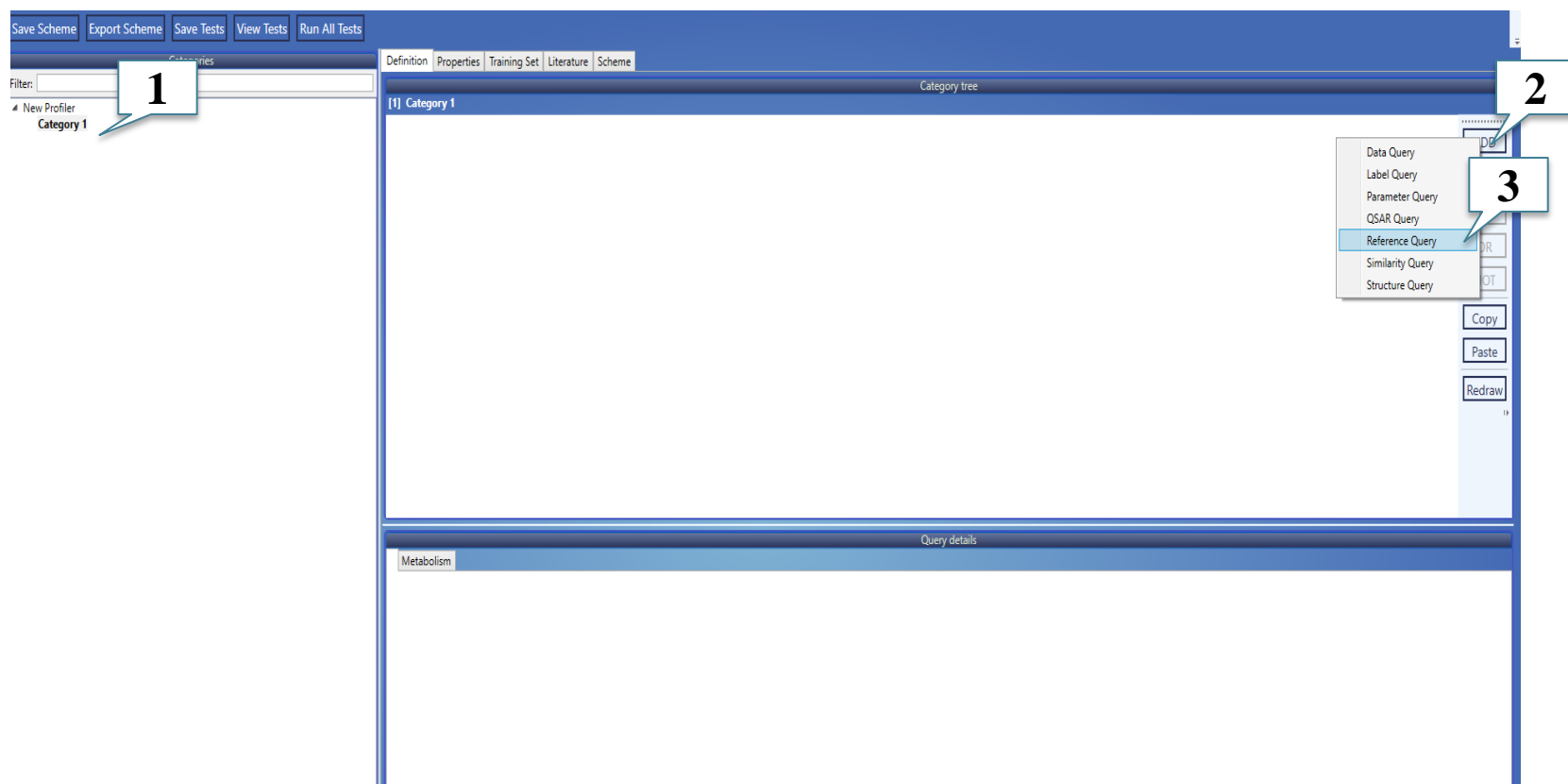
Building of a new profiler

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

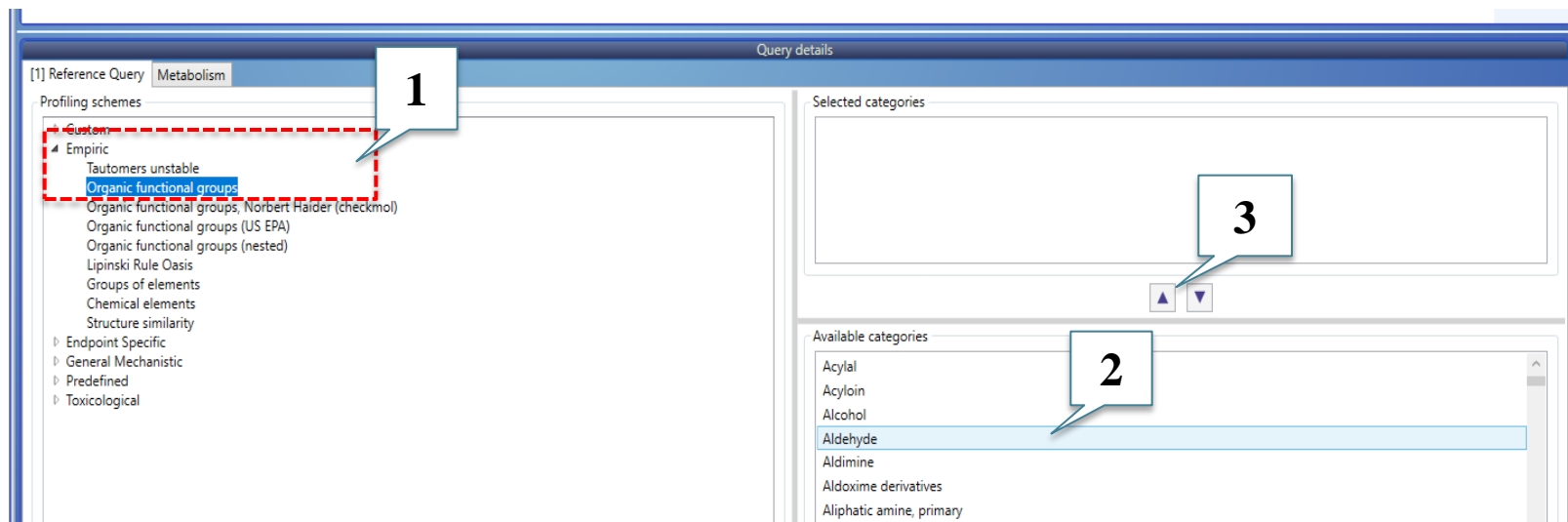
Building the category definition – rule 1



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 interface for building a new profiler. The top menu bar includes 'Save Scheme', 'Export Scheme', 'Save Tests', 'View Tests', and 'Run All Tests'. Below this is a 'Categories' section with a 'Filter:' input field and a tree view showing 'New Profiler' > 'Category 1'. The main workspace is divided into two panels: 'Category tree' and 'Query details'. The 'Category tree' panel shows a single category '1' with a red circle around its icon. The 'Query details' panel shows a list of 'Profiling schemes' under the 'Metabolism' tab, including 'Custom' and 'Empiric'. Under 'Empiric', several schemes are listed, and 'Aldehyde' is selected, highlighted with a red circle. A vertical toolbar on the right side of the 'Category tree' panel contains buttons for 'ADD', 'DEL', 'AND', 'OR', 'NOT', 'Copy', 'Paste', and 'Redraw'.

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 interface. At the top, there are tabs for 'Definition', 'Properties', 'Training Set', 'Literature', and 'Scheme'. Below these is a 'Category tree' section with '[1] Category 1' and two icons labeled '1' and '2'. On the right side, there is a vertical toolbar with buttons: 'ADD', 'DEL', 'AND', 'OR', 'NOT', 'Copy', 'Paste', and 'Redraw'. A callout box labeled '1' points to the 'ADD' button. Below this is the 'Query details' section with tabs for '[2] Reference Query' and 'Metabolism'. It contains two main panels: 'Profiling schemes' and 'Selected categories'. The 'Profiling schemes' panel lists various schemes, with 'Protein binding by OECD' highlighted in red. A callout box labeled '2' points to this item. The 'Selected categories' panel is currently empty. Below it is the 'Available categories' panel, which lists various categories. 'Schiff Base Formers >> Direct Acting Schiff Base Formers' is highlighted in red. A callout box labeled '3' points to this item.

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

Definition Properties Training Set Literature Scheme

[1] Category 1

Category tree

ADD
DEL
AND
OR
NOT
Copy
Paste
Redraw

3

[3] Reference Query Metabolism

Query details

Profiling schemes

- Biodegradation probability (Biowin 1)
- Biodegradation probability (Biowin 2)
- Biodegradation probability (Biowin 5)
- Biodegradation probability (Biowin 6)
- Biodegradation probability (Biowin 7)
- Biodegradation ultimate (Biowin 3)
- Biodegradation primary (Biowin 4)
- Hydrolysis half-life (Kb, pH 7) (Hydrowin)
- Hydrolysis half-life (Kb, pH 8)
- Hydrolysis half-life (Ka, pH 7)
- Hydrolysis half-life (Ka, pH 8)
- Protein binding potency
- Protein binding potency QAS (DPRA 13%)
- Protein binding by OECD**
- Protein binding by QSAR 1.4
- Toxic hazard classification by Cramer
- Toxic hazard classification by Cramer (extended)
- Estrogen Receptor Binding
- DNA binding by OECD

Selected categories

Michael addition

Available categories

(N/A)

- Acylation
- Acylation >> Direct Acylation Involving a Leaving group
- Acylation >> Direct Acylation Involving a Leaving group >> Acetates
- Acylation >> Direct Acylation Involving a Leaving group >> Acyl halides (including benzyl and carbamoyl deriv.)
- Acylation >> Direct Acylation Involving a Leaving group >> Anhydrides

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

Building of a new profiler

Grouping the referential queries

1. To select the three queries keep **Ctrl 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

The screenshot displays the QSAR Toolbox software interface. The main window is titled 'Category tree' and shows a hierarchical structure of queries. A new query is being added, indicated by a red dashed box around a query icon labeled '4'. The 'Add query' dialog box is open, showing the 'SMARTS' search method selected. The 'Contents' panel at the bottom left shows 'Queries' selected. Callout boxes 1 through 4 indicate the steps: 1. Clicking the 'ADD' button in the 'Category tree' window; 2. Clicking on 'Queries' in the 'Contents' panel; 3. The 'SMARTS' search method being selected in the 'Add query' dialog; 4. Clicking the 'OK' button in the 'Add query' dialog.

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, titled 'Category tree', shows a logical flow diagram with nodes labeled 'AND' and 'NOT', and a node labeled '4' circled in red. The bottom panel, titled 'Query details', shows a SMARTS query 'c1ccccc1' and a chemical structure of benzene. Callout boxes '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

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. At the top, there are tabs for 'Definition', 'Properties', 'Training Set', 'Literature', and 'Scheme'. The main window is titled 'Category tree' and shows a hierarchical structure of queries. A callout '1' points to a context menu with options: 'Data Query', 'Label Query', 'Parameter Query', 'QSAR Query', 'Reference Query', 'Similarity Query', and 'Structure Query'. Below this, the 'Query details' panel is visible, showing a list of parameters on the left and configuration fields on the right. A callout '2' points to 'log Kow' in the parameter list. A callout '3' points to the 'Expression' field, which is set to 'between' with values '0' and '0'. The 'Origin' field is set to 'scale: None' and the 'Destination' is 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

New Profiler (Custom) - Profiling Scheme Browser

Save Scheme | Export Scheme | Save Tests | View Tests | Run All Tests

Categories

Filter:

- New Profiler
 - Category 1

Definition | Properties | Training Set | Literature | Scheme

[1] Category 1

```

    graph TD
      N1((1)) --> AND1[AND]
      N2((2)) --> AND1
      N3((3)) --> NOT1[NOT]
      N4((4)) --> NOT2[NOT]
      N5((5)) --> NOT2
      AND1 --> AND2[AND]
      NOT1 --> AND2
      NOT2 --> AND2
    
```

Category tree

ADD
DEL
AND
OR
NOT
Copy
Paste
Redraw

[5] Logical Query | Metabolism

Query details

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

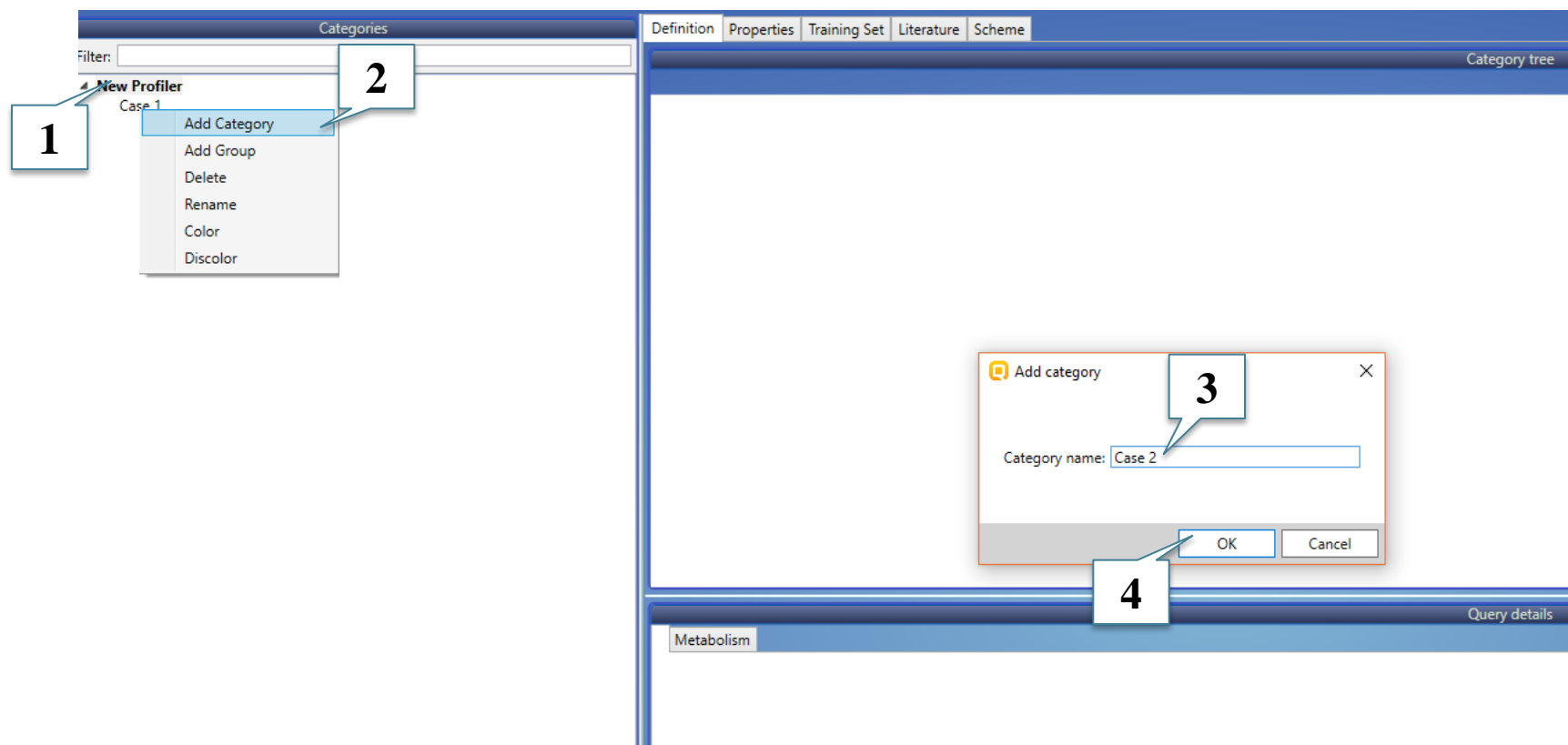
Rename a category

The screenshot shows the 'New Profiler (Custom) - Profiling Scheme Browser' interface. The 'Categories' pane on the left shows a tree structure with 'New Profiler' expanded to 'Category'. A context menu is open over 'Category', with 'Rename' selected. A dialog box titled 'Rename category' is open, with 'Case 1' entered in the 'Category name' field. The 'Query details' pane on the right shows a logical query tree with nodes labeled 'AND' and 'NOT'.

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



1. Right click on the **Profiler name**;
2. Select **Add Category** from the list;
3. Type the name **Case 2**;
4. Click **OK**.

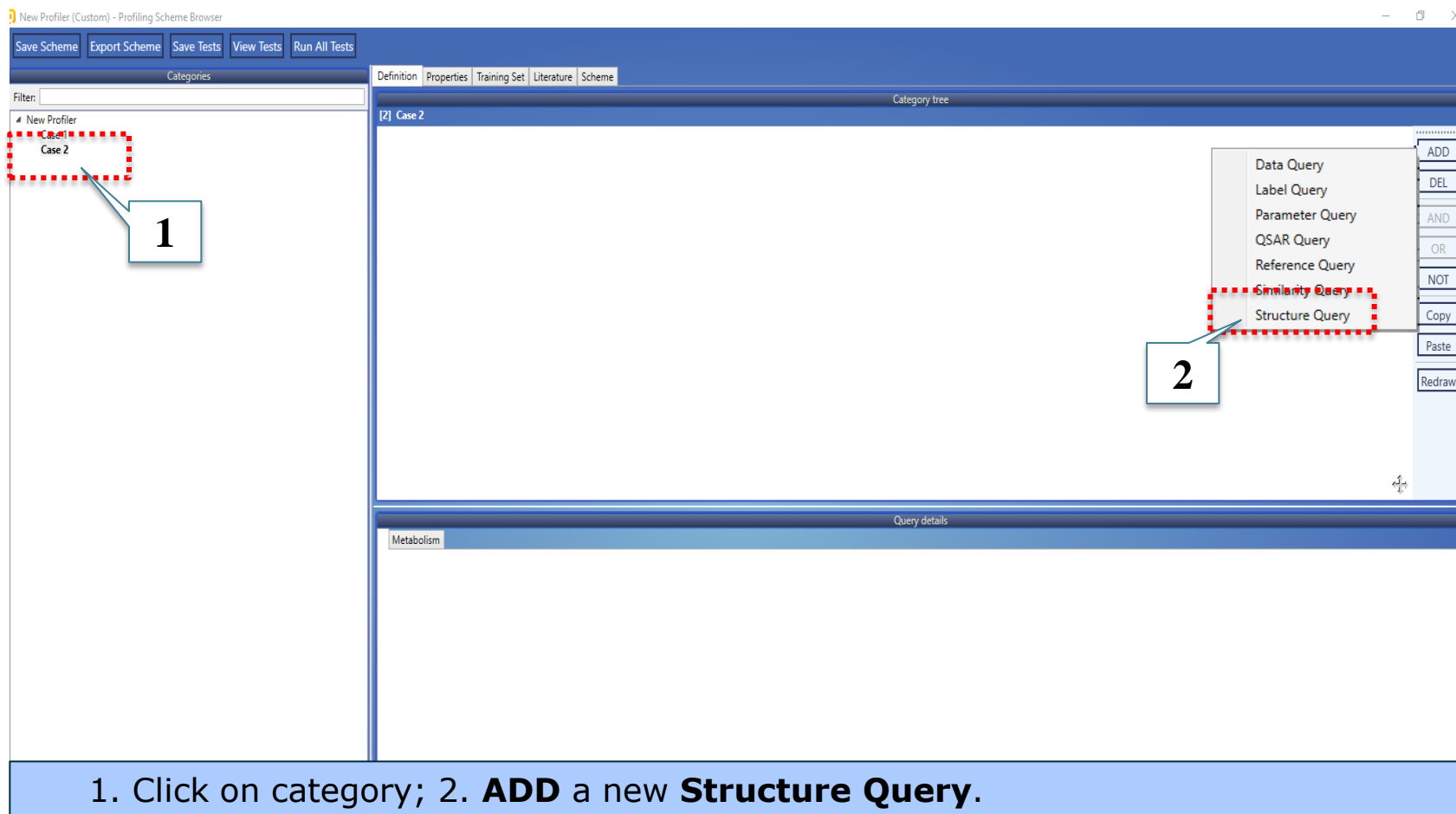
Building of a new profiler

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**.

Select any query or mask for additional details

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.

Building of a new profiler

Search by list of structures – Additional options

The screenshot shows the 'Query details' panel of the QSAR Toolbox. The 'Contents' list includes:

- Search 1: SMARTS
- Search 2: SMARTS
- Search 3: SMARTS
- Search 4: SMARTS
- Search 5: SMARTS
- Search 6: SMARTS
- Search 7: SMARTS

The 'Queries execution mode' is set to 'Any', which is highlighted with a red box and a callout bubble containing the number '1'. The 'SMARTS' field contains the string CC(O)=O. The 'View mode' is 'Facade' and the 'Navigation mode' is 'Cascade'. A chemical structure of acetic acid is displayed in the main window.

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 Tutorial 11 (http://oasis-lmc.org/media/74352/Tutorial_11_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 menu options: Save Scheme, Export Scheme, Save Tests, View Tests, and Run All Tests. Below this is a 'Categories' section with a filter input and a tree view containing 'New Profiler', 'Case 1', 'Case 2', and 'Case 3'. A red dashed box labeled '1' highlights the 'Case 3' icon. To the right, a context menu is open, listing query types: Data Query, Label Query, Parameter Query, QSAR Query, Reference Query, Similarity Query, and Structure Query. The 'Similarity Query' option is highlighted with a red box and labeled '2'. Below the main workspace is the 'Query details' panel, which is currently empty.

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. At the top, there are menu options: 'Save Scheme', 'Export Scheme', 'Save Tests', 'View Tests', and 'Run All Tests'. Below this is a 'Categories' panel on the left with a 'filter:' input field and a tree view containing 'New Profiler', 'Case 1', 'Case 2', and 'Case 3'. The main workspace is titled 'Definition' and shows '[3] Case 3' with a puzzle piece icon and the number '1'. To the right of the workspace are buttons for 'ADD', 'DEL', 'AND', 'OR', 'NOT', 'Copy', 'Paste', and 'Redraw'. Below the workspace is the 'Query details' section, which is currently set to 'Metabolism'. It includes a 'Similarity, %' section with an 'Expression' dropdown set to '=' and a value of '0'. A 'Target' section has a 'SMILES:' input field and an 'Edit' button. A red box highlights the 'Options' button in the 'Query details' section, with a callout box containing the number '1' pointing to it.

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

Building of a new profiler

Search by similarity query

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

- Average
- Combine

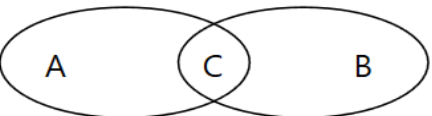
Atom character

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

Formula

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

Description

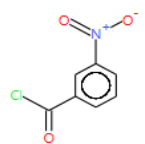


The atom-centered fragment is a topological sphere with center a selected atom and radius specified in **Any atom distance**. For aromatic carbon as a center of the sphere is assumed the aromatic system that contains this atom of concern.

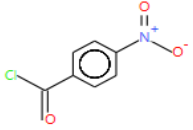
Structure

COC1cc(CC=C)ccc1O Define

Example



A	B	C
2	2	10

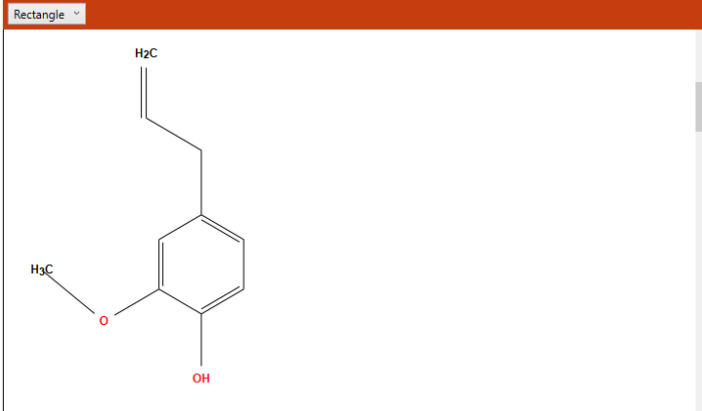


Similarity = 83.3333% Details

<->

2D Editor

Smiles COC1=CC(=CC=C1O)CC=C



OK Cancel

4

Default Help

OK Cancel

1. **Similarity options** set by default; 2. **Define** target structure by pasting SMART for eugenol 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 'Query details' window for 'Case 3'. The 'Similarity %' section is highlighted with a red box and a callout '1'. The 'Expression' dropdown is set to '>' and the 'Similarity %' input field contains '70'. The 'Target' field contains the SMILES string 'COc1cc(C=C)ccc1O'. Below the SMILES string, the chemical structure of Eugenol is shown, which is 4-allyl-2-methoxyphenol. The 'Options' section is also visible, showing 'Mode: Hologram, CombineAllFeatures' and various molecular and atom characteristics.

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, Scheme in order to enrich information supporting queries.

Building of a new profiler

Additional functionalities: Properties section

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 field 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 paste SMILES for YES and NO examples.

Building of a new profiler

Additional functionalities: Literature

Categories

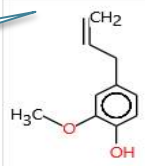
Filter:

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

Visualization of the Literature (.htm) file

Definition Properties Training Set Literature Scheme

Eugenol structural similarity



Similarity options

Similarity options

Measure

- Tanimoto (Jaccard)
- Dice
- Kulczynski-2
- Ochiai(Cosine)
- Vule

Molecular features

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

Options

Calculation

- Fingerprint
- Hologram

Average by features

- Average by features
- Combine all features

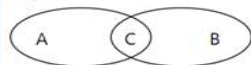
Atom characteristics

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

Formula

$$\frac{c^2}{0.5[(a+b) + (b+c)]}$$

Description



Description: The atom-centered fragment is a topological sphere with center a selected atom and radius specified in **Any atom distance**. For aromatic carbon as a center of the sphere is assumed the aromatic system that contains this atom of concern.

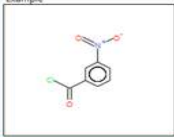
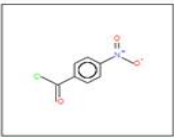
Structure

COc1ccc(C=C)cc1O Define

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 'Training Set' tab in the QSAR Toolbox software. The 'Load' button is highlighted with a red box and a callout '1'. An 'Open' file dialog is overlaid on the software, showing the 'Downloads' folder. The file 'Vinyl pyridines.smi' is selected, highlighted with a red box and a callout '2'. The 'Open' button in the dialog is highlighted with a red box and a callout '3'. To the right of the dialog is the chemical structure of 1-vinylpyridine, which consists of a pyridine ring with a vinyl group (-CH=CH₂) attached to the 1-position.

CAS#	Name	Smiles	Representative h-CLAT DC56/DC84
1337-81-1	1-Vinyl pyridine	C=C1cccn1	Positive

h-CLAT DC86 = Positive (Dendritic cells COLIPA)
h-CLAT DC54 = Positive (Dendritic cells COLIPA)

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

Literature (.htm) file should be collected in a folder and placed in the following directory *C:\Program Files (x86)\Common Files\QSAR Toolbox 4.1\Config\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 dialog box with the message 'Scheme successfully saved on the server.' and an 'OK' button is shown, with the 'OK' button labeled '2'. The 'Close' button (X) of the dialog is labeled '3'. The main window shows a tree view with 'New Profiler', 'Case 1', 'Case 2', and 'Case 3'. The 'Definition' tab is active, showing 'Case 3'. The 'Query details' panel shows a similarity query for 'Metabolism' with a similarity percentage of 70. The 'Target' panel shows the SMILES string 'COc1cc(CC=C)ccc1O' and 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

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, there are buttons for Apply, View, New, and Delete. The left sidebar shows a list of documents and profiling methods. The 'Filter endpoint tree...' panel is visible, and the main results table shows chemical structures and case labels. A red box highlights the 'New Profiler' label in the table, and a red arrow points to the 'New Profiler' checkbox in the 'Profiling methods' list.

	1	2	3	4	5	6
Structure						
Structure info						
Parameters						
Physical Chemical Properties						
Environmental Fate and Transport						
Ecotoxicological Information						
Human Health Hazards						
Profile						
New Profiler	Case 2	Case 2	Case 2 Case 3	Case 2	Case 2	Case 2

Building of a new profiler

Create a new dendroid profiler

How to built 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.