QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

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

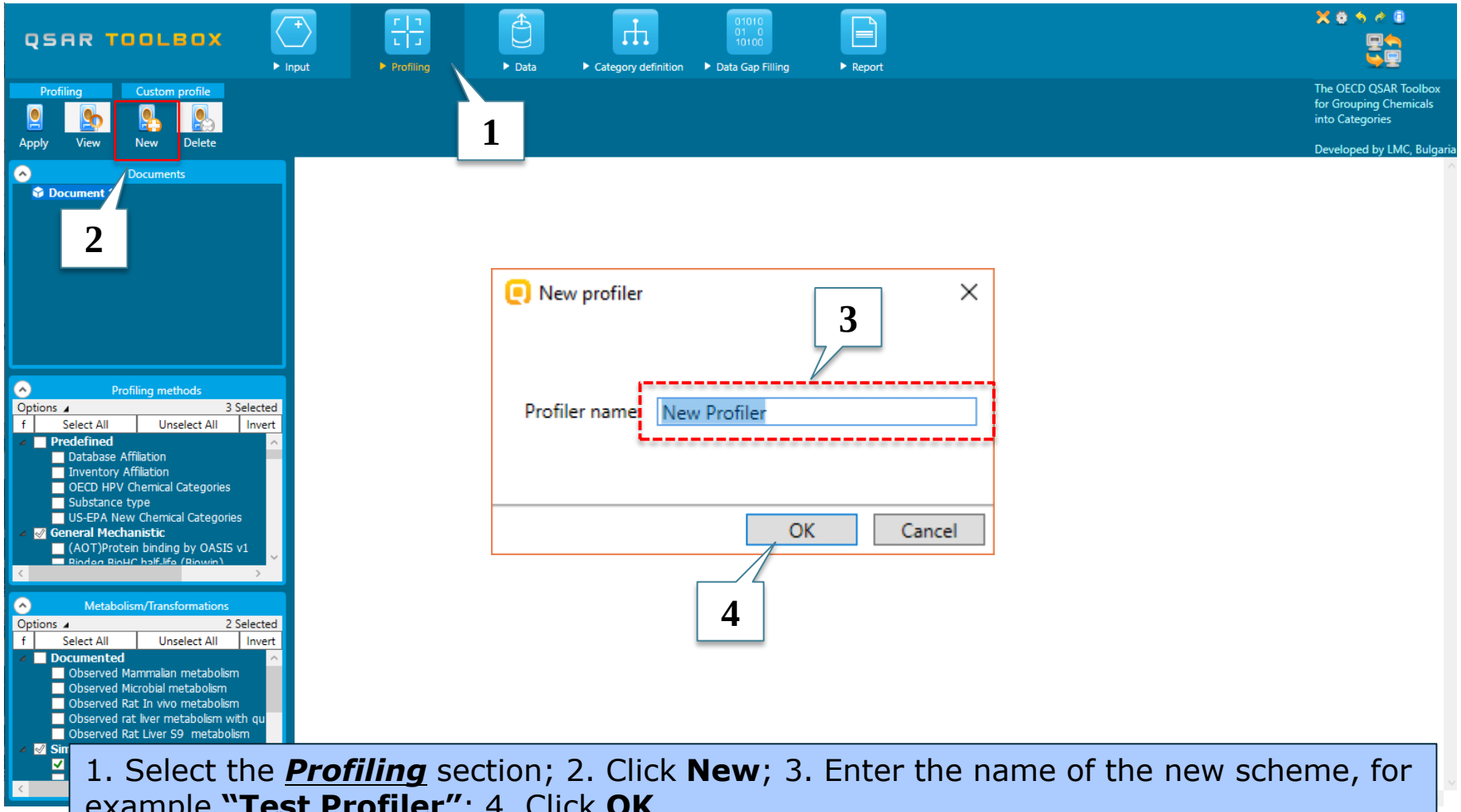
We are going to create a new profiling scheme:

- Open the Toolbox.
- Move to the Profiling module

(see next screen shot).

Building of a new profiler

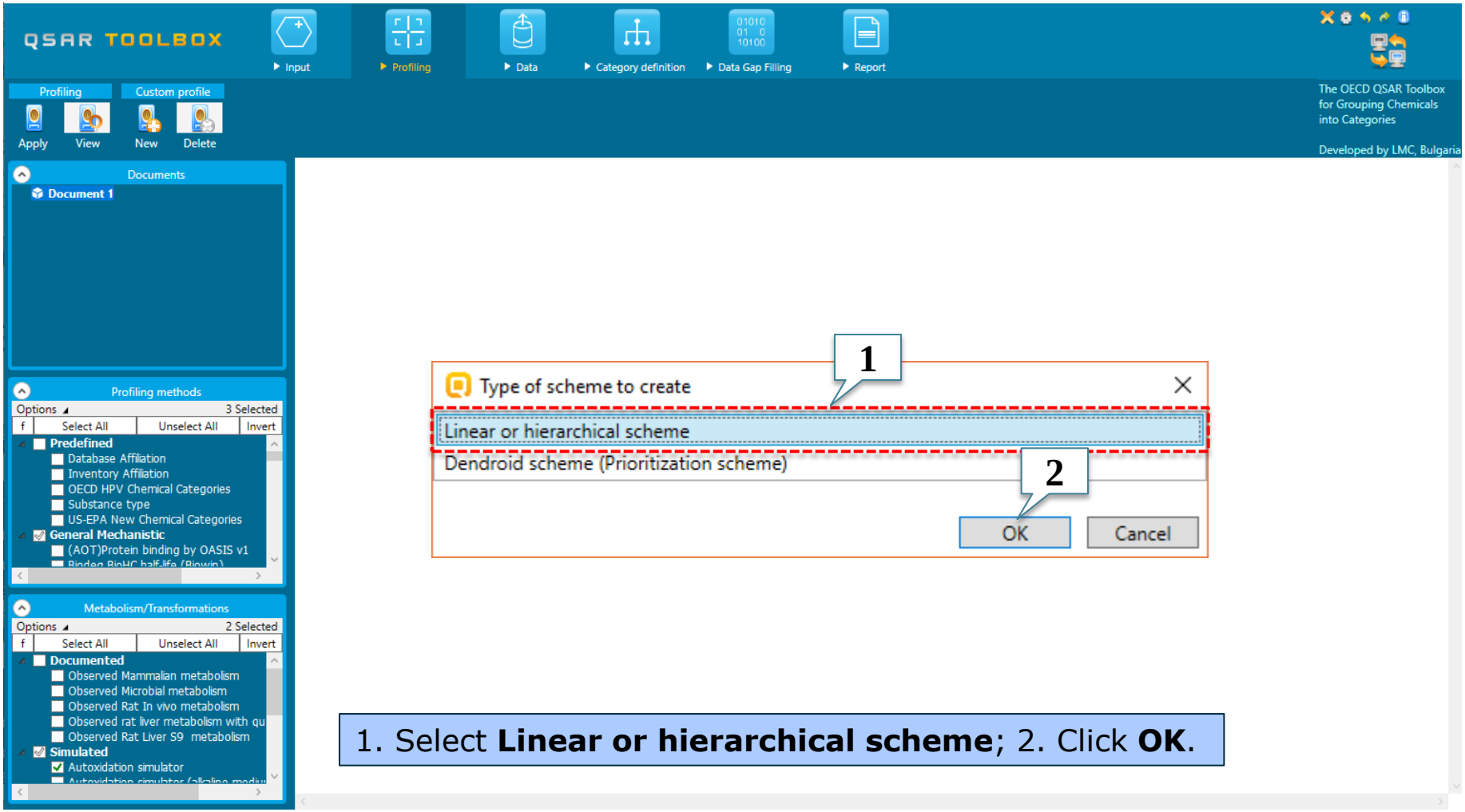
Define the name of the new profiler



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



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

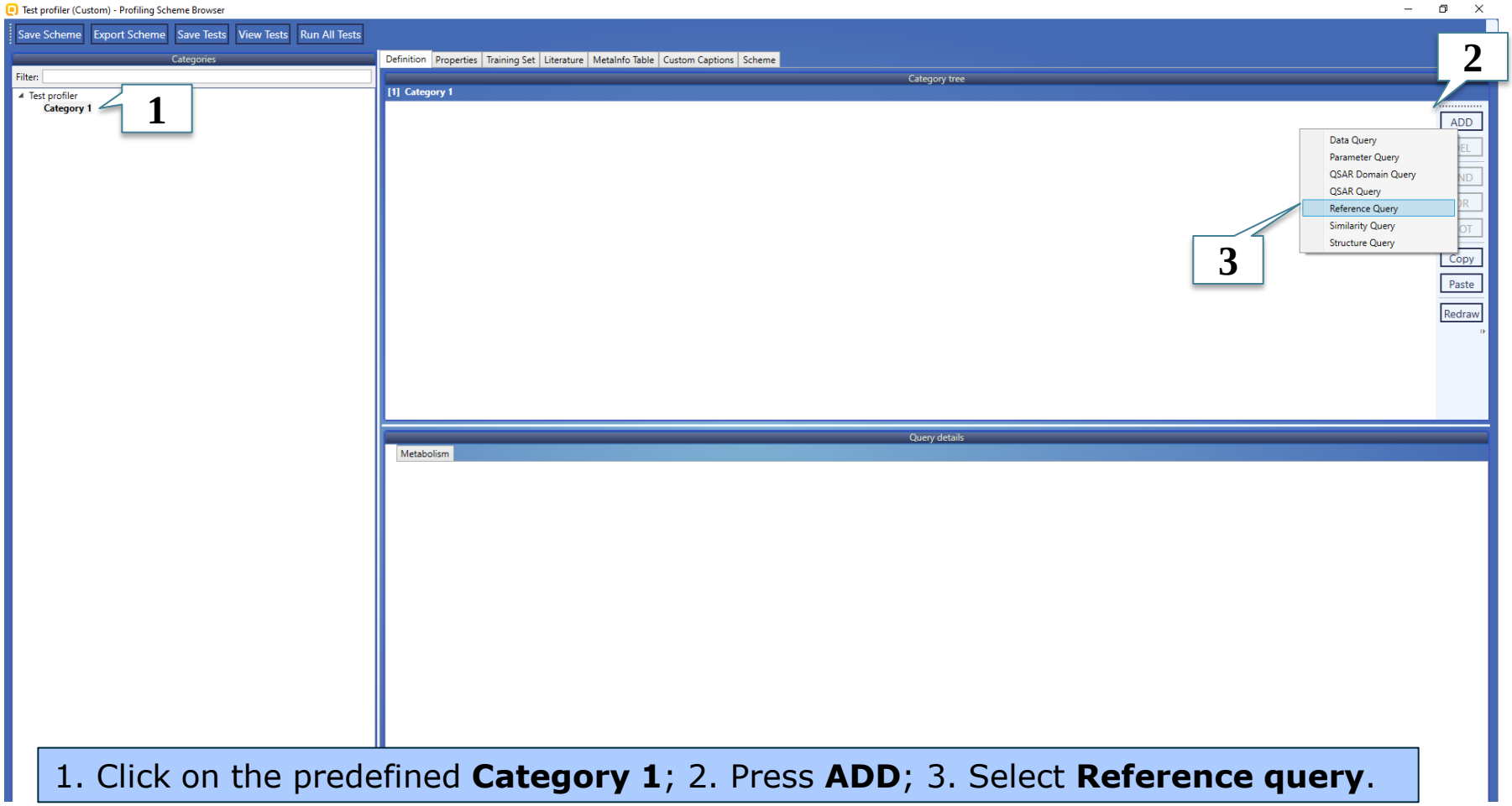
Building of a new profiler

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

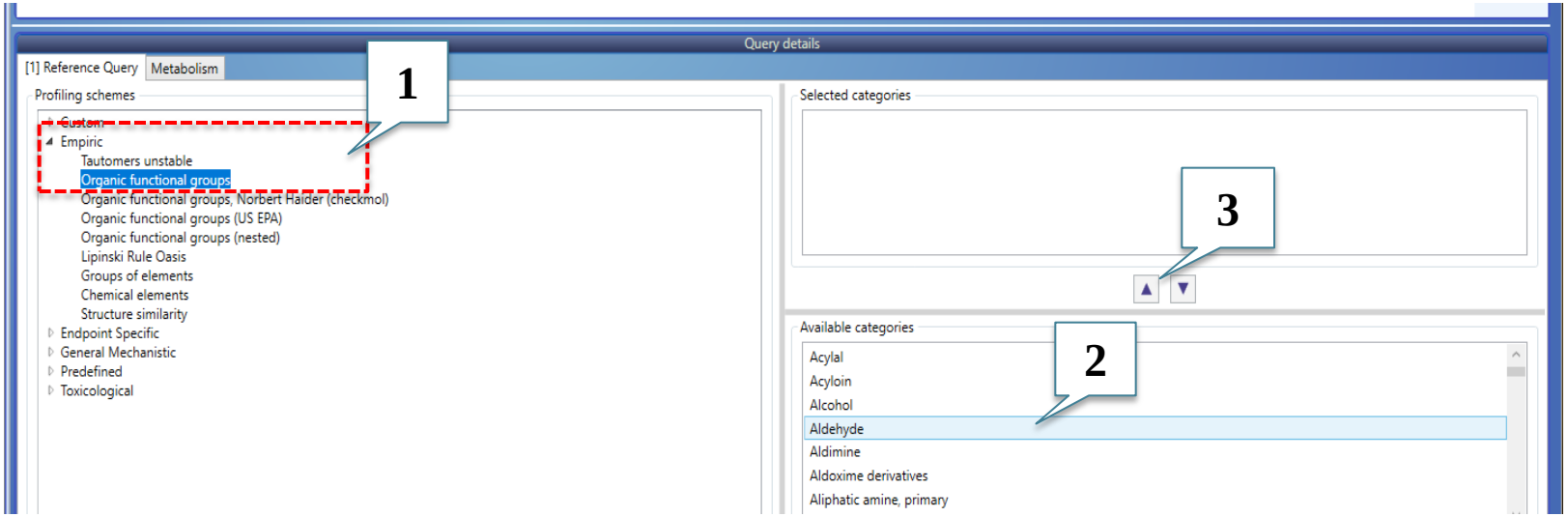
Building the category definition – rule 1



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

Building of a new profiler

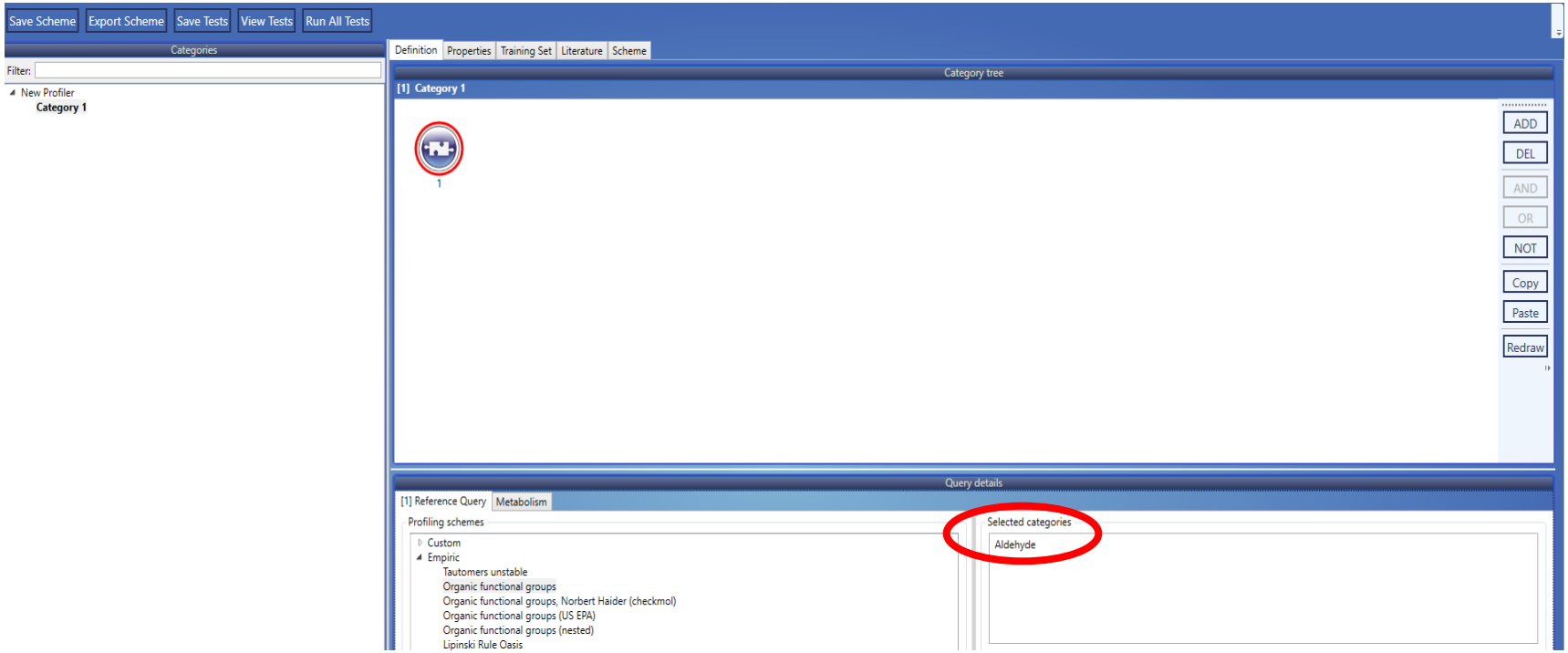
Building the category definition – rule 1



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



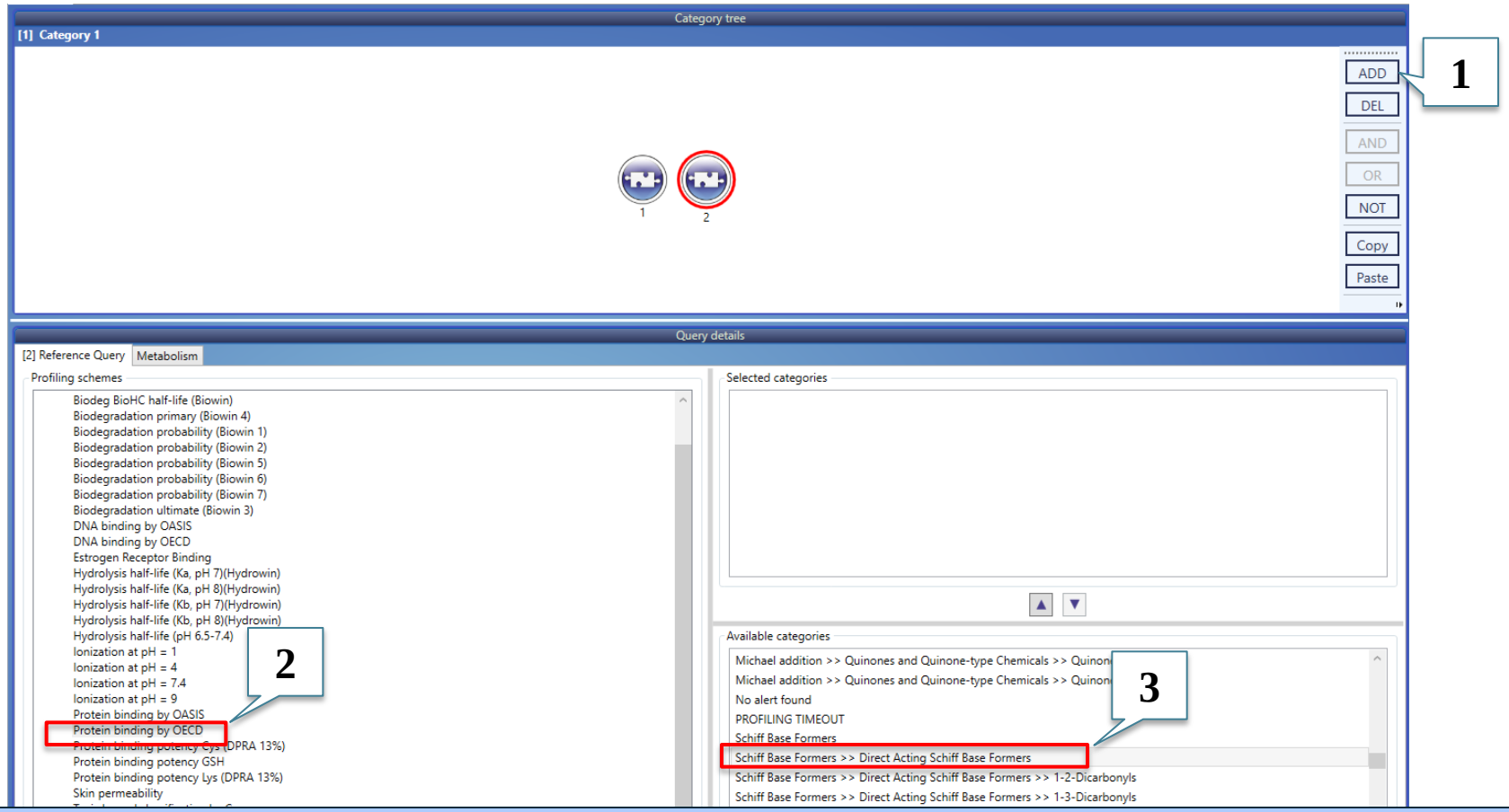
Building of a new profiler
Restriction of the category definition

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

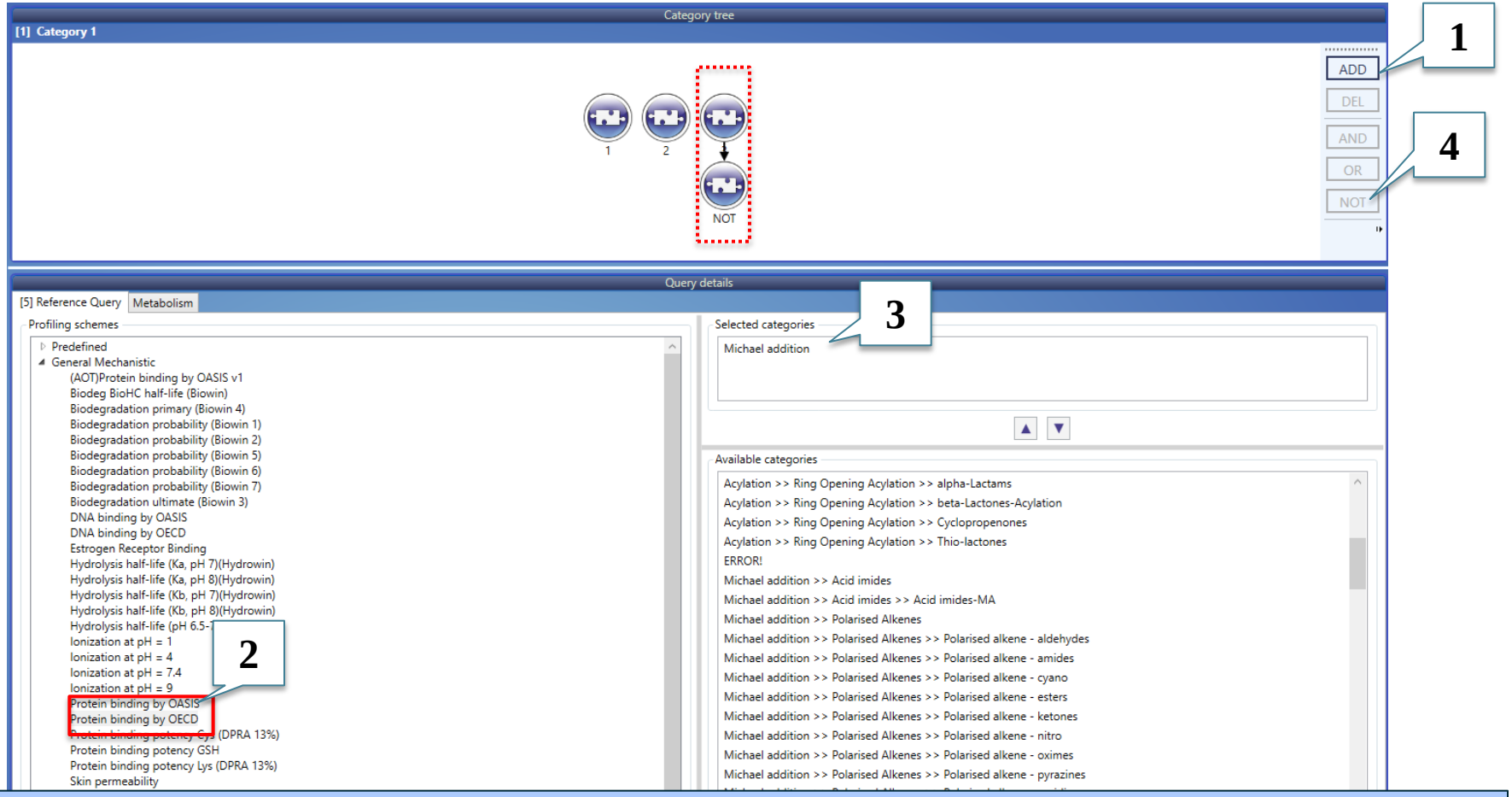
Building the category definition – rule 2



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

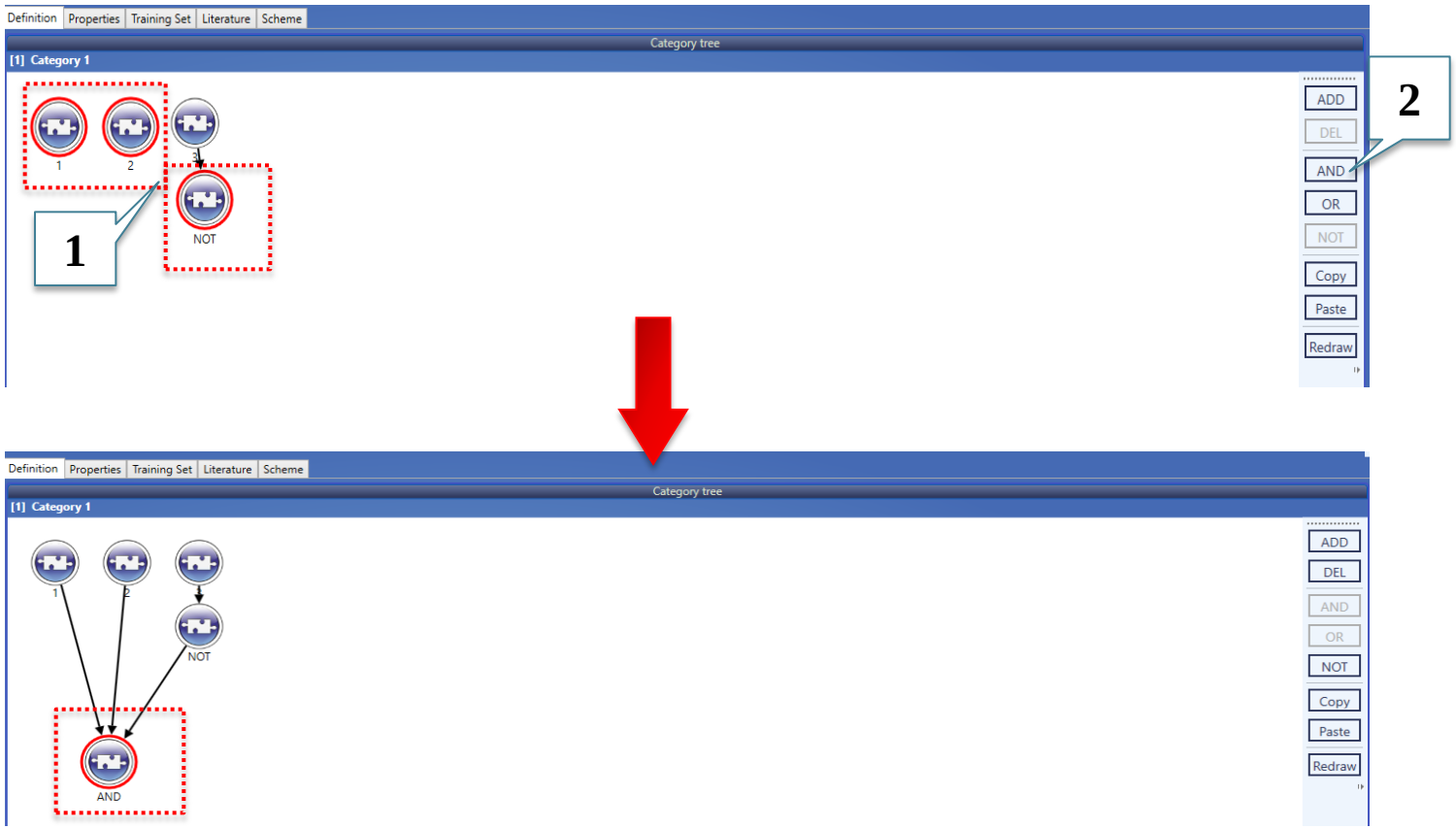
Building the category definition – rule 3



1. **Add** 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



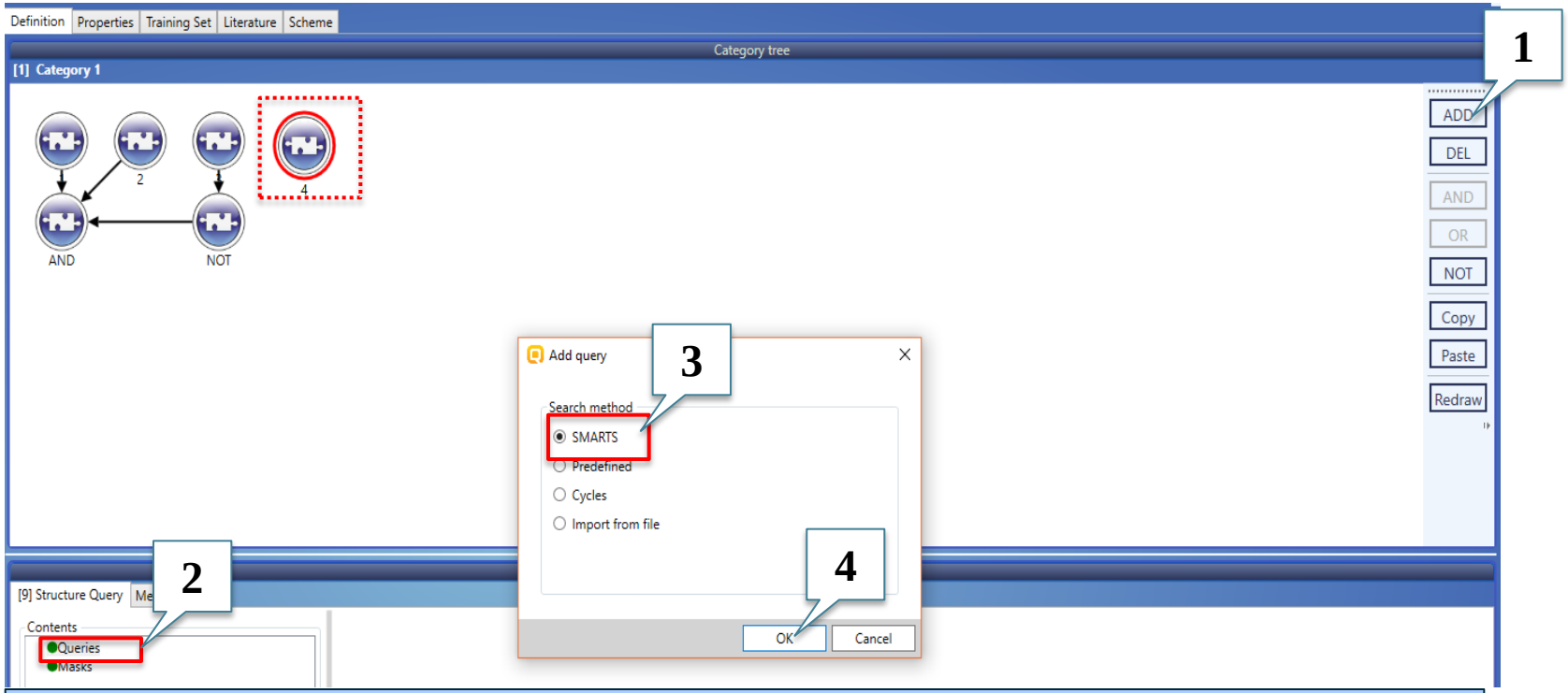
1. To select the three quires keep **Crtl button pressed** and click on each **query** to get **RED** circle; 2. Click **AND.**

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

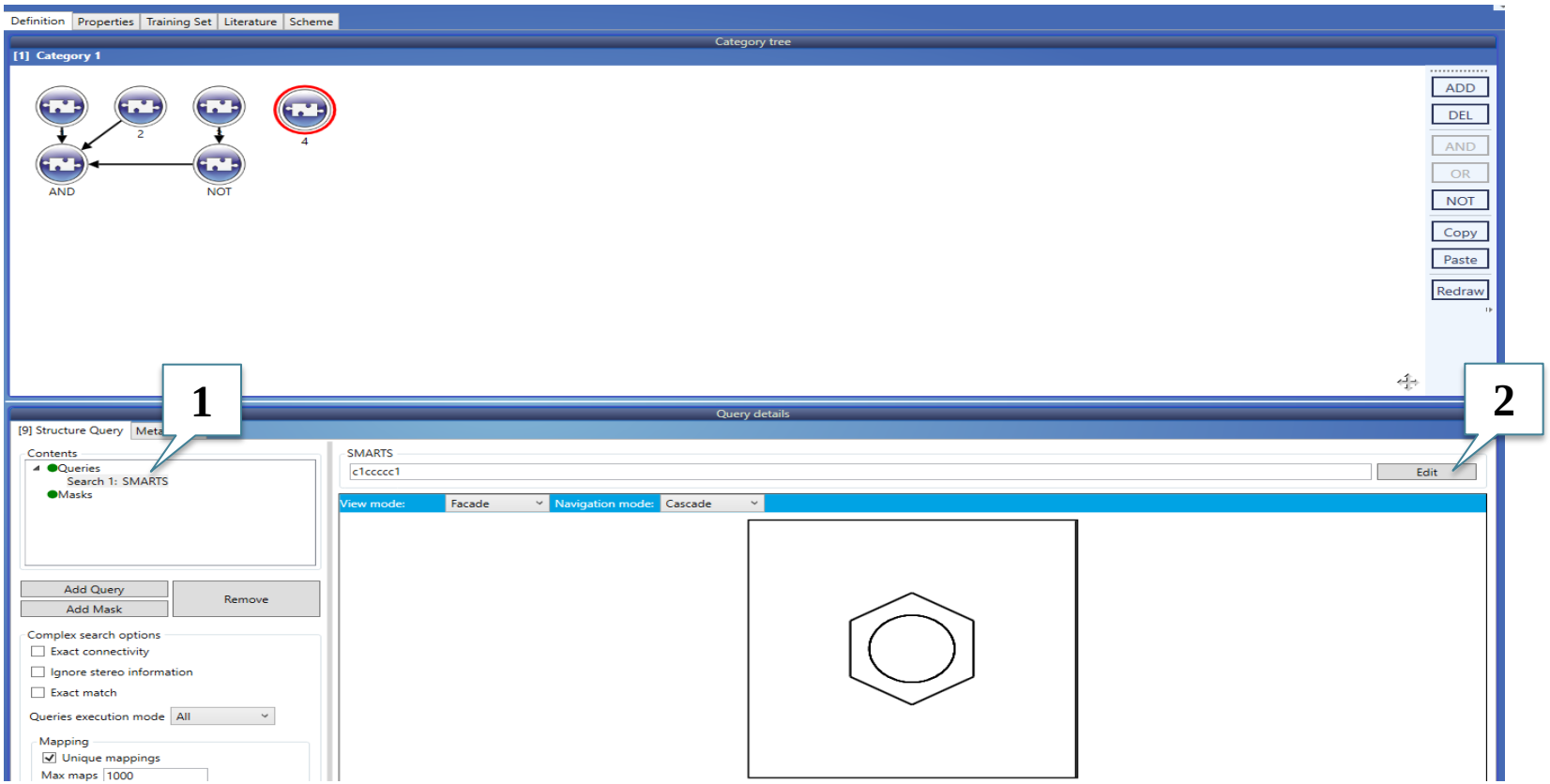
Building the category definition – rule 4



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

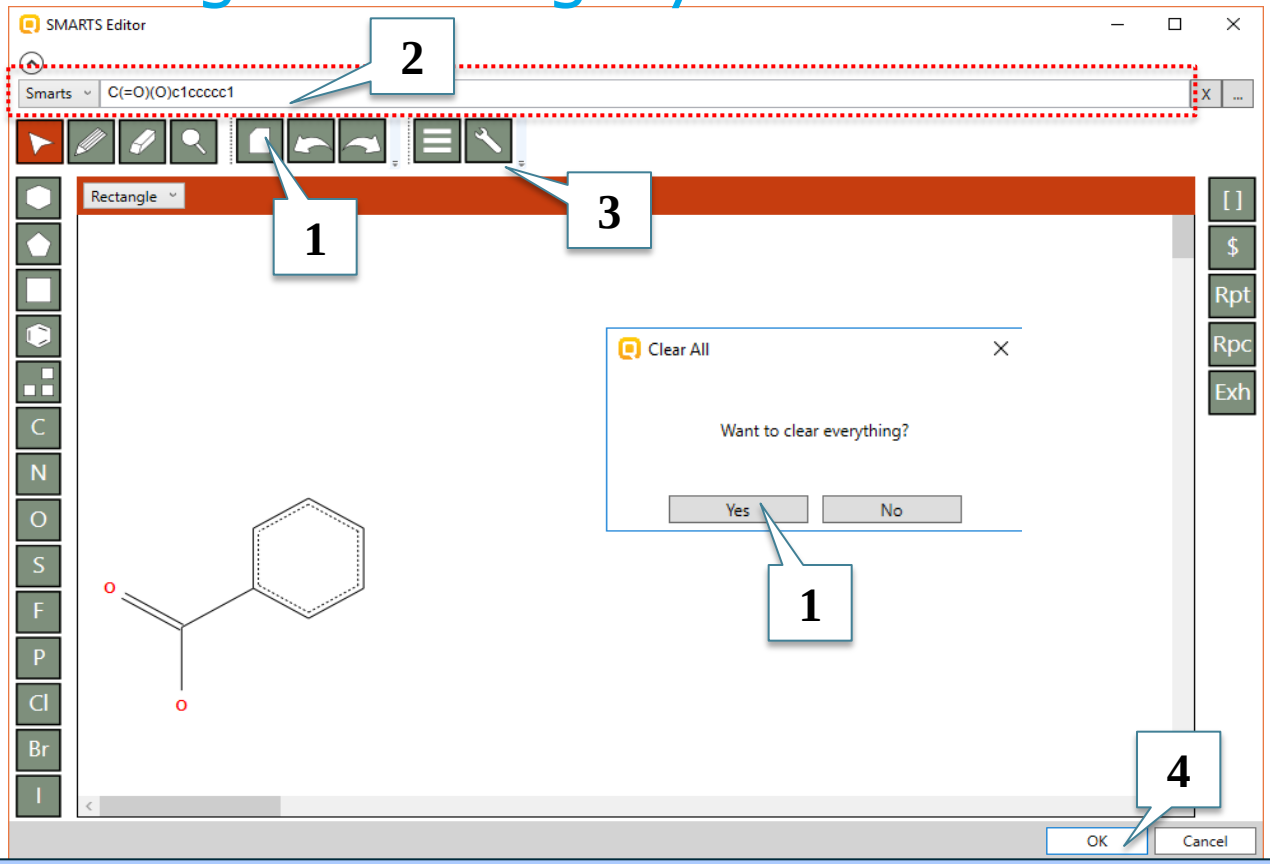
Building the category definition – rule 4



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

Building of a new profiler

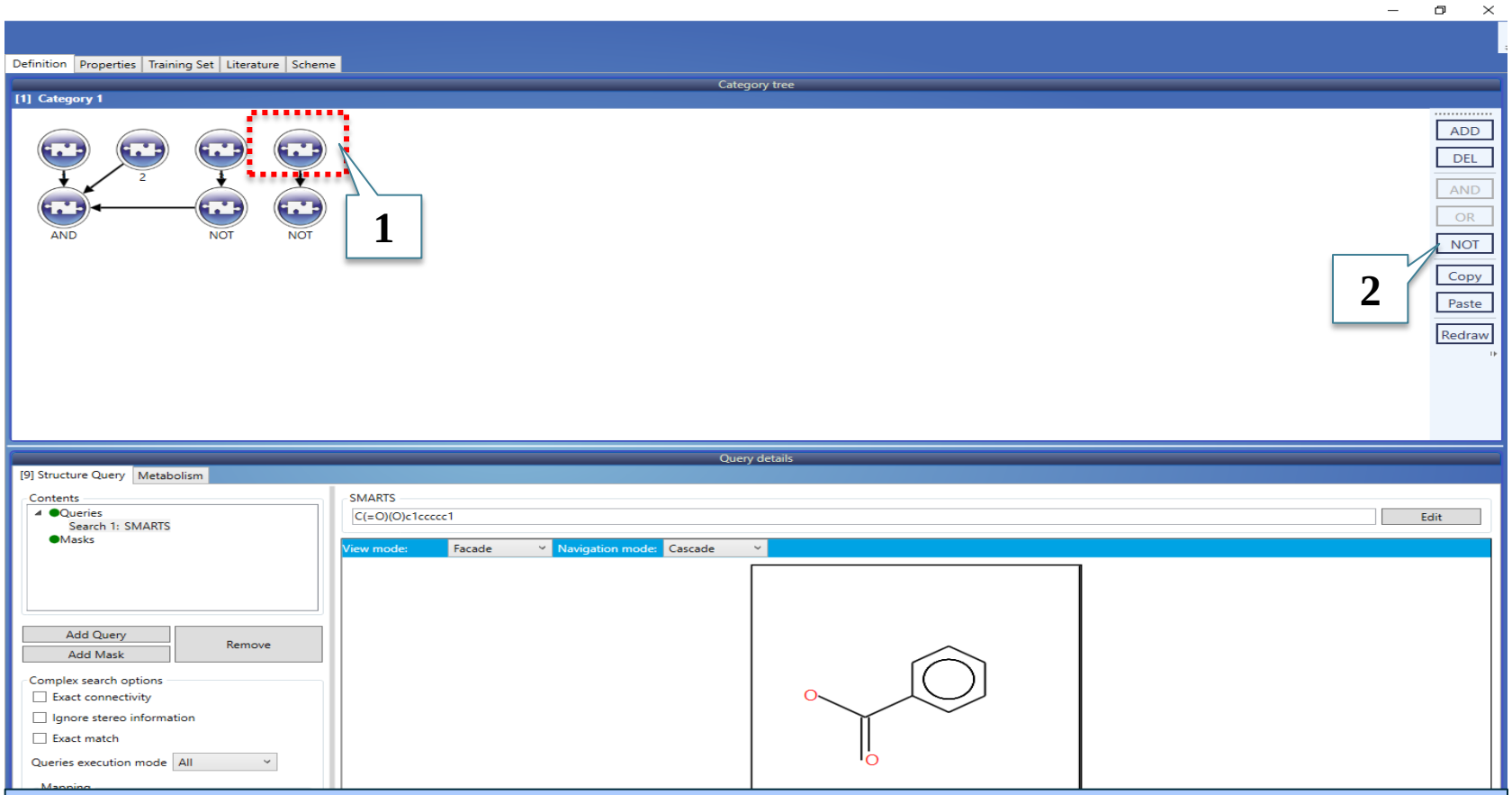
Building the category definition – rule 4



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



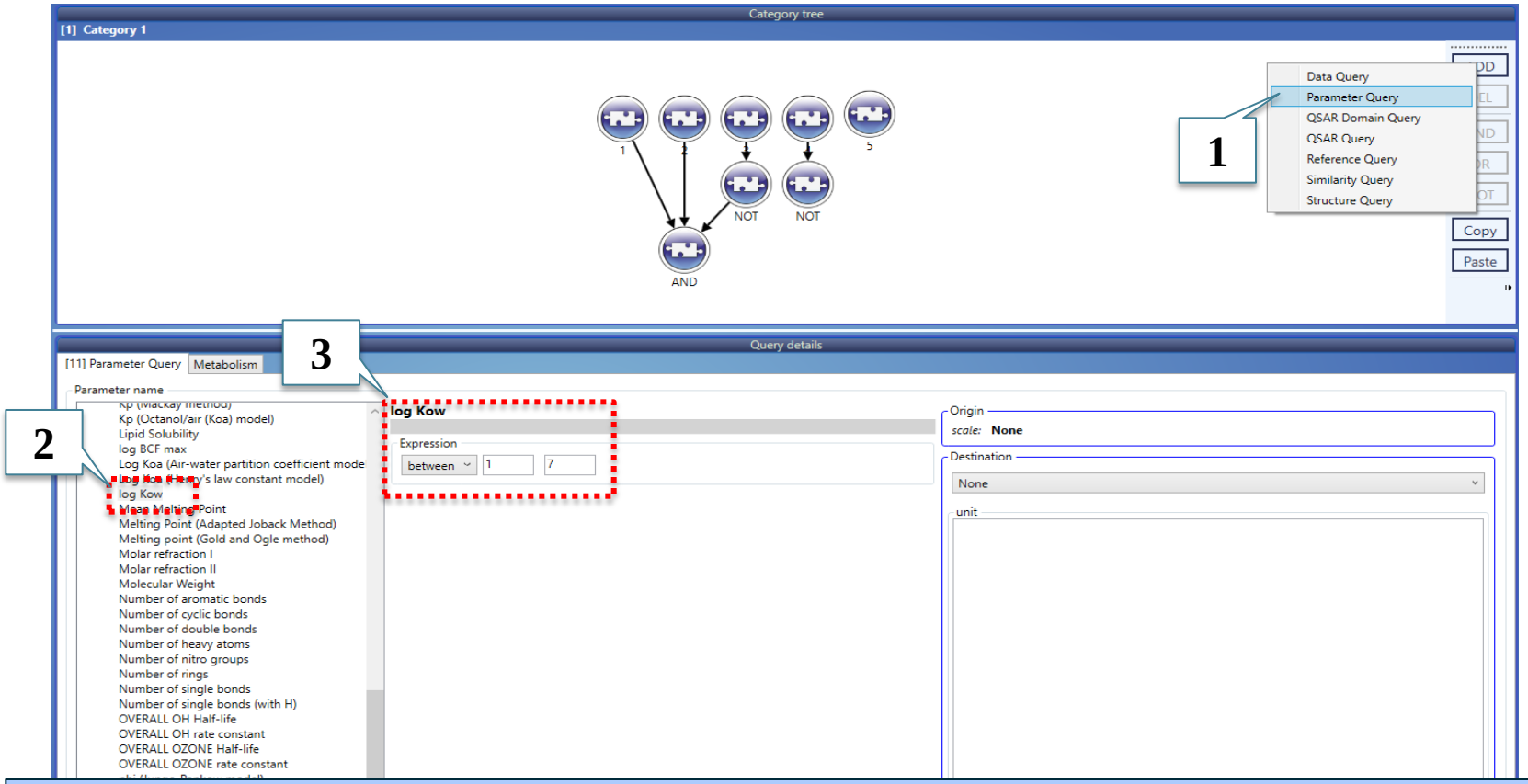
1. Click over the structural query in order to select it (it is **RED**); then 2. Click **NOT** to negate the query.

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

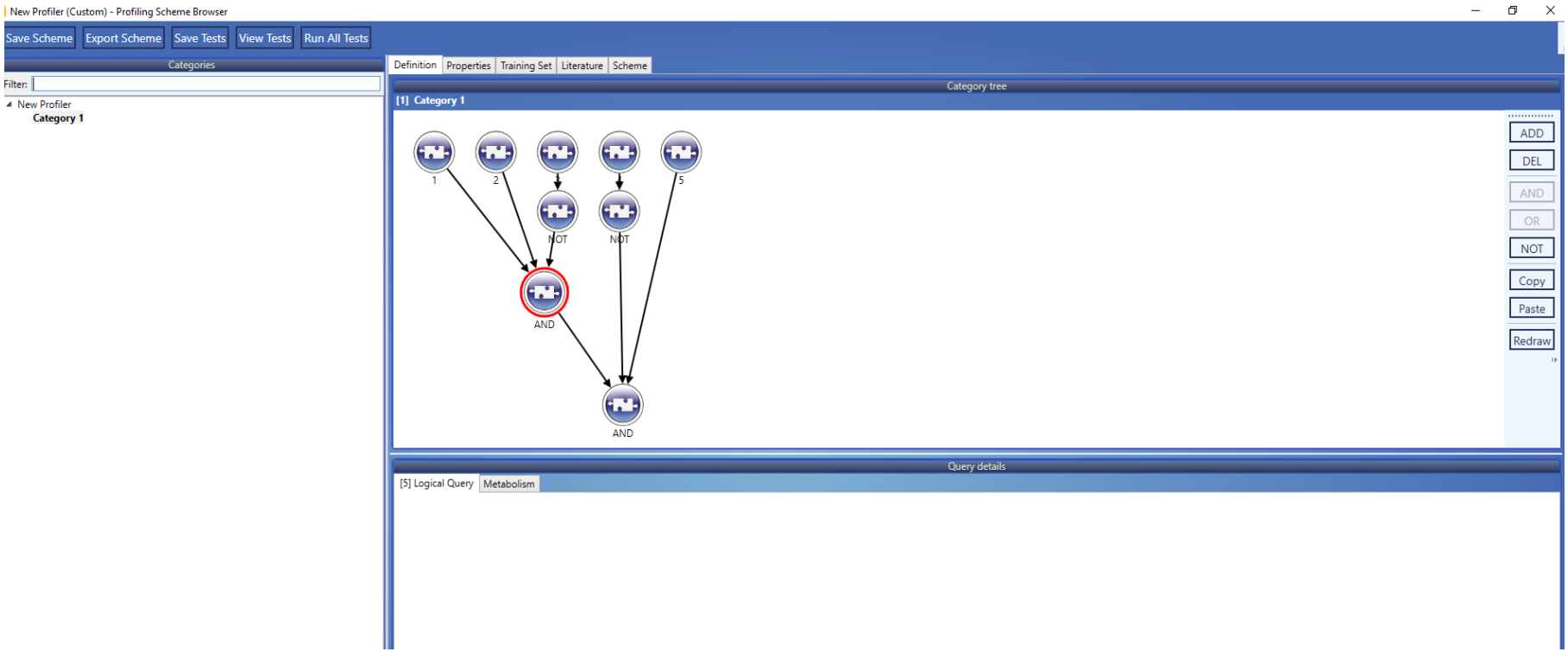


1. ADD new **parametric query**; **2.** From the list with 2D parameters select **log Kow**; 3. Select qualifier **between** and enter the requisite values **(1 to 7).**

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 quieries



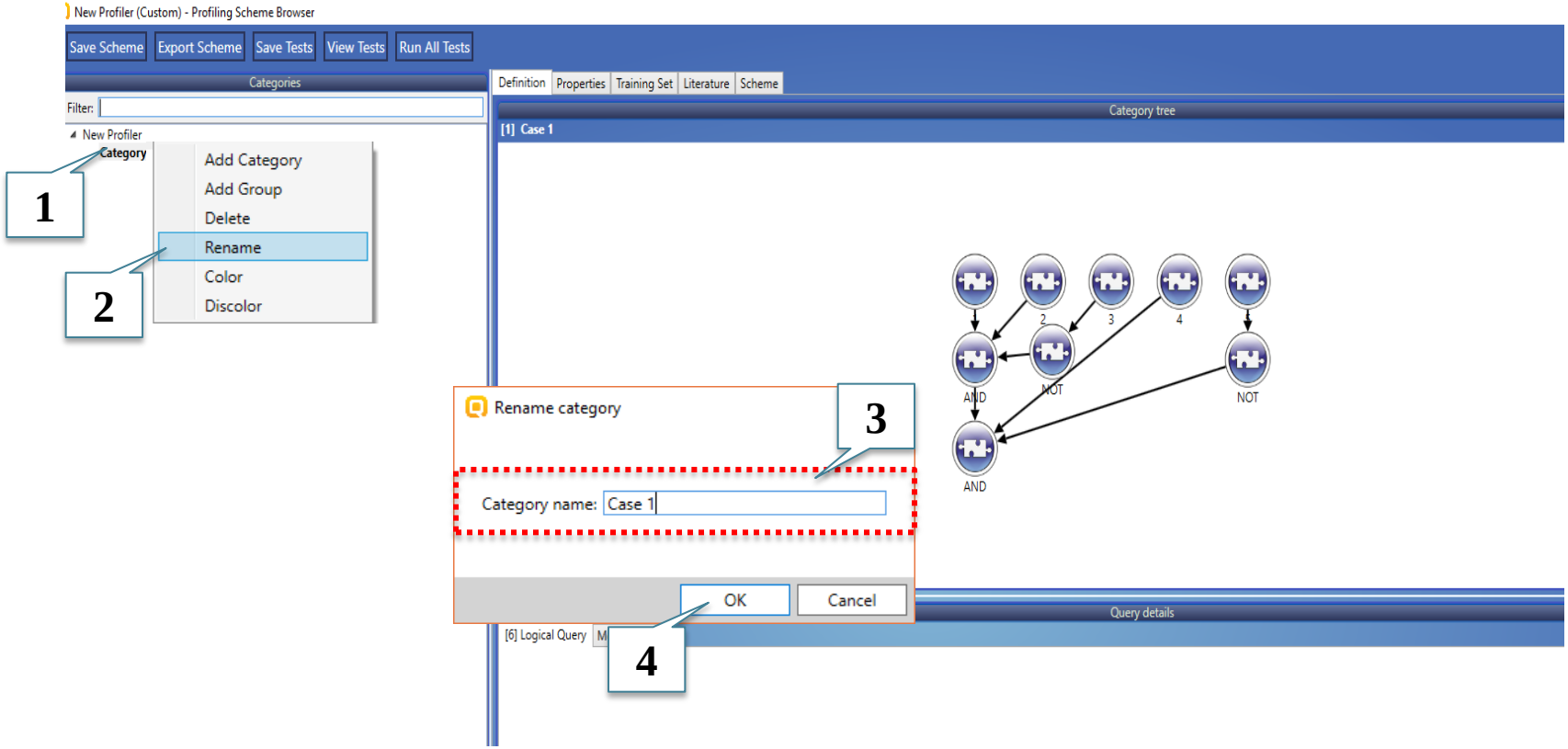
Building of a new profiler

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**.

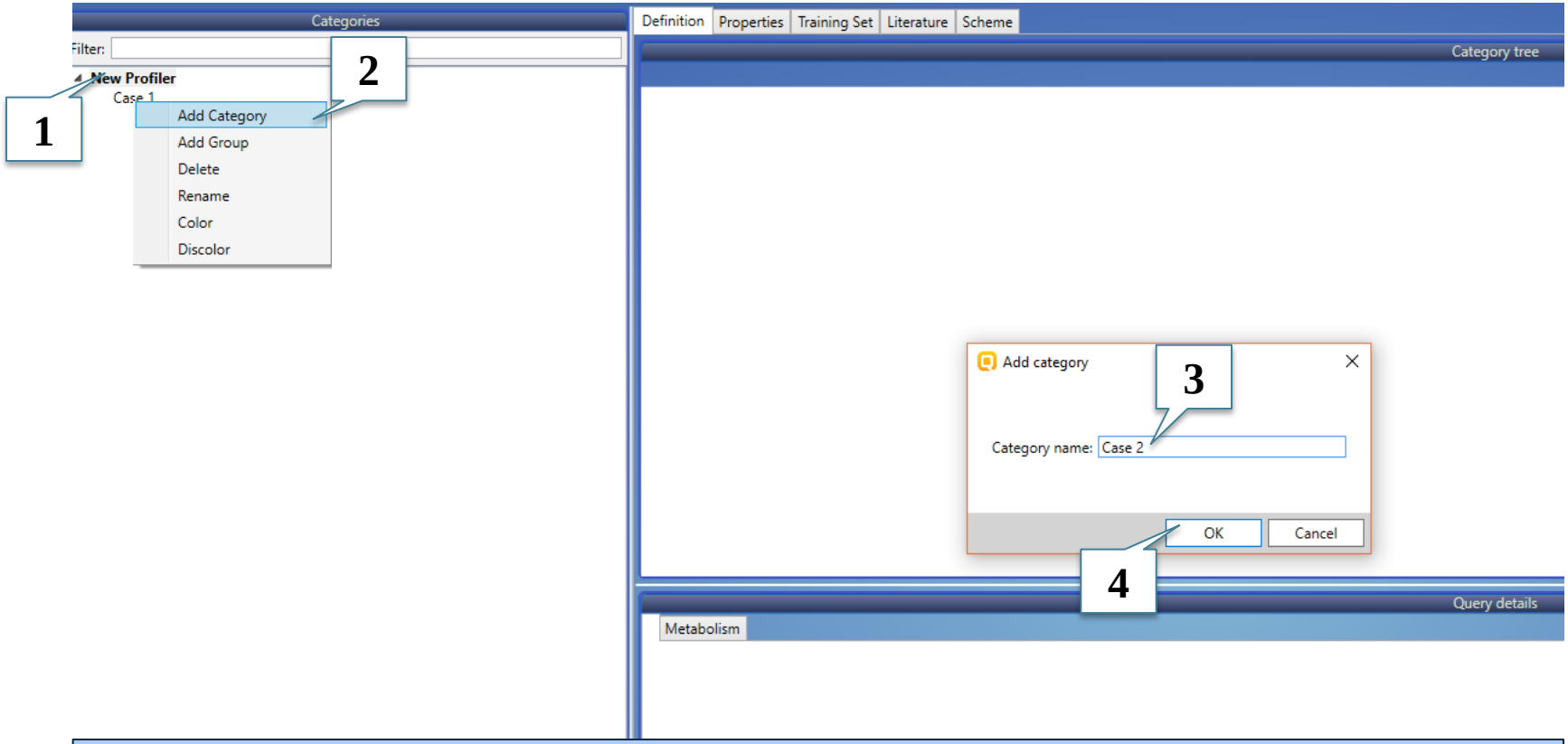
Building of a new profiler

Rename a category



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



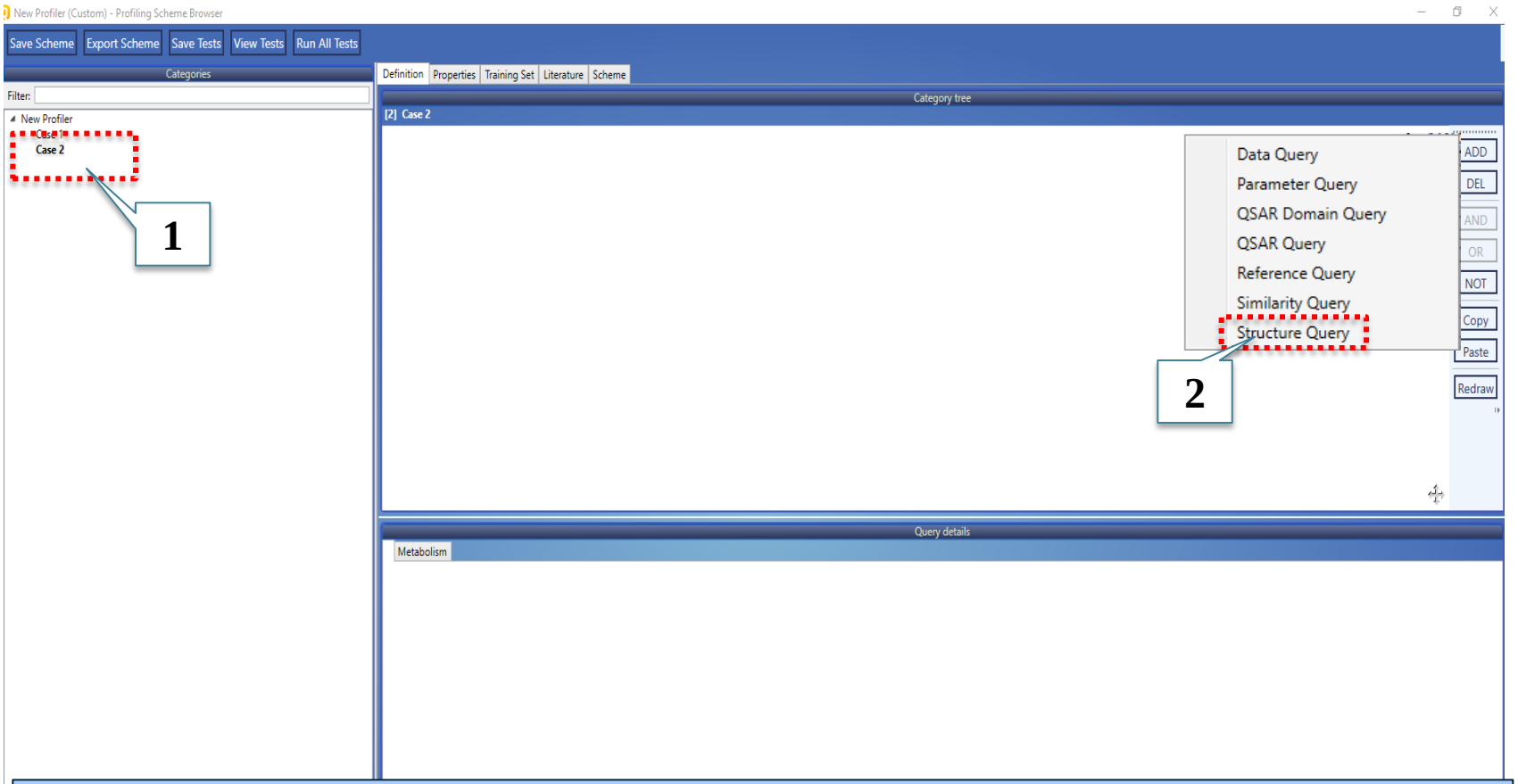
1. Right click on the **Profiler name**; 2. Select **Add Category** from the list; 3. Type the name **Case 2**; 4. Click **OK**.

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

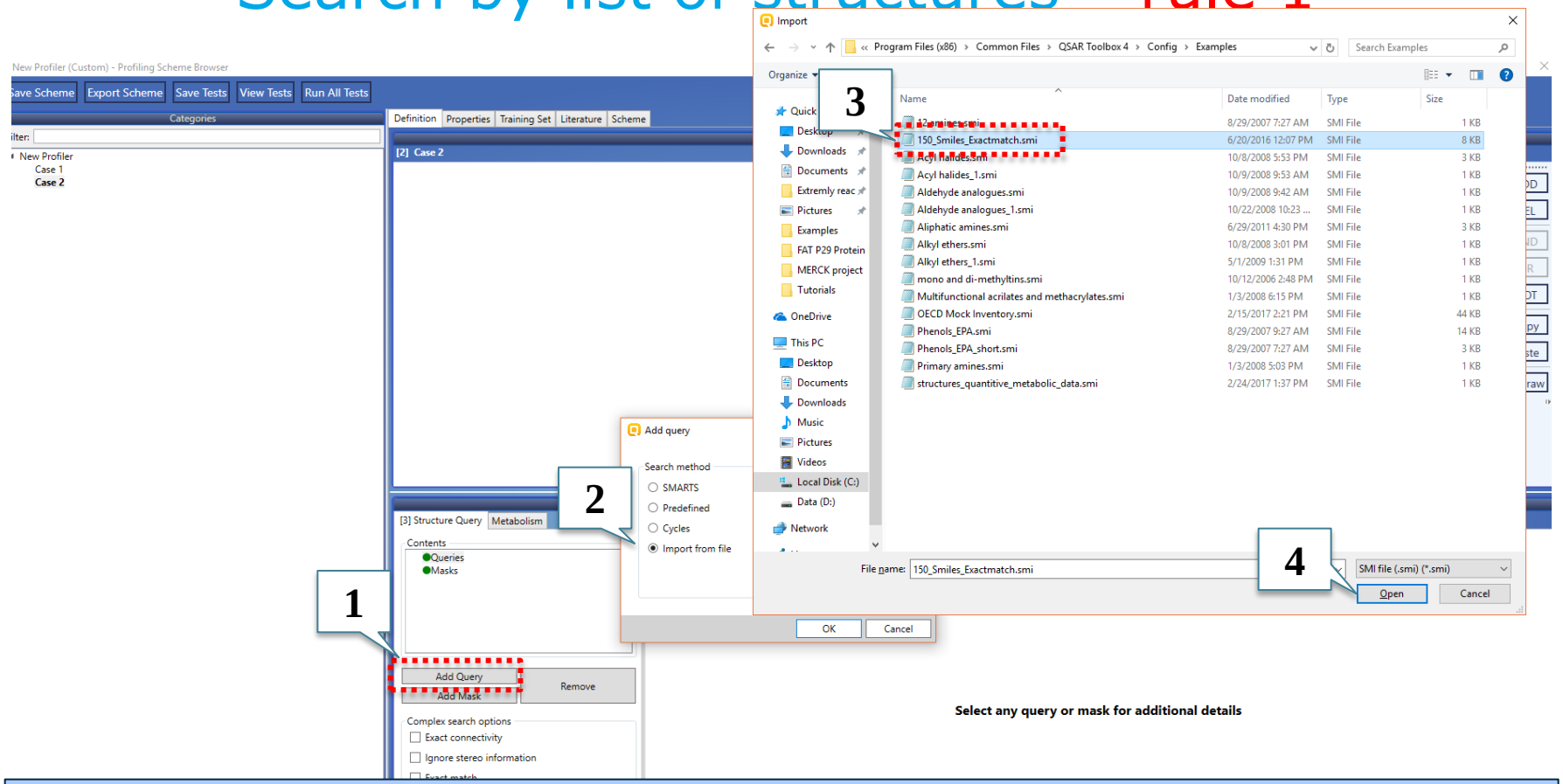
Search by list of structures – rule 1



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

Building of a new profiler

Search by list of structures – rule 1

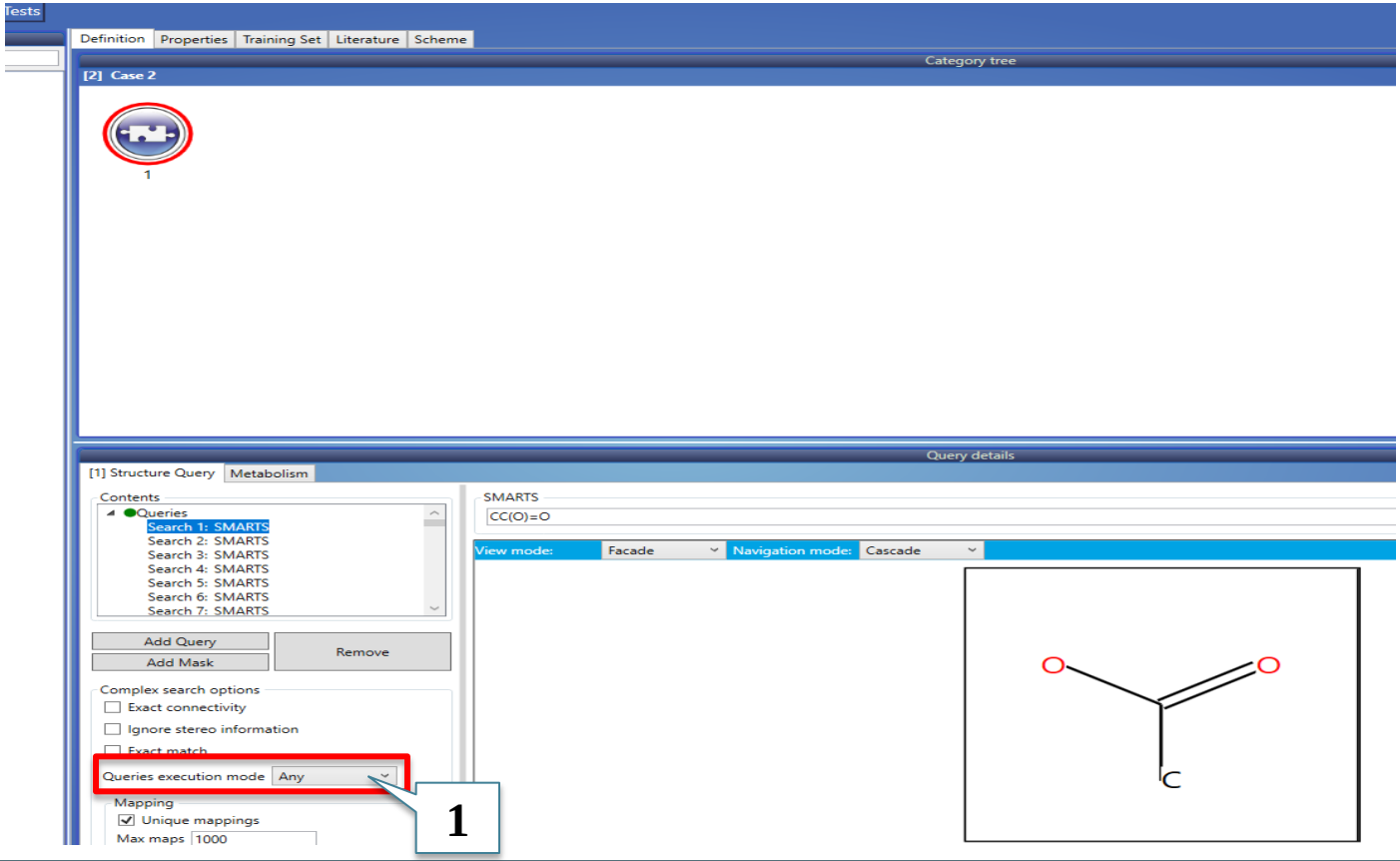


1. Click **Add query**; 2. Select **Import from file**; 3. Load to add a pre-defined **smi** file which contains SMART*; 4. Click **Open**.

* 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

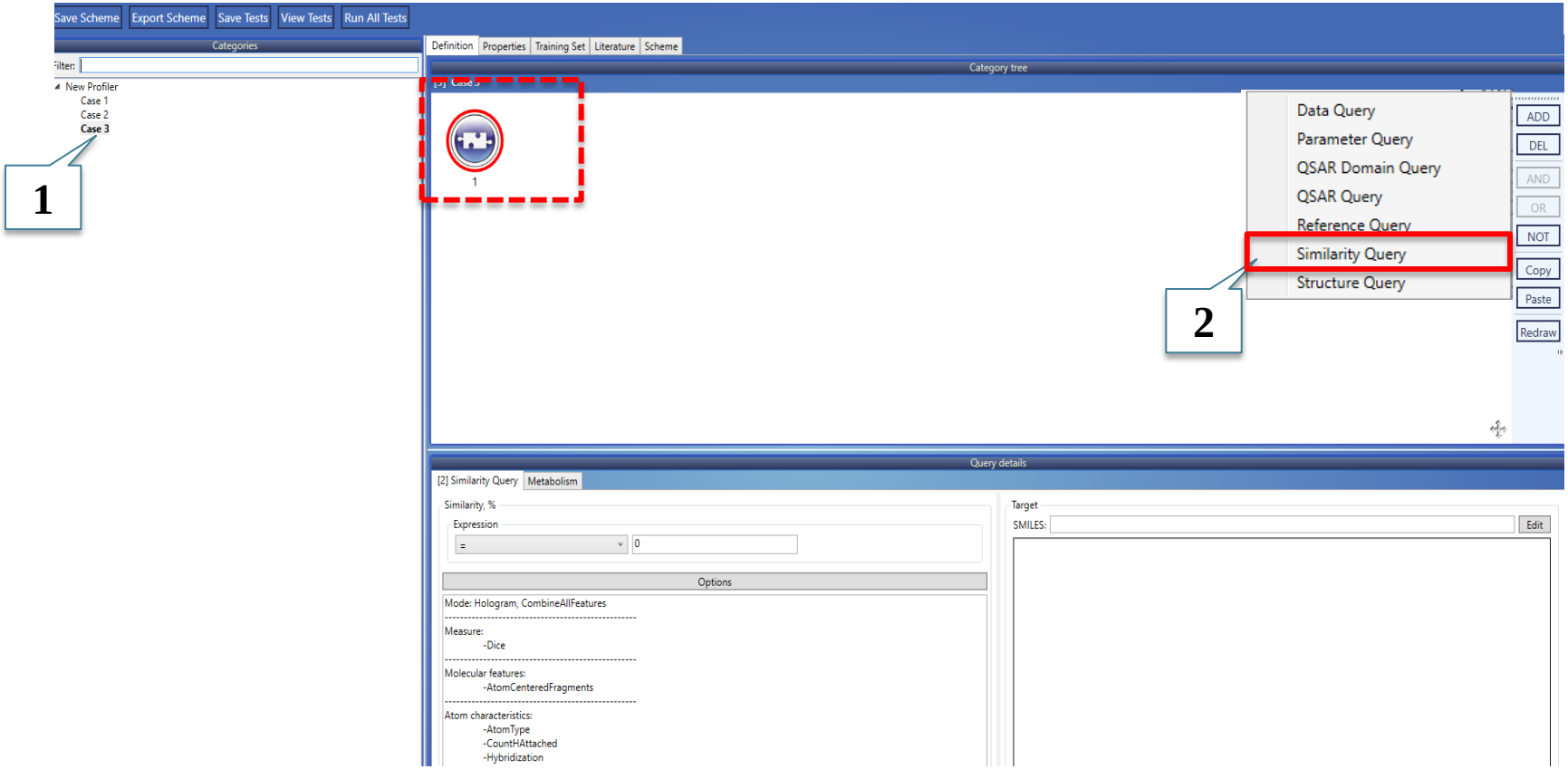


1. Select **Any** from Query execution mode – it means that the SMART in the list are OR-ed and it is enough the target chemical to match at least one of them.

Building of a new profiler
Search by similarity query

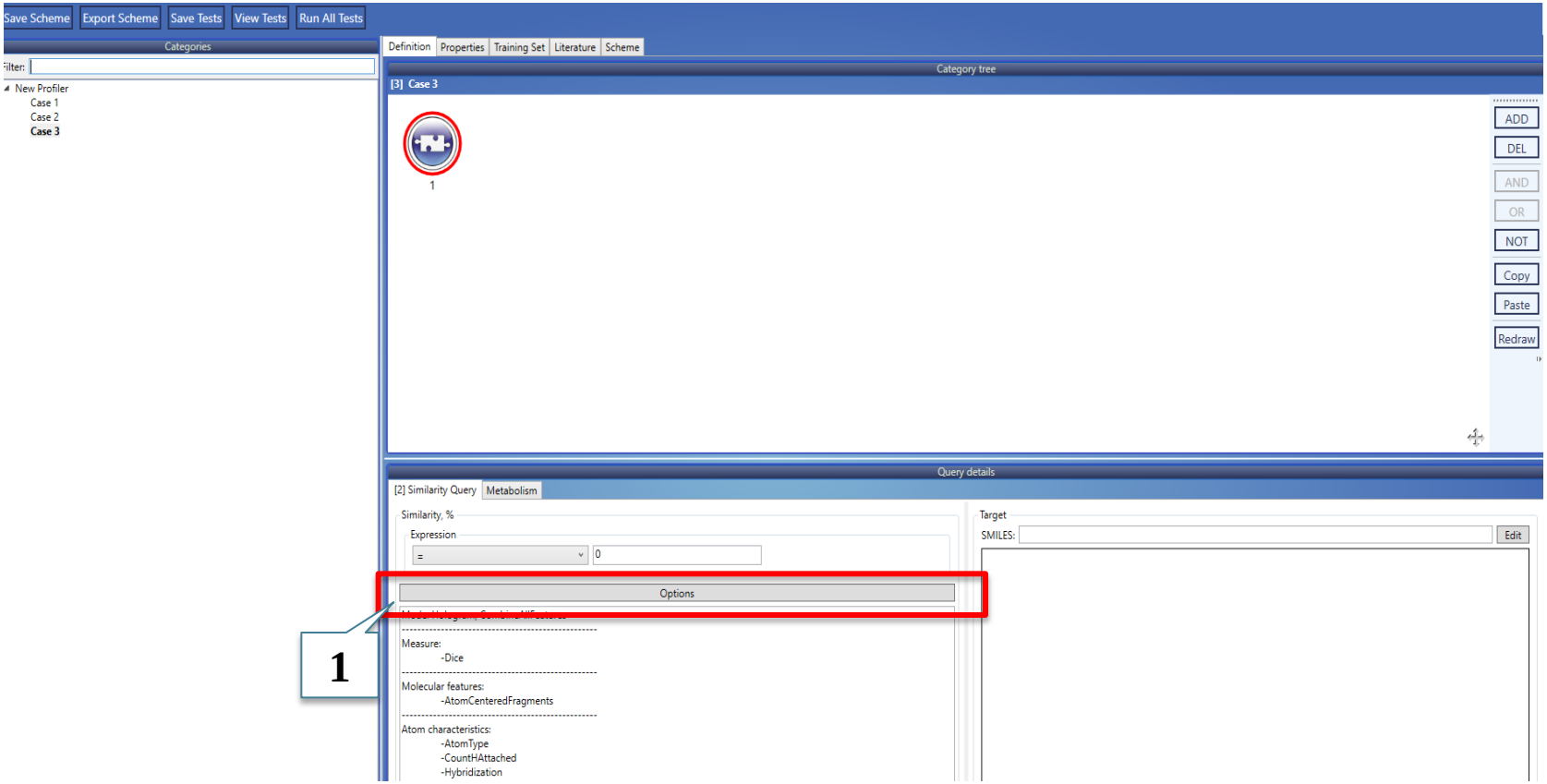
- Similarity query provides possibility to search chemicals similar to a predefined target chemical. In this exercise will search chemical similar to eugenol more than 60%, using default Similarity options – rule 1.
- The Similarity option are explained in details in the following tutorial (https://qsartoolbox.org/wp-content/uploads/2020/02/Tutorial_26_New-optio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

Building of a new profiler
Search by similarity query



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

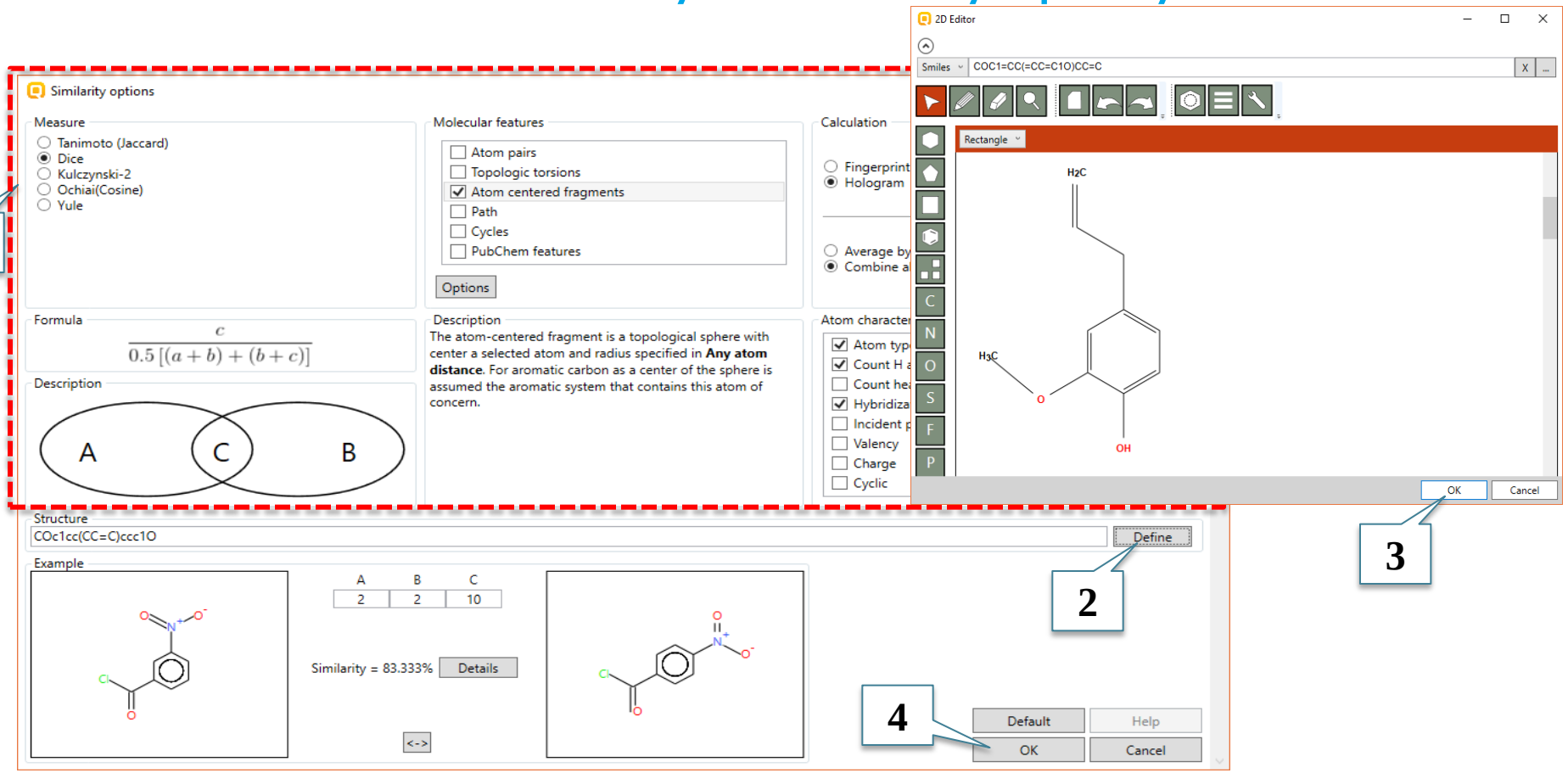
Building of a new profiler
Search by similarity query



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

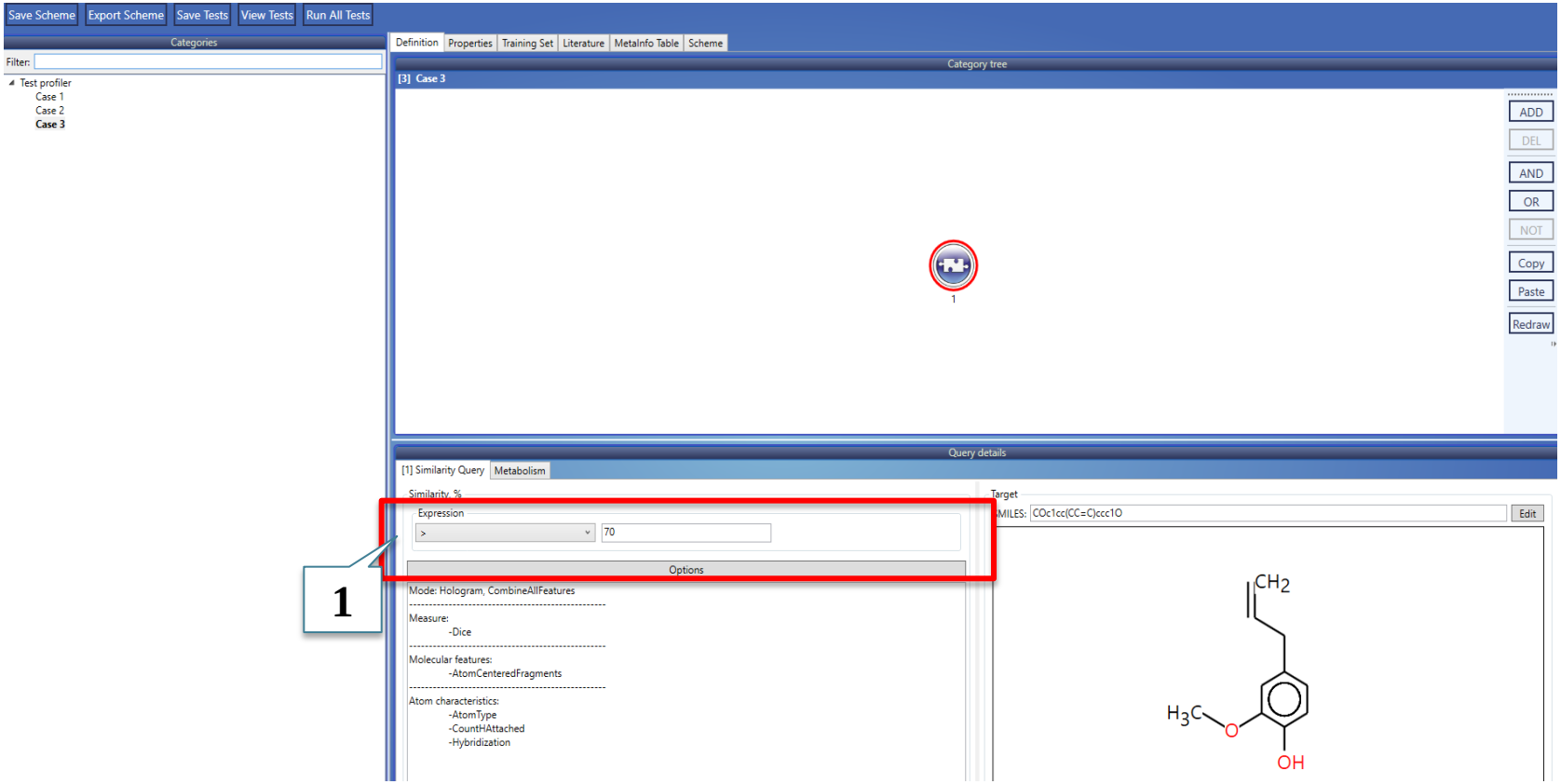
Building of a new profiler

Search by similarity query



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



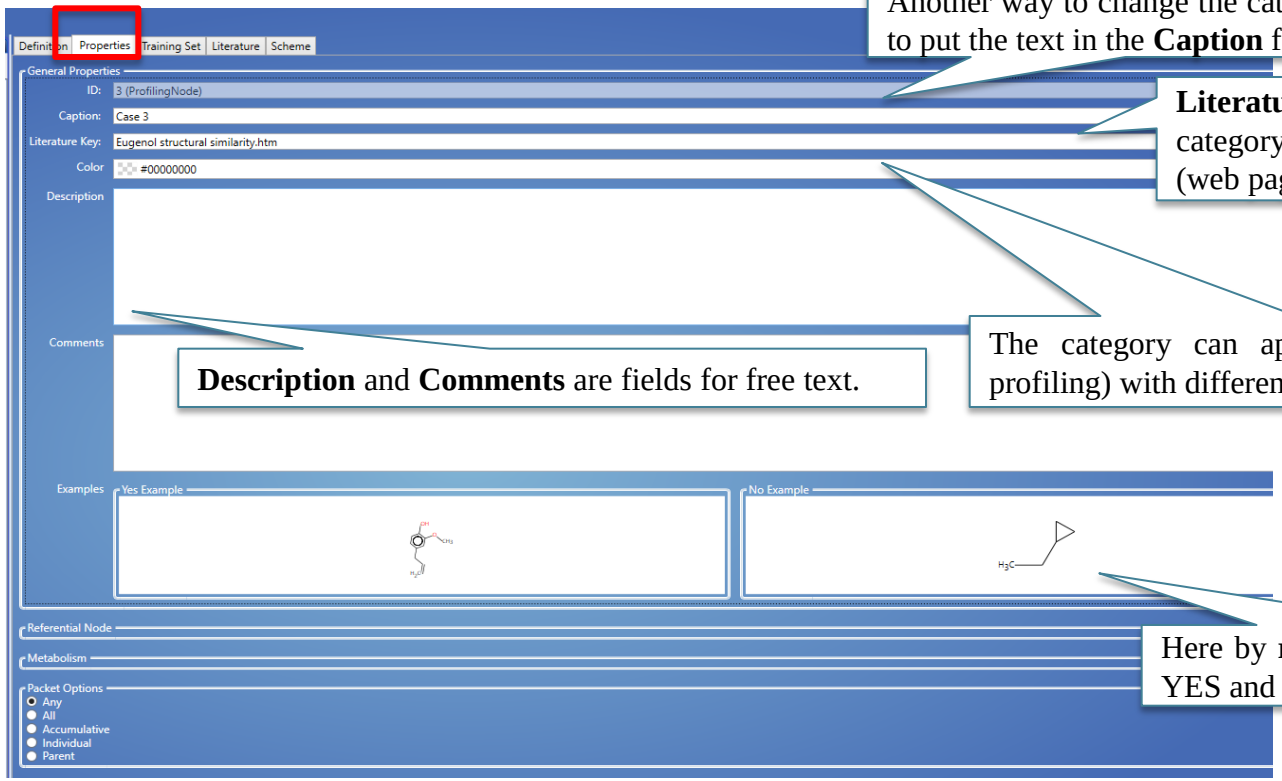
1. Define more than 70% structural similarity between Eugenol and profiled chemicals;

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



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.

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

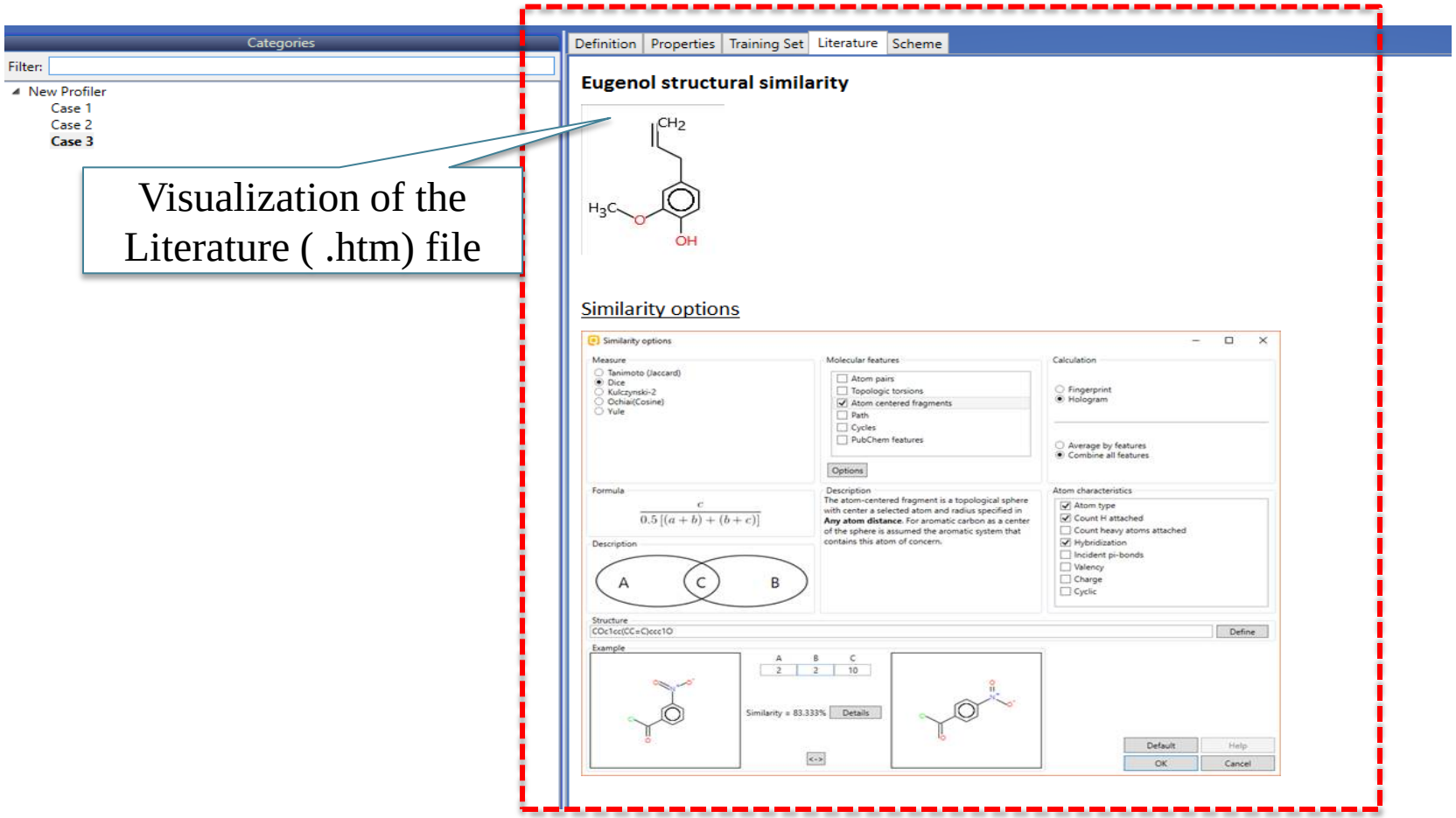
Description and **Comments** are fields for free text.

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

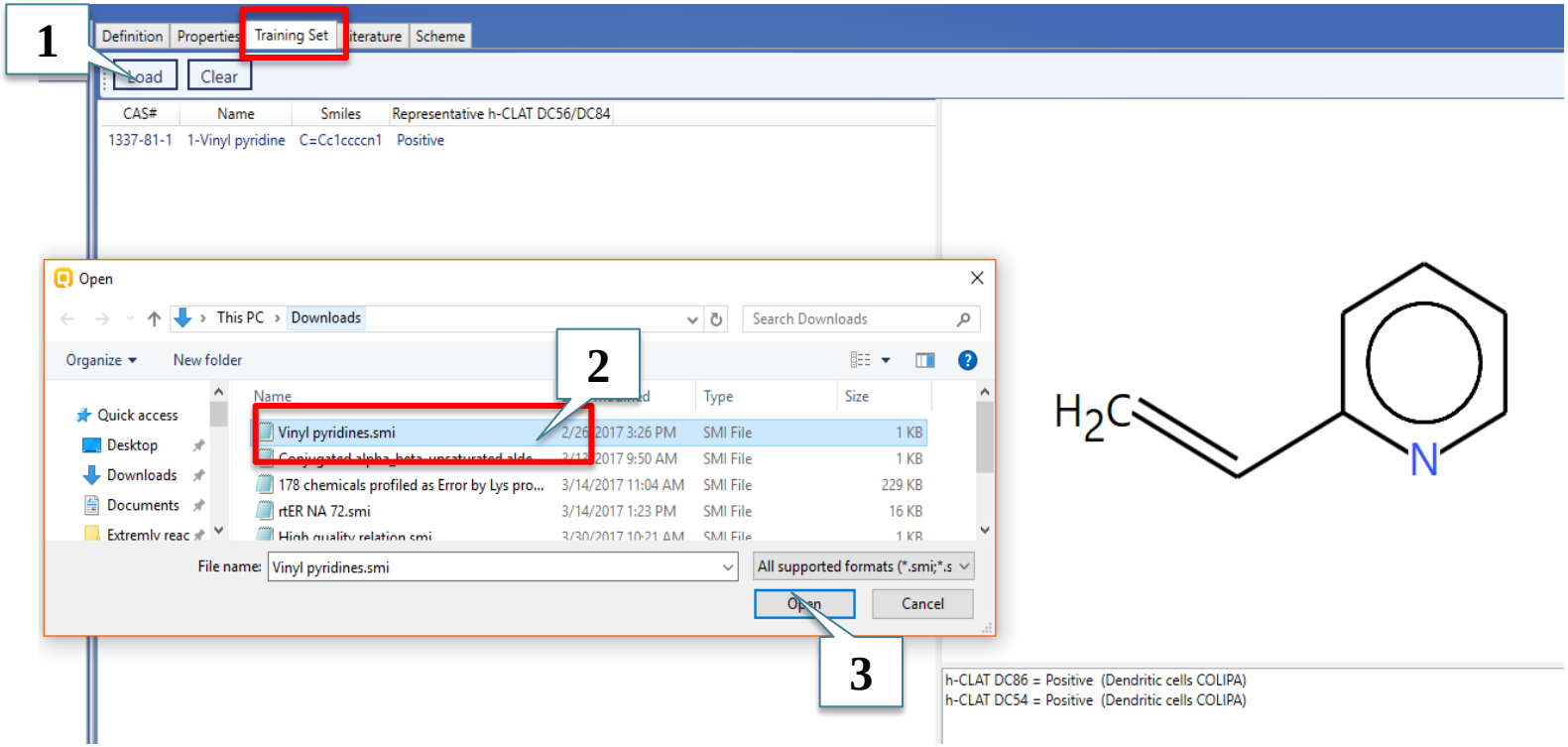
Building of a new profiler

Additional functionalities: Literature

Visualization of the Literature (.htm) file



Building of a new profiler
Additional functionalities: Training set

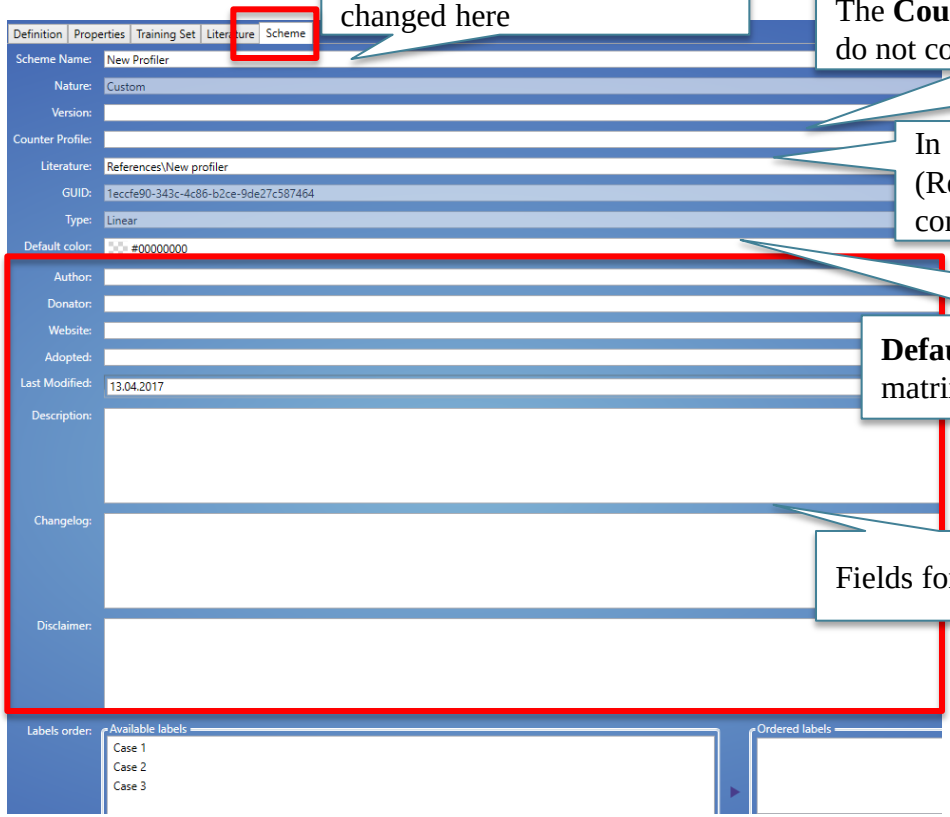


1. Click **Load;** 2. Browse the training set file*; 3. Click **Open**

*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

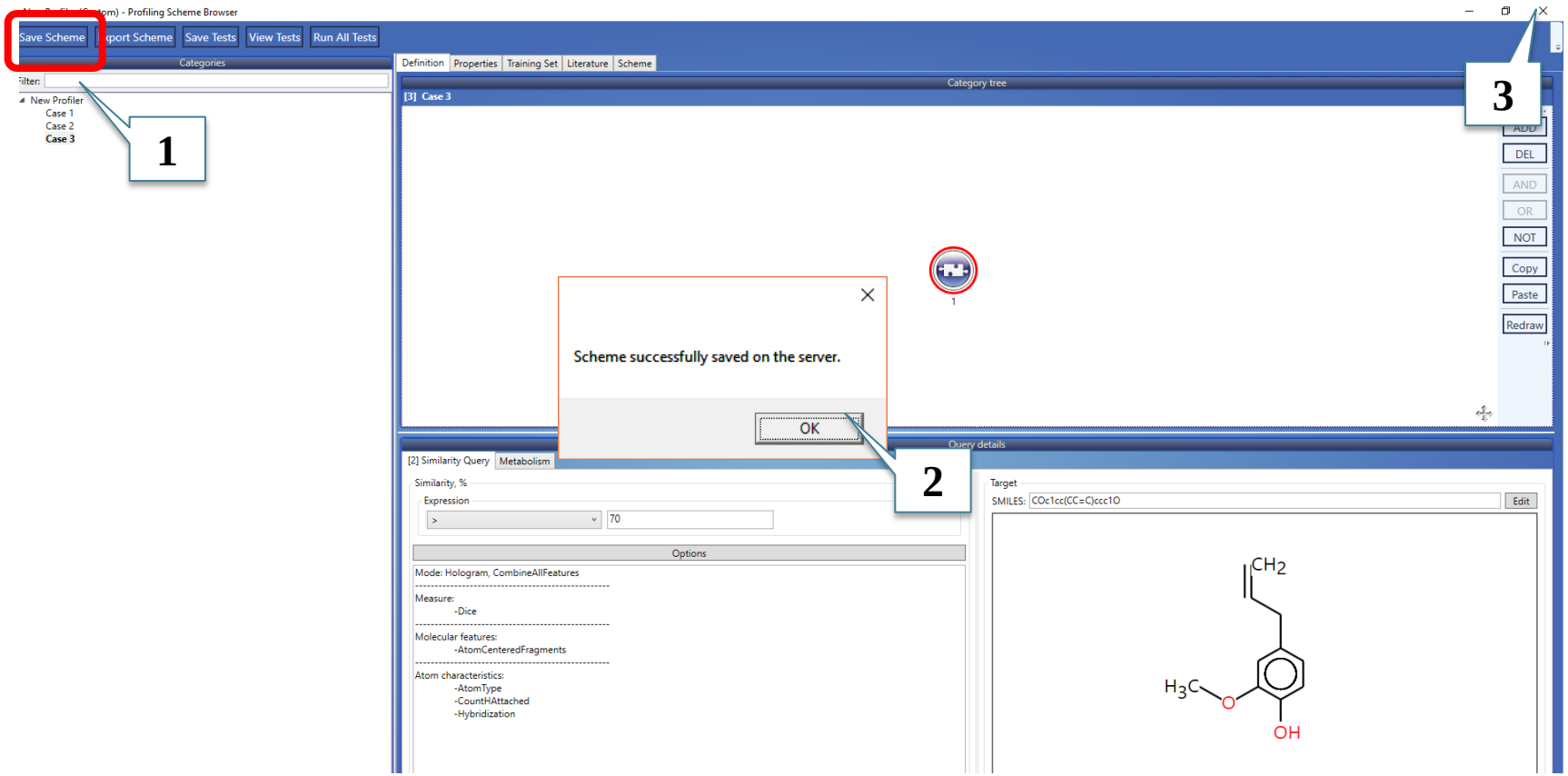
In the **Literature** is needed to put the specific art of the path* (Reference\Name of folder with .htm files) to the folder containing literature files for all categories.

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

Fields for free text

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

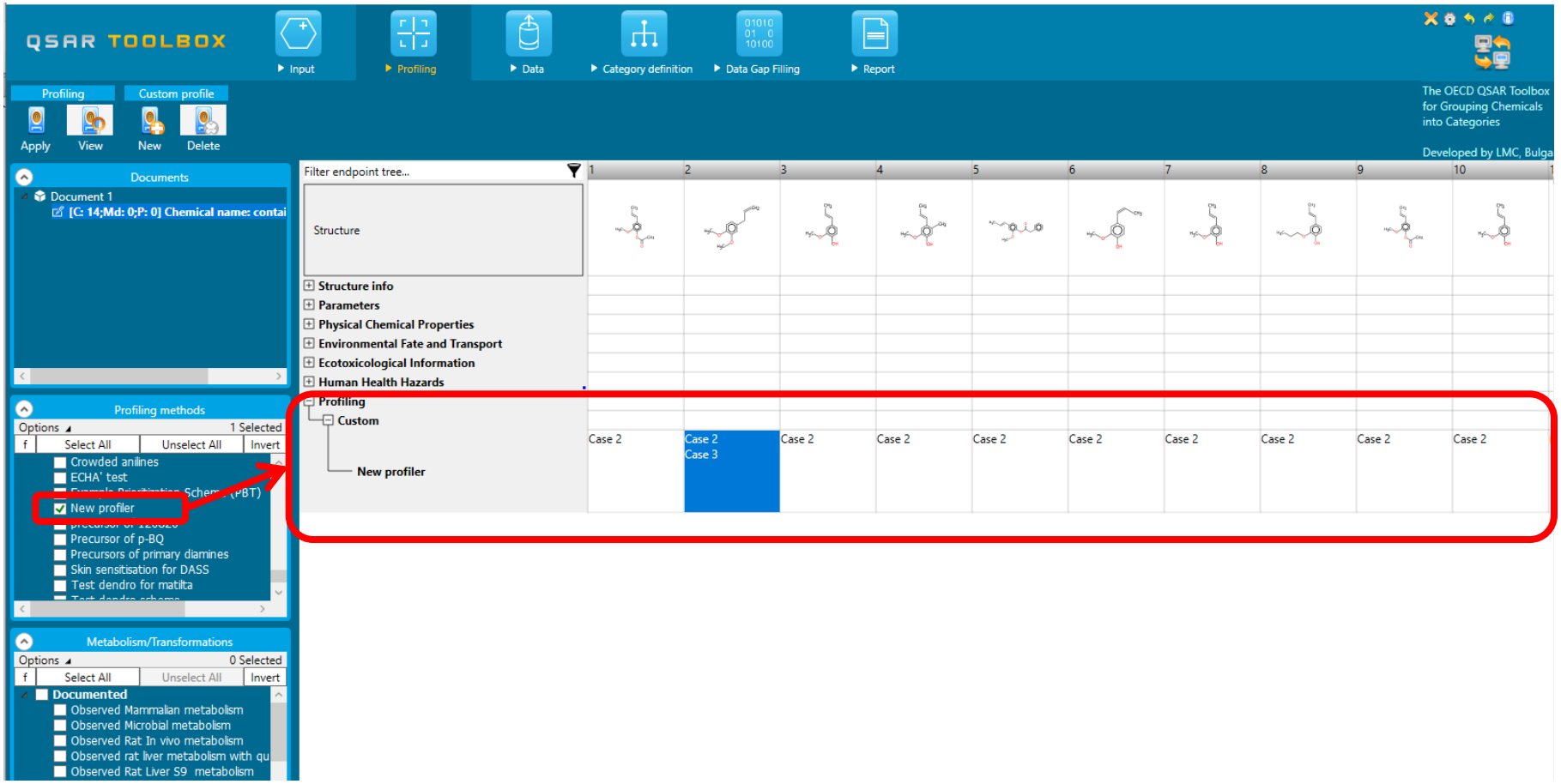


1. Press the **Save scheme** button; 2. Click **OK** to confirm the newly created profiler and 3. Close the window.

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



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.