QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.3.4

Step-by-step example of how to build a category for more than one target chemicals and predict acute toxicity to fish

Outlook

- **Background**
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise 1
- Workflow of the exercise 2

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for evaluating an ad-hoc analogue approach.

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for evaluating an ad-hoc analogue approach.
- By now you are experienced in using the Toolbox so there will be multiple key strokes between screen shots.

Outlook

- Background
- **Objectives**
- Specific Aims
- The exercise
- Workflow of the exercise 1
- Workflow of the exercise 2

Objectives

- To demonstrate how to use the Toolbox to evaluate whether a data gap filling with read-across from potential analogues of target chemicals is robust.

Outlook

- Background
- Objectives
- **Specific Aims**
- The exercise
- Workflow of the exercise 1
- Workflow of the exercise 2

Specific Aims

- To examine the workflow of evaluating an analogue approach.
- To introduce the user to new functionalities within selected modules.
- To explain the rationale behind each step of the exercises.
- To demonstrate with two practical examples how to use the Toolbox to evaluate whether a read-across from a potential analogue to a target chemical is robust.

Outlook

- Background
- Objectives
- Specific Aims
- **The exercise**
- Workflow of the exercise 1
- Workflow of the exercise 2

Exercise

- In this exercise we will perform two examples of ad-hoc read-across for acute toxicity to fish.
- We will do this by first entering the source and target chemicals and analysing the available data for the source chemicals.
- We will then profile the source and target chemicals and evaluate whether the read-across is robust.

Exercise
Side-Bar on the Robustness of a Potential Analogue

- **According to the OECD Guidance on Grouping of Chemicals, the following issues should be taken into account when evaluating the robustness of an analogue approach:**
  - Quality of the experimental result of the source chemical
  - Differences in functionalities in the molecules of the source and target compound (*)
  - Purity and impurity profiles
  - Differences in physical chemical properties
  - Differences in experimental results for other (eco)toxicological endpoints
  - Differences in mode of action (*)
  - Differences in toxicokinetics
- **Some of the issues above (those marked with an *) will be addressed in the current examples with the help of the Toolbox**.

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise 1**
- Workflow of the exercise 2

Workflow of the Exercise 1

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

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise 1**
  - **Chemical Input**
  - Profiling
  - Endpoint

Chemical Input
Overview

- As you know this module provides the user with several means of entering the chemical of interest or the target chemical.

- It is essential to remember that 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.

Chemical Input
Ways of Entering a Chemical

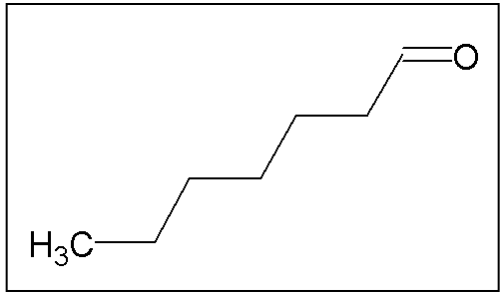
- **Remember there are several ways to enter a target chemical and the most often used are:**
  - CAS#,
  - SMILES (simplified molecular information line entry system) notation, and
  - Drawing the structure.

Chemical Input

Exercise 1

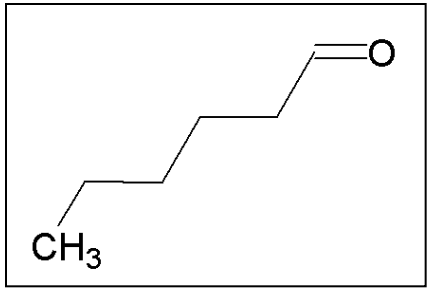
Read-across of acute toxicity to fish from 1-hepatanal 1-hexanal and to 3-ethyl-1-pentanal.

1-heptanal

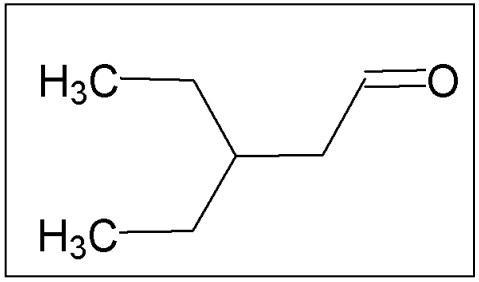


+

1-hexanal



3-ethyl-1-pentanal

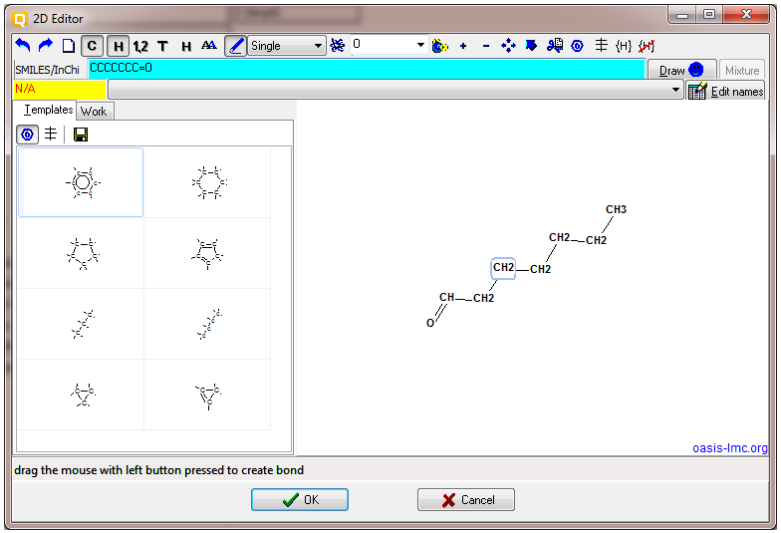


Chemical Input
Input target chemical#1 by SMILES

- In this example, we are entering the structure using the SMILES notation.
- Click on Structure, then
- Enter CCCCCCC=O for n-heptanal on "SMILES/InChi" window.
- The structure is drawn simultaneously while entering the SMILES (see next screen shot).

Chemical Input
Target chemical identity

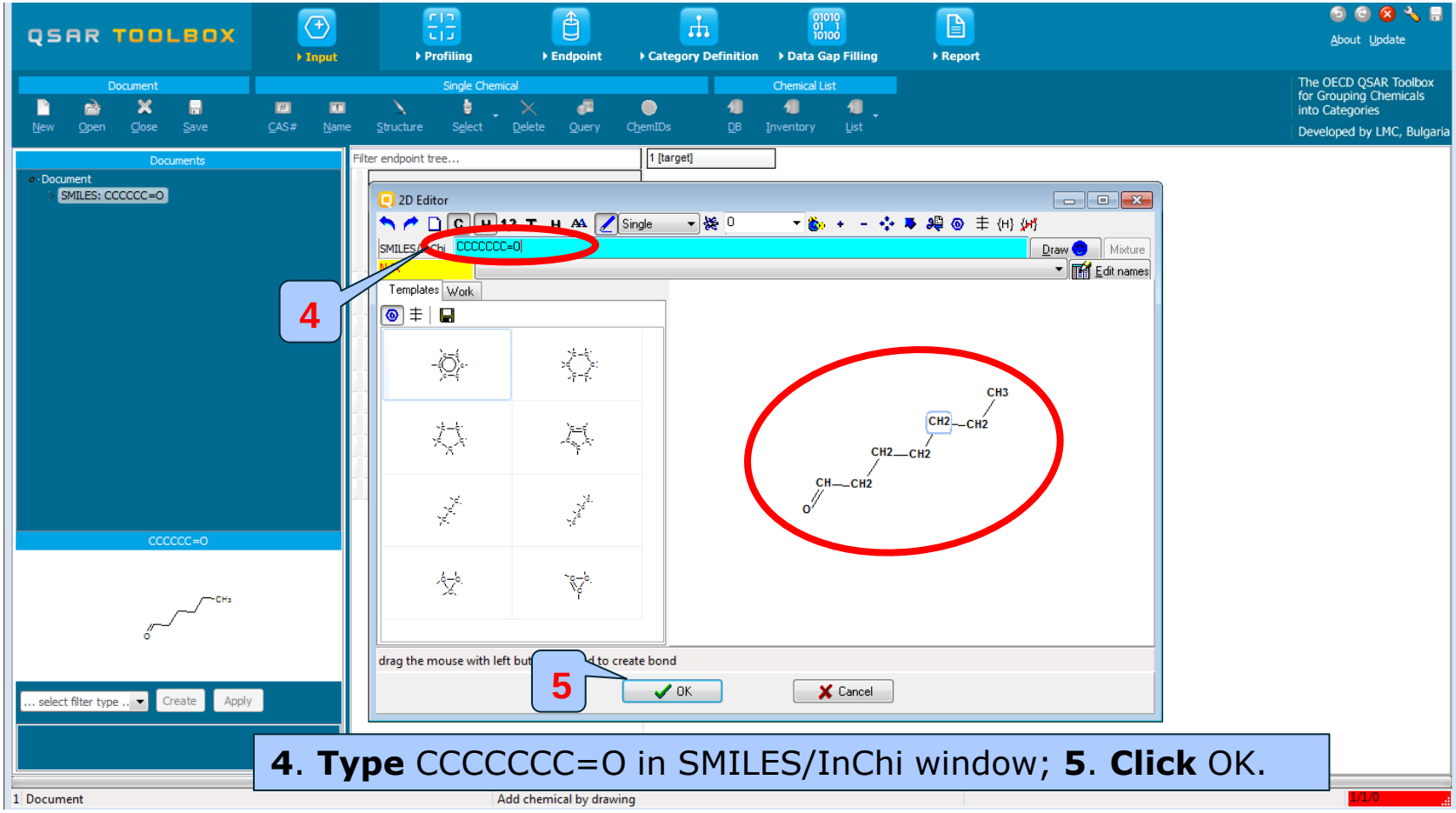
The Toolbox now searches the Toolbox databases and inventories for the presence of the chemical with structure related to the current SMILES notation. It is displayed as a 2D image.



In this case Toolbox found two chemicals answering the required SMILES. This panel displays QA information for presented chemicals. The user can decide which substance is to be retained for the subsequent workflow.

Chemical Input

Input target chemical#1 by SMILES



4. Type CCCCCC=O in SMILES/InChi window; **5. Click** OK.

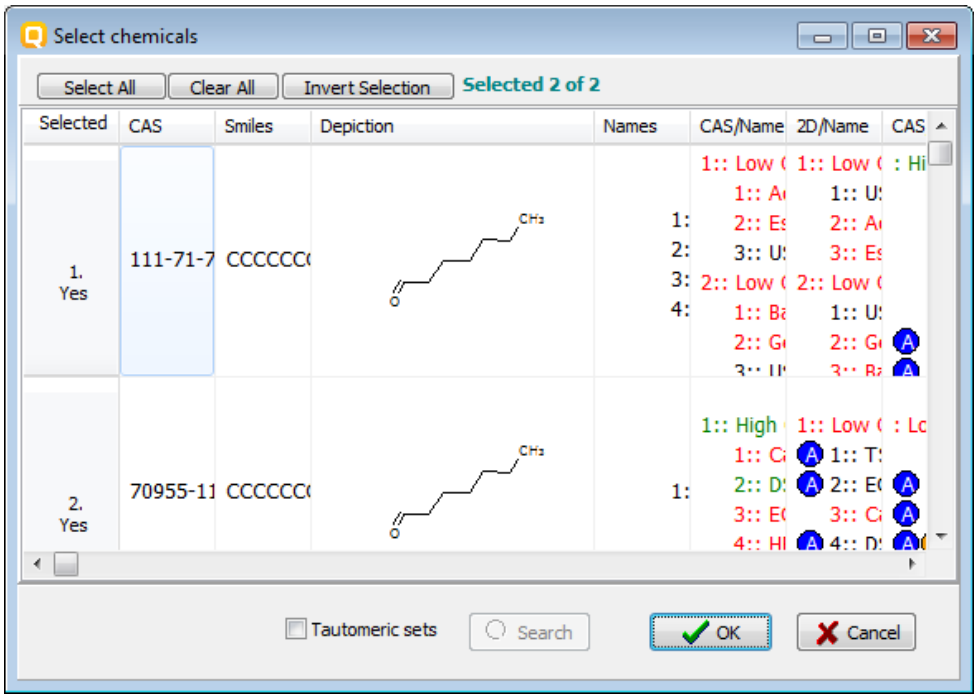
Chemical Input
Input target chemical#1 by SMILES

- The Toolbox now consults its chemical ID database and finds all chemicals with the structure CCCCCCC=O.
- The Toolbox finds three chemicals with the same structure for 1-heptanal but with different CAS numbers and chemical names. Therefore, the Toolbox find two chemicals with different QA relations (CAS-Name; 2D – Name; CAS-2D (see next screen shot).

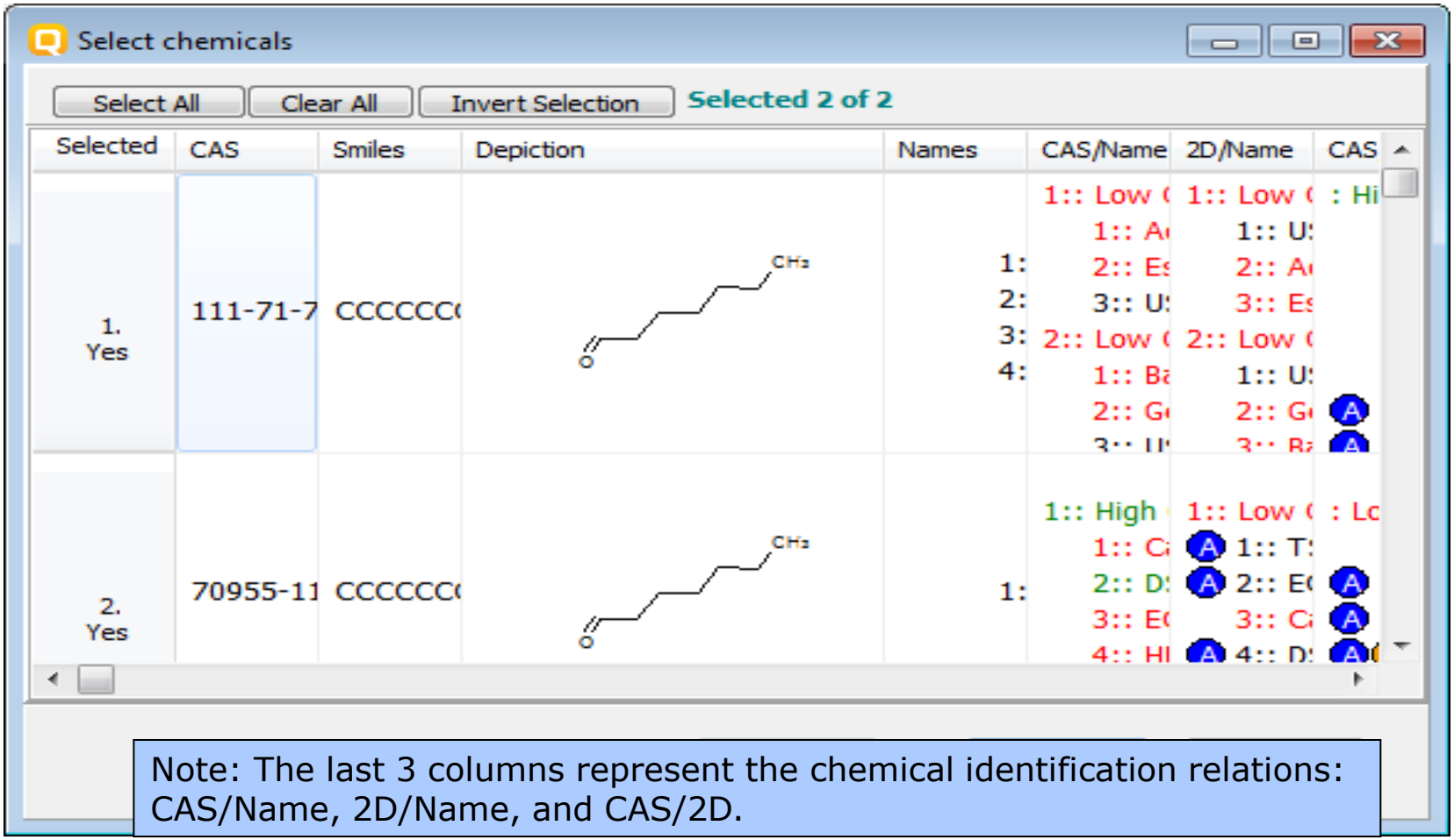
Chemical Input

Input target chemical#1 by SMILES

- The Toolbox finds two chemicals with the same structure and with different QA relations (CAS-Name; 2D –Name; CAS-2D).

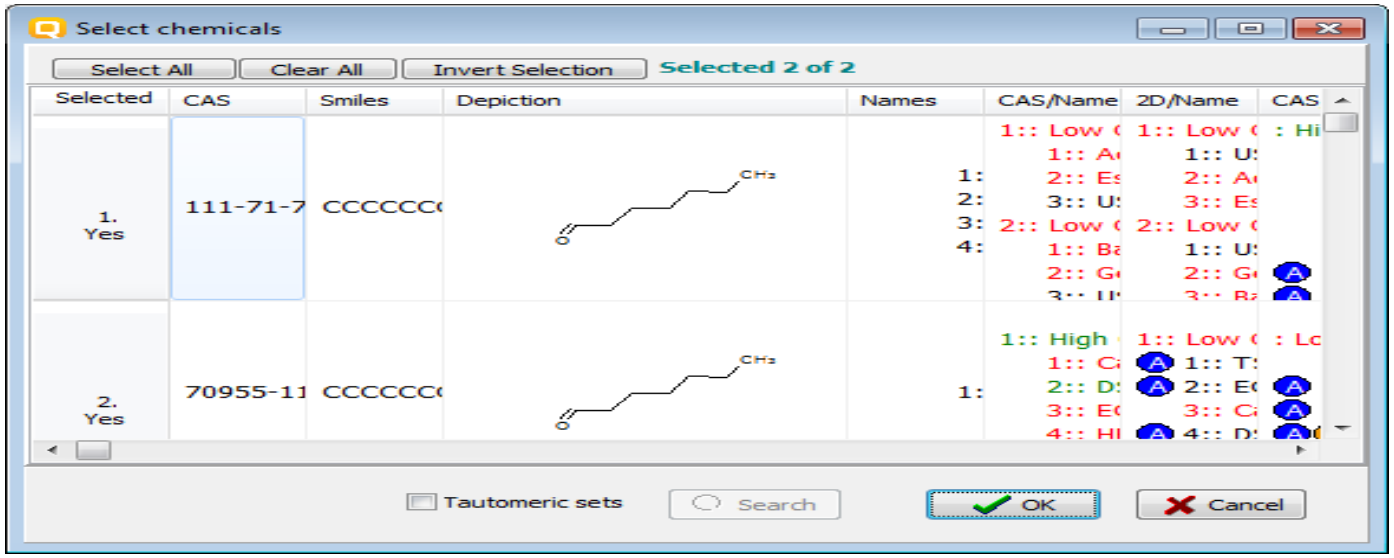


Chemical Input
QA



Note: The last 3 columns represent the chemical identification relations: CAS/Name, 2D/Name, and CAS/2D.

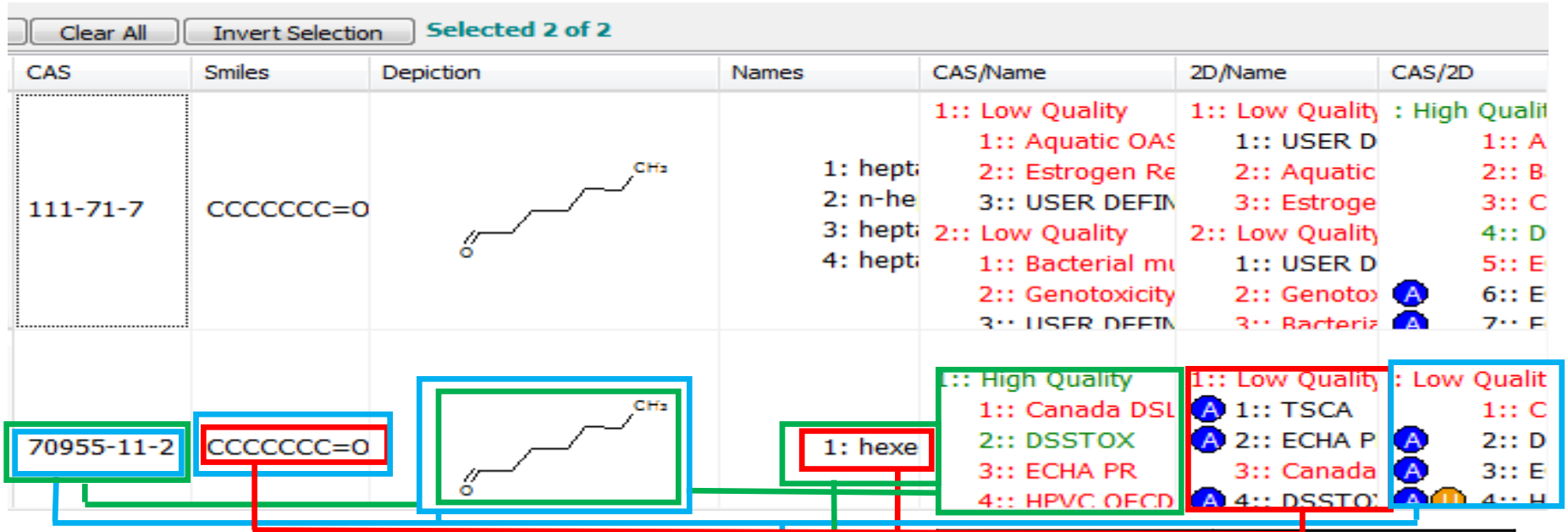
Chemical Input
QA



- The columns represent chemical relations
- The colors represent the quality of relation

Text Color	Evaluated Q
Black	N/A Quality
Red	Low Q
Orange	Moderate Q
Green	High Q
Blue	Conflict

Chemical Input
QA

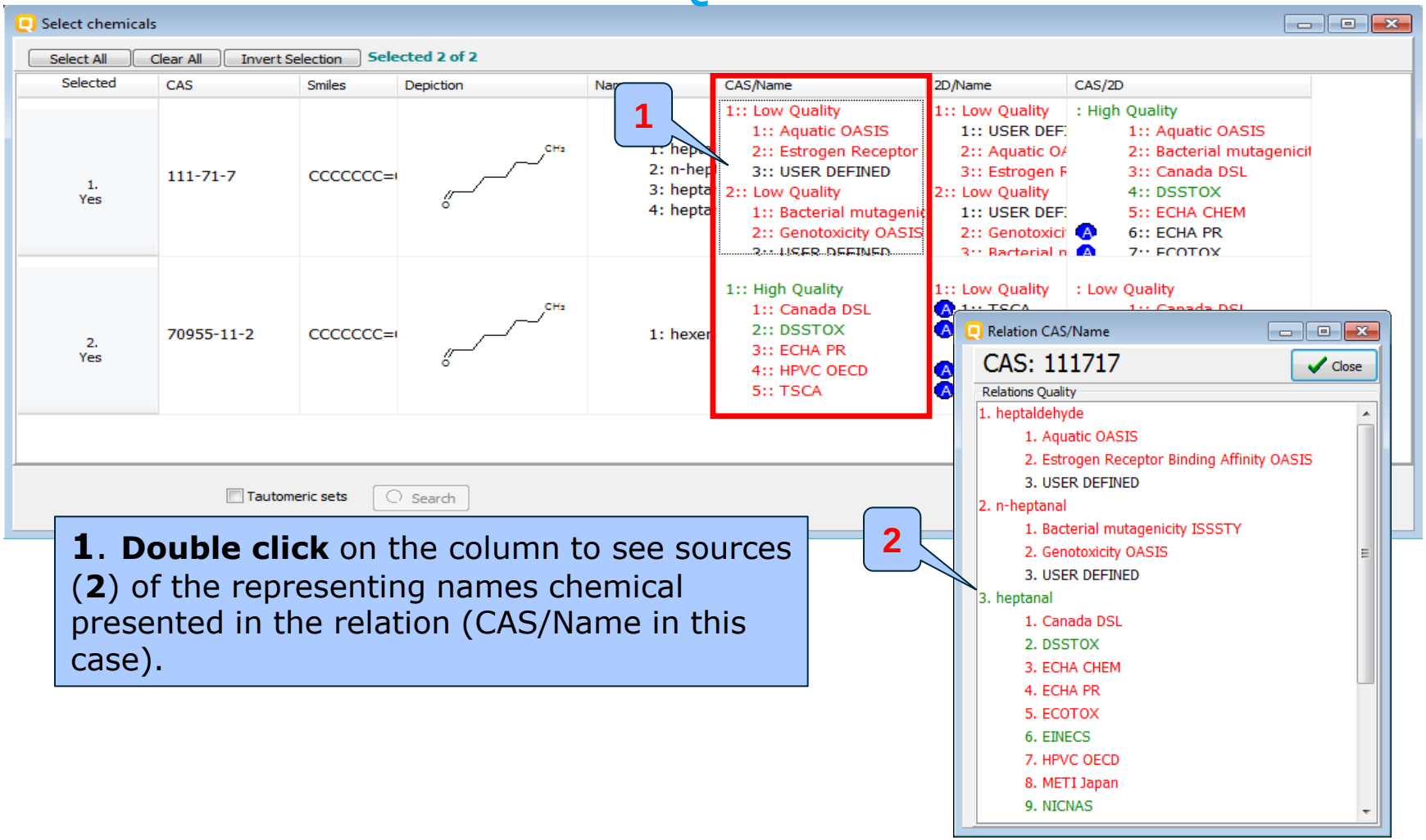


- The columns represent chemical relations
- The colors represent the quality of relation
- Each row represents the quality of relation
- **CAS/Name, Name/2D**, and **CAS/2D** where number corresponds to name numbering

Text Color	Evaluated Q
Black	N/A Quality
Red	Low Q
Orange	Moderate Q
Green	High Q
Blue	Conflict

Chemical Input

QA

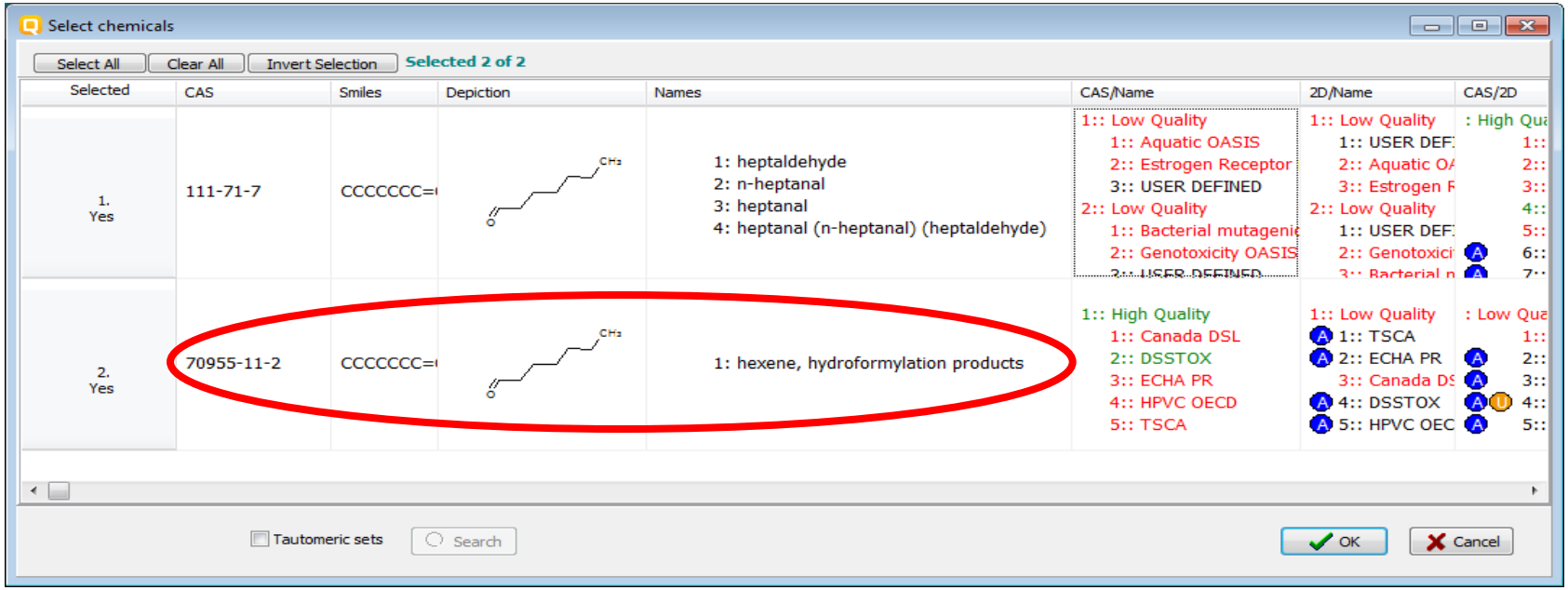


1. **Double click** on the column to see sources (**2**) of the representing names chemical presented in the relation (CAS/Name in this case).

Chemical Input

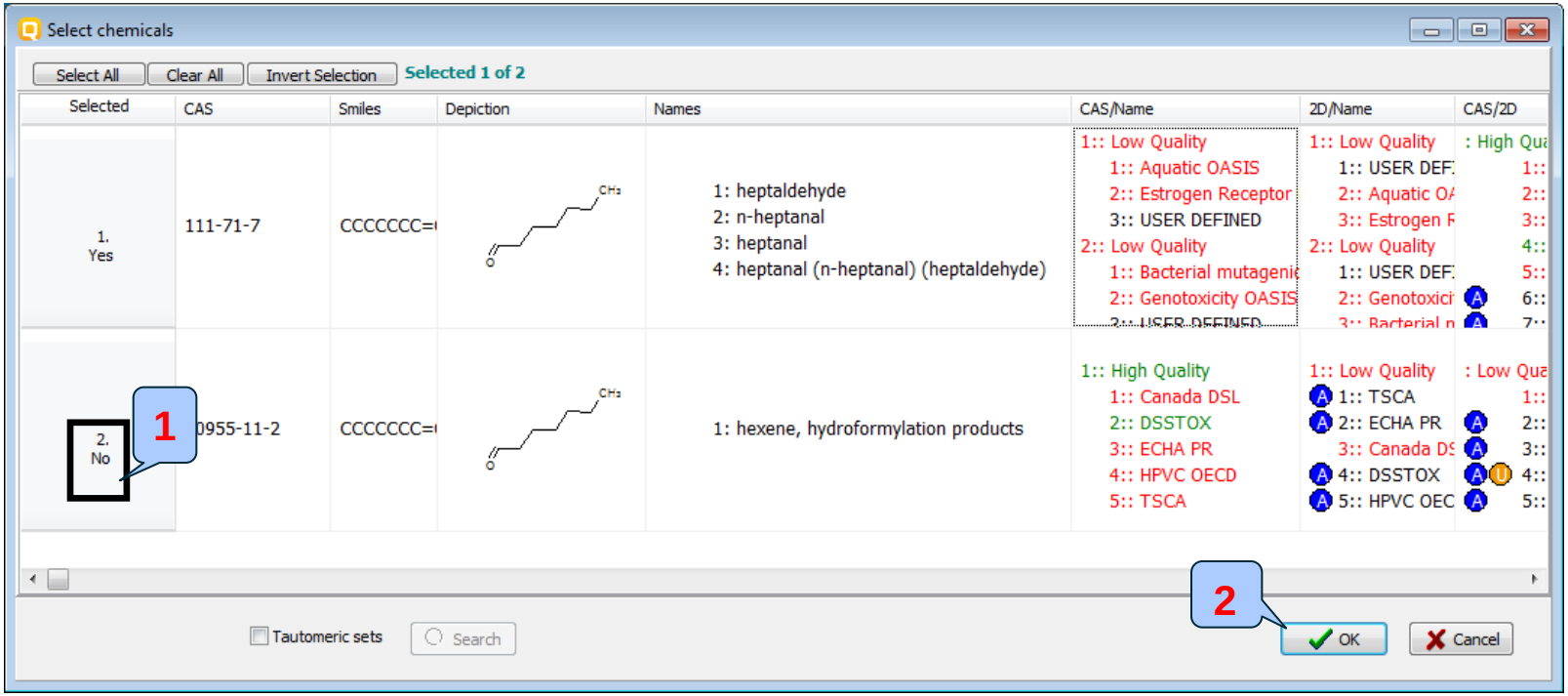
Input target chemical#1 by SMILES

Back to our target chemical, the first one is the actual 1-heptanal while the second one is a mixture containing 1-heptanal. As we are not interested in the mixture this chemical can be removed from the exercise (see next screenshot).



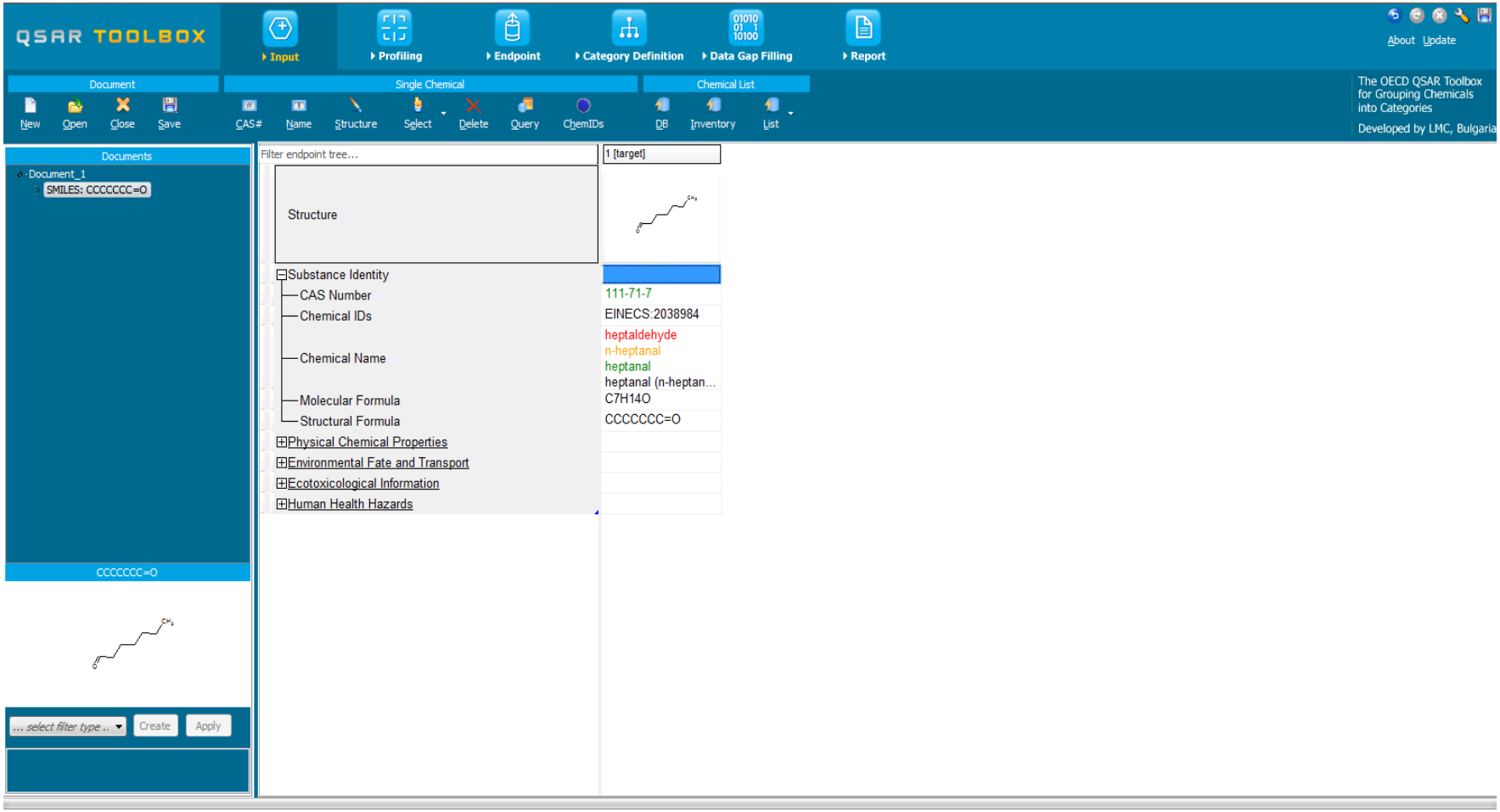
Chemical Input

Input target chemical#1 by SMILES



1. **Click** over the first column with label Yes, then the column become unmarked (labeled with No); **2**. **Click** OK

Chemical Input
Target chemical identity

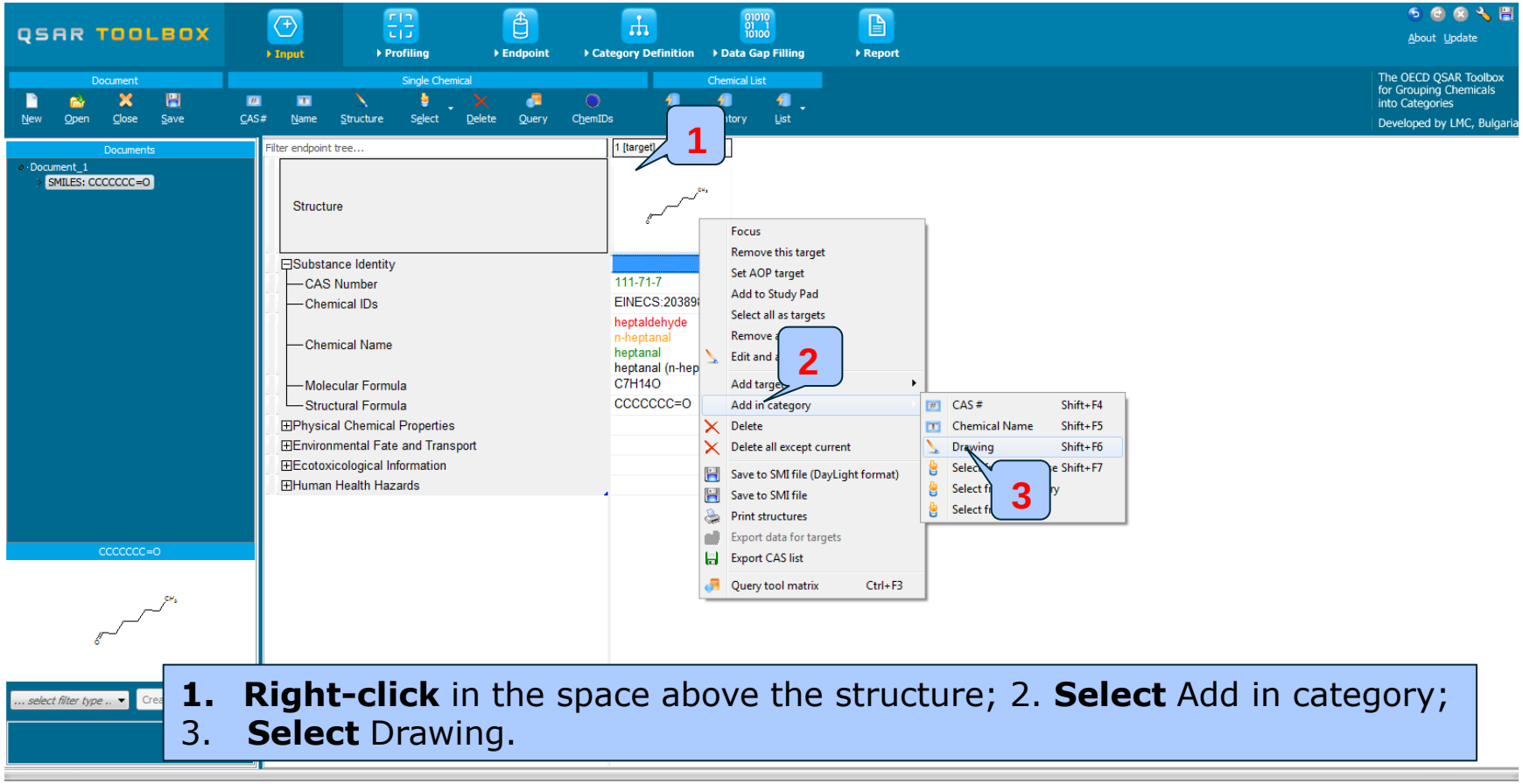


Chemical Input
Exercise 2

- To add additional chemicals by hand into the matrix, right-click above the structure and select "Add target" and then "Structure".
- Enter the SMILES for 1-hexanal: CCCCCC=O and click "OK" (see next screen shot).

Chemical Input

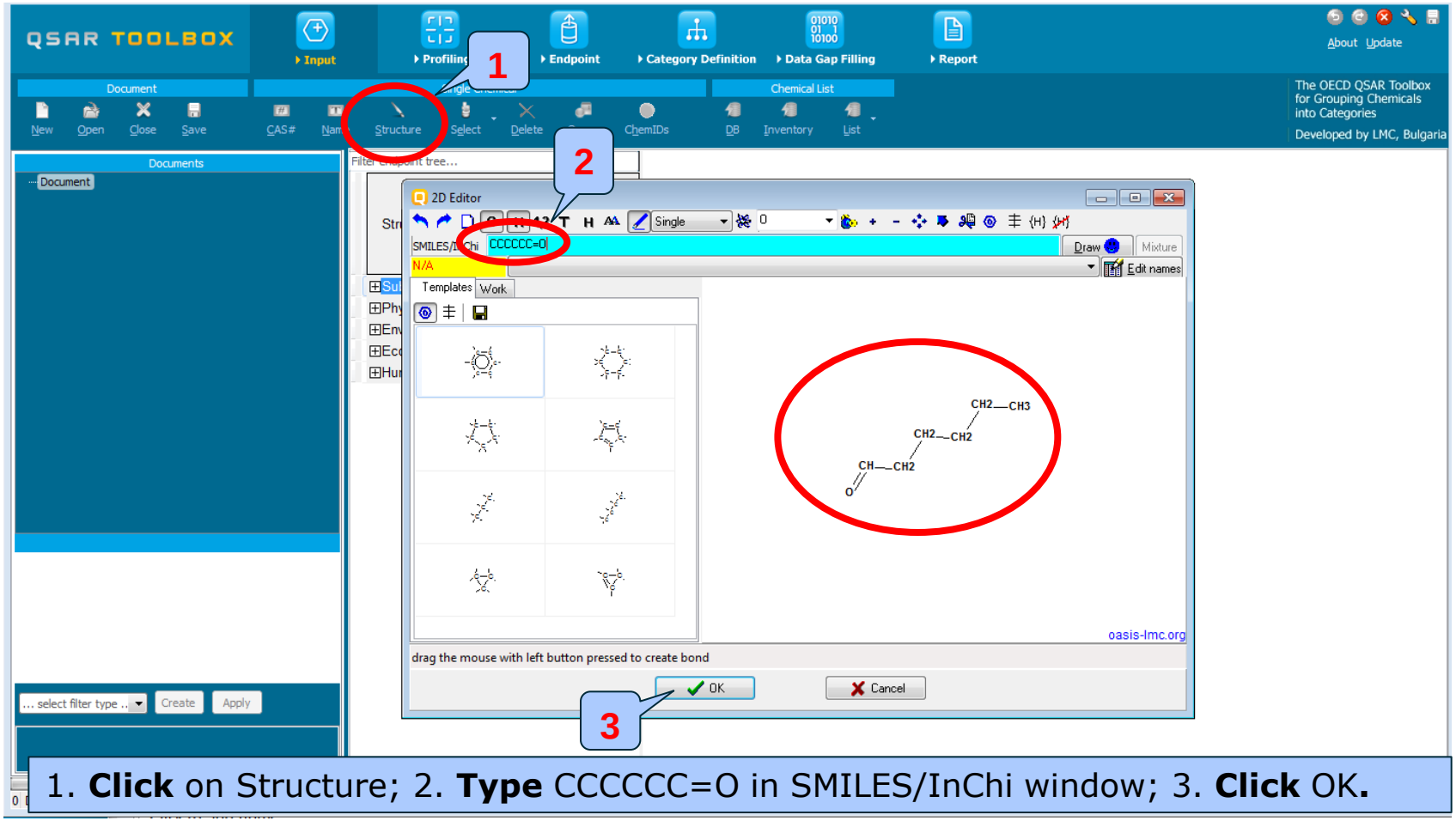
Input target chemical#2 by SMILES



1. **Right-click** in the space above the structure; 2. **Select** Add in category; 3. **Select** Drawing.

Chemical Input

Input target chemical#2 by SMILES

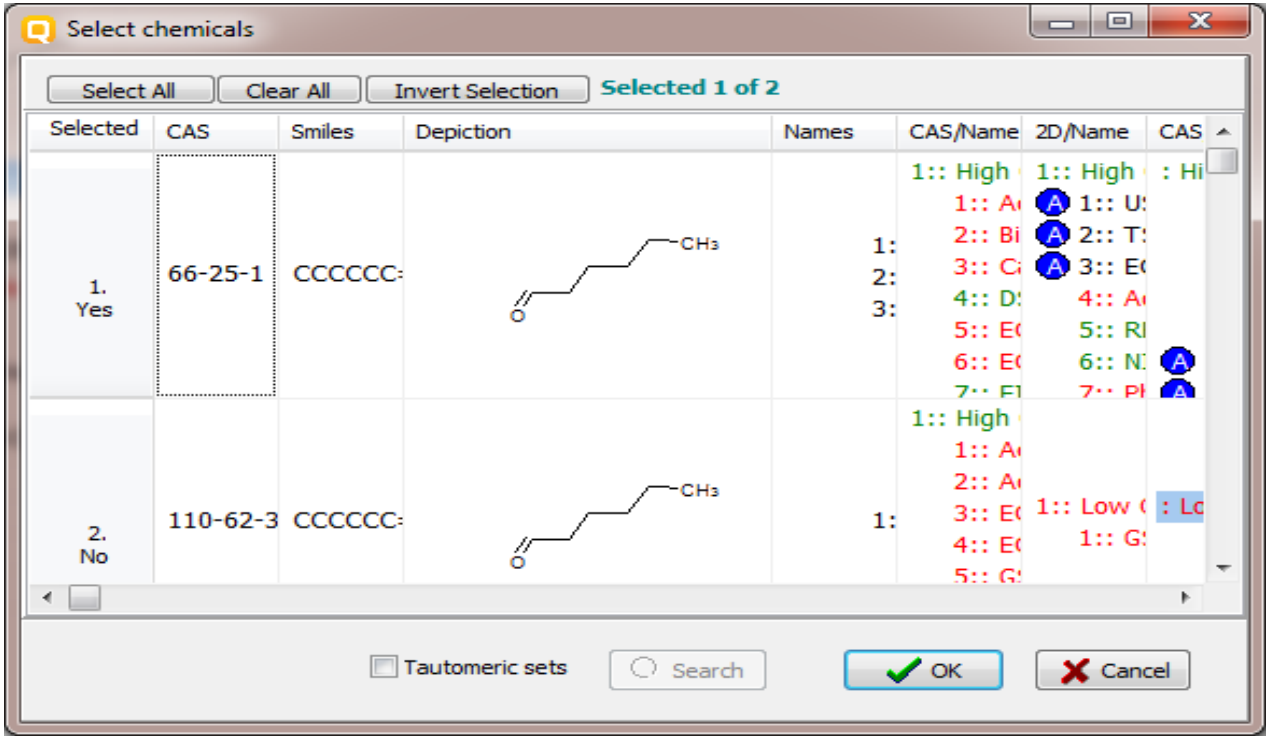


1. **Click** on Structure; 2. **Type** CCCCCC=O in SMILES/InChi window; 3. **Click** OK.

Chemical Input

Input target chemical#2 by SMILES

- The Toolbox finds two chemicals with the same structure and with different QA relations (CAS-Name; 2D –Name; CAS-2D).

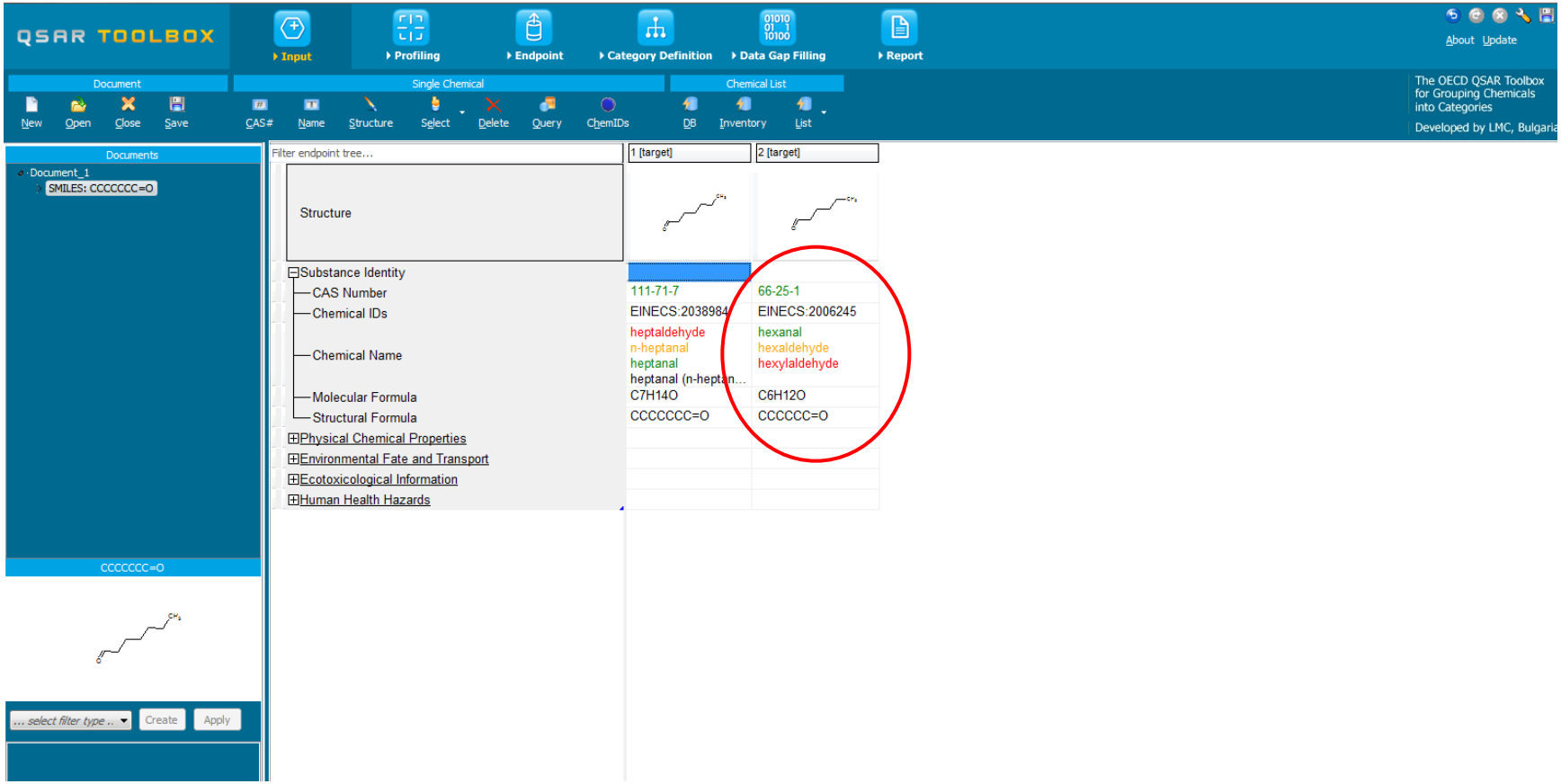


Chemical Input
Target chemical identity

- Click OK to add your target to data matrix

- Click on the box next to "Substance Identity"; this displays the chemical identification information. (see next screen shot).

Chemical Input
Target chemical identity



Chemical Input
Input target chemical#3 by SMILES

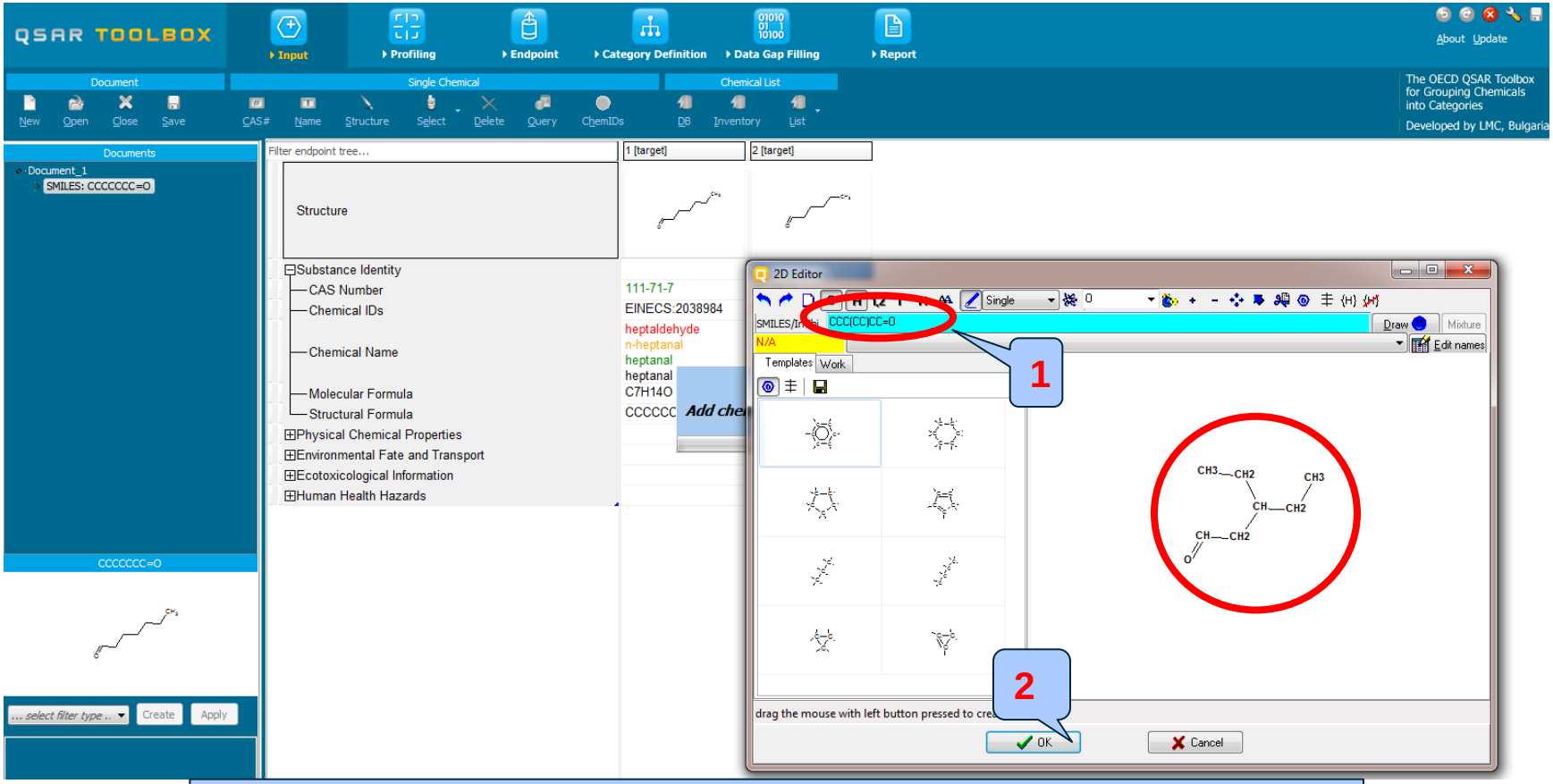
- To add the third chemical by hand into the matrix, right-click above the structure and select "Add in category" and then "Drawing".

- Enter the SMILES for 3-ethyl-1-pentanal: CCC(CC)CC=O and click "OK".

- Your data matrix should now contain your three chemicals (see next screen shot).

Chemical Input

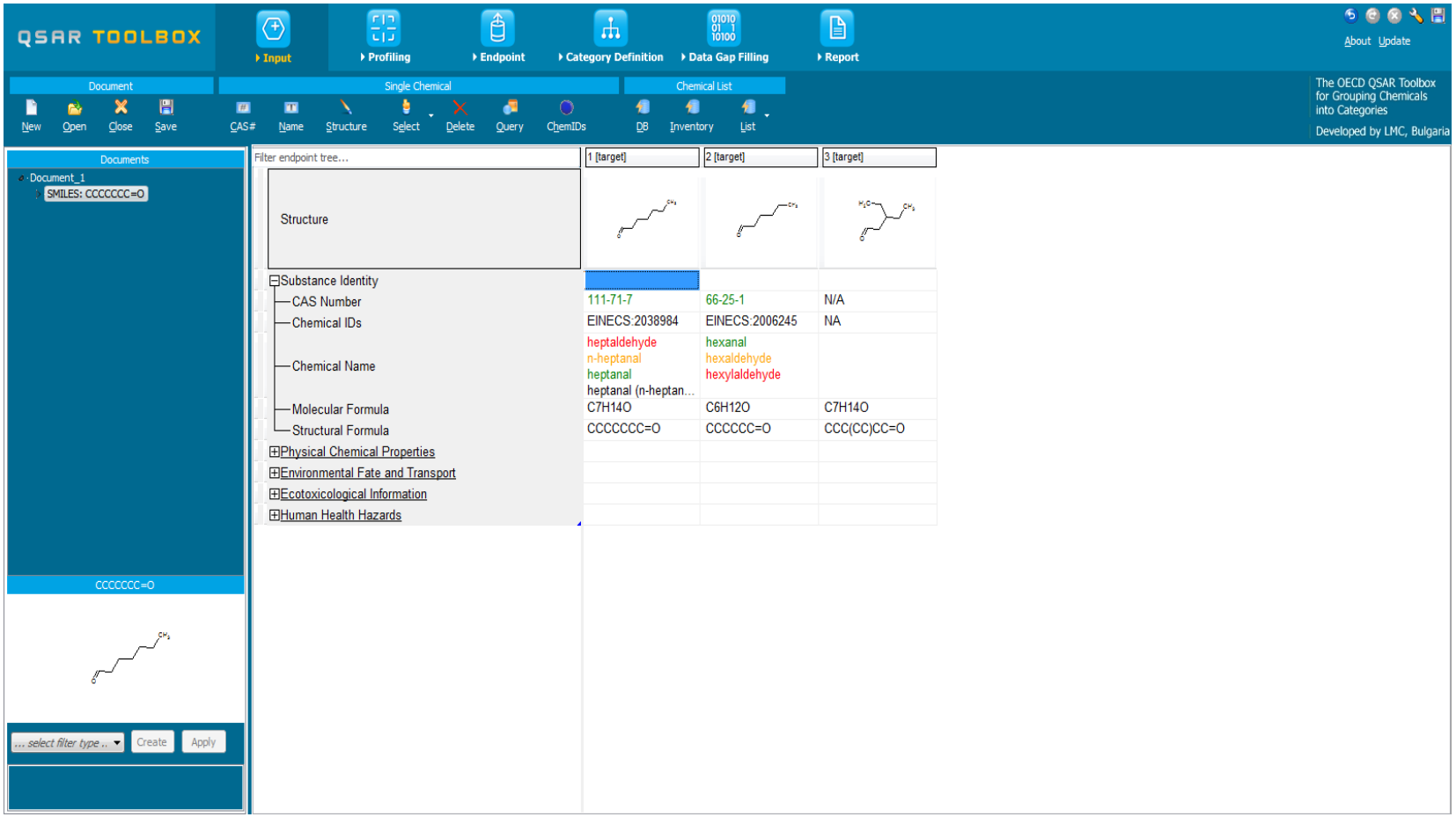
Input target chemical#3 by SMILES



1. **Type** CCC(CC)CC=O in SMILES/InChi window; 2. **Click** OK.

Chemical Input

Input target chemical#3 by SMILES



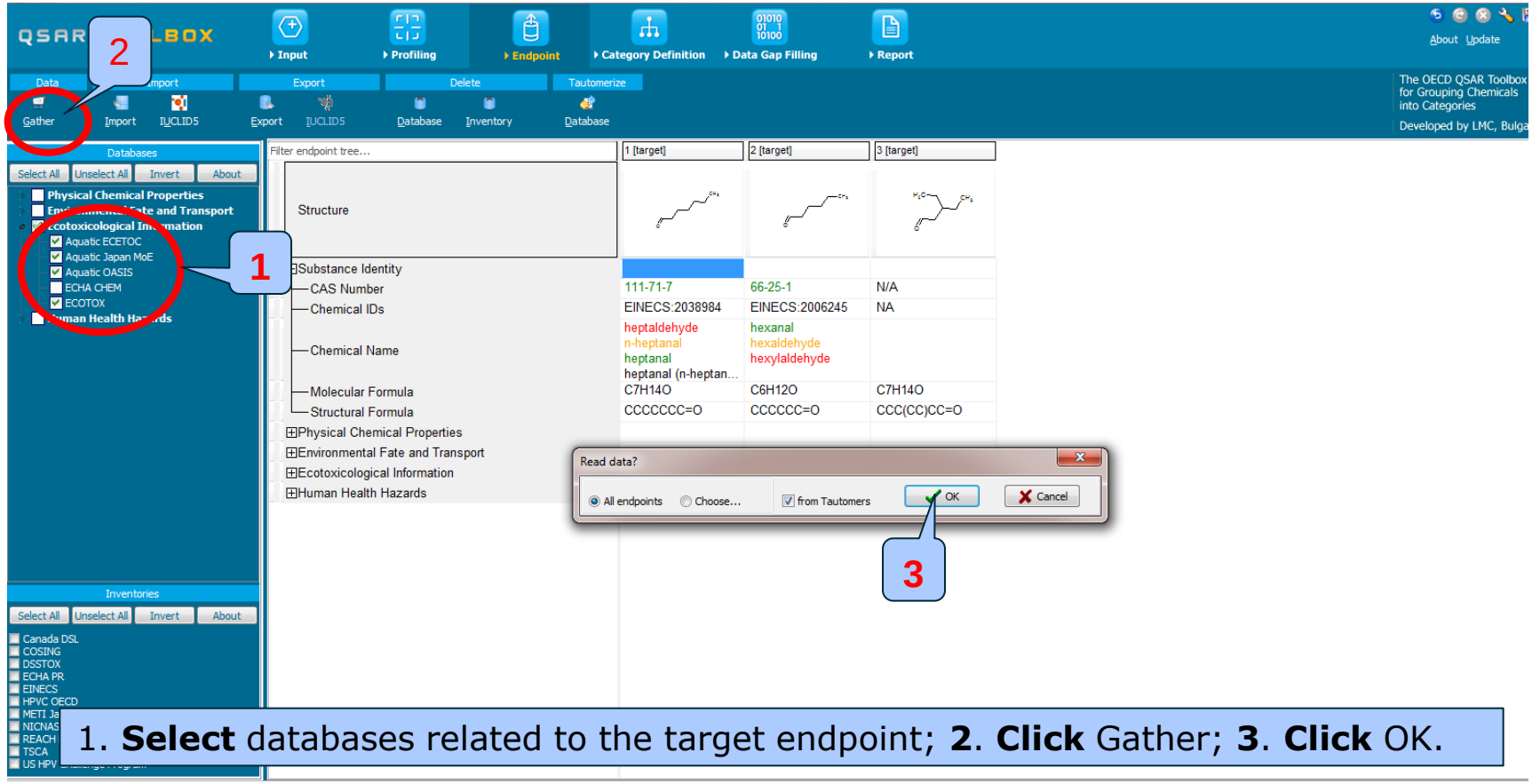
Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise 1**
  - Chemical Input
  - Profiling
  - **Endpoints**

Endpoints

- Move directly to the module "Endpoints".
- Remember, "Endpoints" refer to the electronic process of retrieving fate and toxicity data stored in the Toolbox and it can be gathering in a global fashion or on a more defined basis.
- In this example we only want to retrieve data on toxicity to fish so select the following databases containing information on aquatic toxicity:
  - Aquatic ECETOC
  - Aquatic Japan MoE
  - Aquatic OASIS
  - Aquatic US-EPA ECOTOX
- Click "Gather Data" (see next screen shot).

Endpoints
Gather data

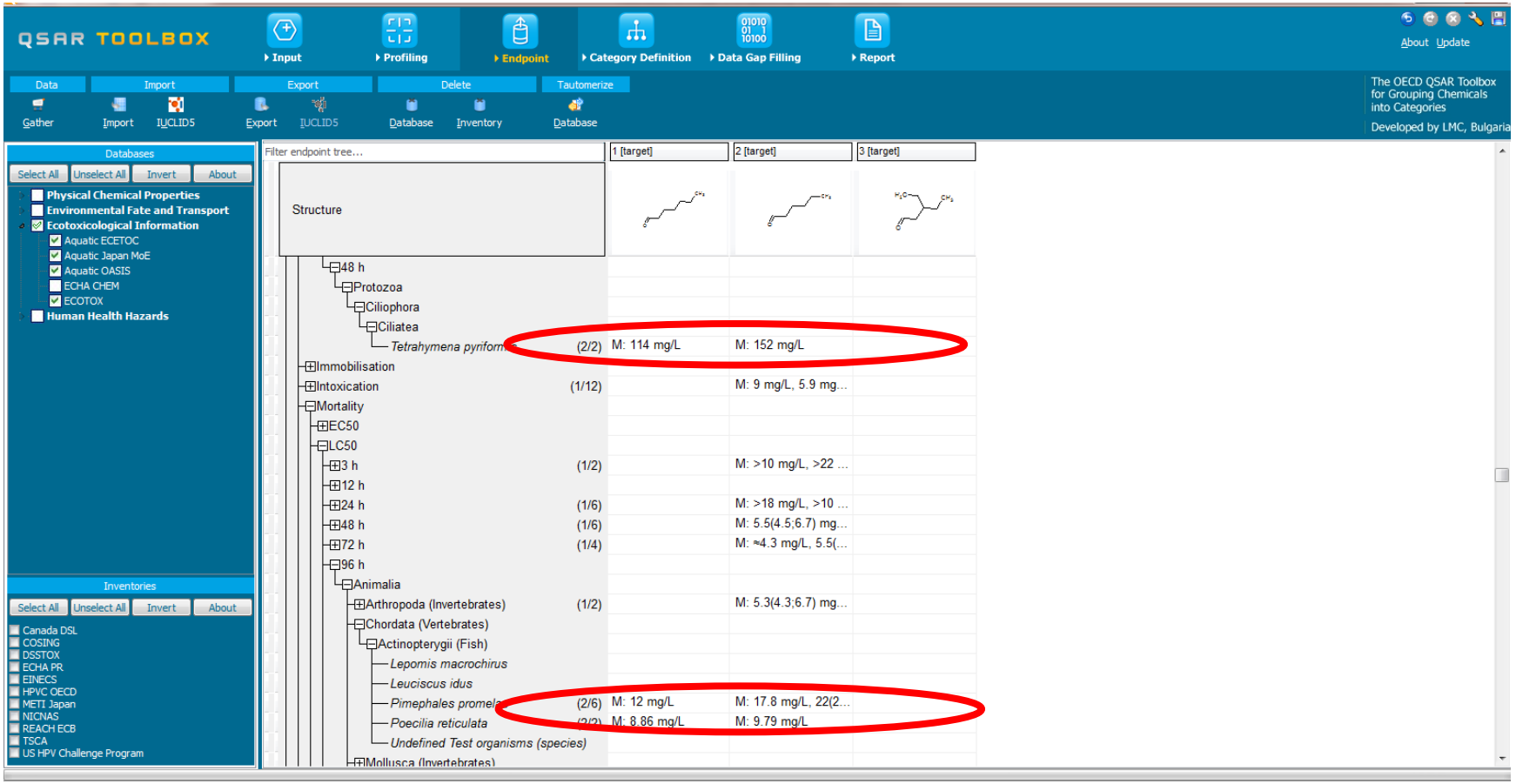


1. **Select** databases related to the target endpoint; **2**. **Click** Gather; **3**. **Click** OK.

Endpoints
Available experimental data

- Results are available for two effects:
  - **Growth** for *Tetrahymena pyriformis* for both n-hexanal and n-heptanal.
  - **Mortality** for two species: *Pimephales promelas* and *Poecilia reticulata* for both n-hexanal and n-heptanal (see next screen shot).
- These can potentially be used for read-across to fill in the data gap for the third target: **3-ethyl-1-pentanal** (e.g. using the lowest available LC50 result).

Endpoints
Available experimental data



Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise 1**
  - Chemical Input
  - **Profiling**
  - Endpoints

Profiling

- Click on "Profiling" to move back (yes back) to the previous module.
- Remember that "Profiling" refers to the process of retrieving information on the target compounds, other than and toxicity data.
- Available information includes likely mechanism(s) of action.
- In this exercise we will use the profiling results to evaluate the robustness of the analogue approach.

Profiling
Profiling the target chemical

- As you remember, the outcome of the profiling determines the most appropriate way to search for analogues.
- For this example the following mechanistic and endpoint specific profiling methods should be selected:

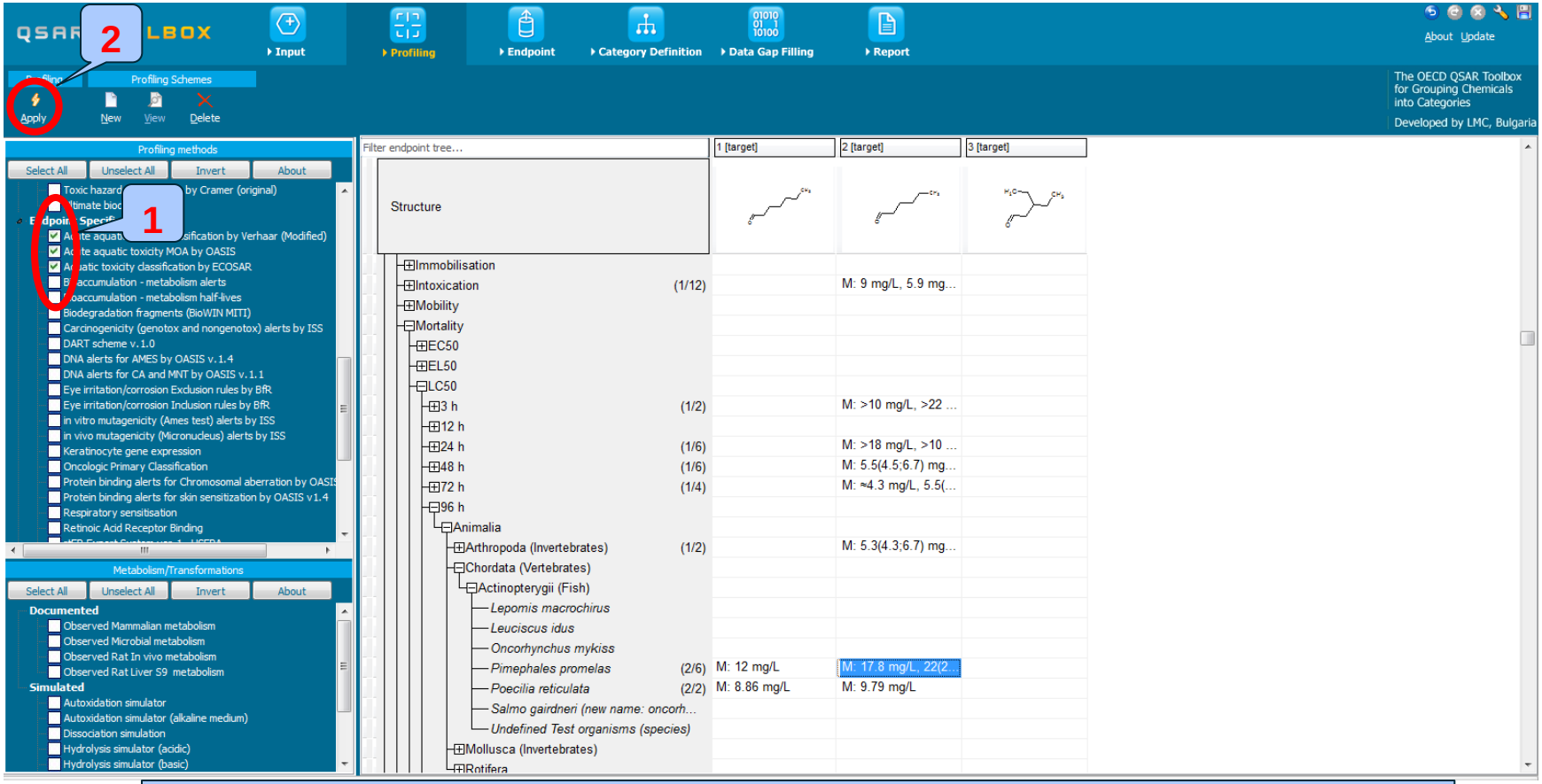
  - Aquatic toxicity classification by ECOSAR

  - Acute aquatic toxicity MOA by OASIS

  - Acute aquatic toxicity classification by Verhaar(Modified)
- Select those 3 "profiling methods" by clicking on the boxes before the names of the profilers before clicking "Apply" (see next screen shot).

Profiling

Profiling the target chemical

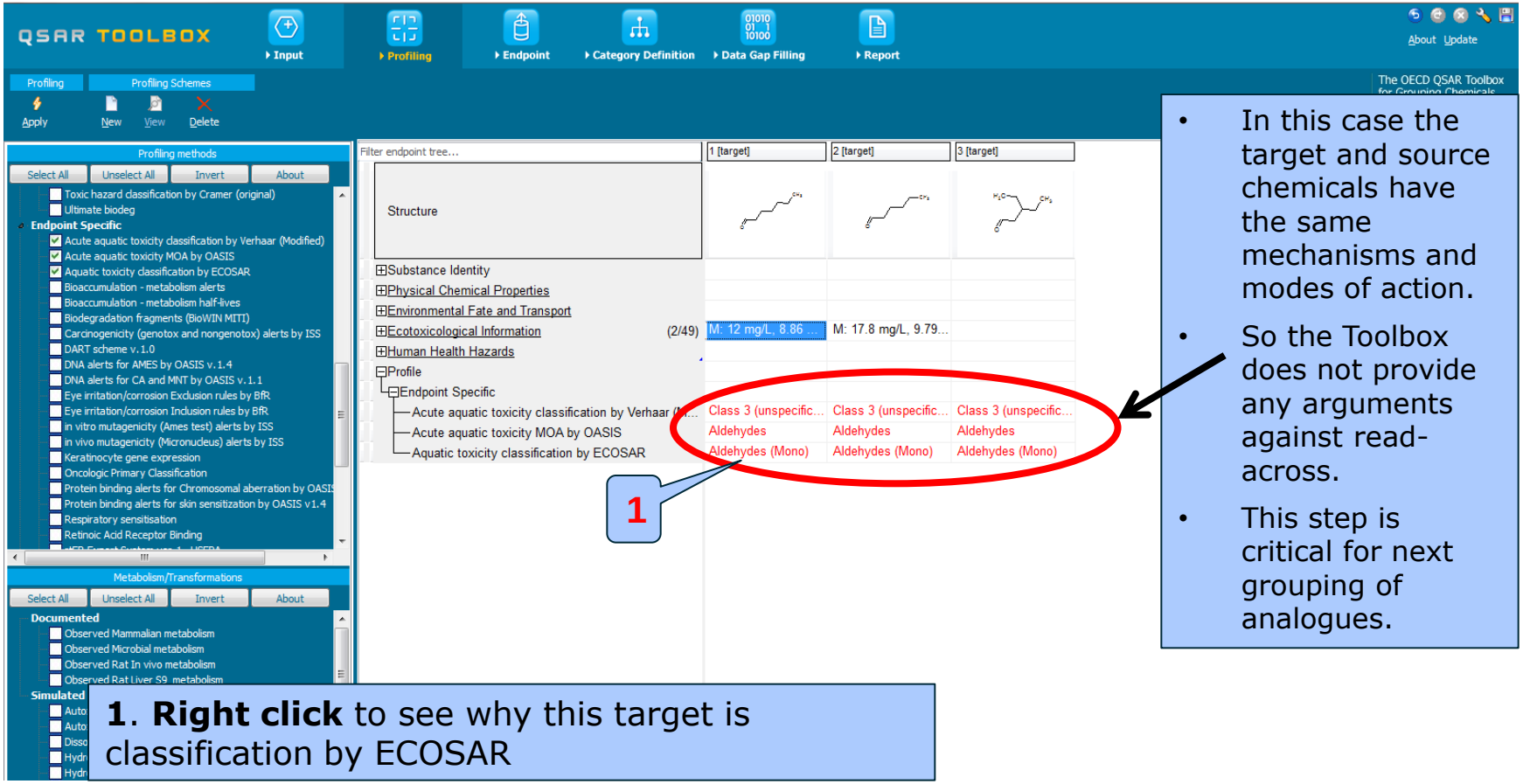


1. **Select** the profilers related to the target endpoint; **2**. **Click** Apply.

Profiling
Profiling the target chemical

- The actual profiling will take several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appeared as a dropdown box under the target chemical.
- The target and source chemicals have the same mechanisms or modes of action relevant for acute aquatic toxicity.
- The Toolbox does not provide any arguments against read-across (see next screen shot).

Profiling
Profiles of the targets



- In this case the target and source chemicals have the same mechanisms and modes of action.
- So the Toolbox does not provide any arguments against read-across.
- This step is critical for next grouping of analogues.

Aquatic toxicity classification by ECOSAR of "n-heptanal"

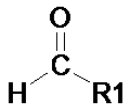
Category definitions

- Aquatic toxicity classification by ECOSAR
  - Acid Halides
  - Acid moiety
  - Acrylamides
  - Acrylates
  - Aldehydes (Mono)
  - Aldehydes (Poly)
  - Aliphatic Amines
  - Alkoxy Silanes
  - Amides
  - Anilines (amino-meta)
  - Anilines (amino-ortho)
  - Anilines (amino-para)
  - Anilines (Hindered)
  - Anilines (Unhindered)
  - Aziridines
  - Benzodioxoles
  - Benzotriazoles
  - Benzoylcyclohexanedione
  - Benzyl Alcohols
  - Benzyl Halides
  - Benzyl Imides
  - Benzyl Nitriles
  - Carbamate Esters
  - Carbamate Esters, Phenyl
  - Carbonyl Ureas
  - Diazoniums, Aromatic
  - Diketones
  - Epoxides, mono acid subst
  - Epoxides, mono
  - Epoxides, Poly
  - Esters (phosphate)
  - Esters (Phosphinates)
  - Esters
  - Esters, Dithiophosphates
  - Esters, Monothiophosphates
  - Halo Acids
  - Halo Alcohols
  - Halo Epoxides
  - Halo Ester
  - Halo Ethers
  - Halo Ketones (2 free H)
  - Halo Nitriles
  - Haloacetamides
  - Haloimides
  - Halopyrdines
  - Hydrazines
  - Hydroquinones
  - Imidazoles
  - Imides
  - Inorganic Compound

Profile Description

ECOSAR Class Definition: Aldehydes (Mono)

The Aldehydes (Mono) class is identified by the following structure:



R1 - attachment must be either an alkyl carbon, aromatic carbon, carbonyl or hydrogen.

The structure can contain only one aldehyde functional group to be classified as Aldehydes (Mono). If a structure contains more than one aldehyde group, it will be classified as Aldehydes (Poly).

If the R1 attachment is an olefinic carbon, acetylenic carbon or allyl group (-CC=C), the structure will be classified as a Vinyl/Allyl Aldehyde. In the current ECOSAR program, structures classified as Vinyl/Allyl Aldehydes are not additionally classified as Aldehydes (Mono).

SMILES String Identifications:
O=C or O=C[H] or O=CH
O=CC … (no other attachments to carbonyl)
O=Cc1ccccc1 … (no other attachments to carbonyl)

Associated ECOSAR Class(s):

Aldehydes (Poly) - If a structure contains more than one aldehyde group, it will be classified as Aldehydes (Poly) instead of Aldehydes (Mono)

Vinyl/Allyl Aldehydes - If the R1 attachment is an olefinic carbon, acetylenic carbon or allyl group (-CC=C), the structure will be classified as a Vinyl/Allyl Aldehyde. In the current ECOSAR program, structures classified as Vinyl/Allyl Aldehydes are not additionally classified as Aldehydes (Mono).

Example Aldehydes (Mono):

CAS No.	Name	SMILES Notation
75-07-0	Acetaldehyde	O=CC
123-72-8	Butanal	O=CCCC
555-16-8	Benzaldehyde, 4-nitro-	O=Cc(ccc(N(=O)(=O))c1)c1
148-53-8	o-Vanillin	O=Cc(c(O)c(OC)cc1)c1
454-89-7	Benzaldehyde, 3-(trifluoromethyl)-	O=Cc(cccc1C(F)(F)F)c1

Profiling
Recap

- You have entered the source and target chemicals being sure of the correct structures.
- You have checked the relevant databases for available experimental results.
- You have profiled the source and target chemicals.
- You have evaluated the robustness of the analogue approach and concluded that the read-across may be acceptable.

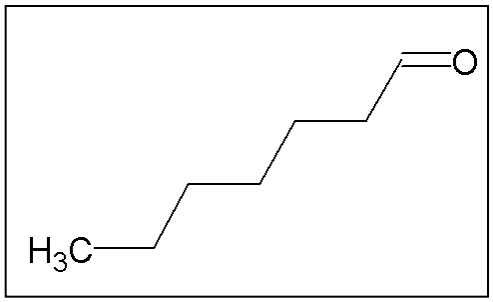
Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise 1
- **Workflow of the exercise 2**
  - **Chemical Input**
  - Profiling
  - Endpoints

Chemical Input
Exercise 2

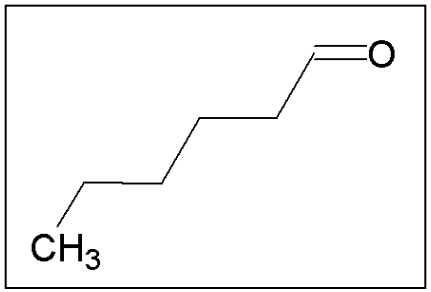
Read-across of acute toxicity to fish from 1-hepatanal and 1-hexanal to 2,5-diene-4-methyl-hexan-1-al.

1-heptanal

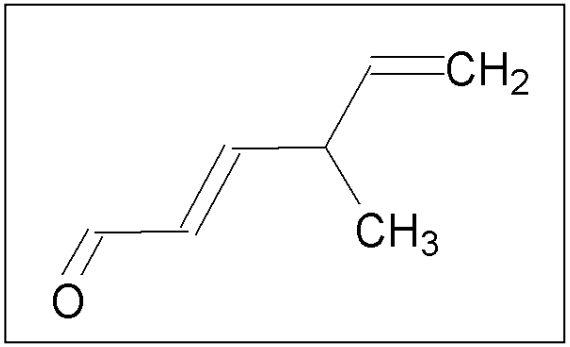


+

1-hexanal



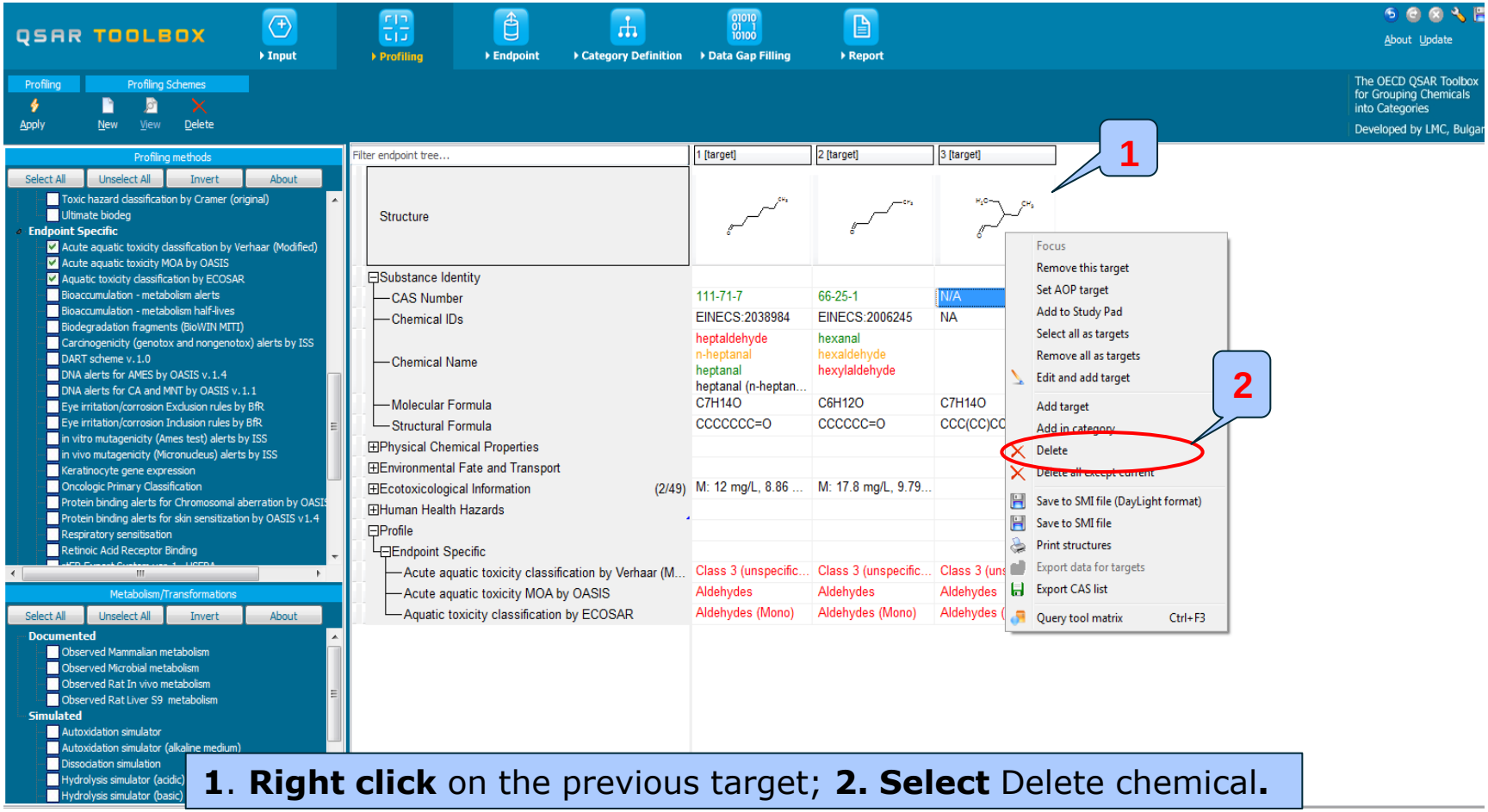
2,5-diene-4-methyl-hexan-1-al



Chemical Input
Exercise 2

- In the second example, we use the same source chemicals and a different target chemical.
- We can therefore simply delete the previous target chemical and enter the identity of the new target chemical.
- Right-click above the structure of chemical 3-ethyl-1-pentanal and select "Delete chemical" (see next screen shot).

Chemical Input
Exercise 2

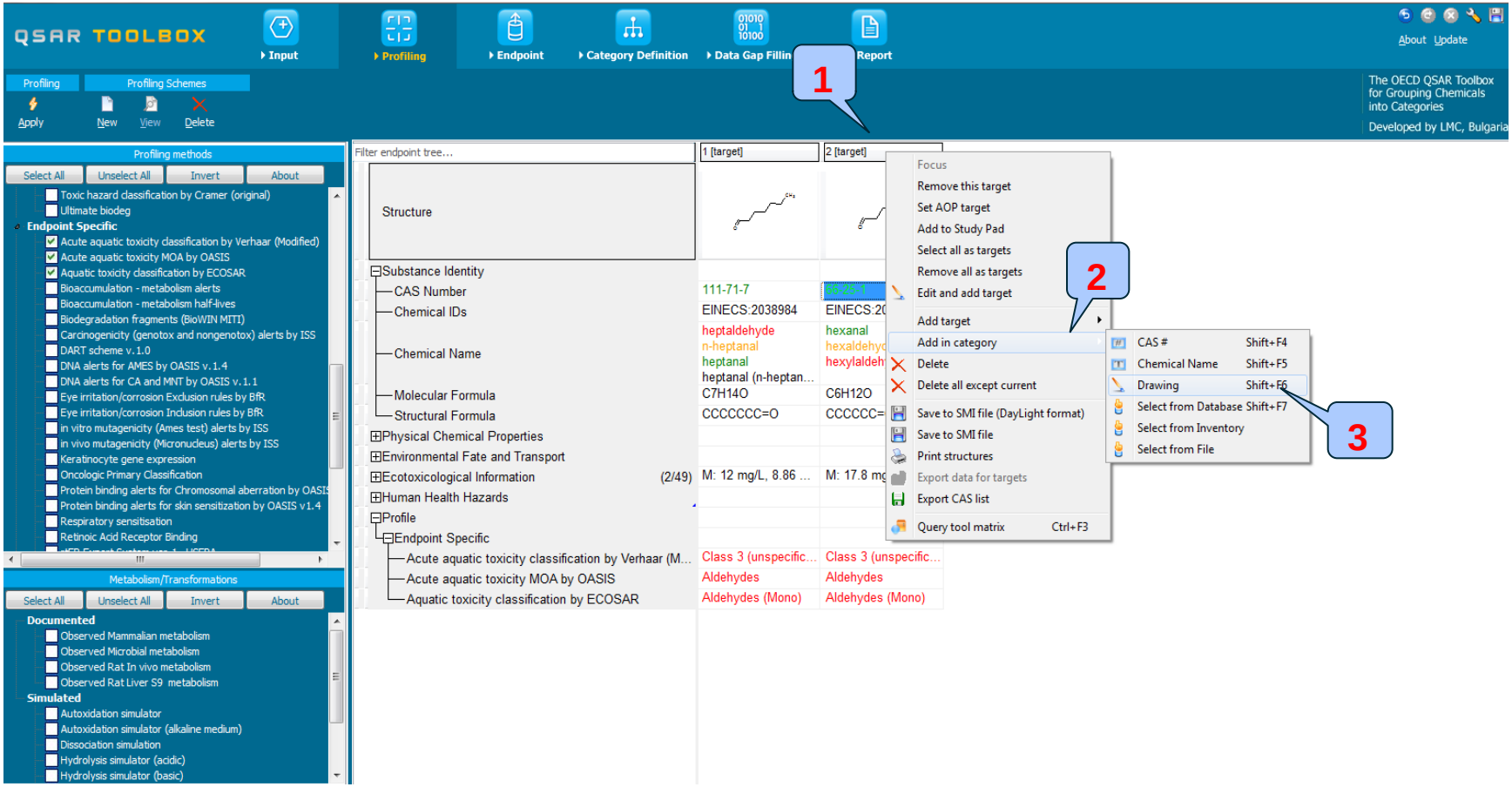


1. Right click on the previous target; **2. Select** Delete chemical**.**

Chemical Input
Exercise 2

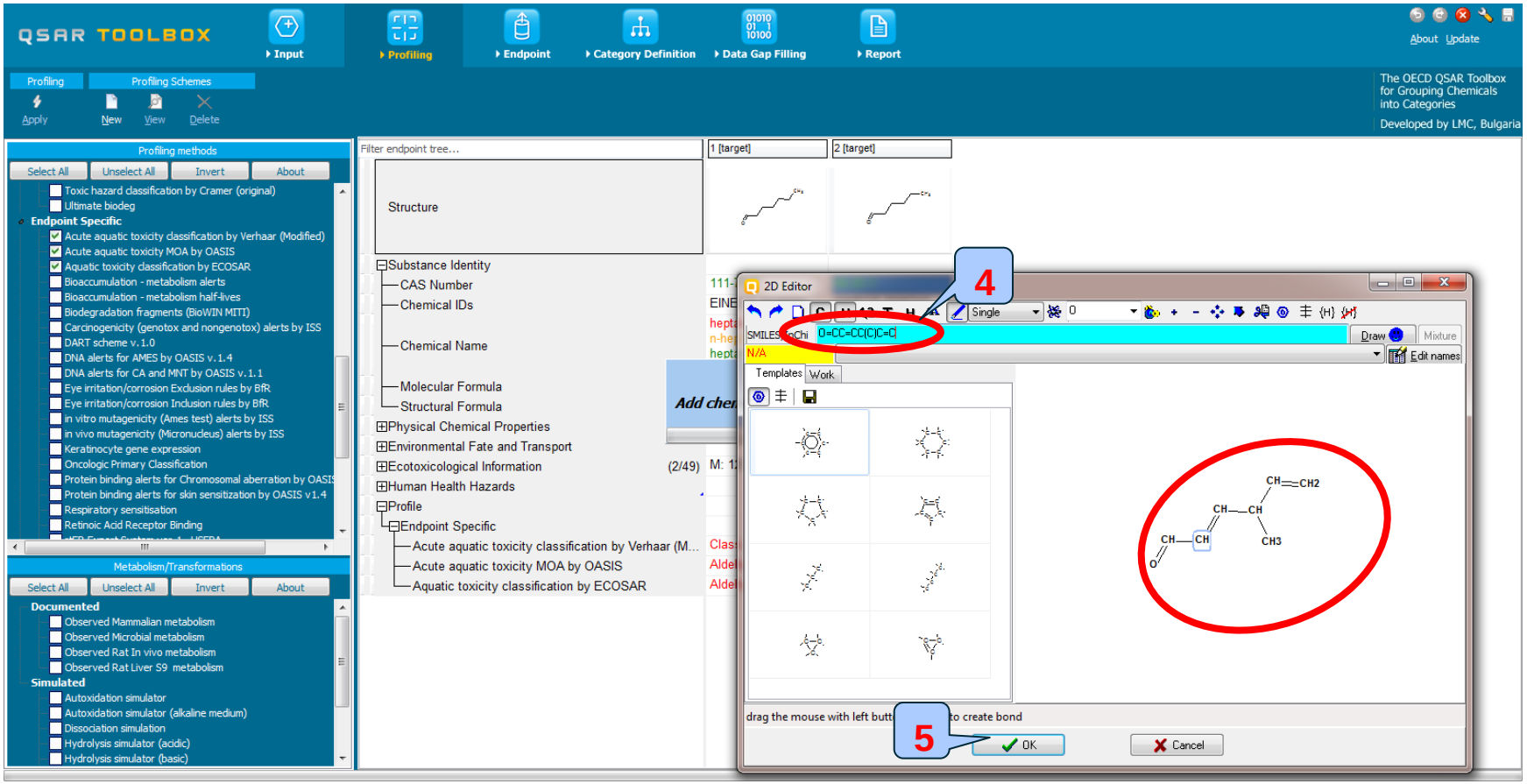
- Add the new target chemical as in the previous exercise.
- To add the third chemical by hand into the matrix, right-click above the structure and select "Add category" and then "Drawing".
- Enter the SMILES for 2,5-diene-4-methyl-hexan-1-al: O=CC=CC(C)C=C and click "OK".

Chemical Input
Exercise 2



1. **Right-click** above the structure; **2**. **Select** Add category; **3**. **Select** Drawing

Chemical Input
Exercise 2



4. **Type** O=CC=CC(C)C=C in SMILES/InChi window; **5**. **Click** OK.

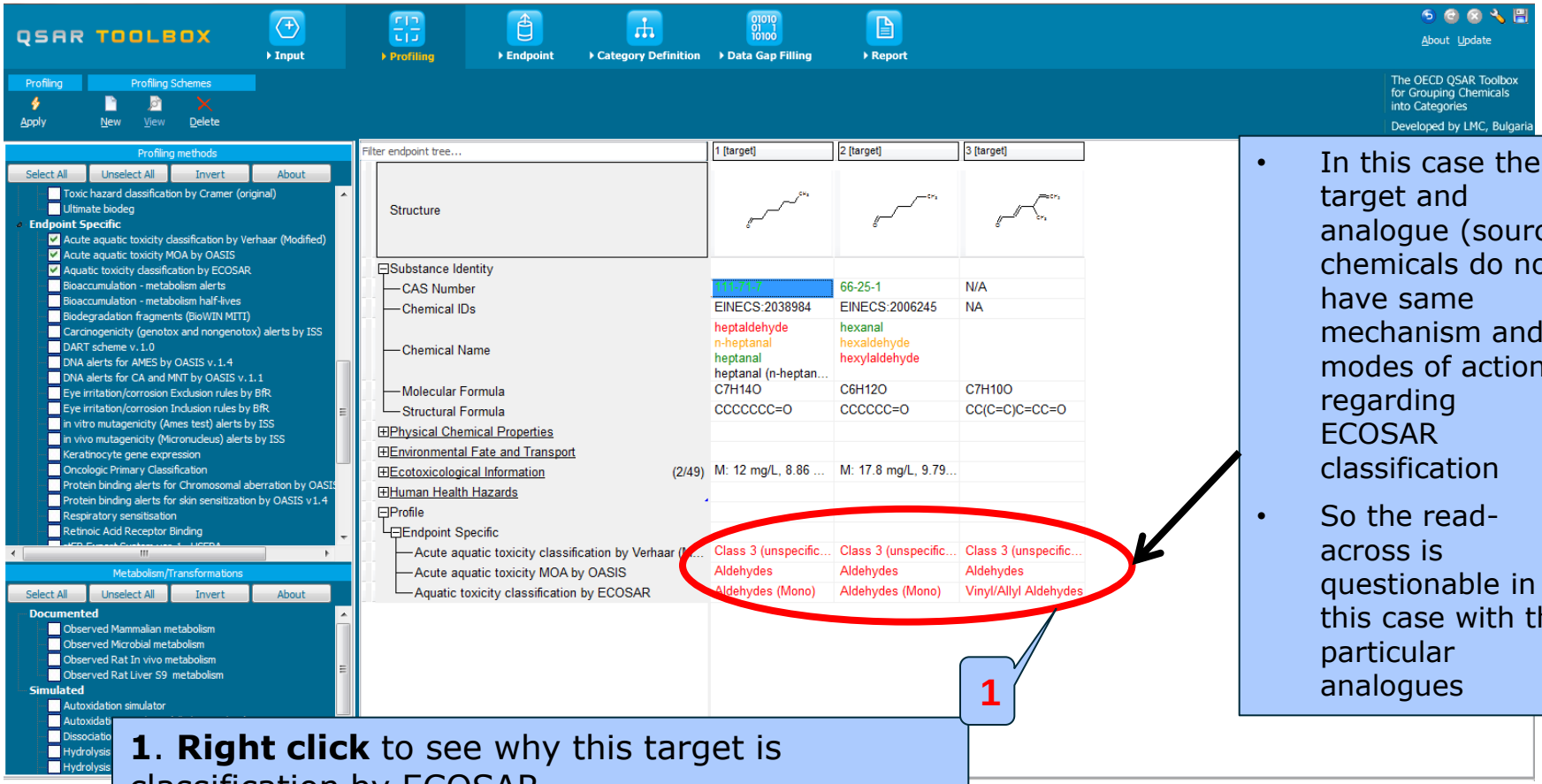
Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise 1
- **Workflow of the exercise 2**
  - Chemical Input
  - **Profiling**
  - Endpoints

Profiling

- In the module profiling, profile the new target chemical with the 3 profilers relevant for aquatic toxicity, in the same way as for the previous example.

Profiling
Profiles of the targets



- In this case the target and analogue (source) chemicals do not have same mechanism and modes of action, regarding ECOSAR classification
- So the read-across is questionable in this case with this particular analogues

1. **Right click** to see why this target is classification by ECOSAR

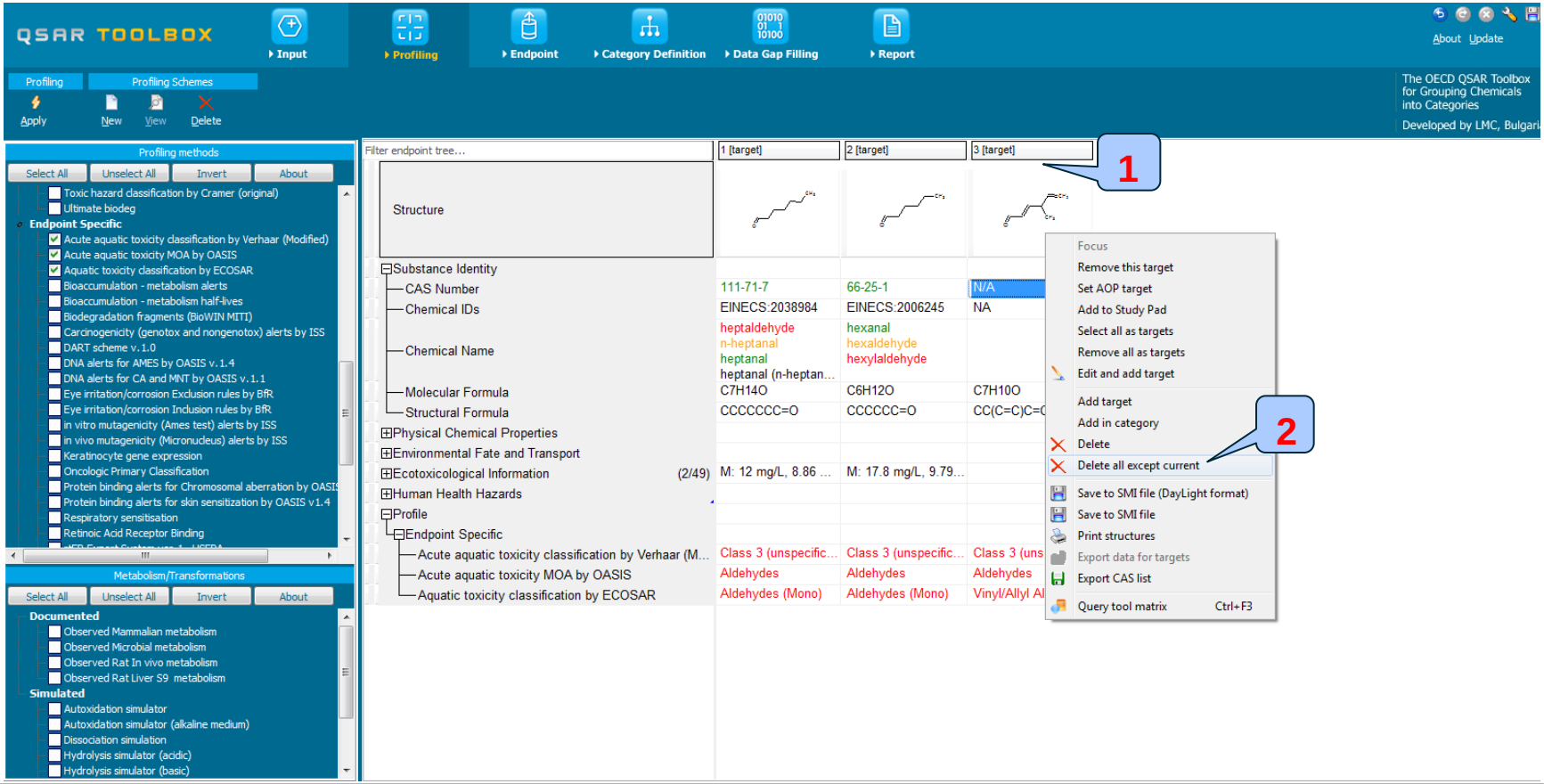
Recap

- You have replaced a target chemical with another target chemical in the data matrix.
- You have profiled the new target chemical.
- You have evaluated the robustness of the analogue approach and concluded that the read-across may not be acceptable by using the current analogue chemicals (source)
- The further workflow is to search for more suitable target analogues

Searching for More Suitable Analogues

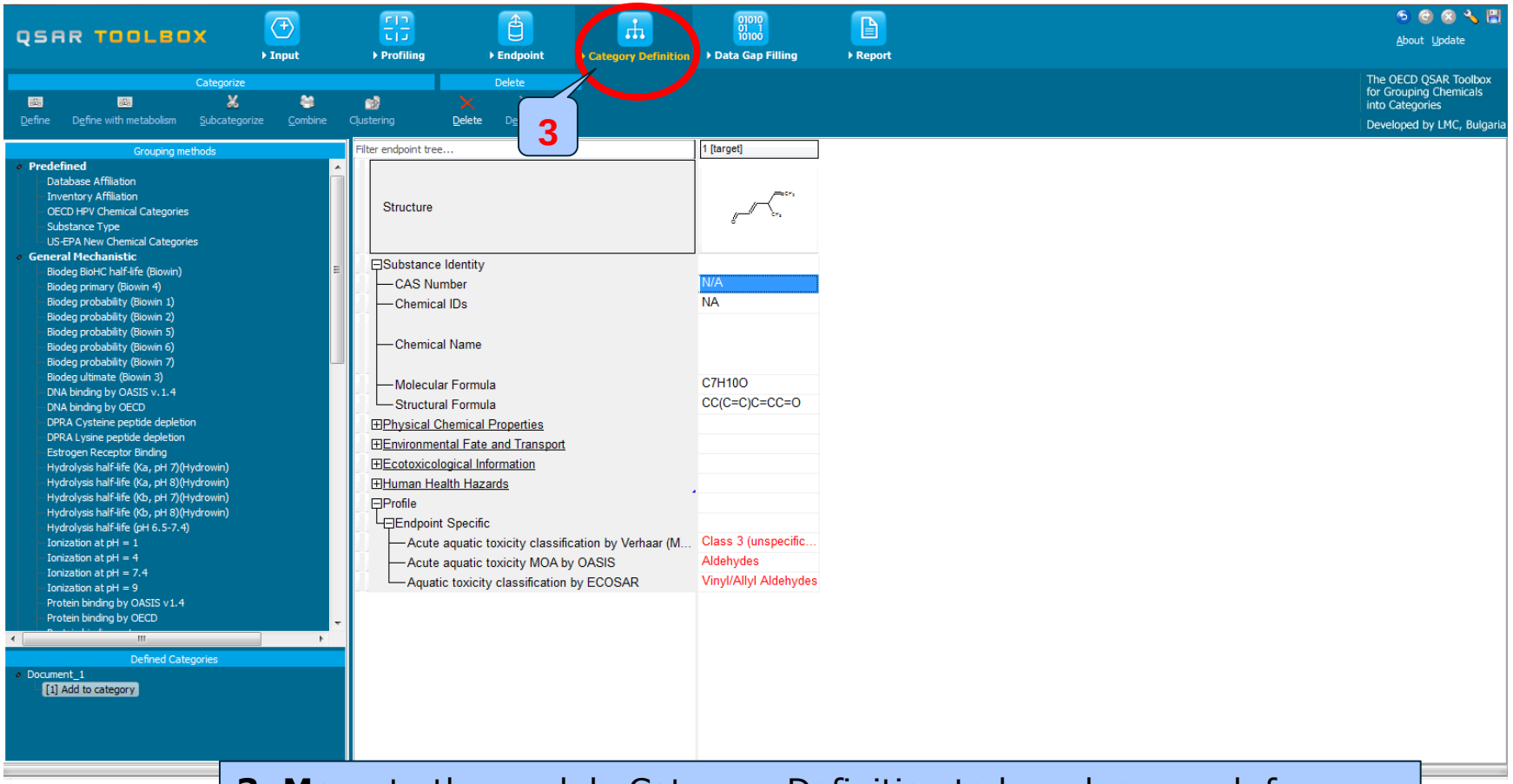
- Before searching for more suitable analogues, delete n-hexanal and n-heptanal from the data matrix by right-clicking above each of them and select "Delete chemical" or right-clicking above the target (2,5-diene-4-methyl-hexan-1-al) and select "Delete all except current" (see next screen shot).
- The aim of the next part of the exercise will be to find analogues which have the same profiling results as the target chemical.

Searching for More Suitable Analogues



1. **Right click** in the space above the target chemical; **2**. **Select** Delete All except current.

Searching for More Suitable Analogues



3. Move to the module Category Definition to launch a search for more suitable analogues.

Outlook

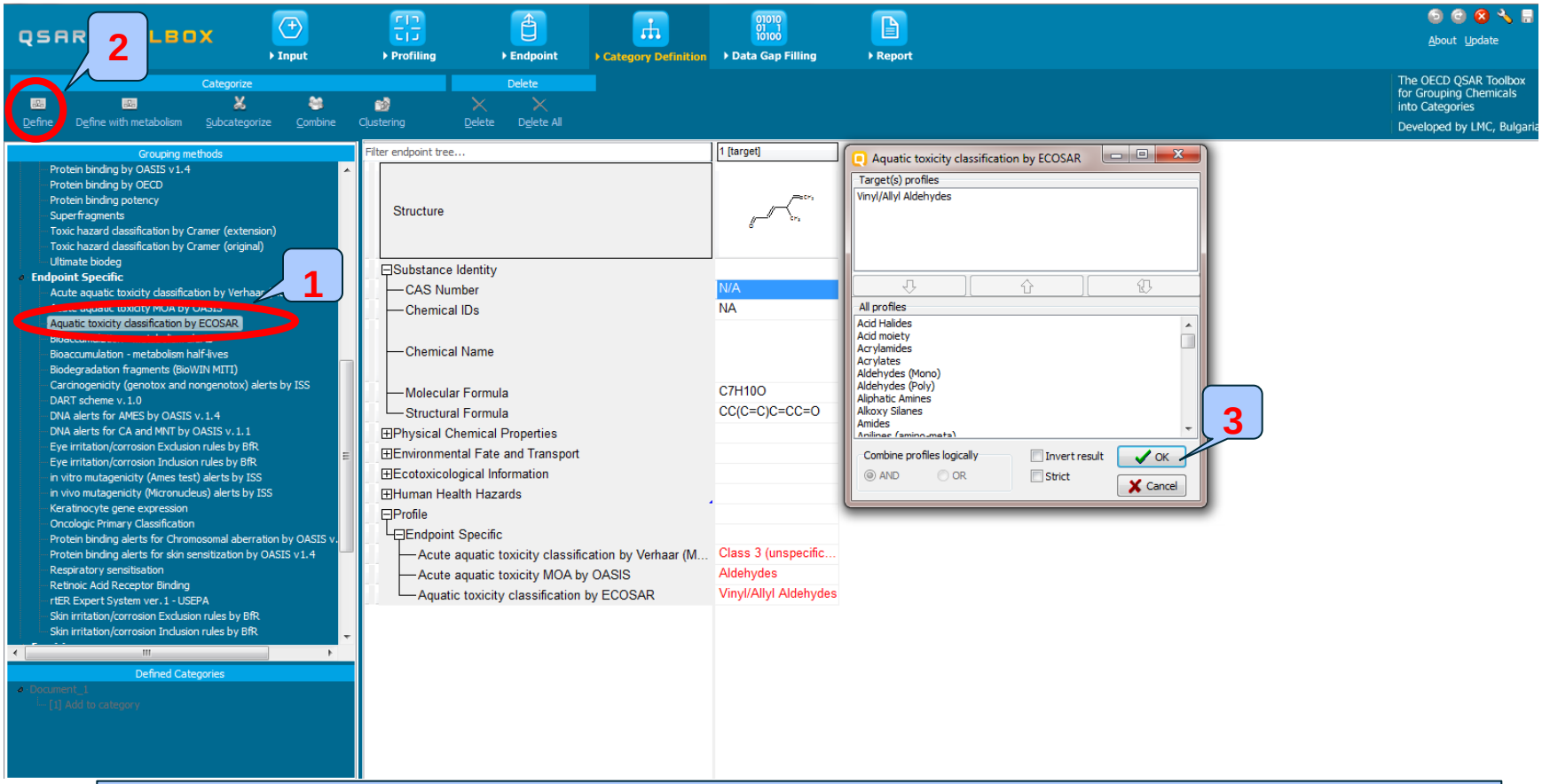
- Background
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise 1
- **Workflow of the exercise 2**
  - Chemical Input
  - Profiling
  - Endpoints
  - **Category definition**

Category definition
Searching for More Suitable Analogues

- Currently it is not possible to query directly by several profiling results in parallel. The user has first to query according to one profiler and then subcategorise the results step-by-step according to other profilers.
- For this example, the user could first select the ECOSAR profiler of the target chemical and query for all the chemicals with the same structural feature in the selected databases (see next screen shot).

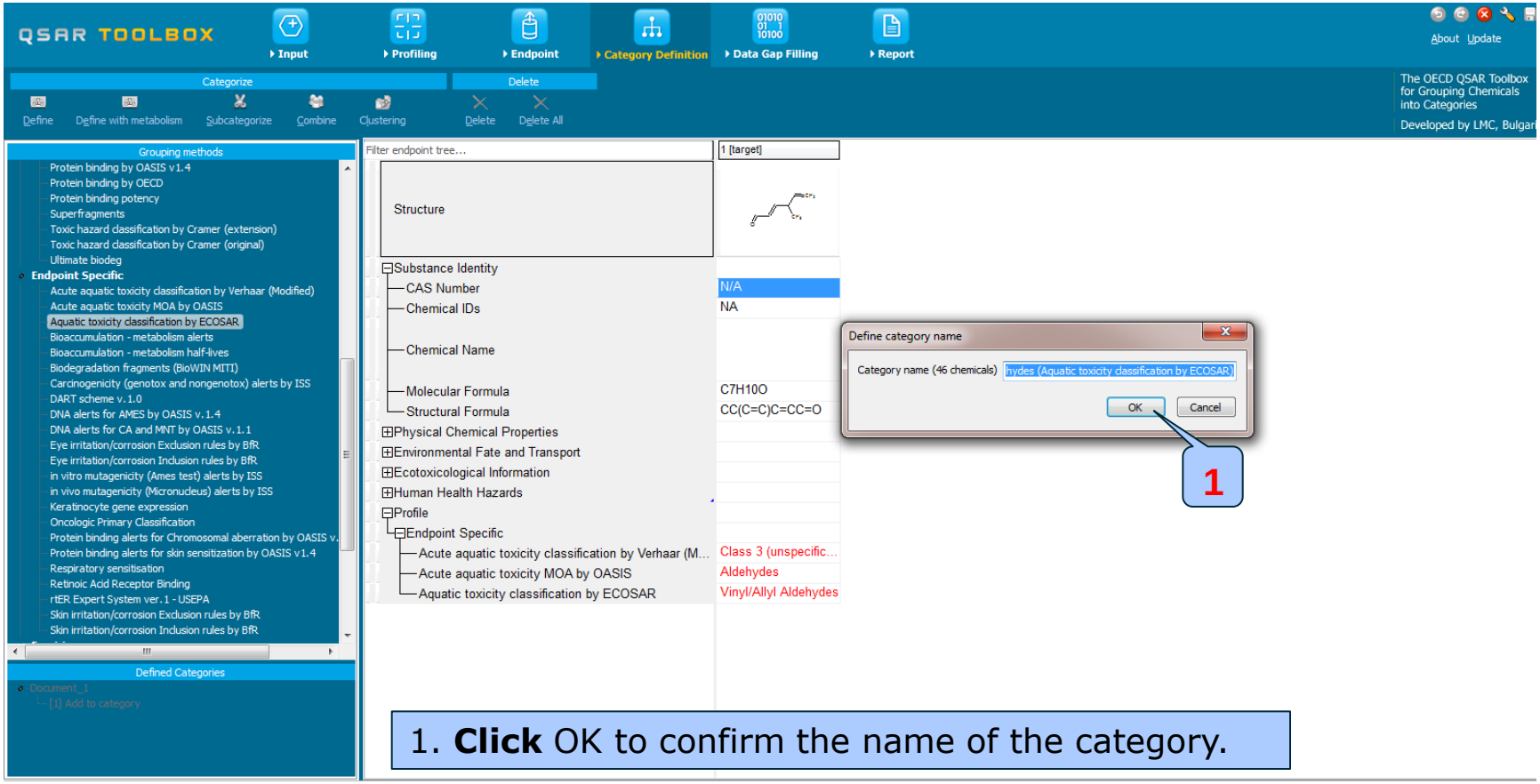
Category definition

Searching for More Suitable Analogues



1. **Highlight** "Aquatic toxicity classification by ECOSAR"; **2**. **Click** Define and confirm the category from classification by ECOSAR profiler; **3**. **Click** OK.

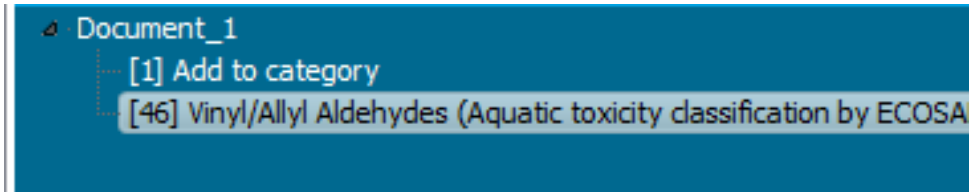
Category definition
Searching for More Suitable Analogues



Category definition
Searching for More Suitable Analogues

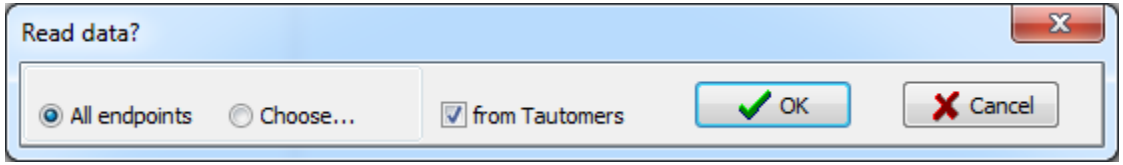
- The Toolbox now identifies all chemicals corresponding to mechanism "Vinyl/Allyl Aldehydes" by Aquatic toxicity classification by ECOSAR listed in the databases selected under "Endpoints".
- 45 analogues are identified. Along with the target they form a mechanistic category used for gap filling.

- The name of the category appears in the "Defined Categories" window, indicating the number of substances belonging to the category.



Category definition
Reading data for Analogues

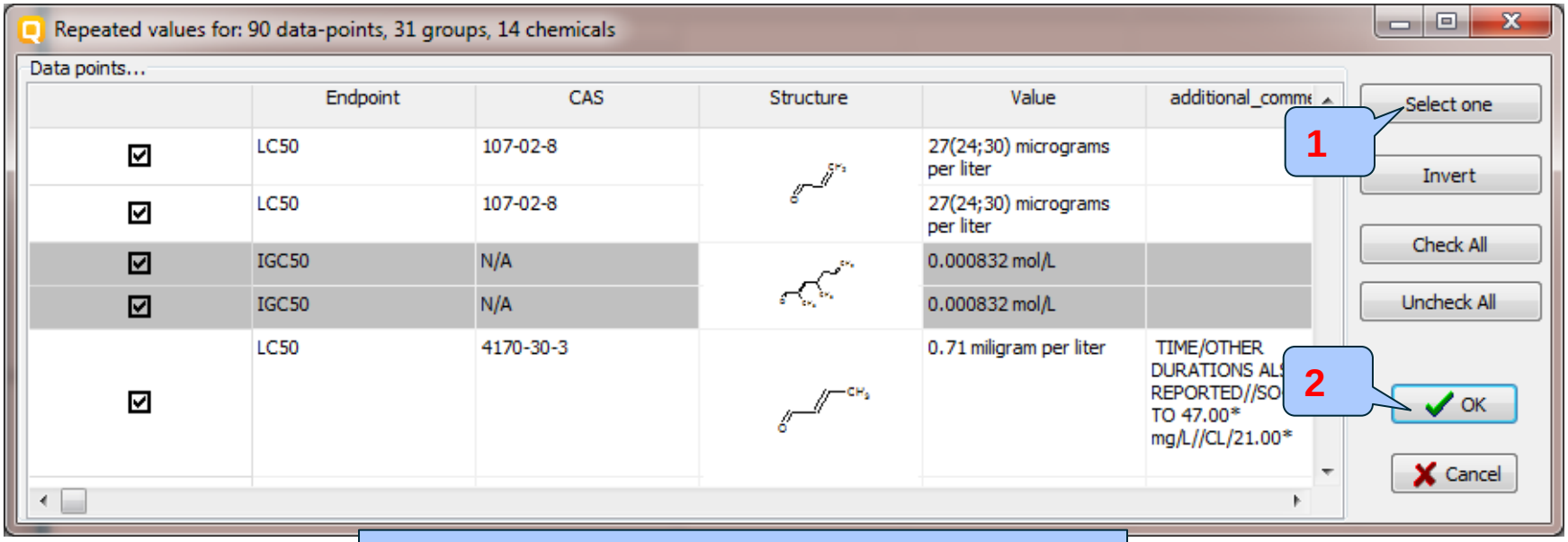
- The Toolbox will now retrieve those chemicals that have the same structural functionality as the target chemical based on ECOSAR profiler (Vinyl/Allyl aldehydes).
- The Toolbox automatically request the user to select the endpoint that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).



- In this example, as only databases are selected that contain information for aquatic toxicity endpoints, both options give the same results.

Category definition
Reading data for Analogues

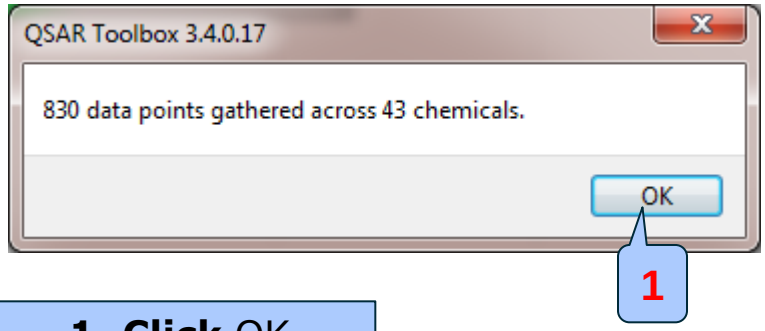
Due to the overlap between the Toolbox databases same data for intersecting chemicals is found simultaneously in more than one database. The data redundancy is identified and the user has the opportunity to select either a single data value or all data values.



1. **Click** Select one and then **2**. **Click** OK.

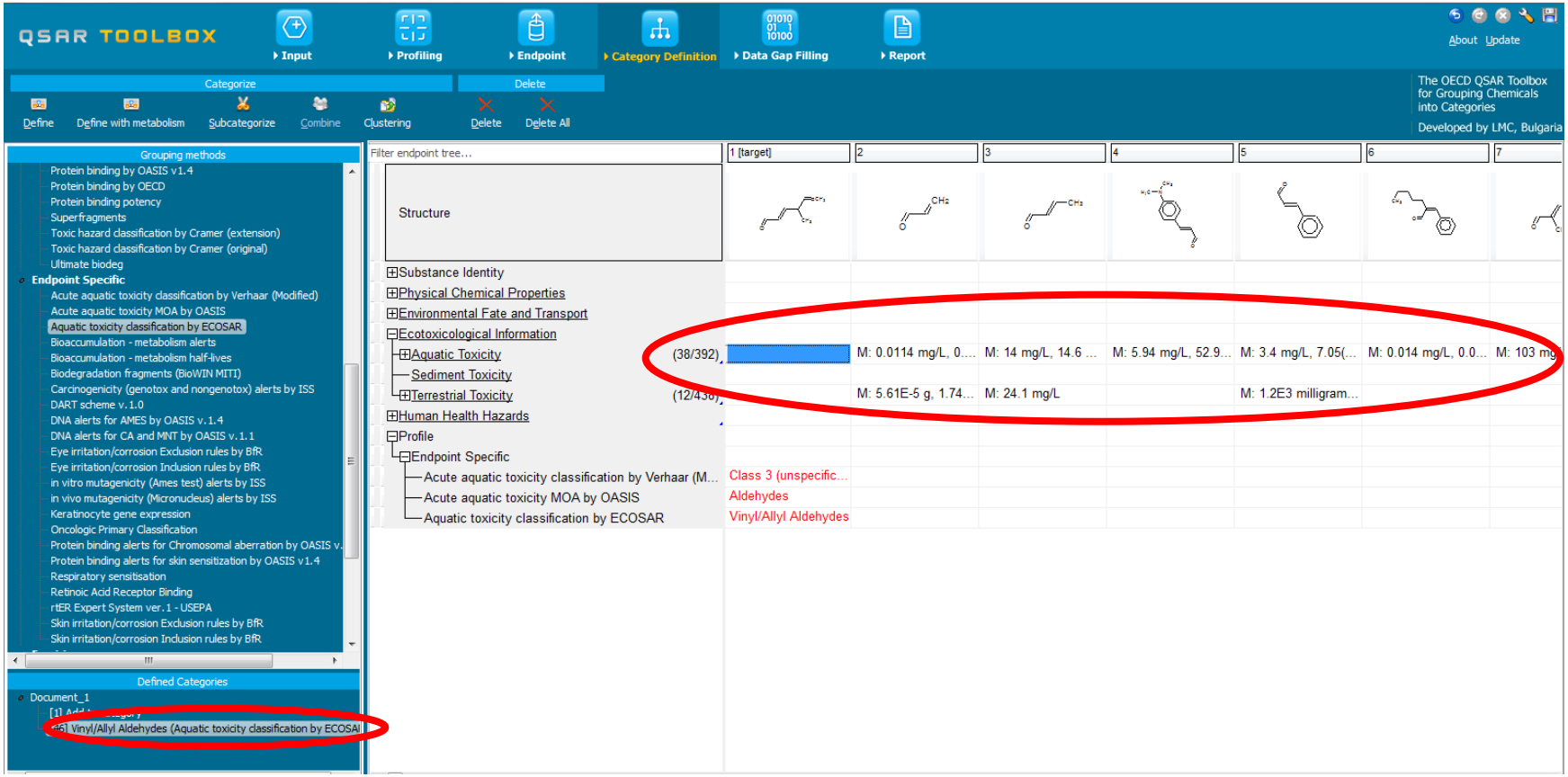
Category definition
Reading data for Analogues

The system automatically gives indication for the number of gather experimental data points



1. **Click** OK

Category Definition
Defined category

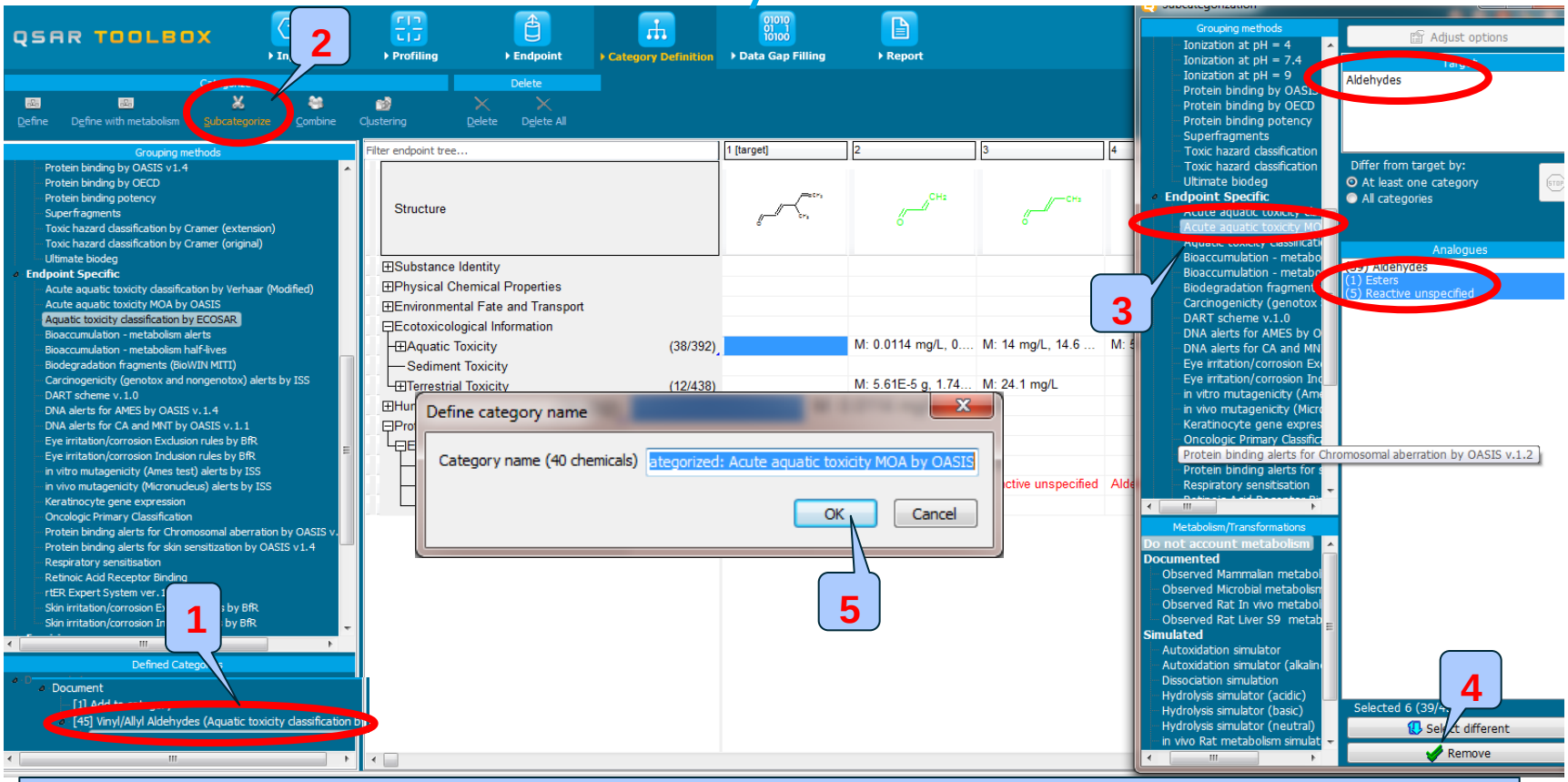


Category Definition
Subcategorisation

- After the available data has been retrieved, the user can then further subcategorize the results according to the following subcategorisations:
  - MOA of action
  - Verhaar classification
- These steps are summarized in the next screen shots.

Category Definition

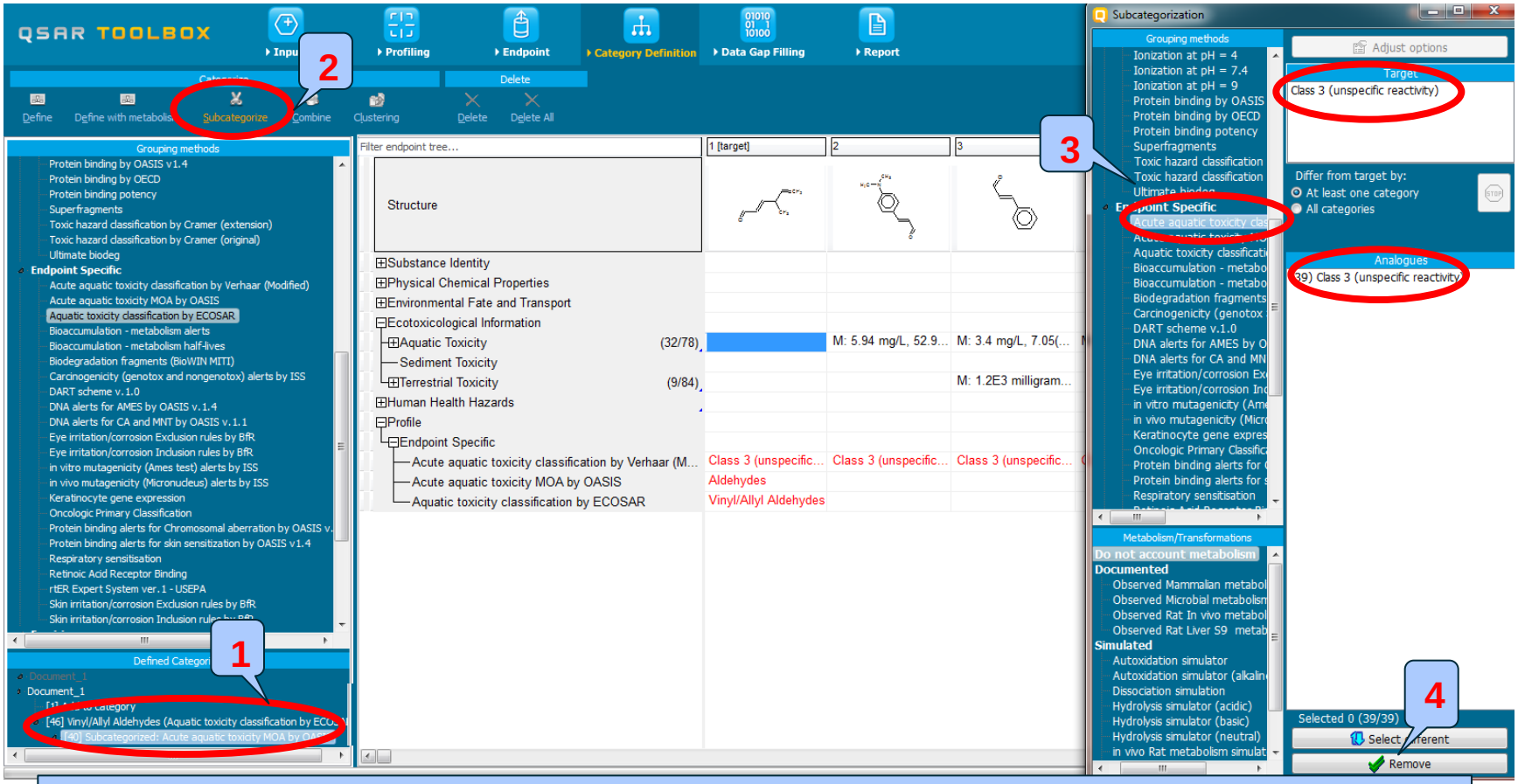
Subcategorisation by Acute aquatic toxicity MOA by OASIS



1. Select current category; **2. Click** Subcategorize; **3. Select** Acute aquatic toxicity MOA by OASIS profiler; **4.** Remove dissimilar chemicals and **5.** Confirm new category by **clicking** OK.

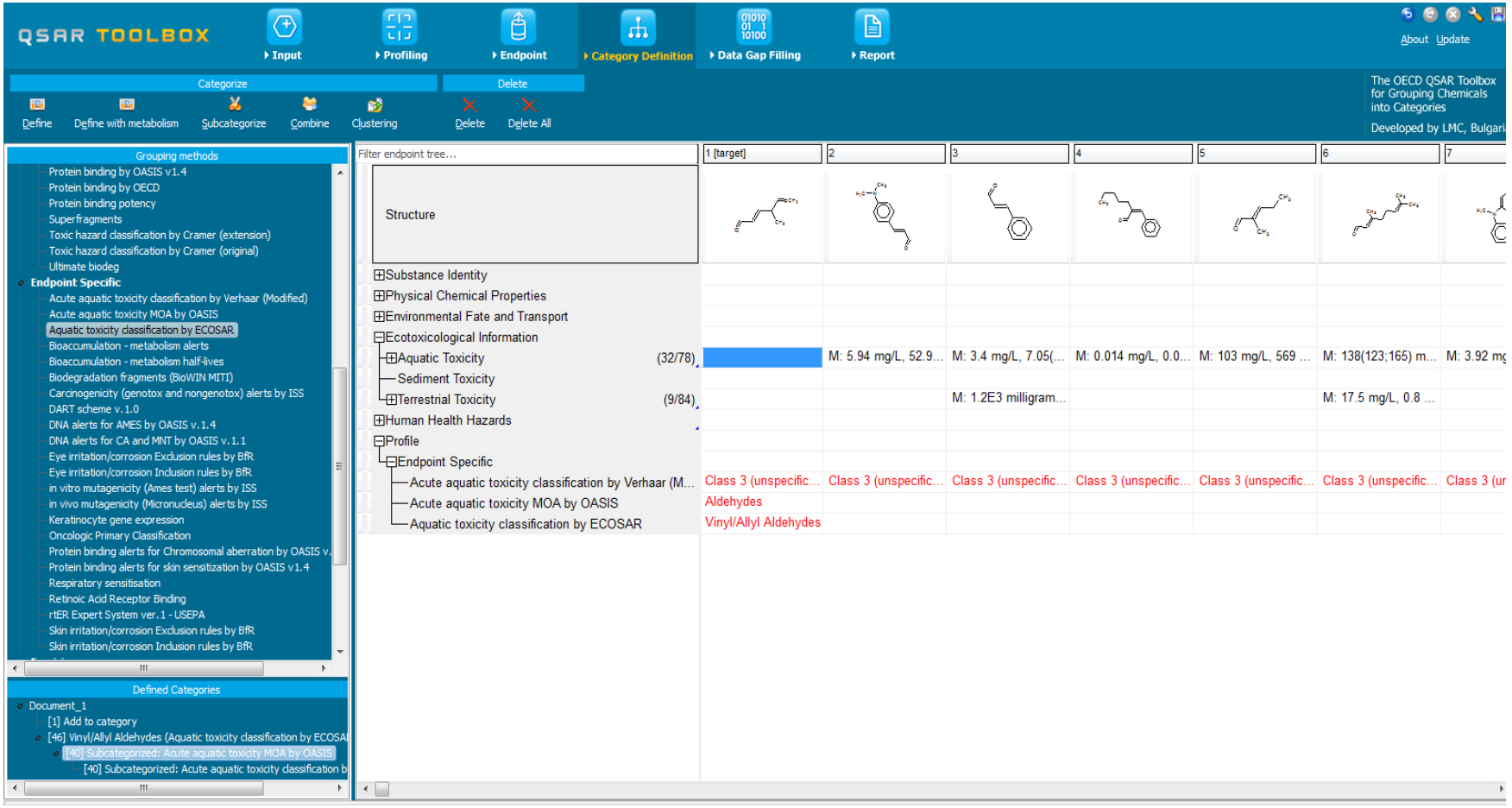
Category Definition

Subcategorisation by Acute aquatic classification by Verhaar(Modified)



1. **Select** category with 39 analaogues; **2**. **Click** Subcategorise; **3**. **Select** Verhaar profiler; Note all analogues are in the same category as the target chemical so no further action is required.

Category Definition
Results after subcategorisation



Category Definition
Interpretation of the results

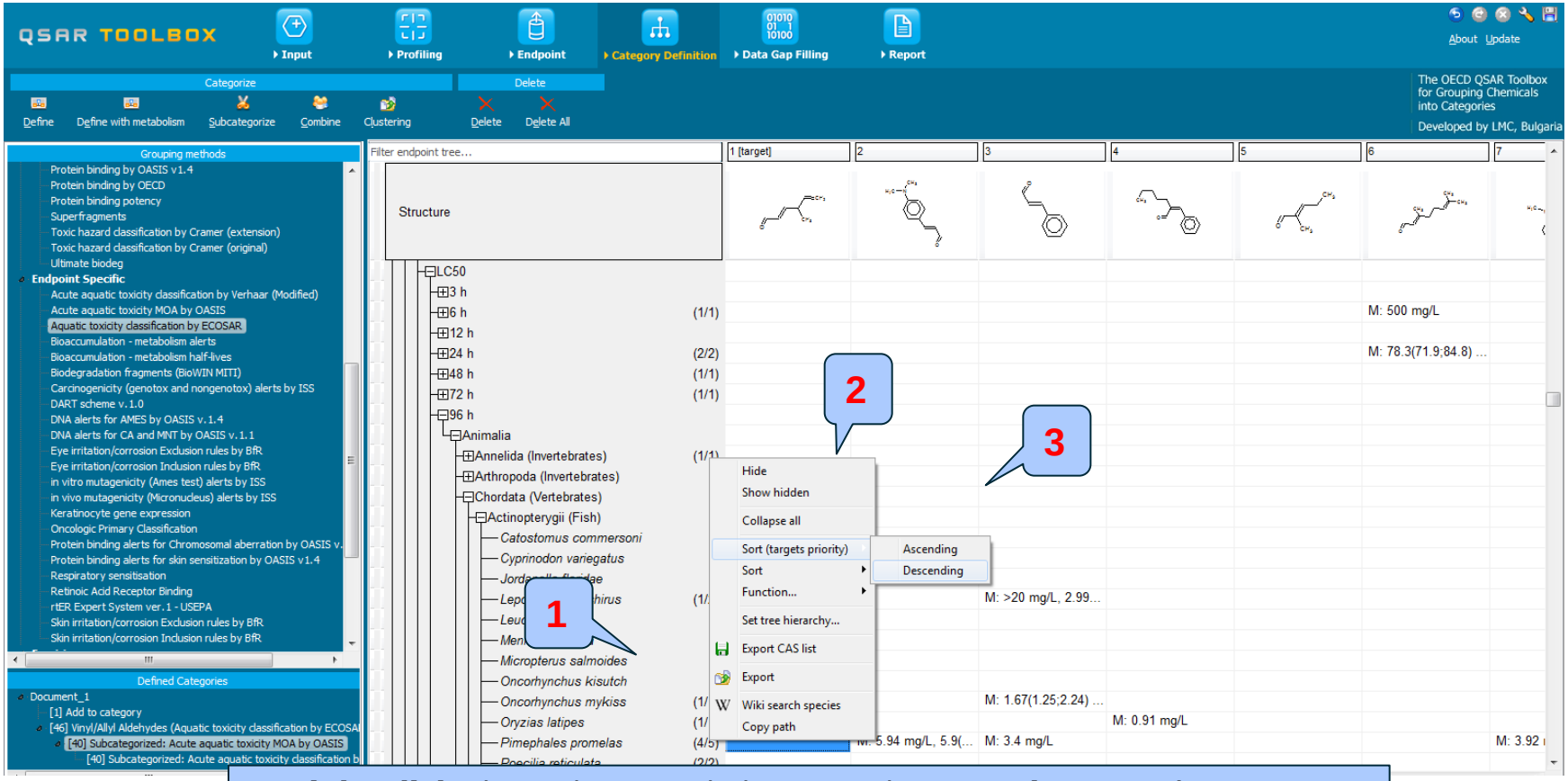
- Following the above-described subcategorisation exercise, 39 chemicals are left in the category. All have same mechanisms of action.
- The result is a group of chemicals which are classified as Vinyl/Allyl class by ECOSAR category and have same mode of action according to the MOA profiler.
- For 4 chemicals, experimental results for acute toxicity to fish are available- 4 chemicals have 96h-LC50 results from 3.4 to 7.29 mg/l for *Pimephales promelas;* 2 chemicals have 96h-LC50 results from 7.62 to 9.81mg/l for *Poecilia reticulata*; 1 chemical has 96h-LC50 0.91mg/l for *Oryzias latipes*.

  (see next two screen shots)

Category Definition

Interpretation of the results

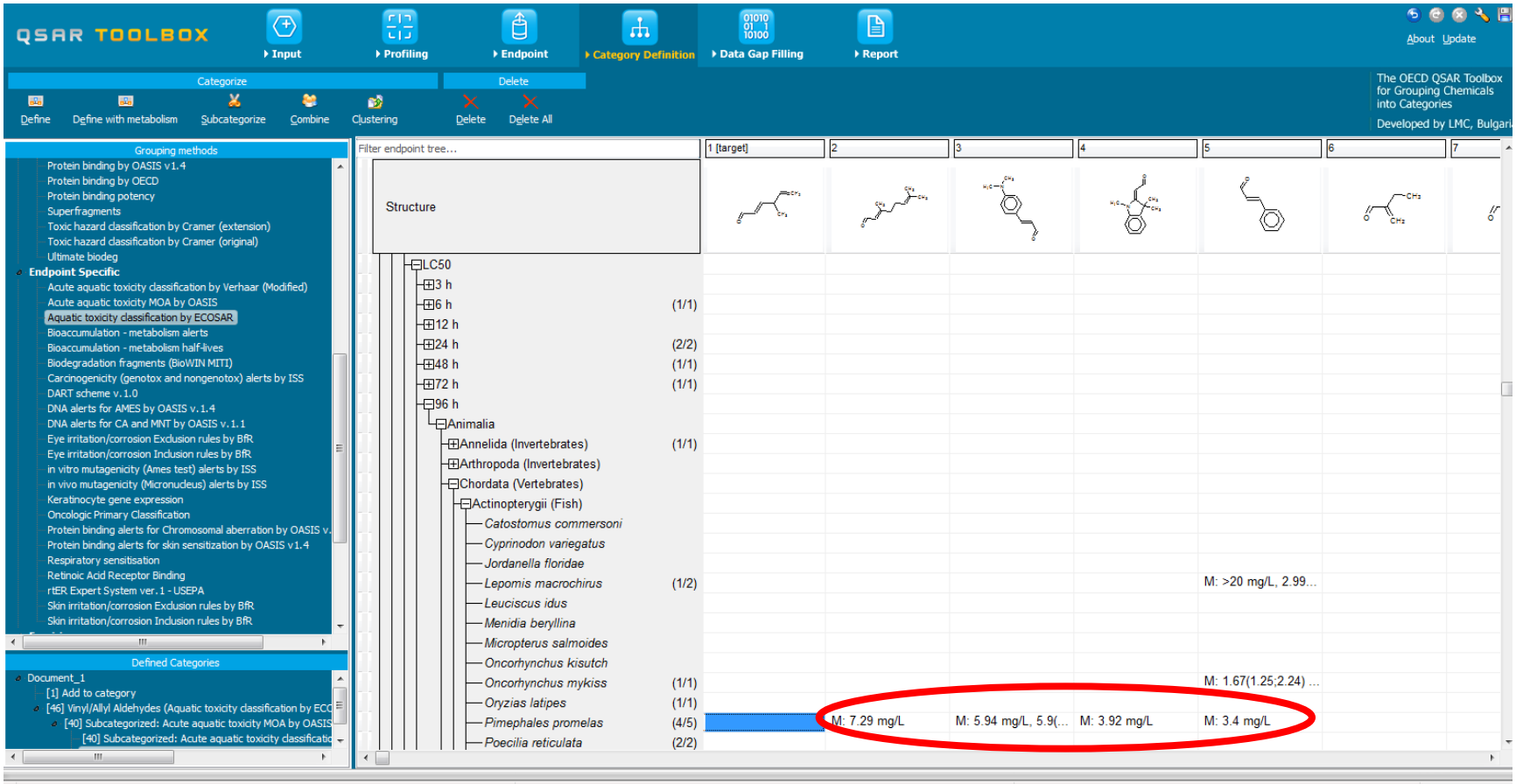
96h-LC50 for *Pimephales promelas*



1. Right click above the Pimephales promelas; **2. Select** Sort (targets priority), then **3.** Descending.

Category Definition
Interpretation of the results

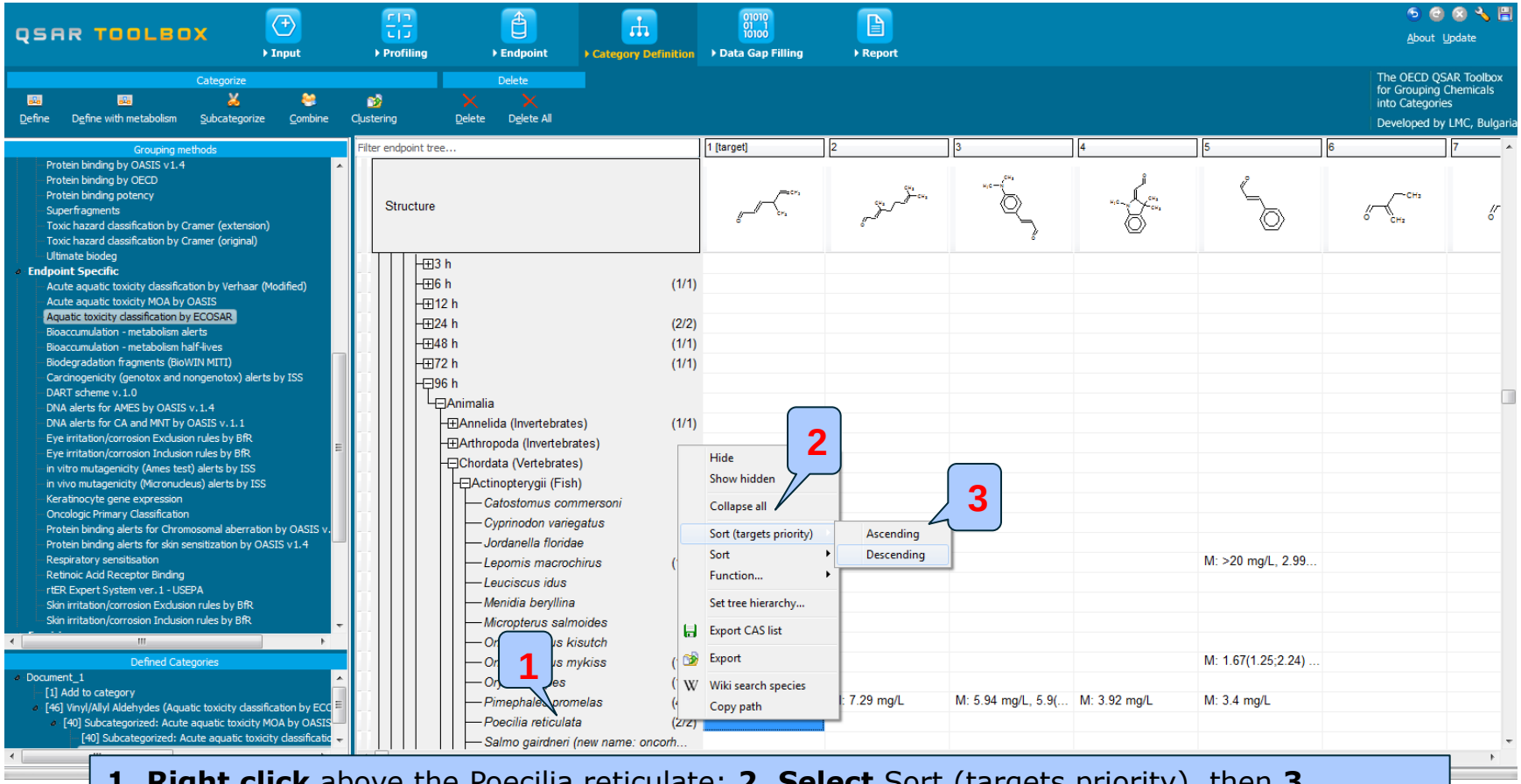
96h-LC50 from 3.4 to 7.29 mg/l for *Pimephales promelas*



Category Definition

Interpretation of the results

96h-LC50 results for *Poecilia reticulata*

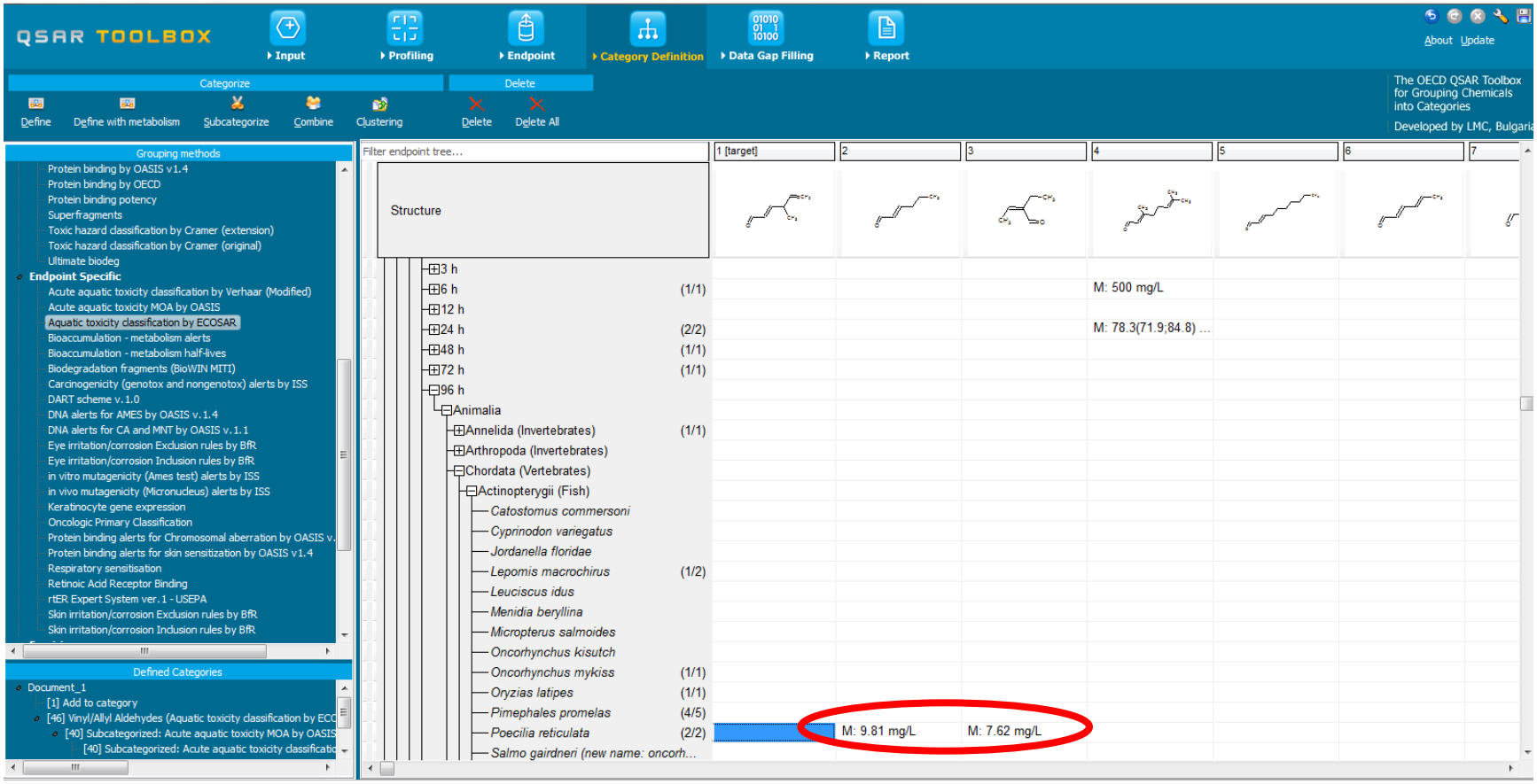


1. Right click above the Poecilia reticulate; **2. Select** Sort (targets priority), then **3.** Descending**.**

Category Definition

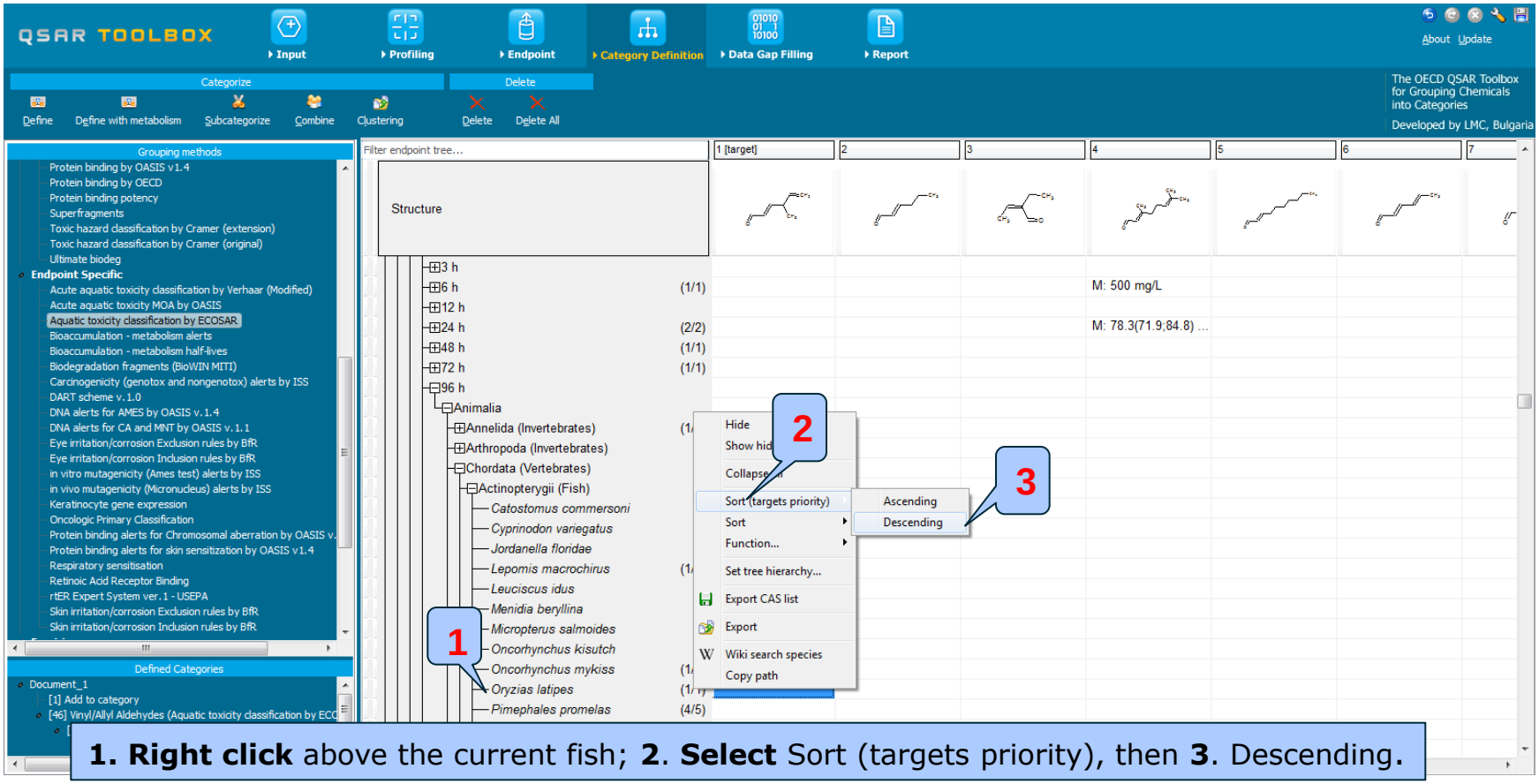
Interpretation of the results

96h-LC50 results from 7.6 to 9.8 mg/l for *Poecilia reticulata.*



Category Definition
Interpretation of the results

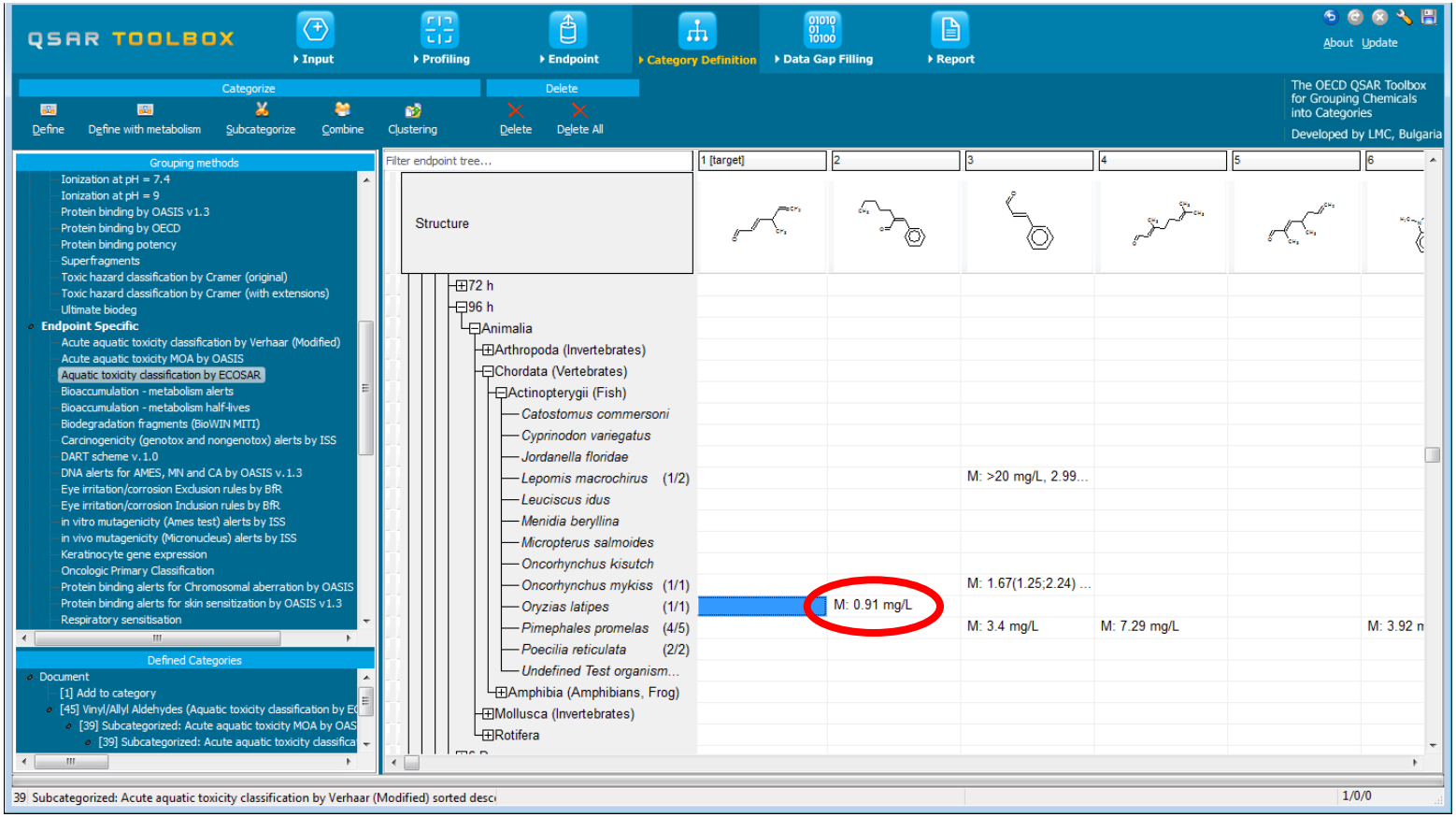
96h-LC50 results for *Oryzias latipes*



1. Right click above the current fish; **2. Select** Sort (targets priority), then **3.** Descending.

Category Definition
Interpretation of the results

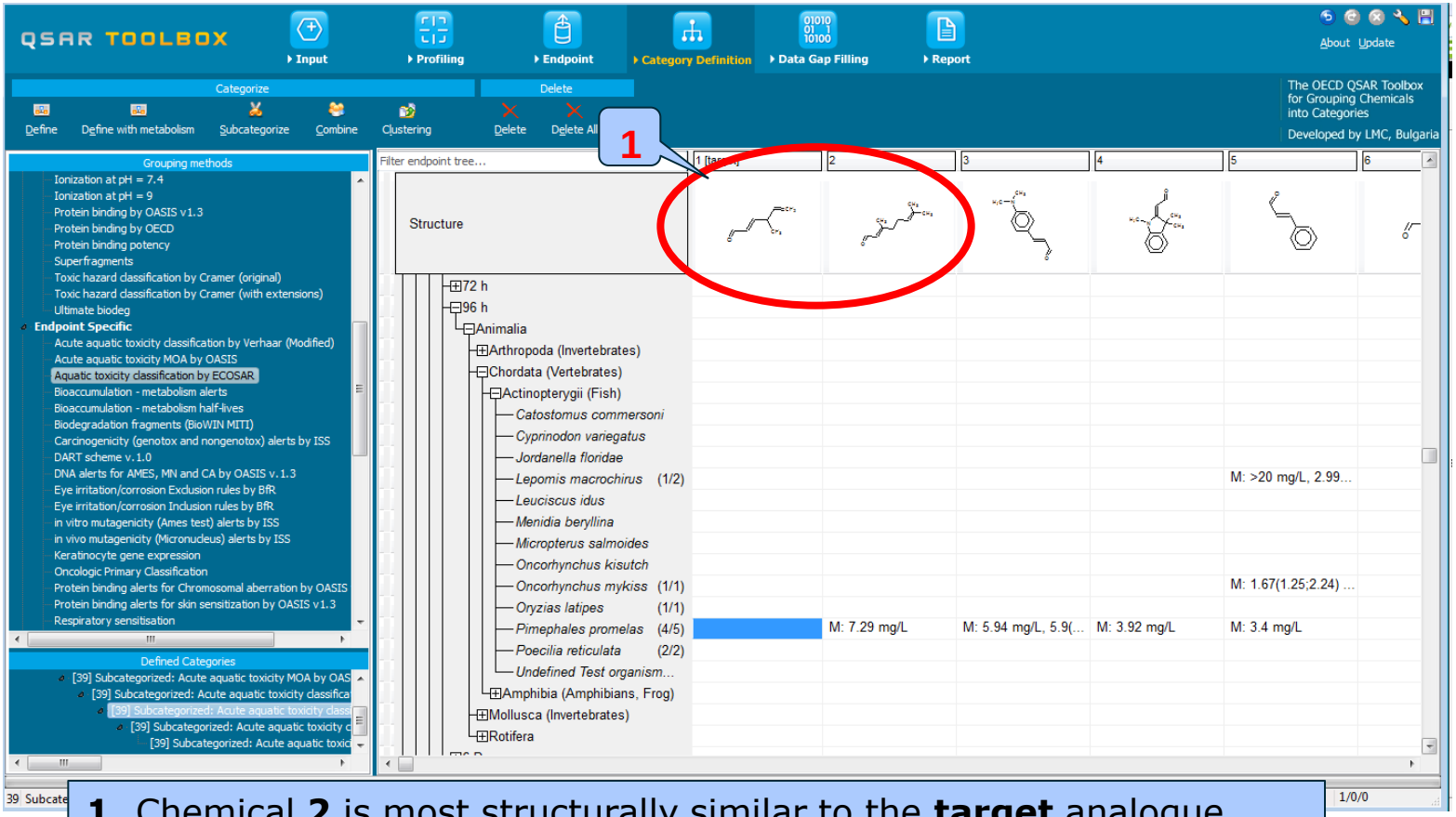
96h-LC50 results for *Oryzias latipes*



Category Definition
Interpretation of the results

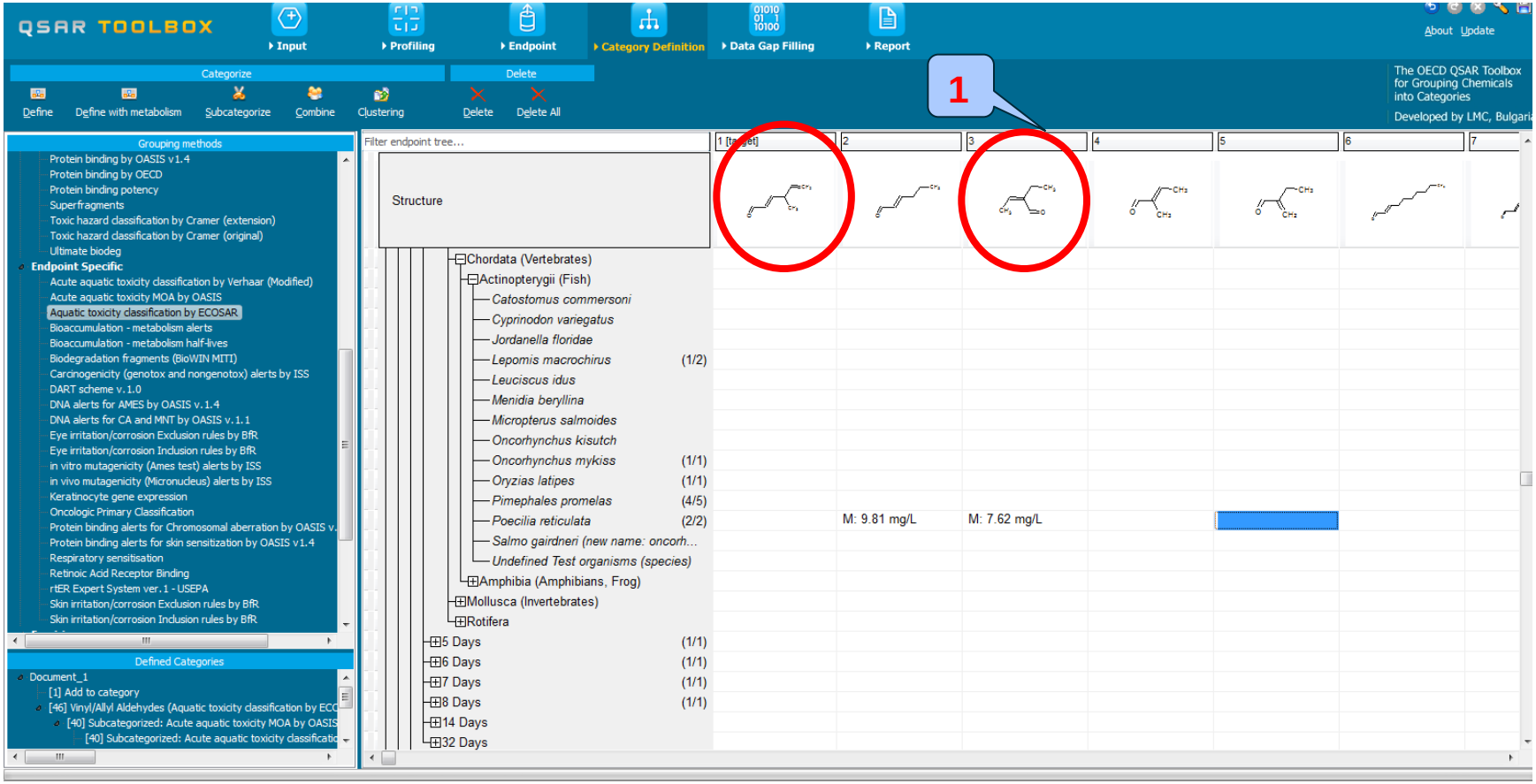
- Further visual analysis of the structures (see next two screen shots) could indicate that the following results are most suitable for read-across:
- *Pimephales promelas* : (E)-3,7-Dimethyl-2,6-octadienal - 96hLC50 = 7.3 mg/l
- *Poecilia reticulata* : 2-Ethyl-2-butenal - 96hLC50 = 7.6 mg/l
- Indeed those chemical are structurally most similar based on branching and functional groups in the molecule.

Category Definition
Interpretation of the results



1. Chemical **2** is most structurally similar to the **target** analogue**.**

Category Definition
Interpretation of the results



1. Chemical **3** is most structurally similar to the **target** analogue**.**

Category Definition
Recap

- You have searched for suitable analogues having the same profile than the target compound by successive subcategorisation with 3 profilers.
- You have chosen the most suitable candidates to be used for read-across based on a visual analysis of their molecular structure.

Report

- Remember the report module (not reviewed in this exercise) allows you to generate a report on the predictions performed with the Toolbox. This module contains predefined report templates as well as a template editor with which users can define their own user defined templates. The report can then be printed or saved in different formats.
- The obtained prediction could be saved as a file and loaded later on in the system (see Tutorial 5).

Congratulation

- You have used some more functions of the Toolbox and changed up the workflow to address new issues.
- By now you should feel comfortable moving the curser around the basic screens for each one the modules.
- Continue to using the Toolbox and you speed and confidence will increase sharply.