QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.1

Step-by-step example for building QSAR model

Outlook

- **Background**
- Objectives
- The exercise
- Workflow of the exercise

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for building a QSAR model for predicting aquatic toxicity.

- By now you have some experience in using the Toolbox so there will be multiple key strokes between screen shots.

Outlook

- Background
- **Objectives**
- The exercise
- Workflow of the exercise

Objectives

- **This presentation demonstrates building a QSAR model for predicting acute toxicity of aldehydes to *Tetrahymena pyriformis*. The presentation addresses specifically:**
  - predicting acute toxicity for a target chemical;
  - building a QSAR model based on the prediction;
  - applying the model to other aldehydes;
  - exporting the predictions to a file;

Outlook

- Background
- Objectives
- **The exercise**
- Workflow of the exercise

The Exercise

- **This exercise includes the following steps:**
  - select a target chemical – Furfural, CAS 98-01-1;
  - extract available experimental results;
  - search for analogues;
  - estimate the 48h-IGC50 for *Tetrahymena pyriformis* by using trend analysis;
  - improve the data set by either:
    - subcategorizing by "Protein binding" mechanisms, or
    - assessing the difference between outliers and the target chemical
  - evaluate and save the model;
  - use the model to display its training set, visualize its applicability domain and perform predictions.

Outlook

- Background
- Objectives
- The exercise
- **Workflow of the exercise**

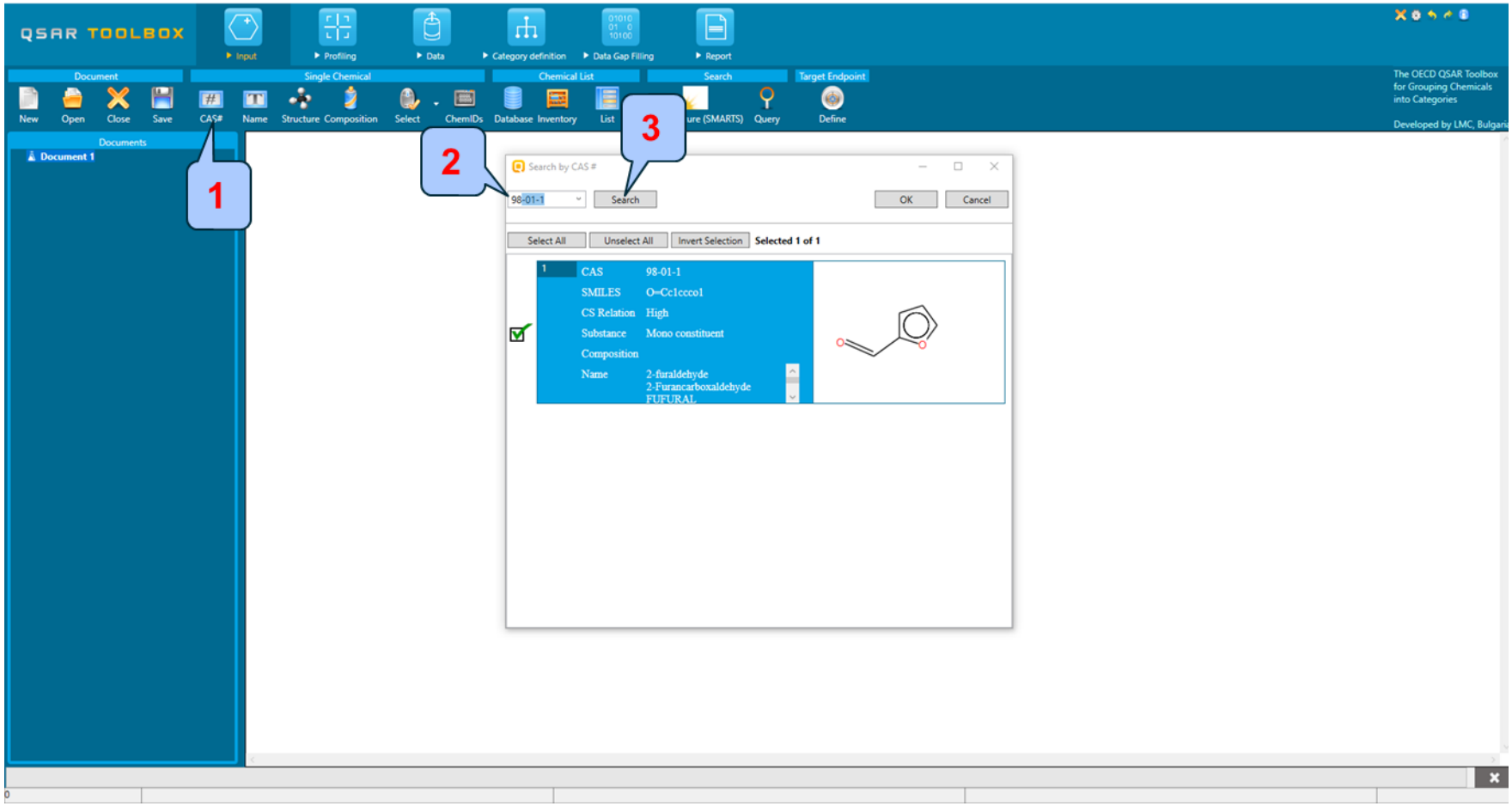
Workflow of the exercise

- **Remember the Toolbox has 6 modules which are used in a sequential workflow:**
  - Input
  - Profiling
  - Data
  - Category Definition
  - Data Gap Filling
  - Report

Outlook

- Background
- Objectives
- The exercise
- **Workflow of the exercise**
  - **Input**

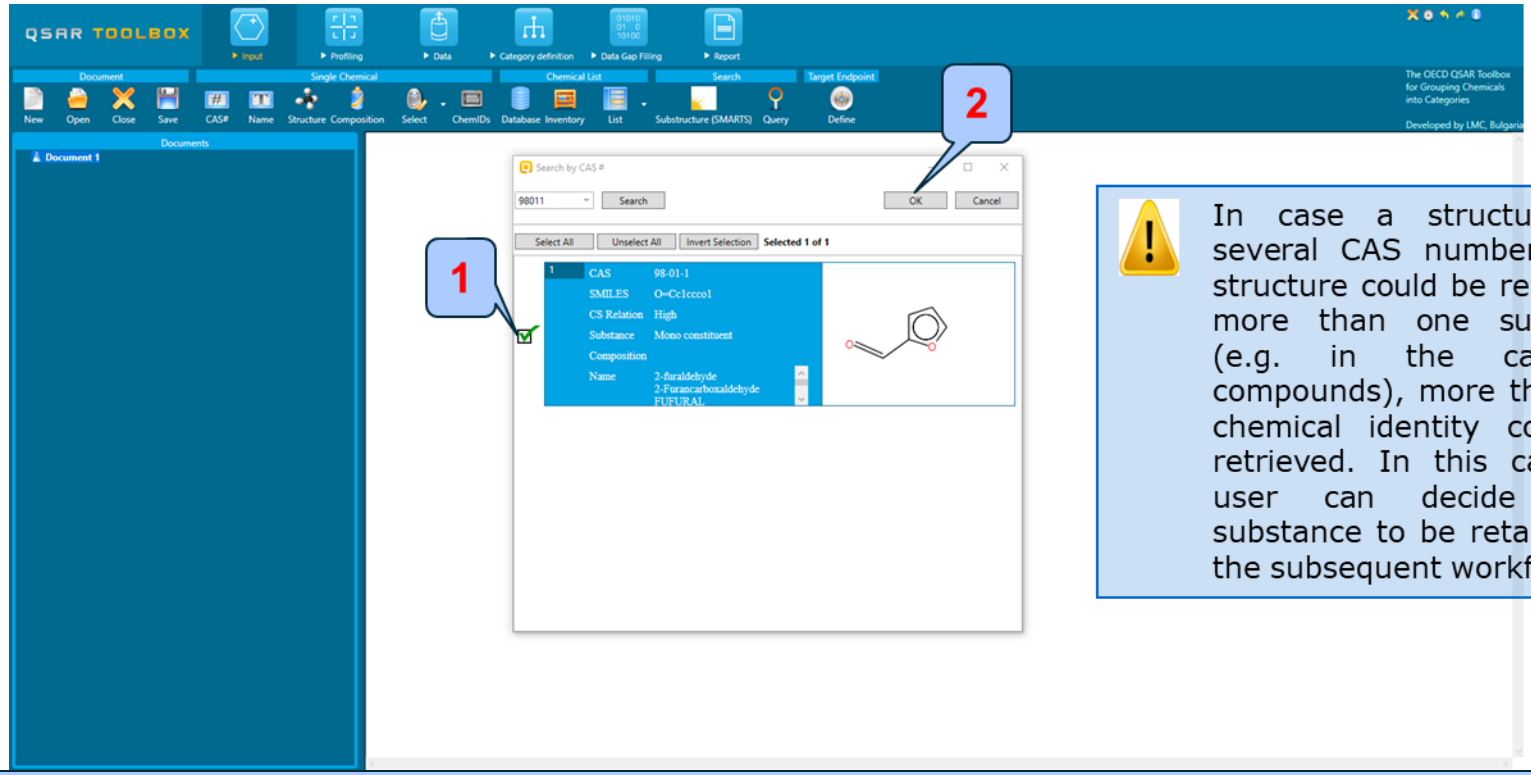
Input



1. **Click** on CAS# 2. **Enter** CAS# 98-01-1; 3. **Click** Search;

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 CAS number. It is displayed as a 2D image. Note it is unselected by default.



In case a structure has several CAS numbers or a structure could be related to more than one substance (e.g. in the case of compounds), more than one chemical identity could be retrieved. In this case the user can decide which substance to be retained for the subsequent workflow.

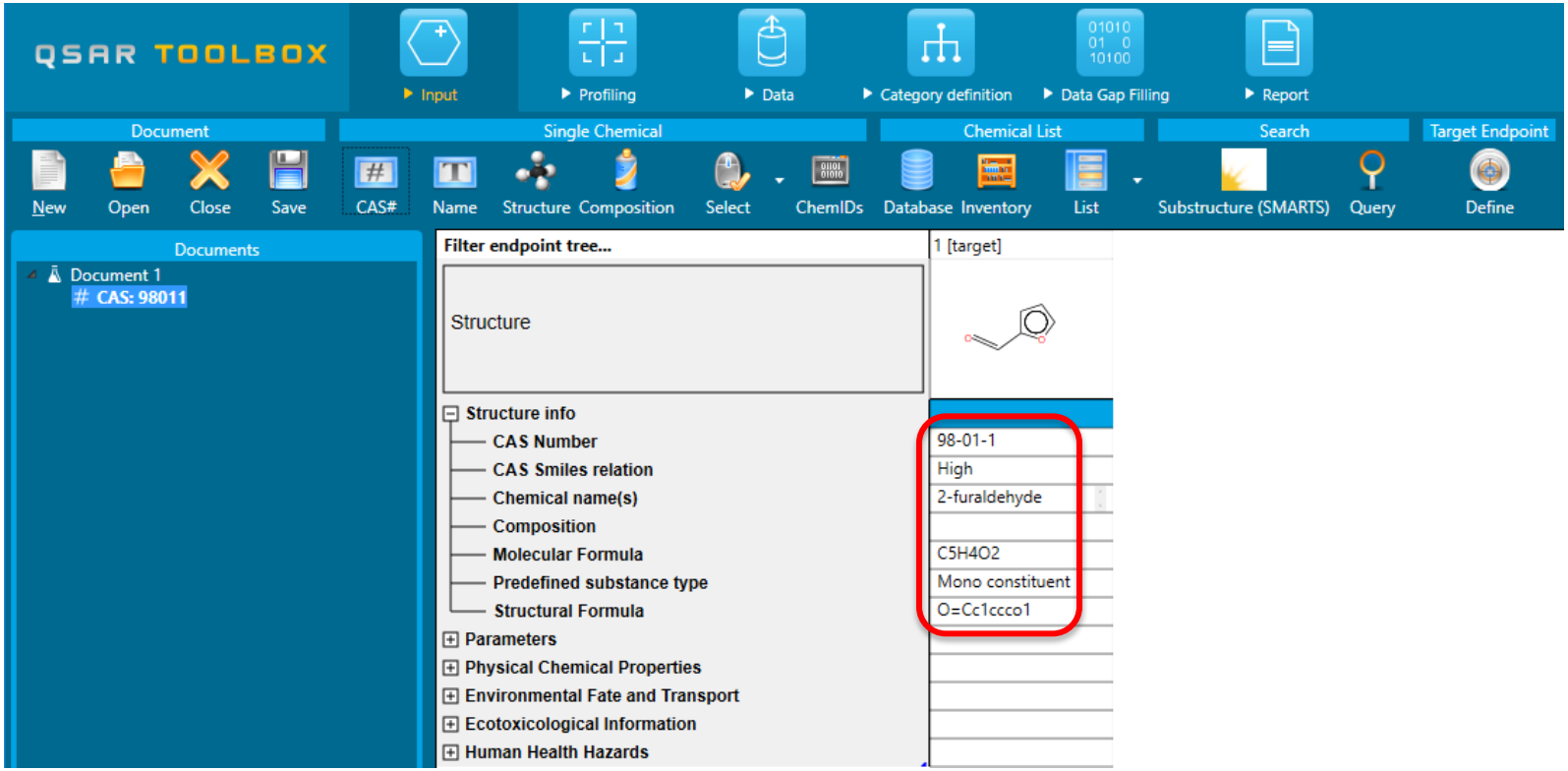
1. **Mark** desired chemical (in case there is only one chemical it is marked by default);
2. **Click** OK to add chemical in data matrix;

Input
Target chemical identity

- Target chemical is displayed on the data matrix.
- To see chemical identification click on the box next to "Structure info" (see next screen shot).

Chemical Input

Target chemical identity



Outlook

- Background
- Objectives
- The exercise
- **Workflow of the exercise**
  - Input
  - **Profiling**

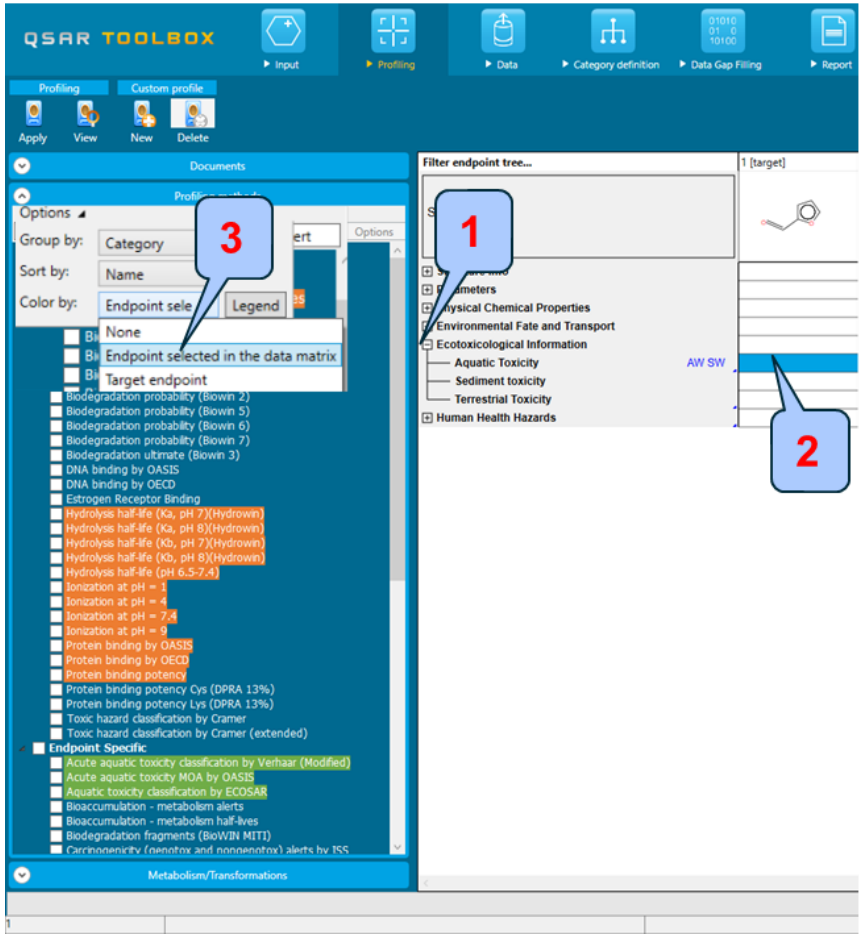
Profiling

Profiling the target chemical

- Select the "Profiling methods" related to the target endpoint
- This selects (a green check mark appears) or deselects (green check disappears) profilers.
- To help the user to choose suitable profiling methods, a new feature has been developed – see next slide

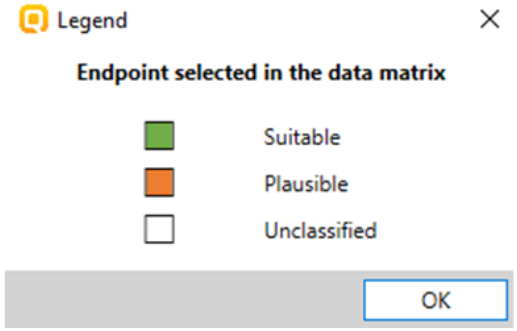
Profiling

Profiling the target chemical



1. **Click** on the box to open the nodes of the tree;
2. **Mark the** box in the data matrix related to the target endpoint of chemical;
3. From Profiling / Profiling methods / Options / **Color by** (pop-up menu) select **Endpoint selected in the data matrix**

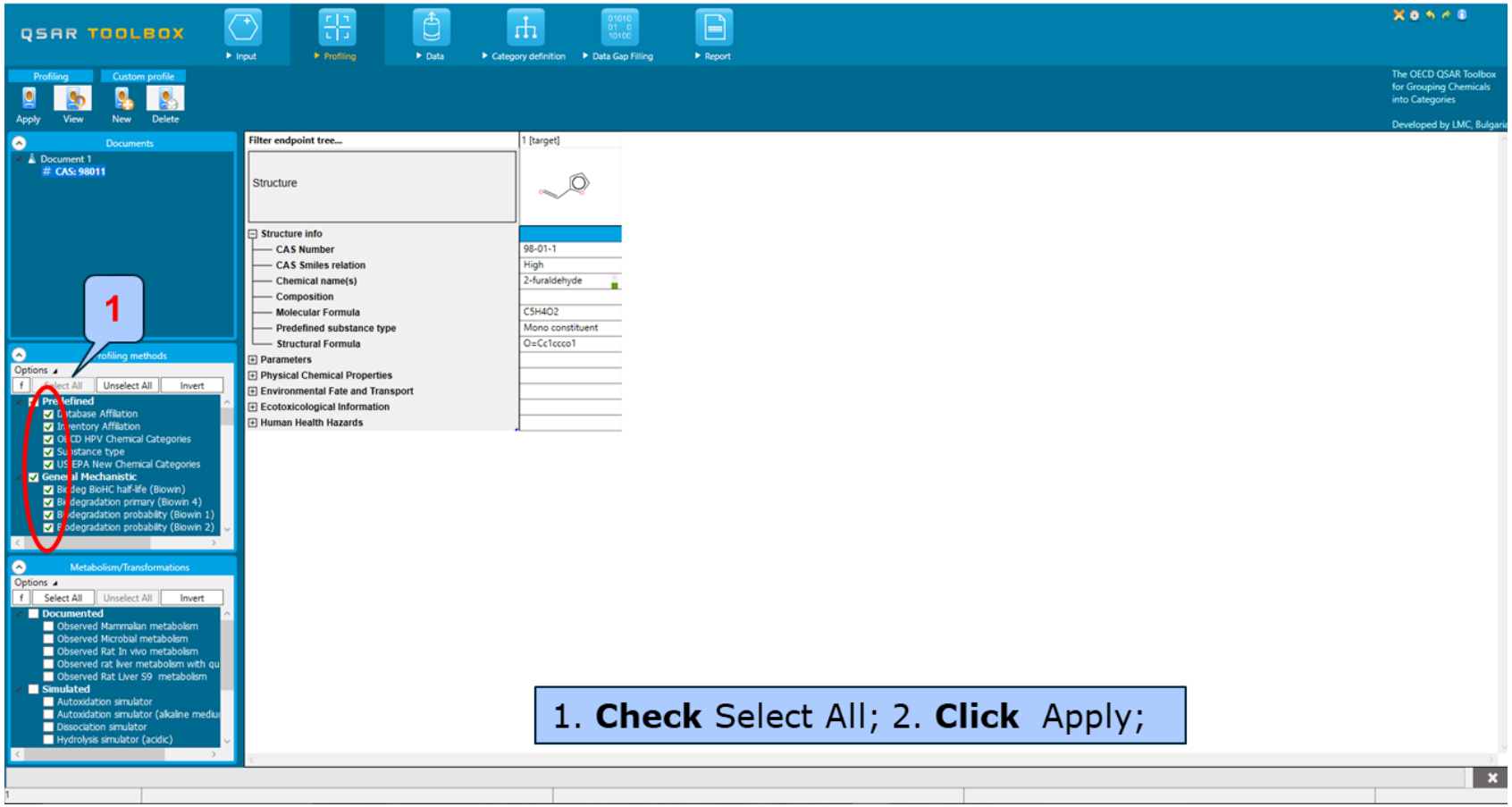
According to the legend profilers related to the selected endpoint are highlighted in **Green** or **Orange**



Profiling

Profiling the target chemical

For this example, select all profilers.



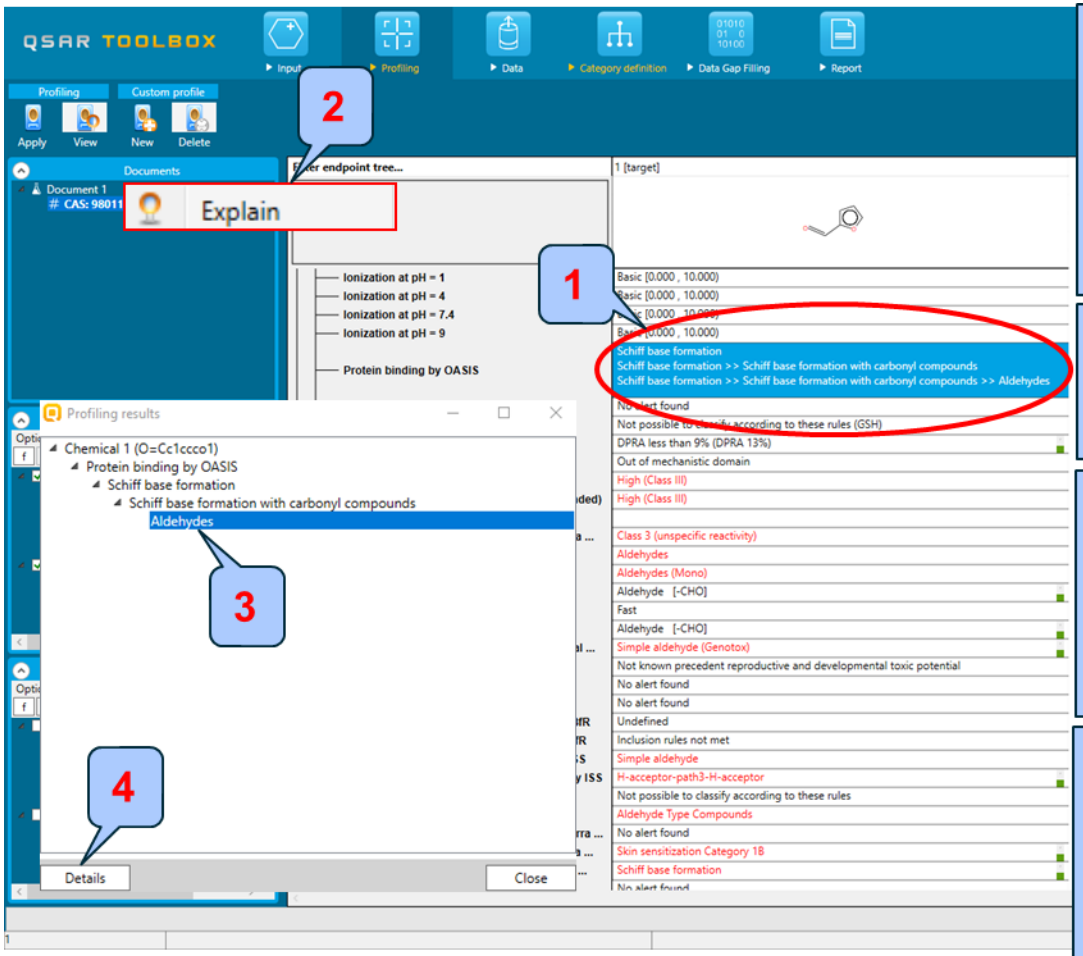
Profiling

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 (see next screen shot).
- Green-white rectangles in some result boxes indicate there is more than one profiling result and the field needs to be expanded.

Profiling

Profiles of "Furfural"



In this case there is structural evidence that the target could interact to DNA and proteins, it has also mode of action and it is as aldehyde. This step is critical for next grouping of analogues.

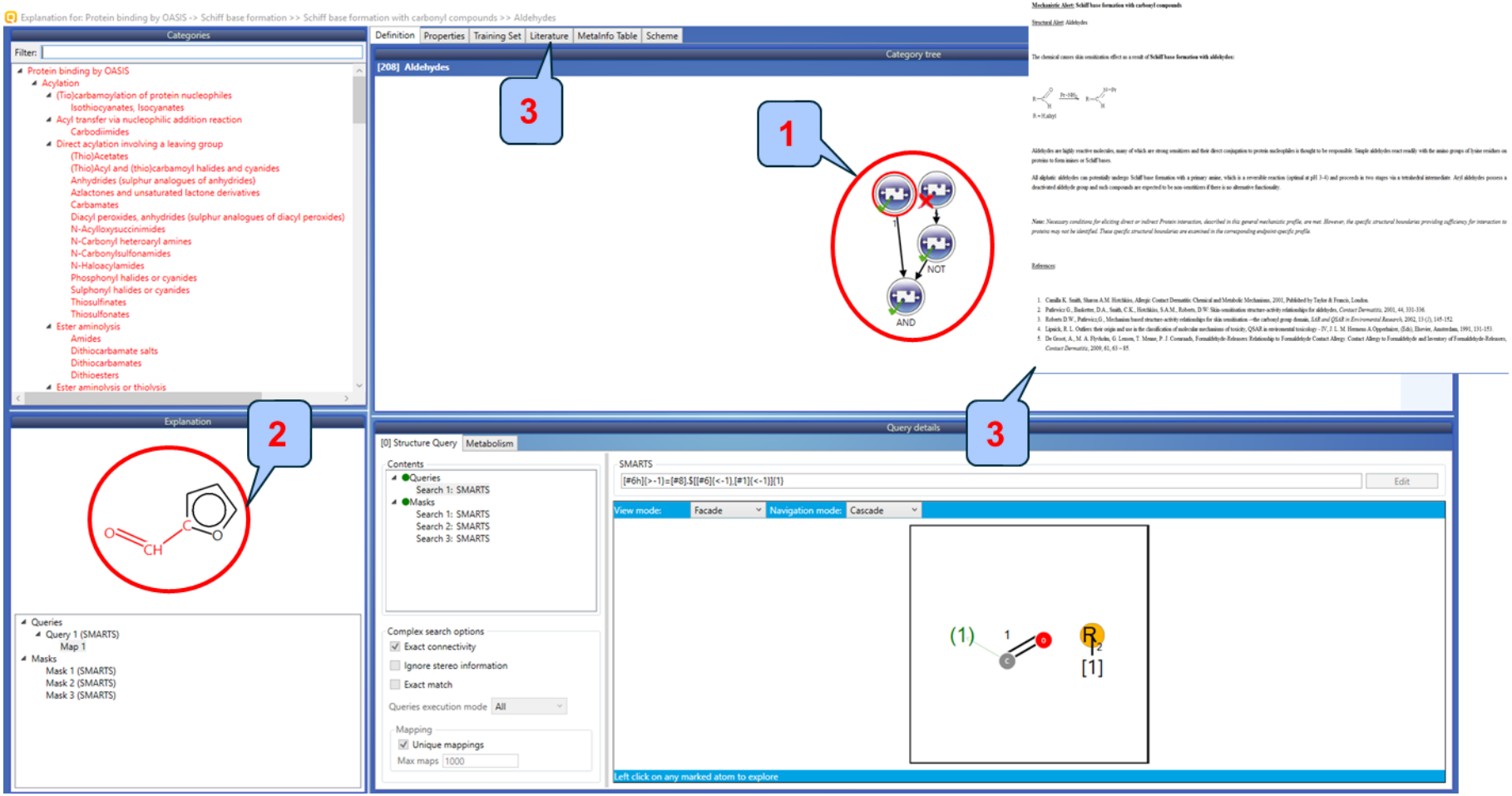
1. **Right click** to see why the target is Protein binder by "Protein binding by OASIS";

The "Protein binding by OASIS" profiler has hierarchical structure consisting of three levels: Structural alert, Mechanistic alert and Mechanistic domain

2. **Left click** on the "Explain";
3. From the list of the profiling results **Click** on the structural alert "Aldehydes";
4. **Click** Details;

Profiling

Profiles of "Furfural"



1. Structural boundary of the category; **2.** Definition of the used common fragments;
3. Mechanistic justification of the category (Literature tab)

Outlook

- Background
- Objectives
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - **Data**

Data

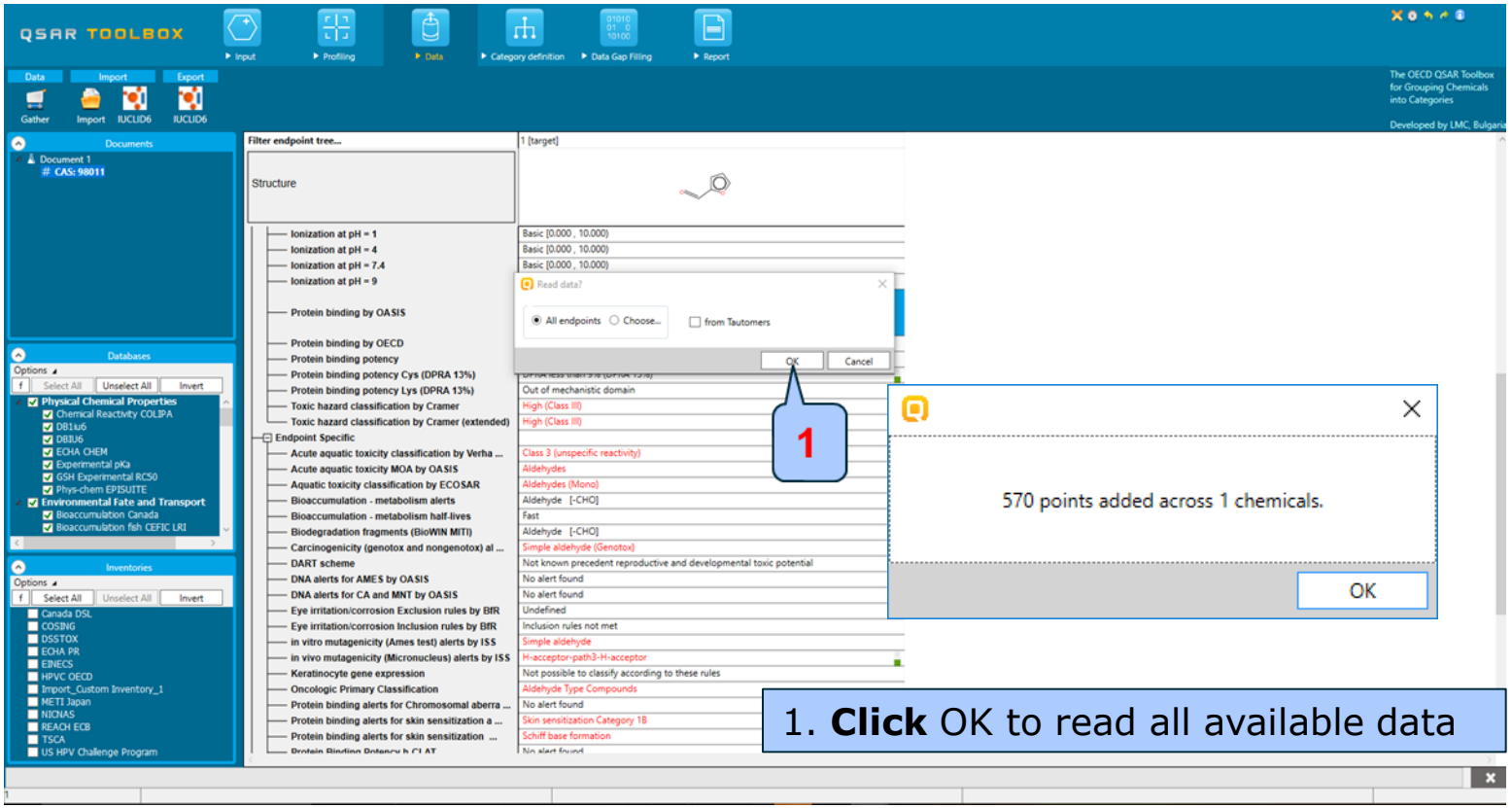
Extracting endpoint values

A new functionality for specifying bases containing data with desired endpoint is available (similar to this one for suitable profiling methods). In our case we will use all databases.

1. **Go to Data**
2. **Select** all databases
3. **Click** Gather

Data
Process of collecting data

Toxicity information on the target chemical will be electronically collected from the selected datasets.

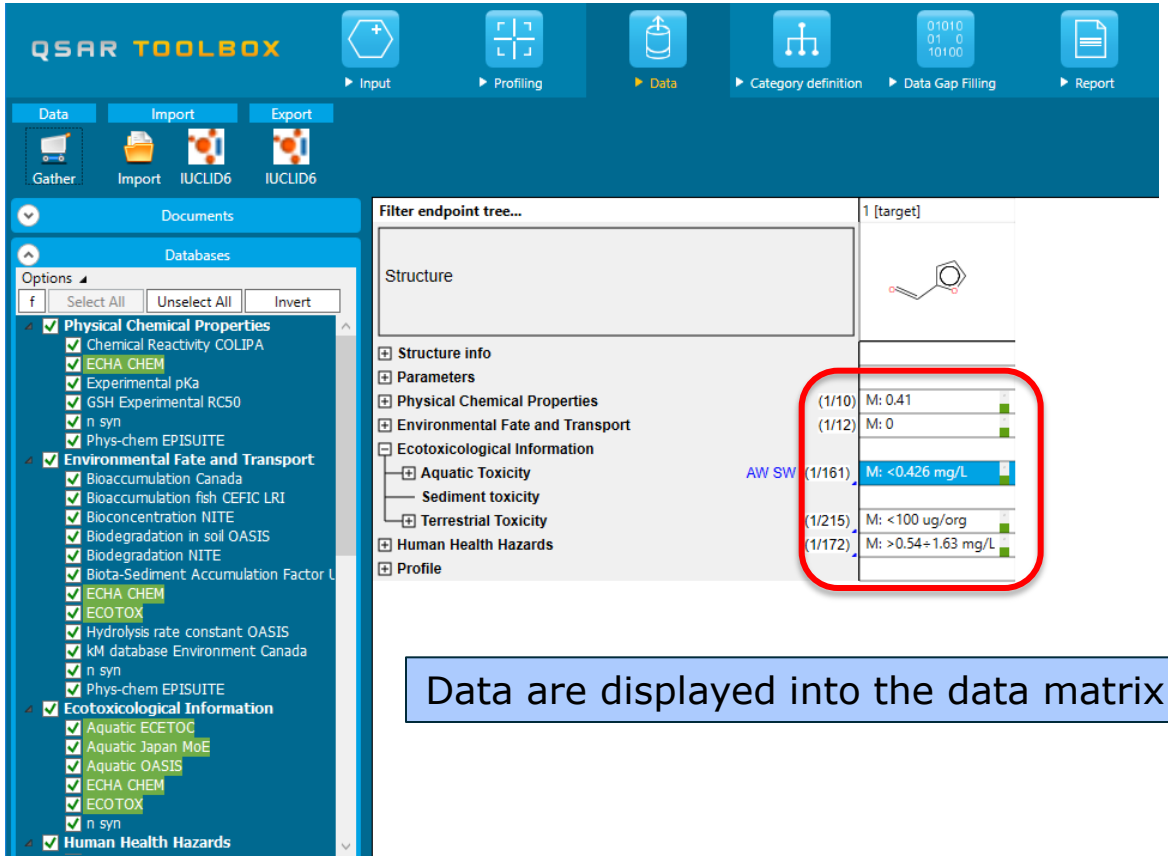


A window with "Read data?" appears. Now the user could choose (via radio button) to collect "All" or "Endpoint specific" data. In our case collect "All".

Data

Read data for analogues

In this example, an insert window appears stating that there were found 570 data points available for the target chemical appears. Click OK.



Data are displayed into the data matrix

Outlook

- Background
- Objectives
- The exercise
- **Workflow of the exercise**
  - Chemical Input
  - Profiling
  - Data
  - **Category definition**

Category definition
Target endpoint

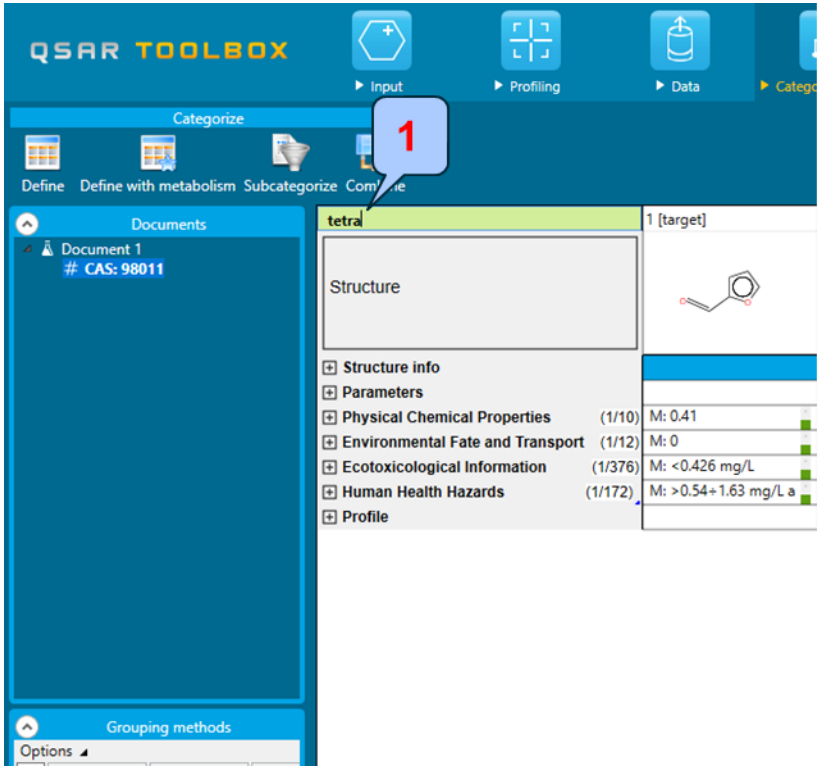
- In this exercise we will build a QSAR model to estimate the following endpoint:

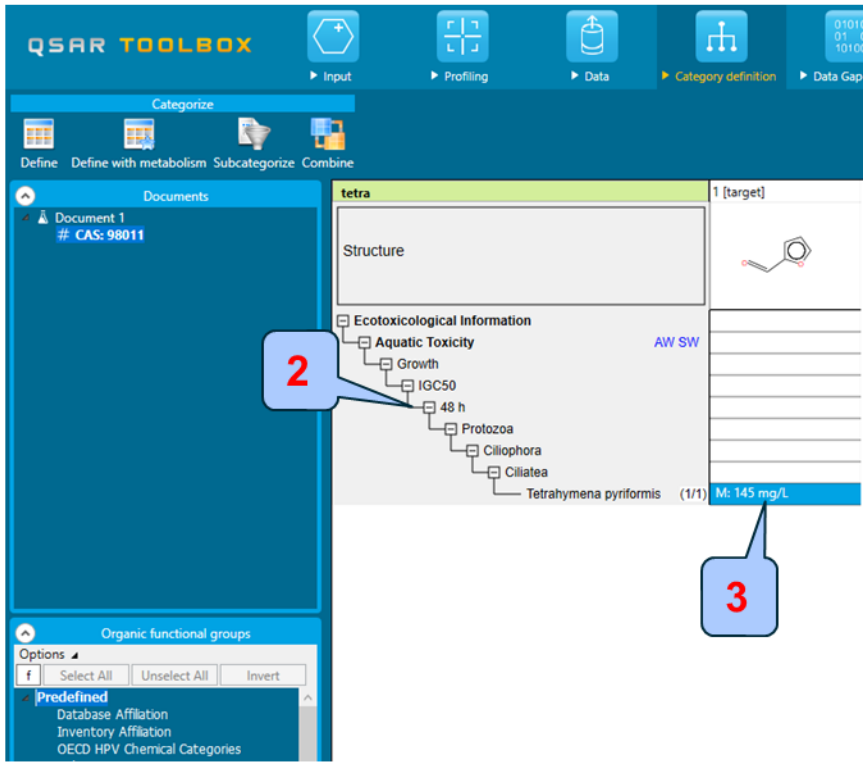
  Ecotoxicological Information#Aquatic Toxicity#Growth#IGC50#48h#Protozoa#Ciliophora#Ciliatea#Tetrahymena pyriformis

- The initial search for analogues is based on structural similarity by US-EPA categorization

Category definition

Navigate to the target endpoint





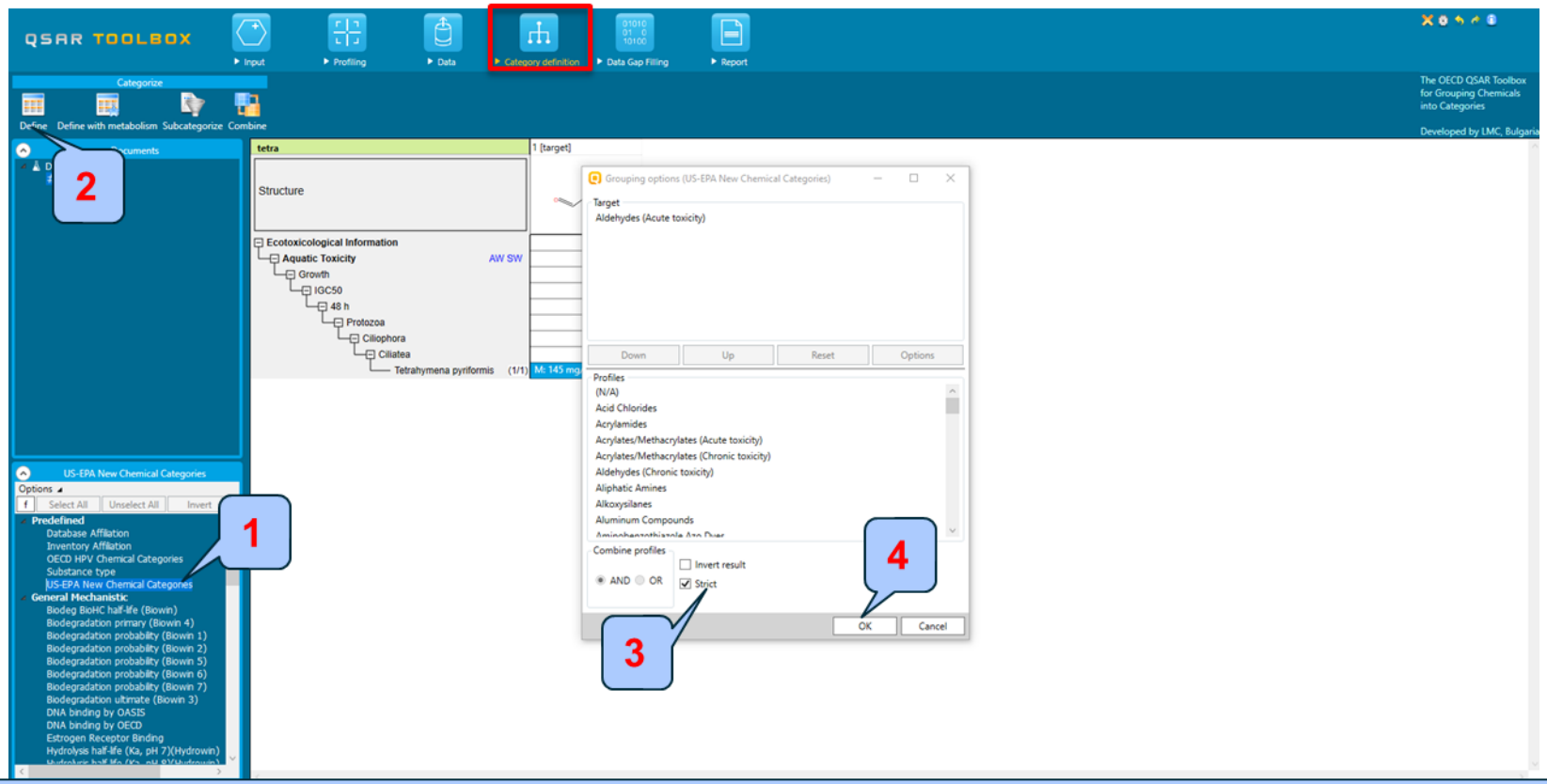
1. Use Filter (in green) to type "Tetra" in the empty field and c**lick** Enter; 2. This will open nodes of data matrix to the target endpoint; 3. **Highlight** the cell that will be filled in (in this case we will reproduce the observed data);

Category definition
Defining US-EPA category

- The initial search for analogues is based on structural similarity, of US EPA categorization
- Select US-EPA New Chemical Category
- Click Define (see next screen shot)

Category definition

Defining US-EPA category



1. **Highlight** "US-EPA New Chemical Categories"; 2. **Click** Define; 3. **Put** a tick in the Strict box (see next screen shot); 4. **Click** OK to confirm the category **Aldehydes (Acute toxicity);**

Category definition

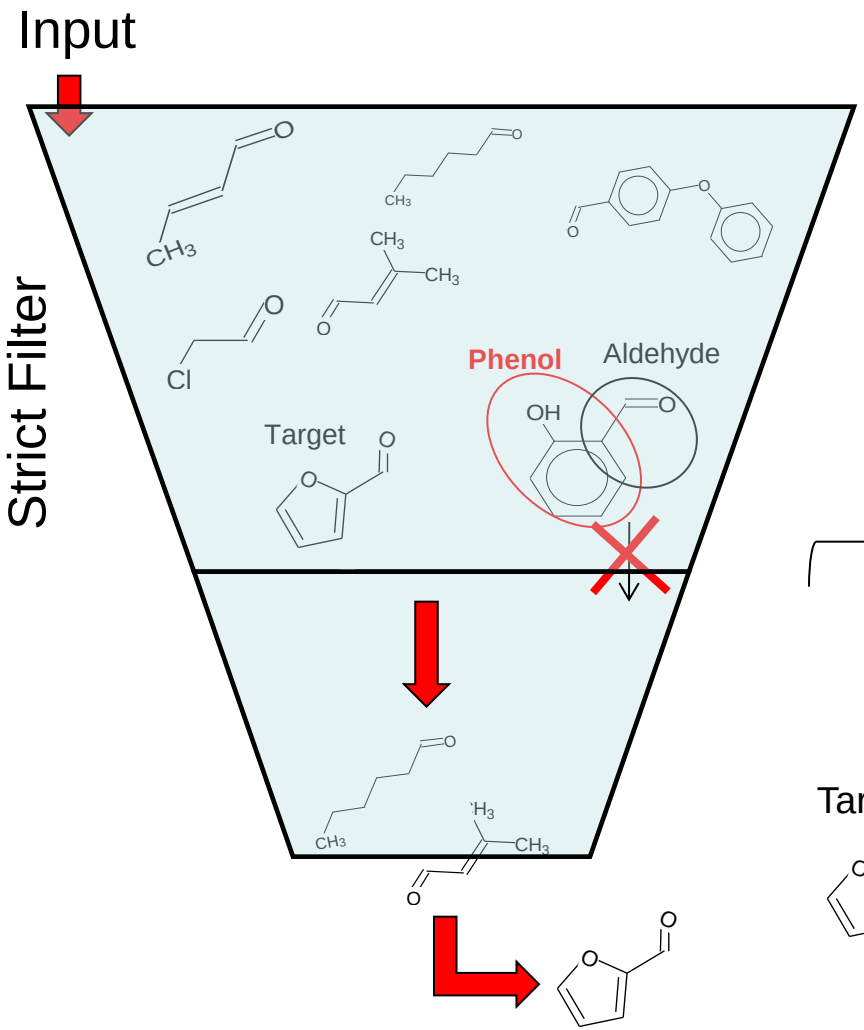
Defining US-EPA category strict functionality

- The **Strict** functionality means that the software will group analogues having **ONLY** the categories of the target and will exclude the analogues having any other categories according to the profiler used in the grouping method.

- For example, if the profiling for the target results in *Aldehydes (Acute toxicity)* **ONLY** according to US-EPA category, the group of analogues will include *Aldehydes (Acute toxicity)* **ONLY**. (See next screen shot)

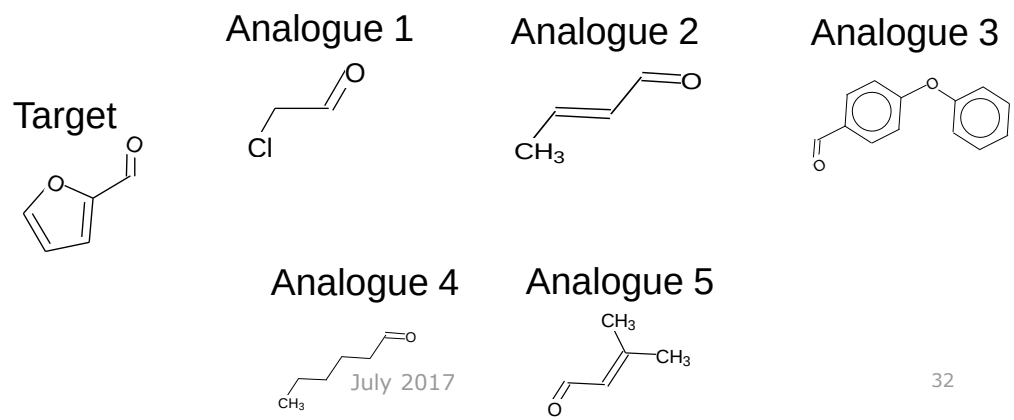
Category definition

Defining US-EPA category strict functionality



The target and analogues have *Aldehydes* **ONLY** according to US-EPA category

Defined Category



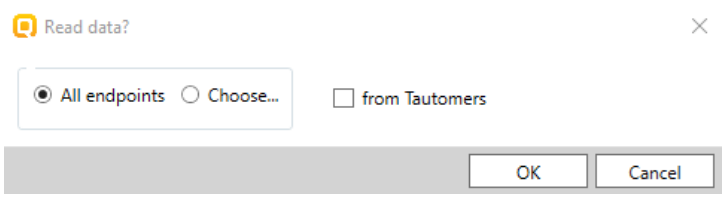
Category definition
Analogues

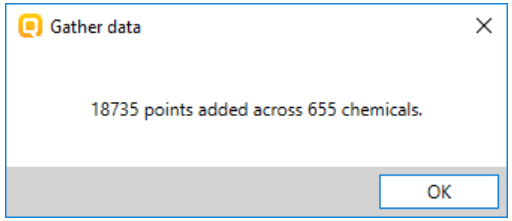
- The Toolbox now identifies all chemicals corresponding to *Aldehydes (Acute toxicity)* by US-EPA listed in the databases selected under "Data".

- 665 analogues including the target chemical are identified; they form a mechanistic category "**Aldehydes (Acute toxicity)**", which will be used for gap filling.

Category definition
Reading data for Analogues

- The Toolbox will now retrieve those chemicals that have the same structural alert as the target
- 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 bellow)

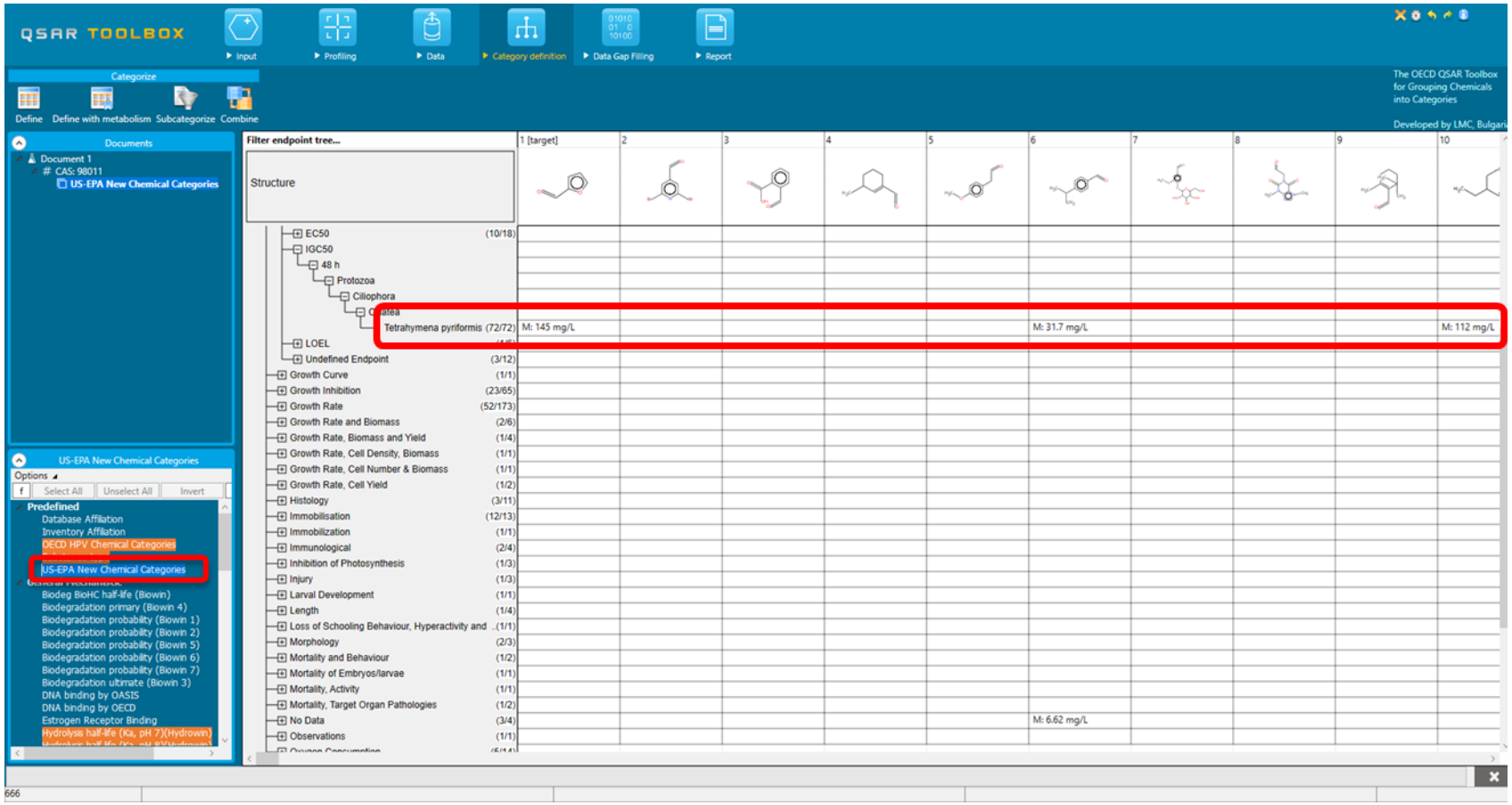




Category definition

Summary information for Analogues

After a message for number of data collected the experimental results for the target and analogues are inserted into the matrix.

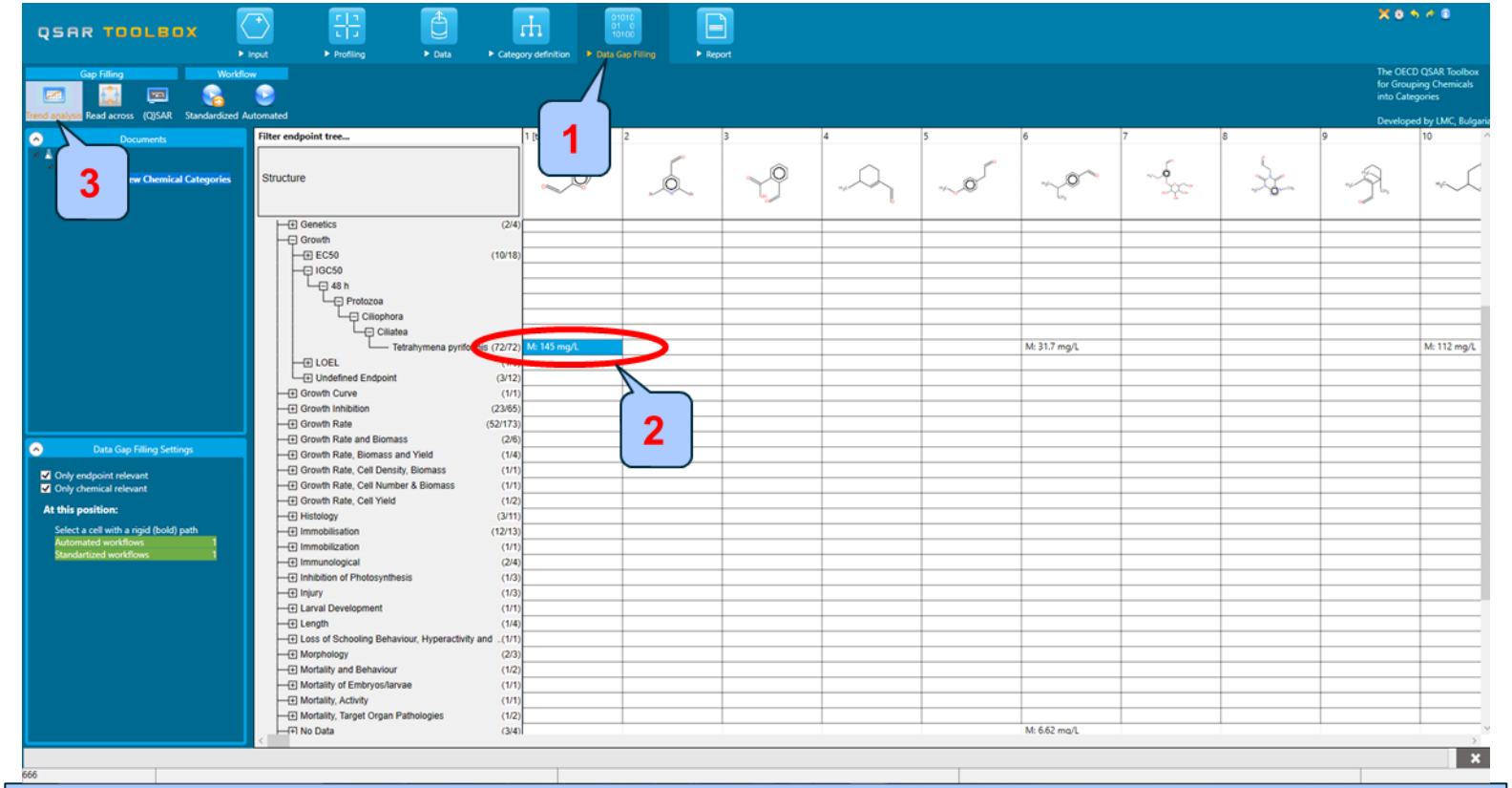


Outlook

- Background
- Objectives
- The exercise
- **Workflow of the exercise**
  - Chemical Input
  - Profiling
  - Data
  - Category definition
  - **Data gap filling**

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

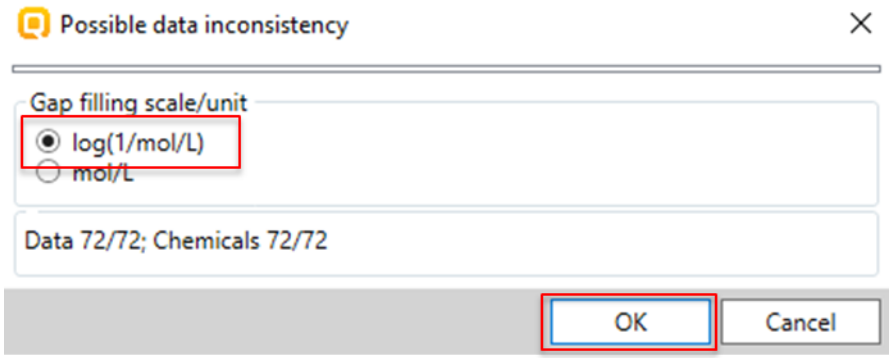
Apply Trend analysis



1. **Go to** Data Gap Filling 2. **Highlight** the Data gap corresponding to IGC50, *Tetrahymena pyriformis* under the target chemical; 3. **Select** Trend analysis;

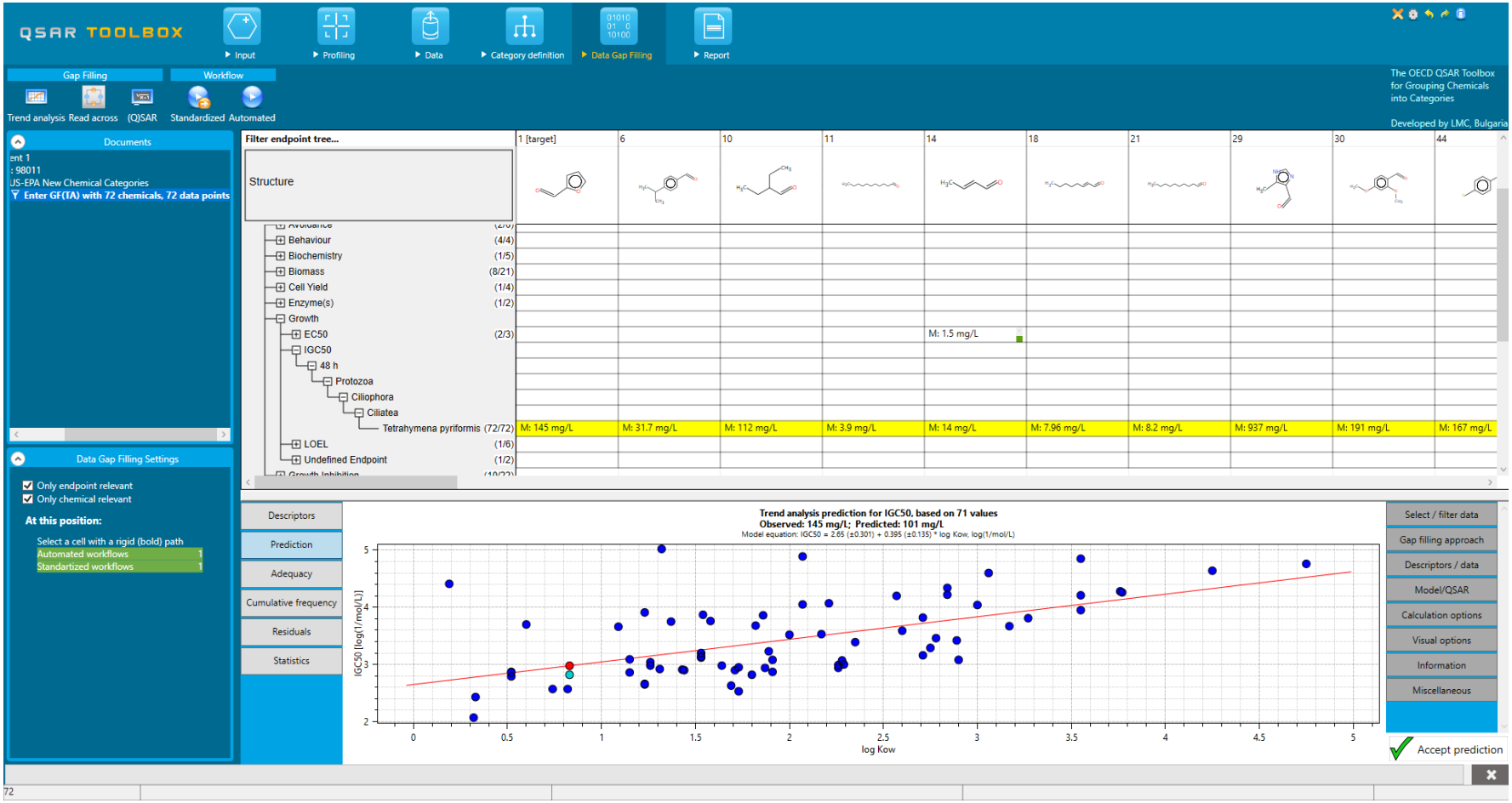
Data Gap Filling (IGC 50 48h of *T. pyriformis*) Apply Trend analysis

- A message for possible data inconsistency appears
- It is recommended the log(1/mol/L) scale to be chosen



- The resulting plot can be seen on next screen shot

Data Gap Filling
(IGC 50 48h of *T. pyriformis*)



Data Gap Filling (IGC 50 48h of *T. pyriformis*)

Interpreting dots on the graph

- The resulting plot outlines the experimental results of all analogues (Y axis) according to a descriptor (X axis) with LogKow being the default descriptor (see previous screen shot).
- The **RED** dot represents the predicted value for target chemical.
- The **BLUE** dots represent the experimental results available for the analogues.
- The **LIGHT BLUE** dots (see the following screen shots) represent analogues belonging to different subcategories.

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

An accurate analysis of data set

- In this example, the mechanistic properties of the analogues are consistent.
- Subcategorization can be performed based on protein binding mechanisms. This is the second stage of analogue search - requiring the same interaction mechanism.
- Acute effects are associated with covalent interaction of chemicals within cell proteins, i.e. with protein binding.
- Chemicals with a different protein binding mechanism / reactions compared to the target chemical will be removed.

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

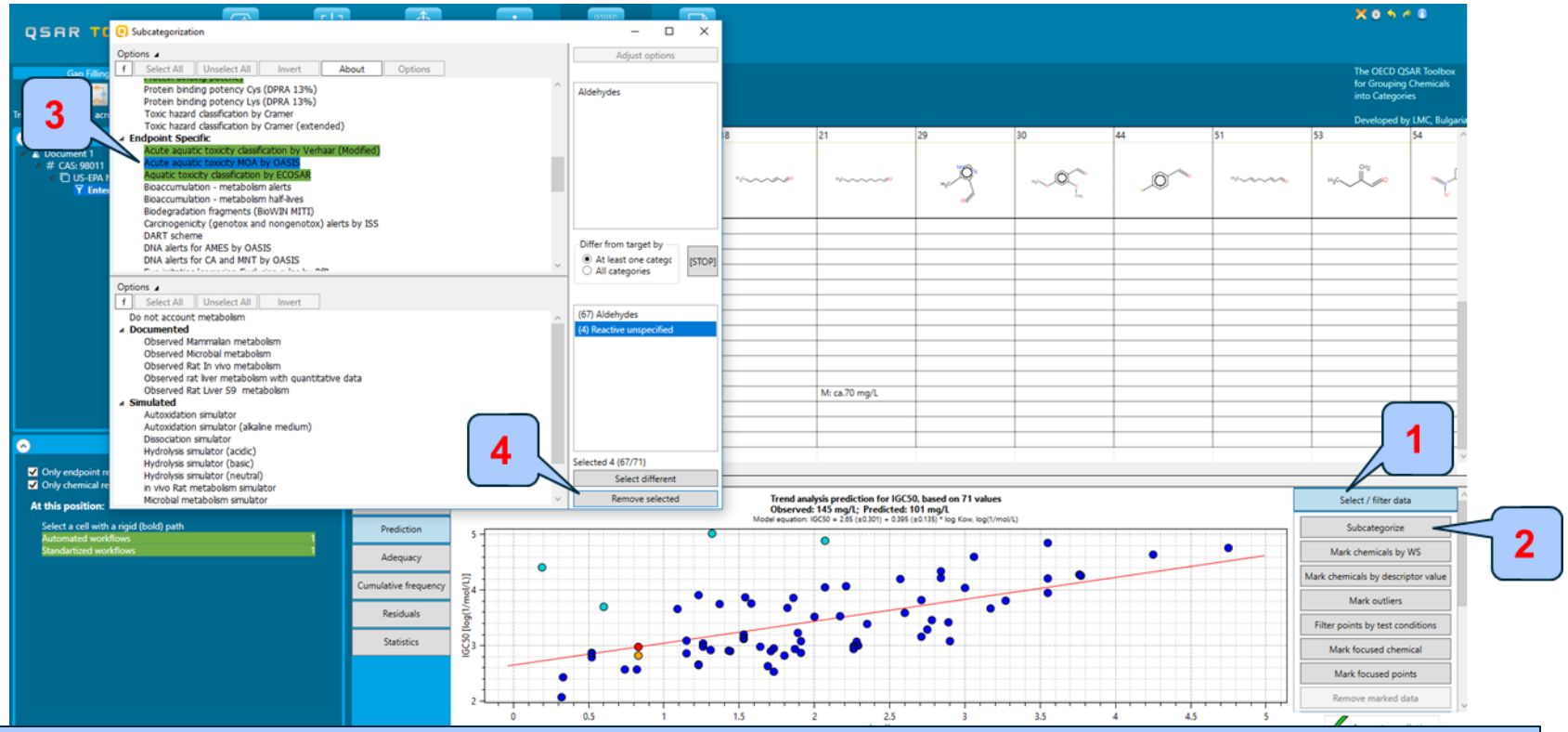
Subcategorisation

- After the available data has been retrieved, the user can then further subcategorize the results according to the following endpoint-specific subcategorizations:

  - Acute aquatic toxicity MOA by OASIS
  - Protein binding by OASIS
  - Aquatic toxicity classification by ECOSAR

- These steps are summarized in the next screen shots.

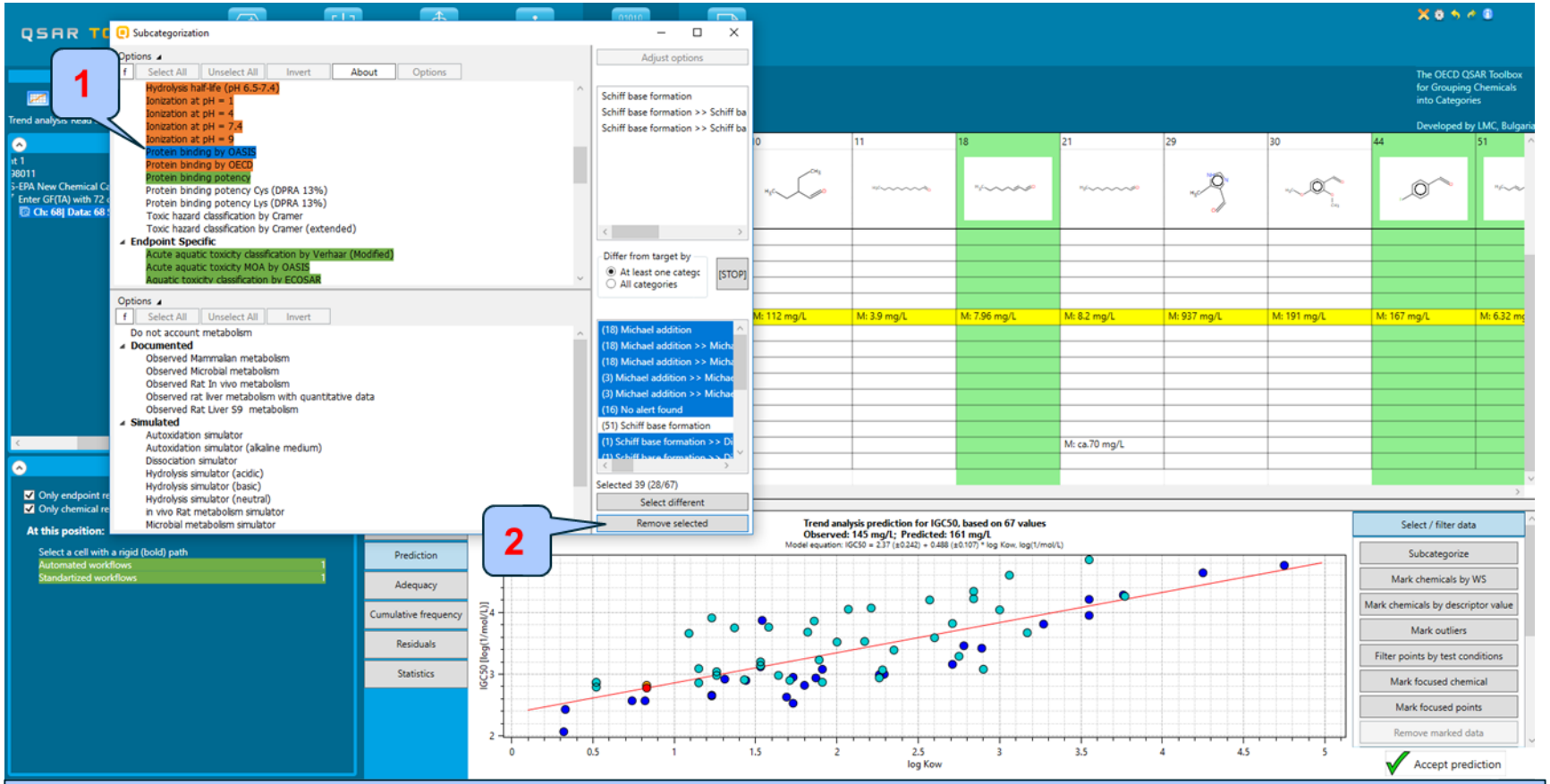
Data Gap Filling
(IGC 50 48h of *T. pyriformis*)
Subcategorization 1: Acute aquatic toxicity MOA by OASIS



1. **Click** Select / filter data; 2. **Select** Subcategorize; 3. **Select** "Acute aquatic toxicity MOA by OASIS" (note there is the same suggestion of appropriate for subcategorization profiles and metabolic simulators); 4. **Click** "Remove selected" to eliminate dissimilar to the target chemicals

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

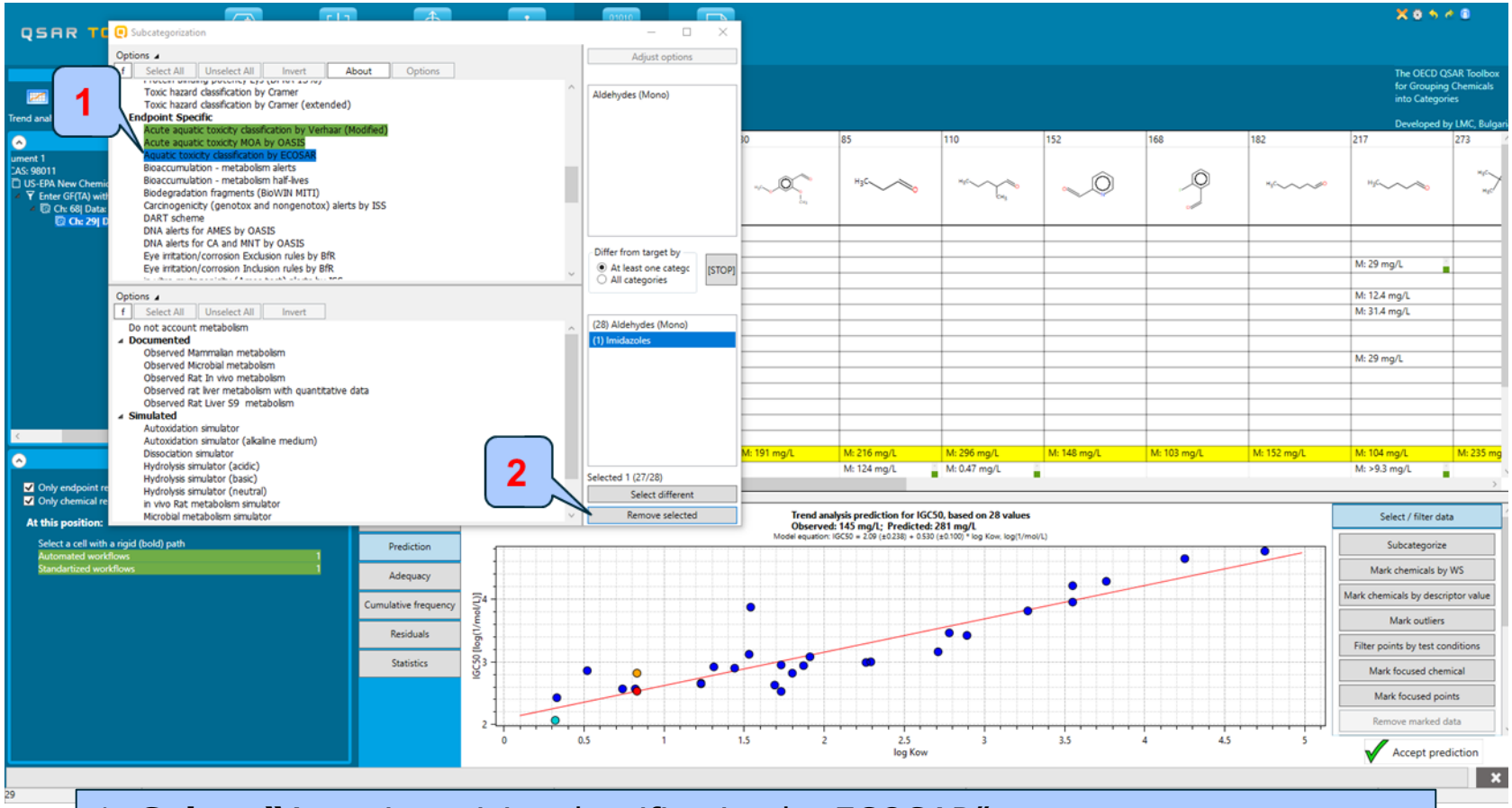
Subcategorization 2:Protein binding by OASIS



1. **Select** "Protein binding by OASIS";
2. **Click** "Remove selected" to eliminate dissimilar to the target chemicals.

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

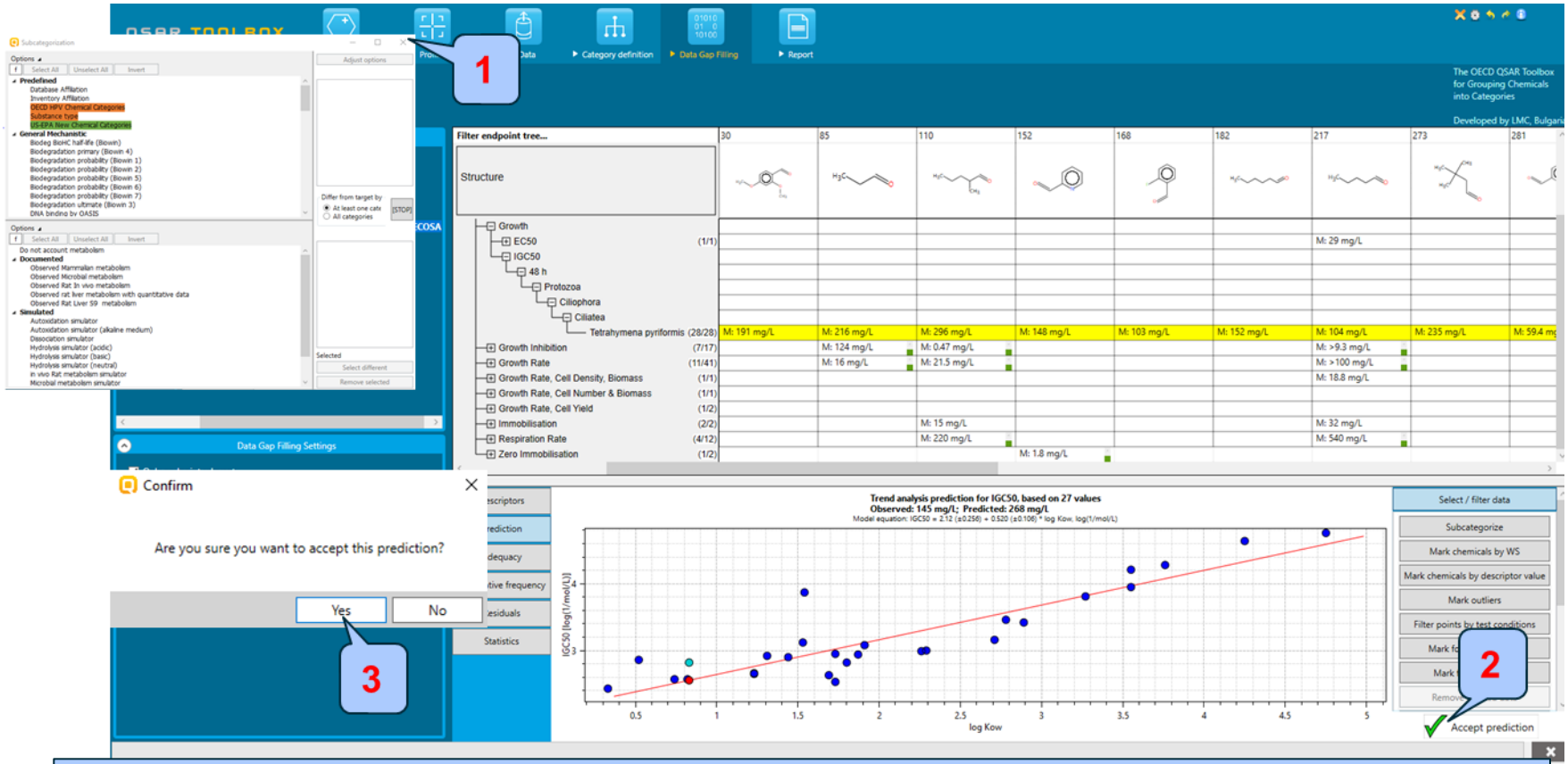
Subcategorization 3: Aquatic toxicity classification by ECOSAR



1. **Select** "Aquatic toxicity classification by ECOSAR";
2. **Click** "Remove selected" to eliminate dissimilar to the target chemicals;

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

Results after subcategorisation



1. Close "Subcategorization" window 2. **Click** "Accept prediction"; 3. **Click** "Yes" ("No" allows to continue with the subcategorization);

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

Evaluation of the model

- To assess the model accuracy use:
  - Adequacy (predictions after leave-one-out)
  - Statistics
  - Cumulative frequency
  - Residuals
- See next four screen shots

Data Gap Filling
(IGC 50 48h of *T. pyriformis*)
Evaluation of the model - Adequacy

1. Position on the last level of document tree **2. Click** "Adequacy";

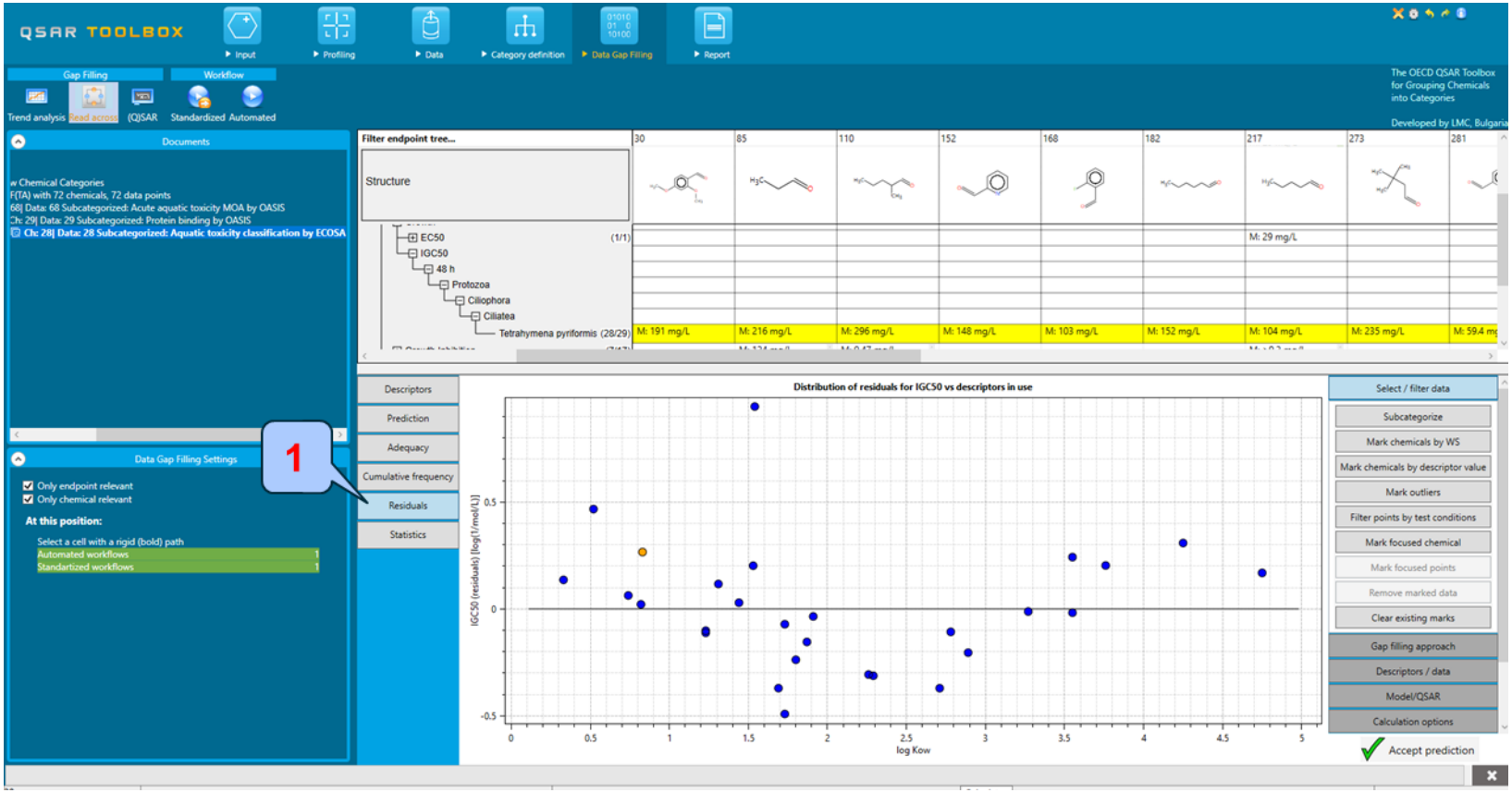
Data Gap Filling (IGC 50 48h of *T. pyriformis*)

Evaluation of the model - Cumulative frequency

1. **Click** "Cumulative frequency; The residuals abs (obs-predicted) for 95% of analogues are comparable with the variation of experimental data.

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

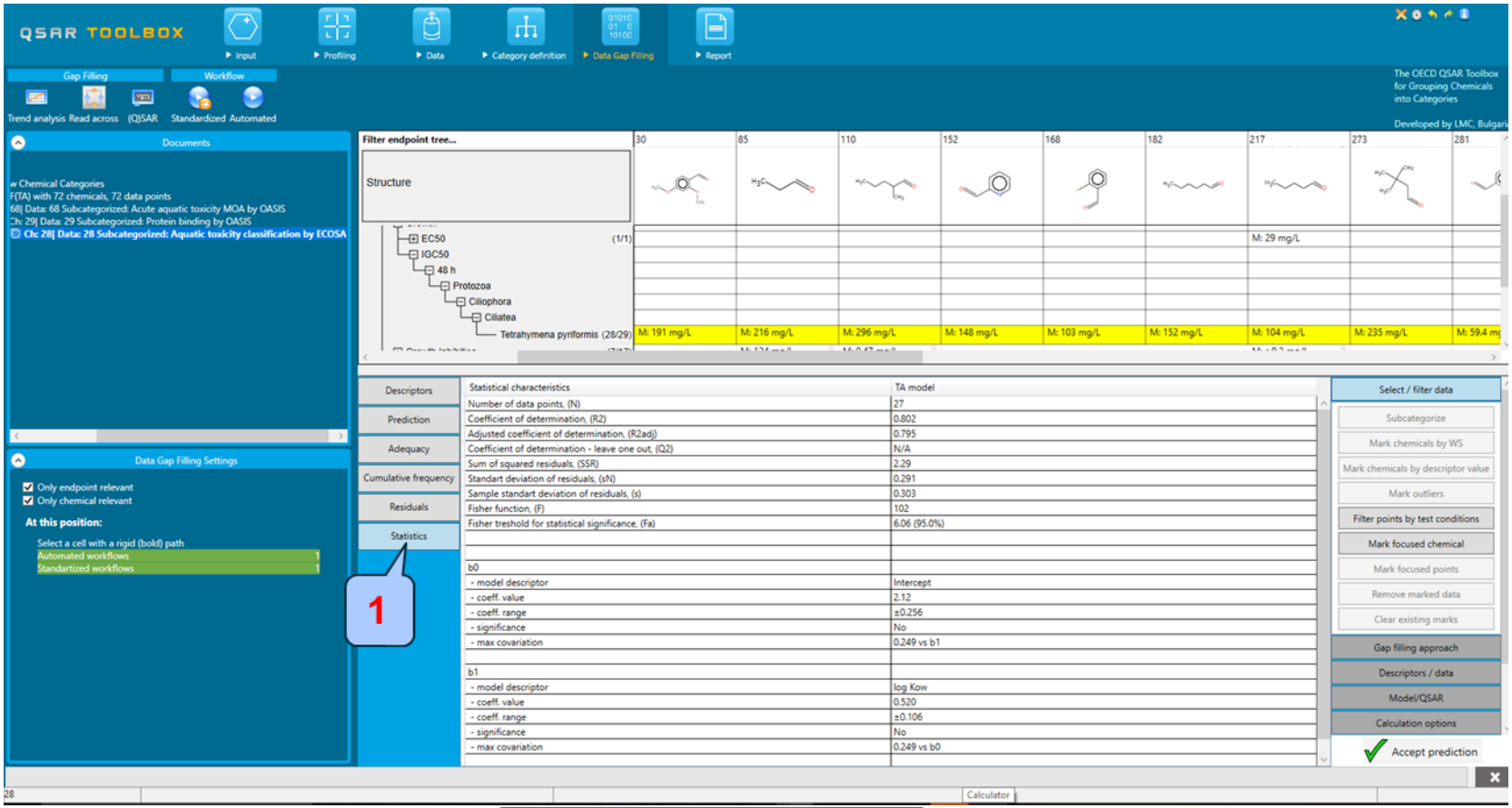
Evaluation of the model - Residuals



1. **Click** "Residuals";

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

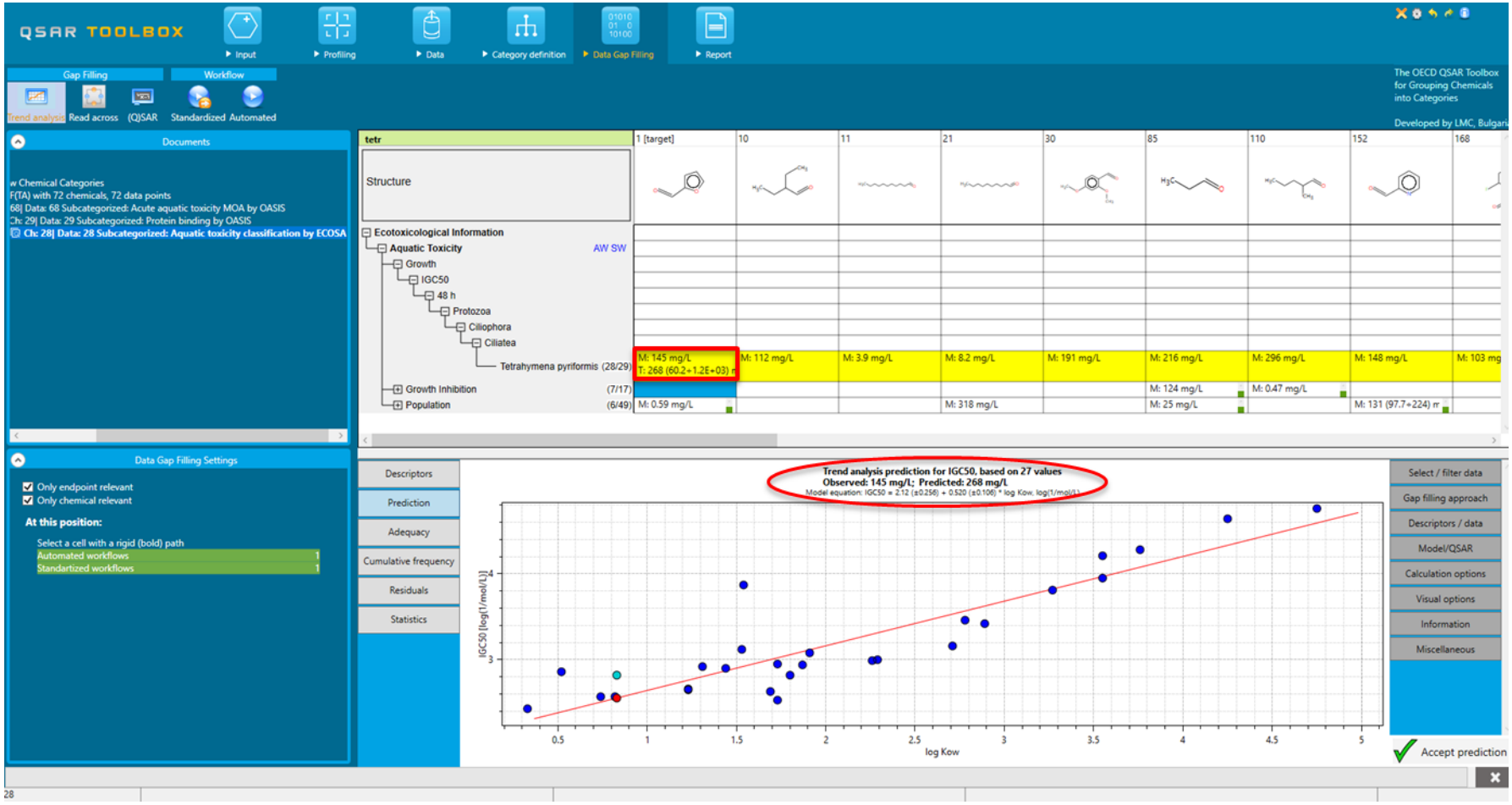
Evaluation of the model - Statistics



1. **Click** "Statistics";

Data Gap Filling (IGC 50 48h of *T. pyriformis*)

Results after subcategorisation

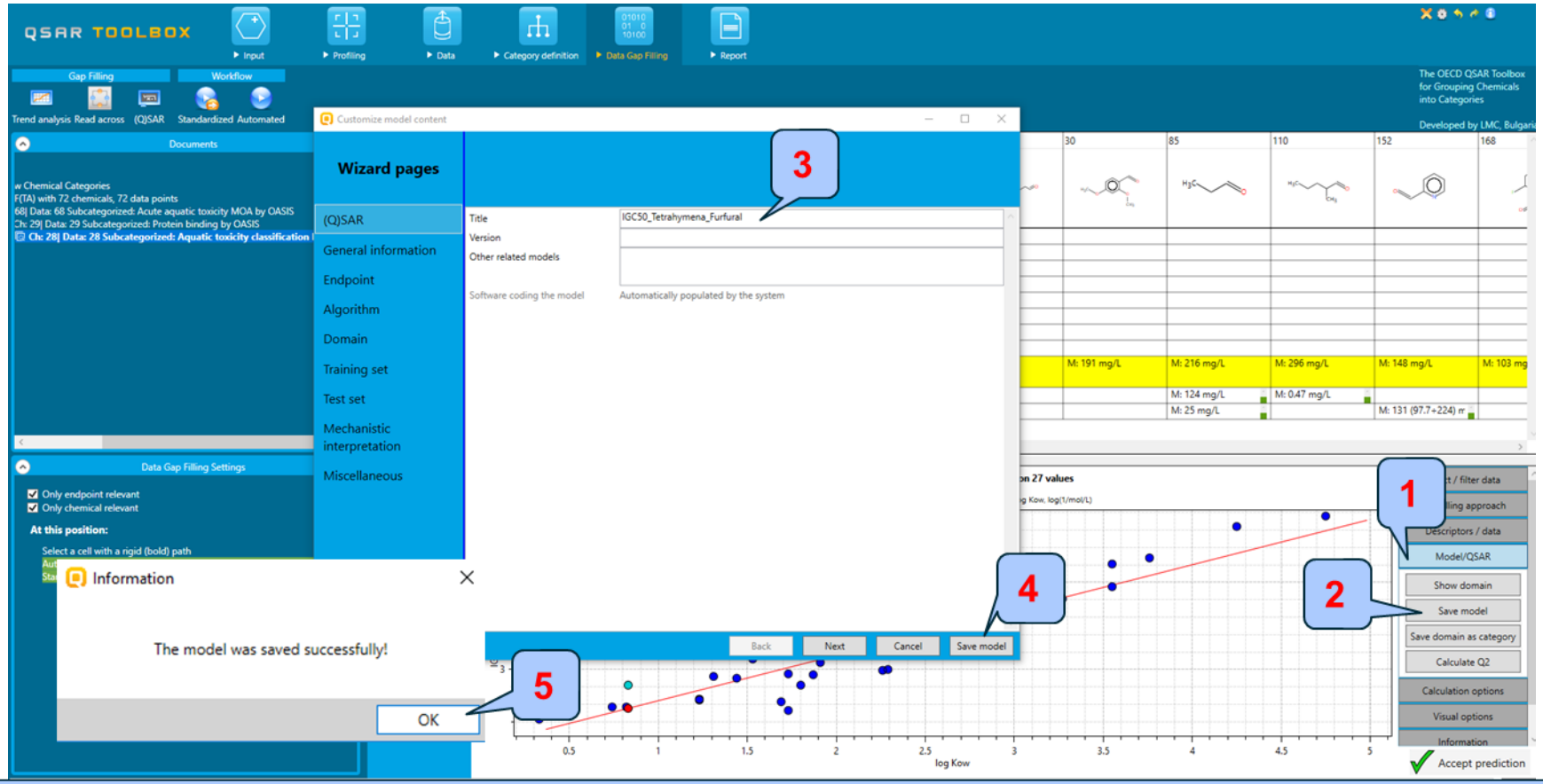


Data Gap Filling (IGC 50 48h of *T. pyriformis*)

Save the derived QSAR model

- To save the new regression model follow these steps:
  - Go to the last row on the Document tree
  - Click on “Model/QSAR”
  - Select Save model
  - Enter the model name and fill editable fields if necessary
  - Click on OK

Data Gap Filling
(IGC 50 48h of *T. pyriformis*)
Save the derived QSAR model



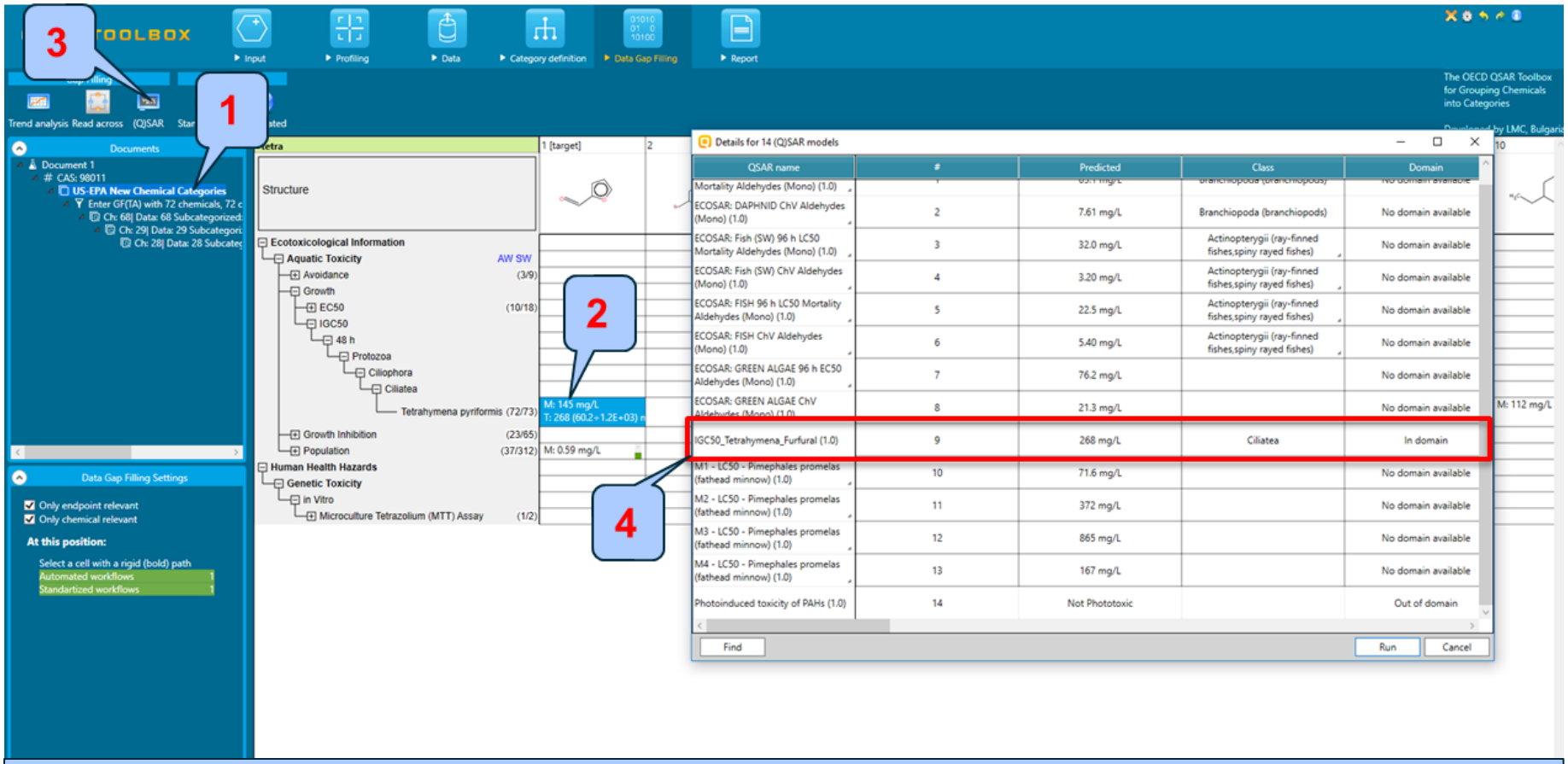
1. **Click** "Model/QSAR"; 2. **Select** "Save model"; 3. **Type** Name of the model and fill fields in the Wizard if necessary (Use Next/Back buttons to navigate within it); 4. **Click** "Save model"; 5. **Click** OK on the message;

Outlook

- Background
- Objectives
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - Data
  - Category definition
  - Data gap filling
    - **QSAR model**

Data Gap Filling

How to see the derived QSAR?



1. Select a non-Gap filling list from the document tree; 2. Note the accepted prediction is inserted into data matrix 3. **Click** "(Q)SAR"; 4. The derived QSAR is listed in the panel with Relevant (Q)SAR models.

Data Gap Filling

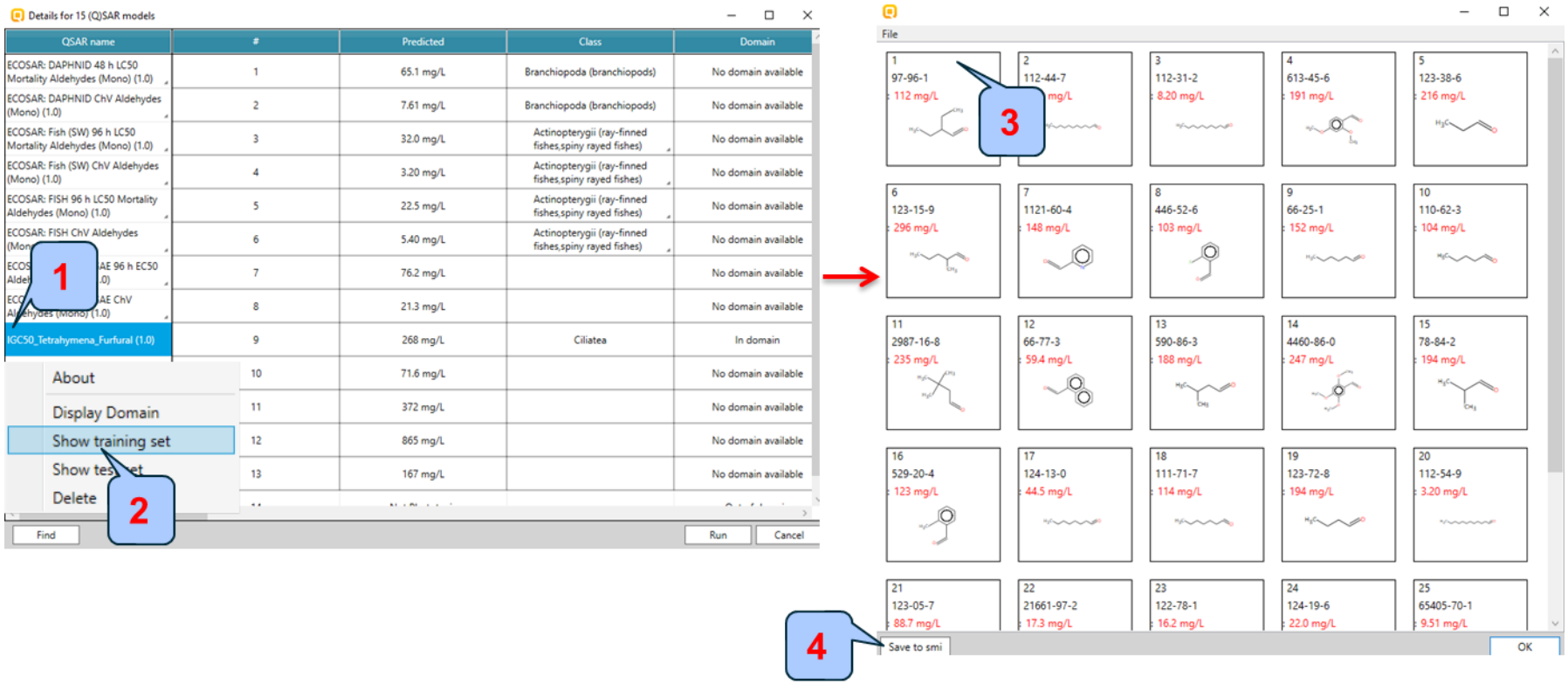
How to see the derived QSAR?

As seen in the next five screen shots the derived model can be used to:

- Visualize training set of the model:
- Visualize the domain of the model:
- Visualize whether a chemical is in the applicability domain of the model:
- Enter in Data Gap filling
- Perform predictions for:
  - Selected chemical
  - All chemicals (in the matrix)
  - Chemicals in domain:

Data Gap Filling

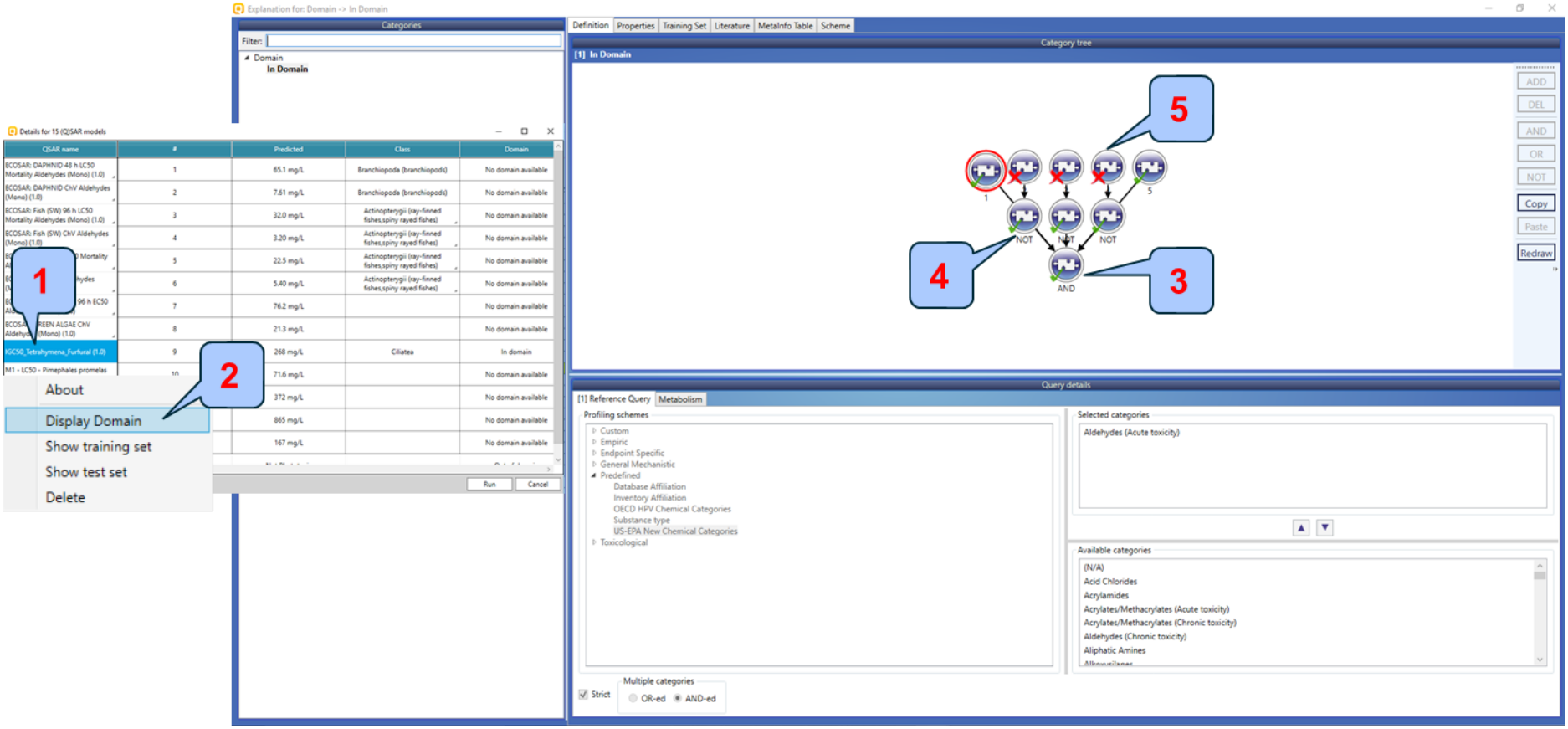
Visualisation of the training set



1. **Right Click** on the derived QSAR model; 2. **Select** Show training Set; 3. Note the experimental data is displayed under CAS# of each chemical; 4. The training set can be saved as *.smi file.

Data Gap Filling

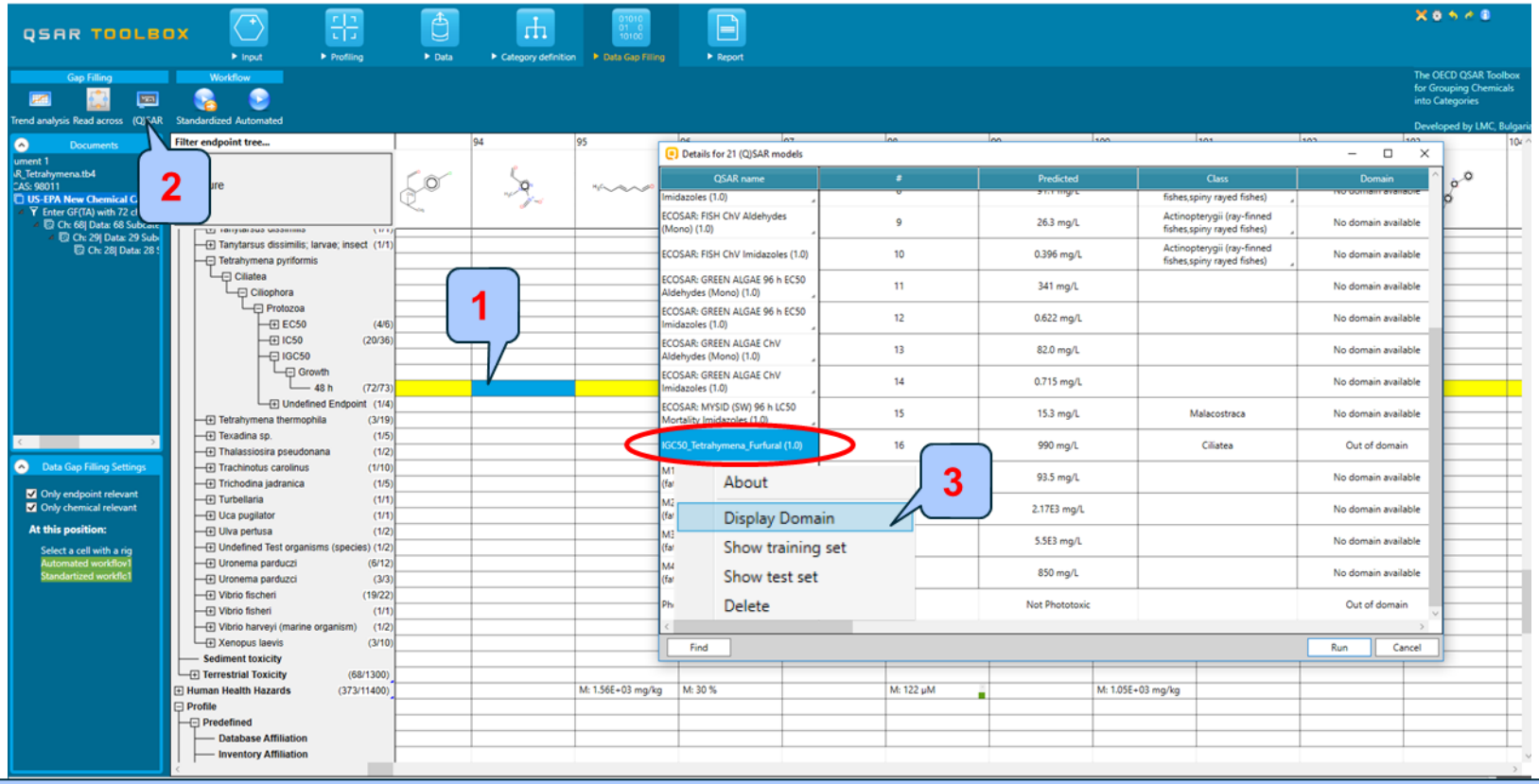
Visualisation of model domain



1. Right click on the derived QSAR model; 2. **Select** "Display Domain"; 3. Note the boundaries of the domain are combined logically; 4. If the chemical answers the query of the domain then the current query is a labelled with **GREEN** tick; 5. Otherwise is labelled with **RED** cross.

Data Gap Filling

Visualisation whether a chemical is in the domain of the model



1. **Highlight** the cell of one of the analogues (e.g., chemical # 94 in the data matrix; 2. Click on "(Q)SAR"; 3. **Right click** above the model; 3. **Left click** on Display domain (see next screen shot).

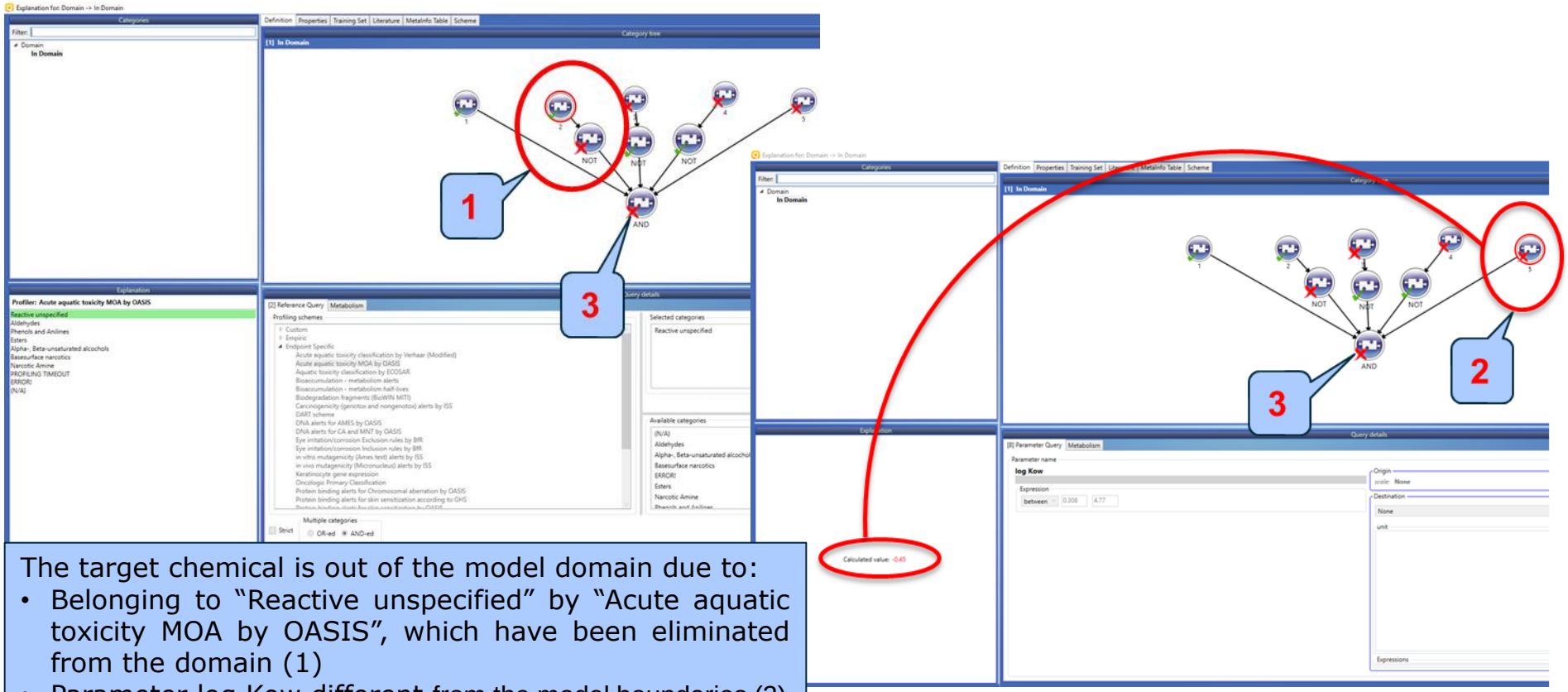
Data Gap Filling

Visualisation whether a chemical is in the domain of the model

- The chemical is an aldehyde as required by US-EPA categorization group (boundary 1 on next screen shot).
- The chemical is an aldehyde as required by Acute aquatic toxicity MOA by OASIS group (boundary 2).
- It can react with protein by Schiff-base formation and should not belong to any of the eliminated mechanistic domains according to Protein binding by OASIS (boundary 3):
  - Michael addition (α,β-Aldehydes, Conjugated systems with electron withdrawing groups)
  - SNAr (Activated aryl and heteroaryl compounds)
  - Schiff base formation (Bis aldehydes, Di-substituted α,β-unsaturated aldehydes and Aromatic carbonyl compounds)
- The chemical is an aldehyde as required by Aquatic toxicity classification by ECOSAR (boundary 4).
- Another requirement is Log Kow to be >=0.308 and <= 4.77 (boundary 5):

Data Gap Filling

Visualisation whether a chemical is in the domain of the model



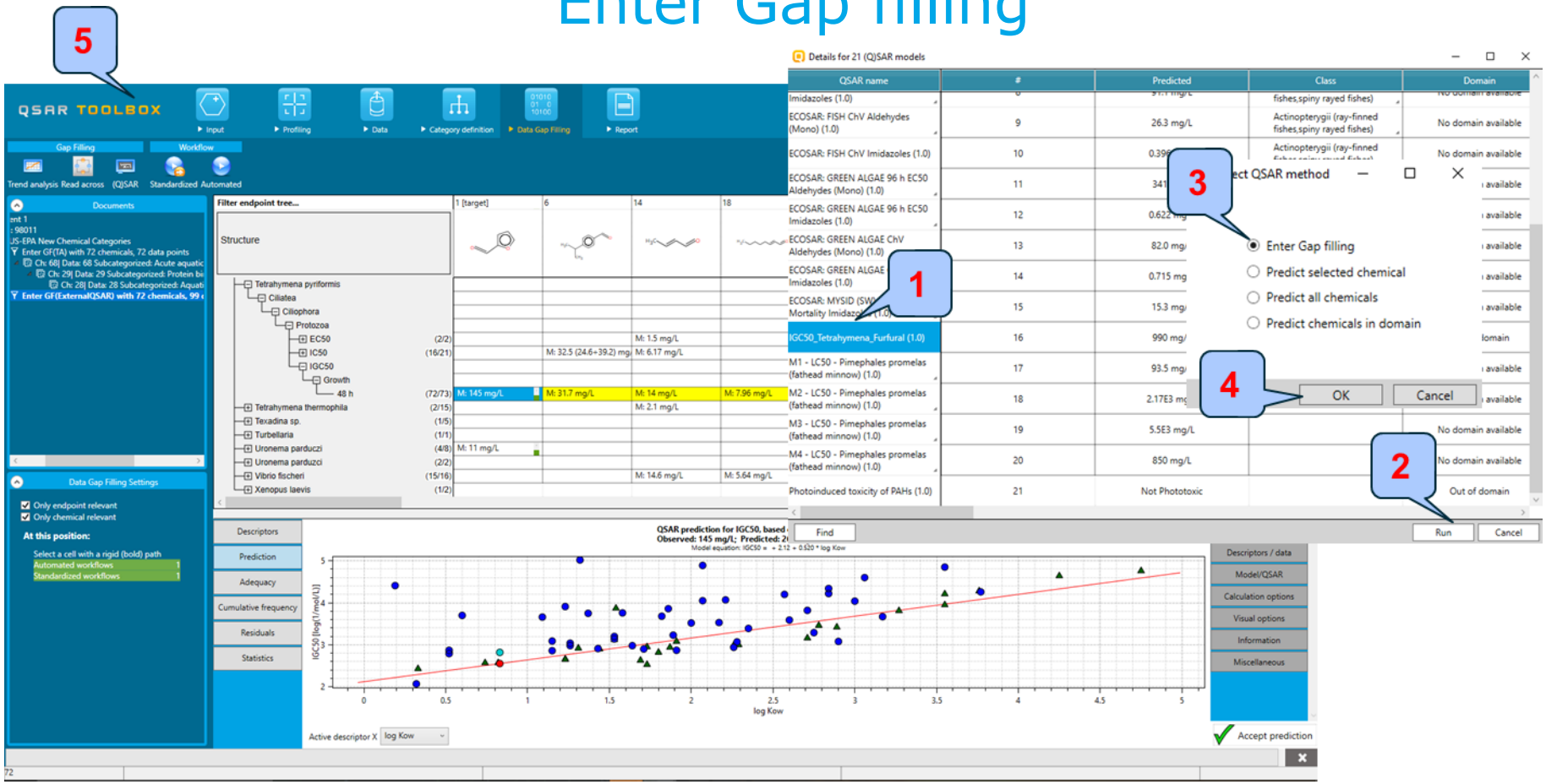
The target chemical is out of the model domain due to:

- Belonging to "Reactive unspecified" by "Acute aquatic toxicity MOA by OASIS", which have been eliminated from the domain (1)
- Parameter log Kow different from the model boundaries (2)

The definitive designation for belonging or not to the domain is the collectible boundary (3) which is red crossed in case of "Out of domain" (green checked in case of "In domain")

Data Gap Filling

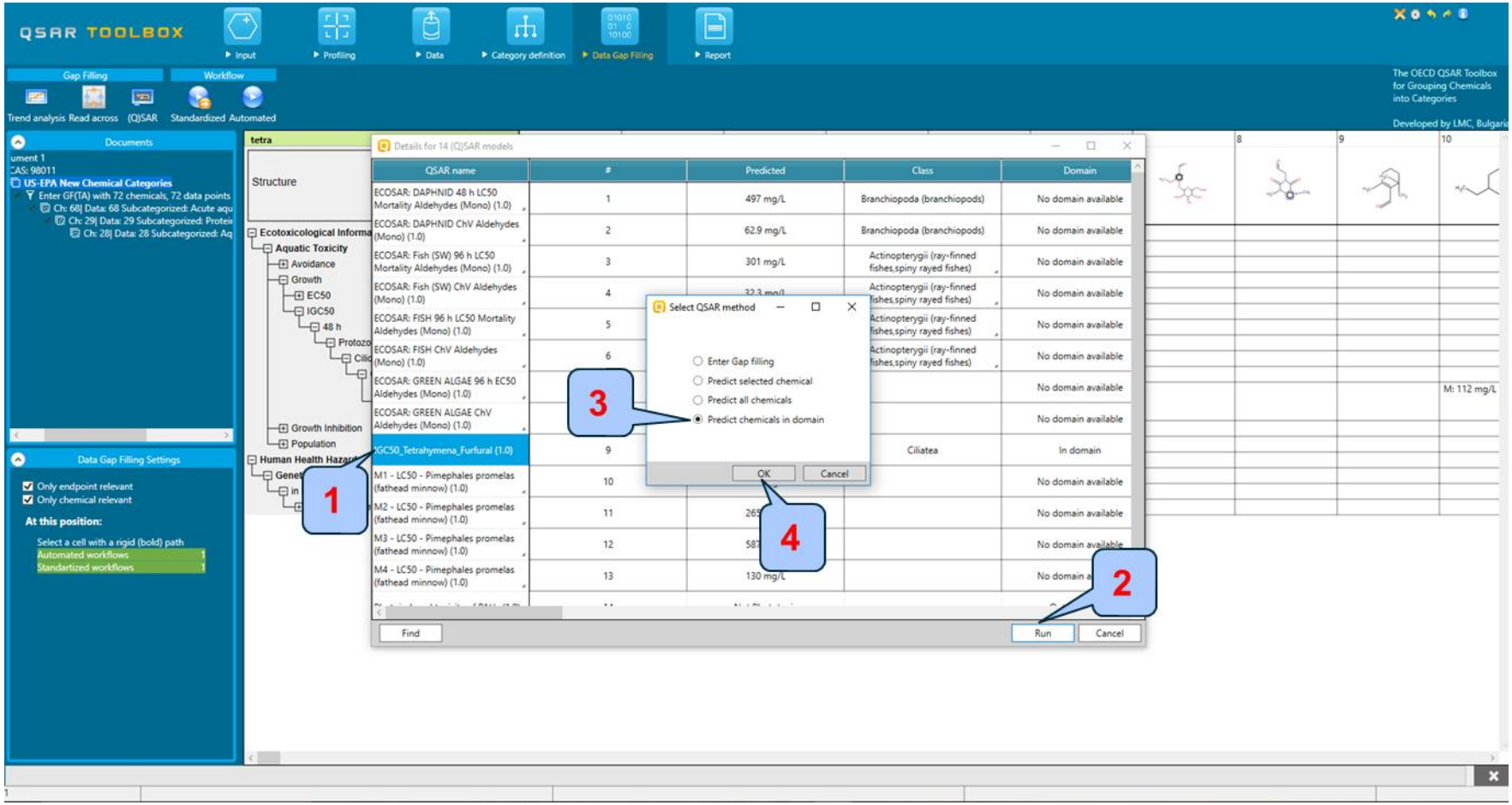
Enter Gap filling



Go to target chemical and call (Q)SAR**;**
1. **Mark** the model; 2. **Click** Run; 3. **Select** Enter Gap filling; 4. **Click** OK; 5. You are in Gap filling and can operate;

Data Gap Filling

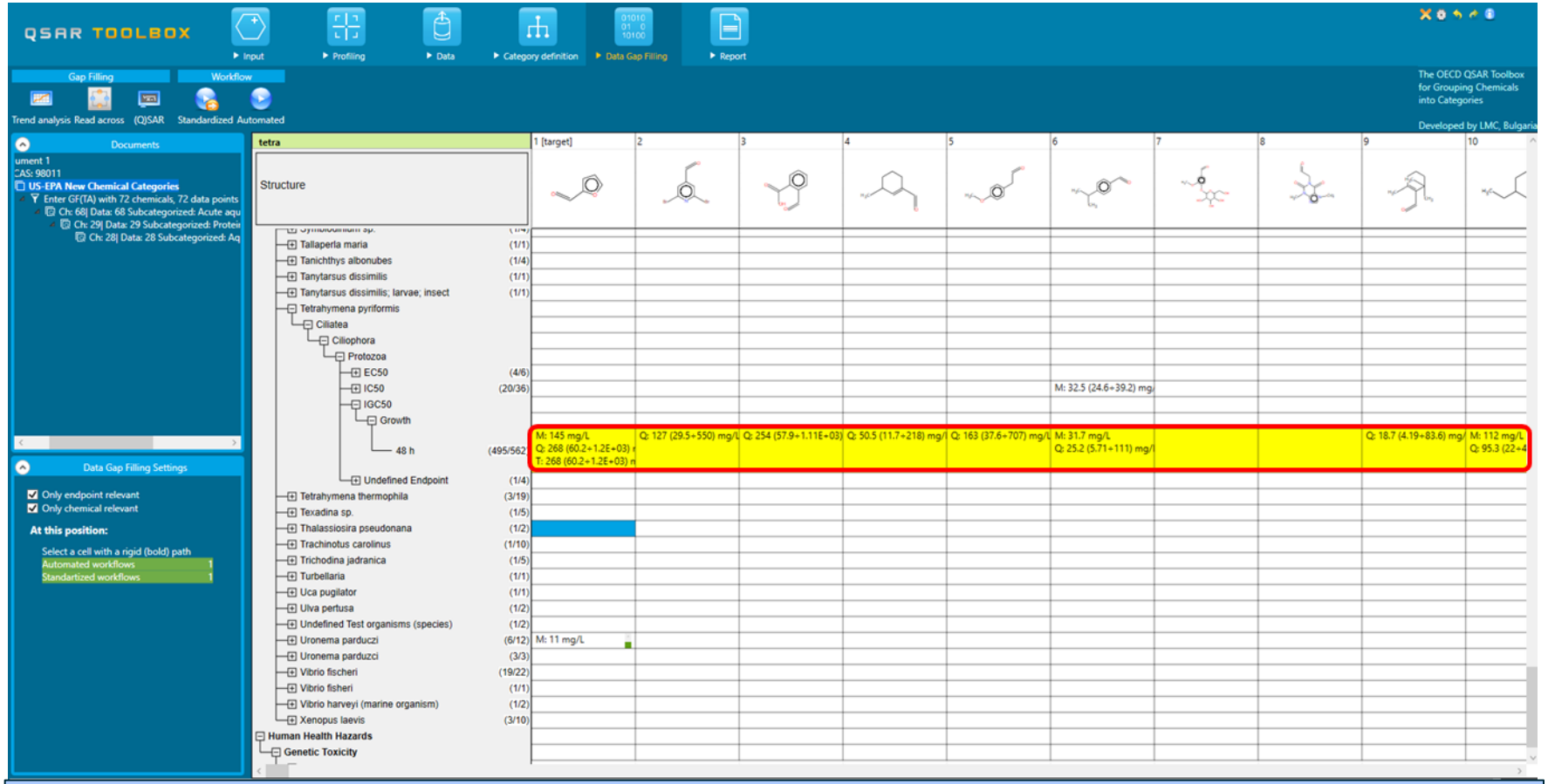
Perform prediction for chemicals in domain (for selected chemical and all chemicals - analogically)



1. **Mark** the model; 2. **Click** Run**;** 3. **Select** Predict Chemicals in domain; 4. **Click** OK;

Data Gap Filling

Perform prediction for chemicals in domain



The process of applying the model is indicated by status bar on the bottom of the window. The predictions are placed on the matrix. Note there are different signs for the origin of the data: M for experimental data, T for result of Trend analysis, Q for originated from QSAR data.

Outlook

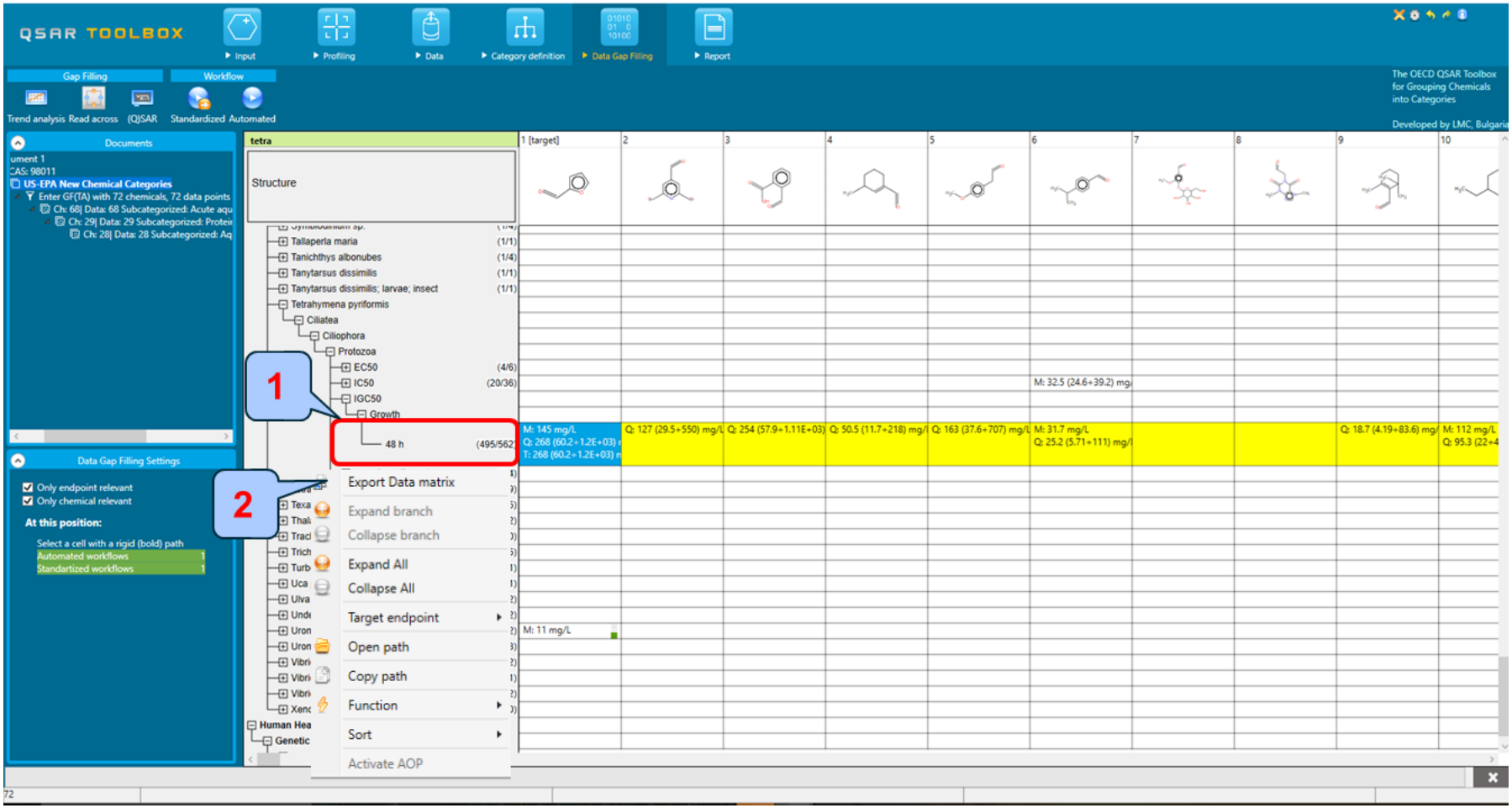
- Background
- Objectives
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - Data
  - Category definition
  - Data gap filling
  - QSAR model
  - **Export QSAR prediction**

Export QSAR results

- The predictions for the chemicals in the matrix can be exported into a file.

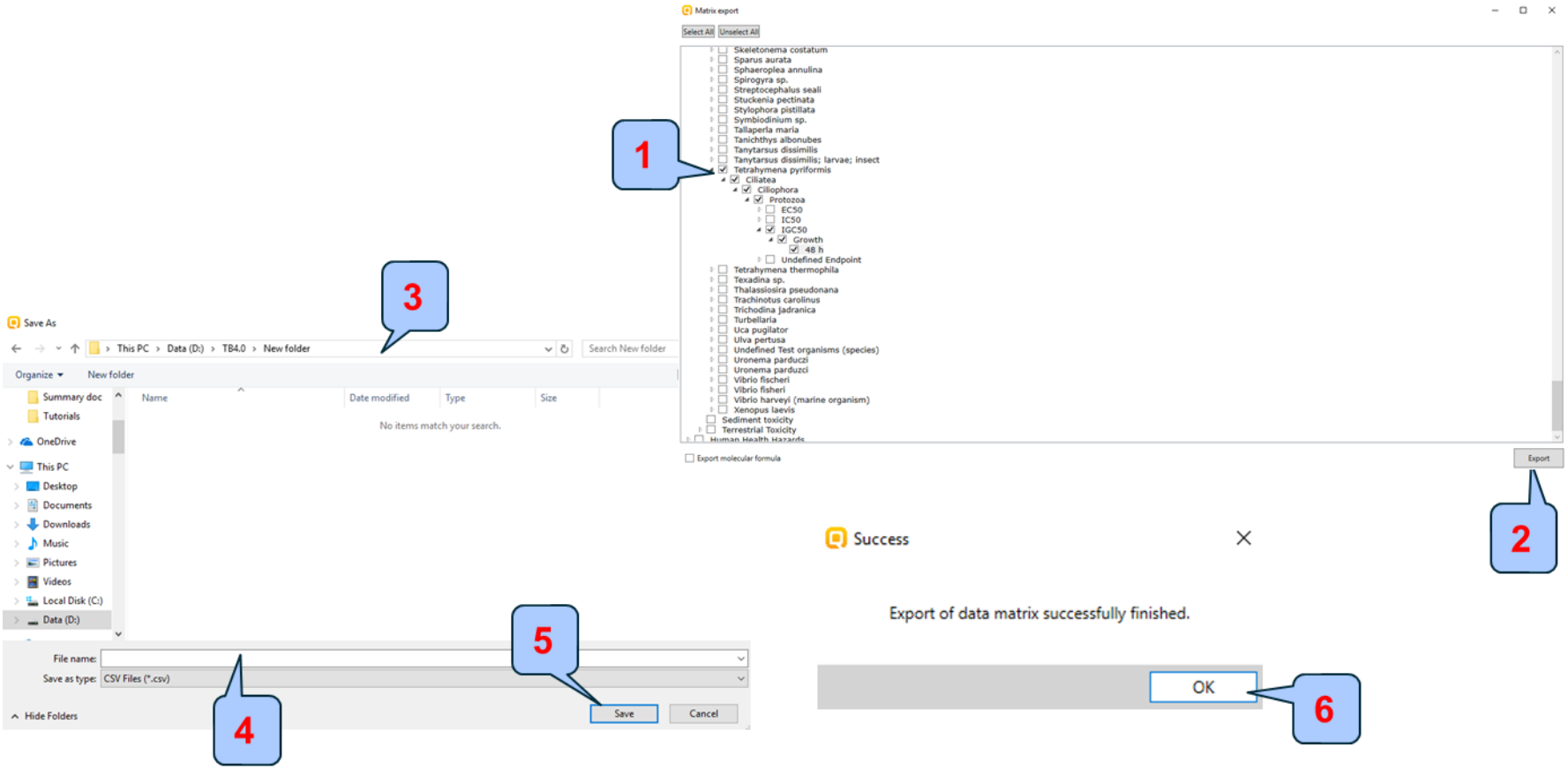
- In the Endpoint tree right click on Tetrahymena pyriformis (for the endpoint IGC50 48h for Tetrahymena pyriformis) and select Export Data matrix from the context menu (see next three screen shots).

Export QSAR results



1. **Right click** on the row of endpoint tree associated with predictions from the QSAR model; 2. **Select** Export Data matrix (see next screen shot).

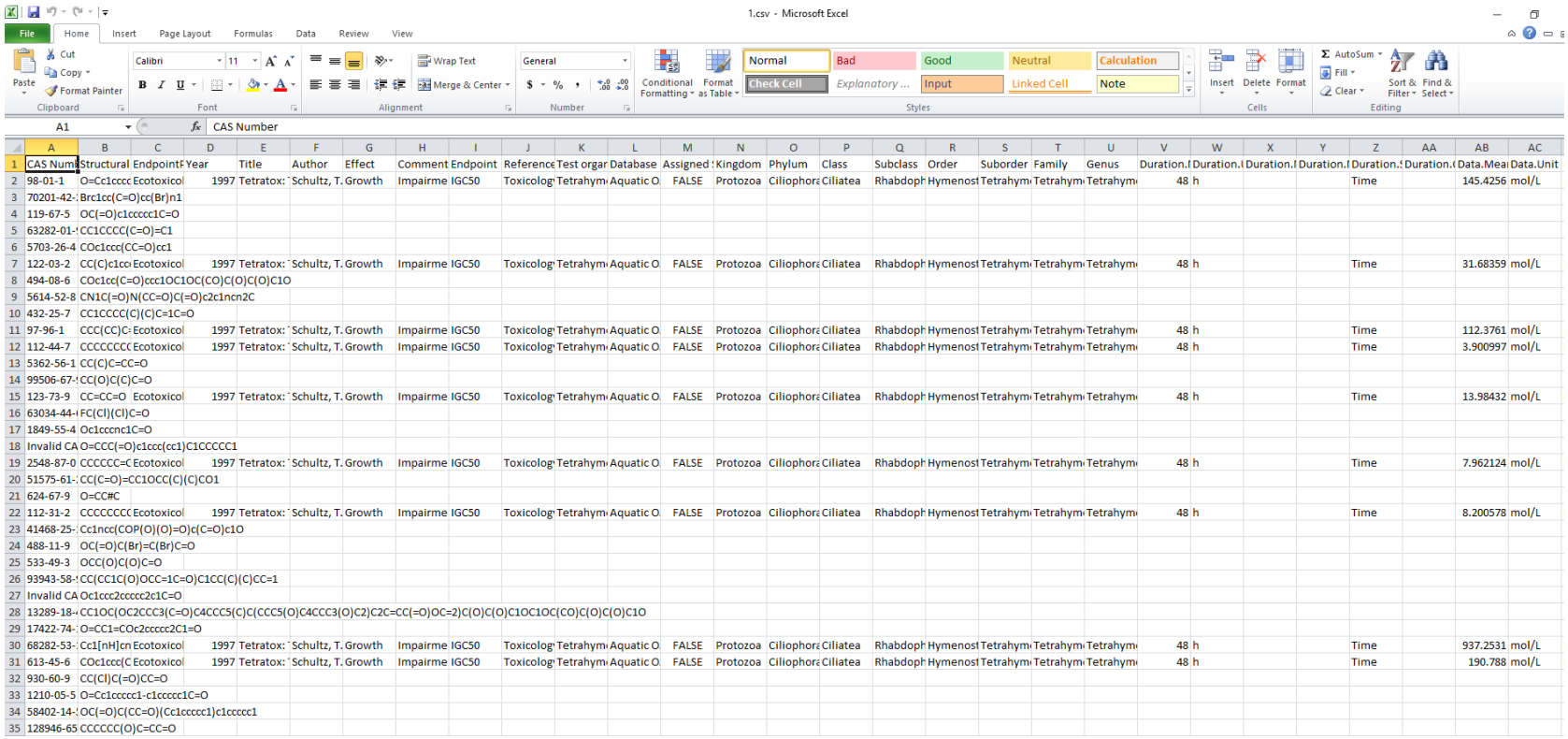
Export QSAR results



1. The nodes from the tree associated with QSAR predictions which will be exported are labelled with check marks; 2. **Click** Export; 3. **Browse** to **save** the folder on your PC; 4. **Give** a name of the file; 5. **Click** Save; 6. **Click** OK when the file is exported.

Export QSAR results

The resulting file in *.csv format can be opened via Microsoft Excel and further analysed.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	CAS Numb	Structural	Endpoint F	Year	Title	Author	Effect	Comment	Endpoint	Reference	Test organ	Database	Assigned	Kingdom	Phylum	Class	Subclass	Order	Suborder	Family	Genus	Duration.(	Duration.(	Duration.(	Duration.(	Duration.:	Duration.(	Data.Mea	Data.Unit
2	98-01-1	O=Cc1ccc	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		145.4256	mol/L
3	70201-42-	Brc1cc(C=O)cc(Br)n1																											
4	119-67-5	OC(=O)c1ccccc1C=O																											
5	63282-01-	CC1CCCC(C=O)=C1																											
6	5703-26-4	COc1ccc(CC=O)cc1																											
7	122-03-2	CC(C)c1cc	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		31.68359	mol/L
8	494-08-6	COc1cc(C=O)ccc1OC1OC(CO)C(O)C(O)C1O																											
9	5614-52-8	CN1C(=O)N(CC=O)C(=O)c2c1ncn2C																											
10	432-25-7	CC1CCCC(C)(C)C=1C=O																											
11	97-96-1	CCC(CC)C	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		112.3761	mol/L
12	112-44-7	CCCCCCC	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		3.900997	mol/L
13	5362-56-1	CC(C)C=CC=O																											
14	99506-67-	CC(O)C(C)C=O																											
15	123-73-9	CC=CC=O	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		13.98432	mol/L
16	63034-44-	FC(Cl)(Cl)C=O																											
17	1849-55-4	Oc1ccccnc1C=O																											
18	Invalid CA	O=CCC(=O)c1ccc(cc1)C1CCCCC1																											
19	2548-87-0	CCCCCC=C	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		7.962124	mol/L
20	51575-61-	CC(C=O)=CC1OCC(C)(C)CO1																											
21	624-67-9	O=CC#C																											
22	112-31-2	CCCCCCC	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		8.200578	mol/L
23	41468-25-	Cc1ncc(COP(O)(O)=O)c(C=O)c1O																											
24	488-11-9	OC(=O)C(Br)=C(Br)C=O																											
25	533-49-3	OCC(O)C(O)C=O																											
26	93943-58-	CC(CC1C(O)OCC=1C=O)C1CC(C)(C)CC=1																											
27	Invalid CA	Oc1ccc2ccccc2c1C=O																											
28	13289-18-	CC1OC(OC2CCC3(C=O)C4CCC5(C)C(CCC5(O)C4CCC3(O)C2)C2C=CC(=O)OC=2)C(O)C(O)C1OC1OC(CO)C(O)C(O)C1O																											
29	17422-74-	O=CC1=COc2ccccc2C1=O																											
30	68282-53-	Cc1[nH]cn	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		937.2531	mol/L
31	613-45-6	COc1ccc(C	Ecotoxicol	1997	Tetratox:	Schultz, T.	Growth	Impairme	IGC50	Toxicolog	Tetrahym	Aquatic O	FALSE	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48	h			Time		190.788	mol/L
32	930-60-9	CC(Cl)C(=O)CC=O																											
33	1210-05-5	O=Cc1ccccc1-c1ccccc1C=O																											
34	58402-14-	OC(=O)C(CC=O)(Cc1ccccc1)c1ccccc1																											
35	128946-65	CCCCCC(O)C=CC=O																											

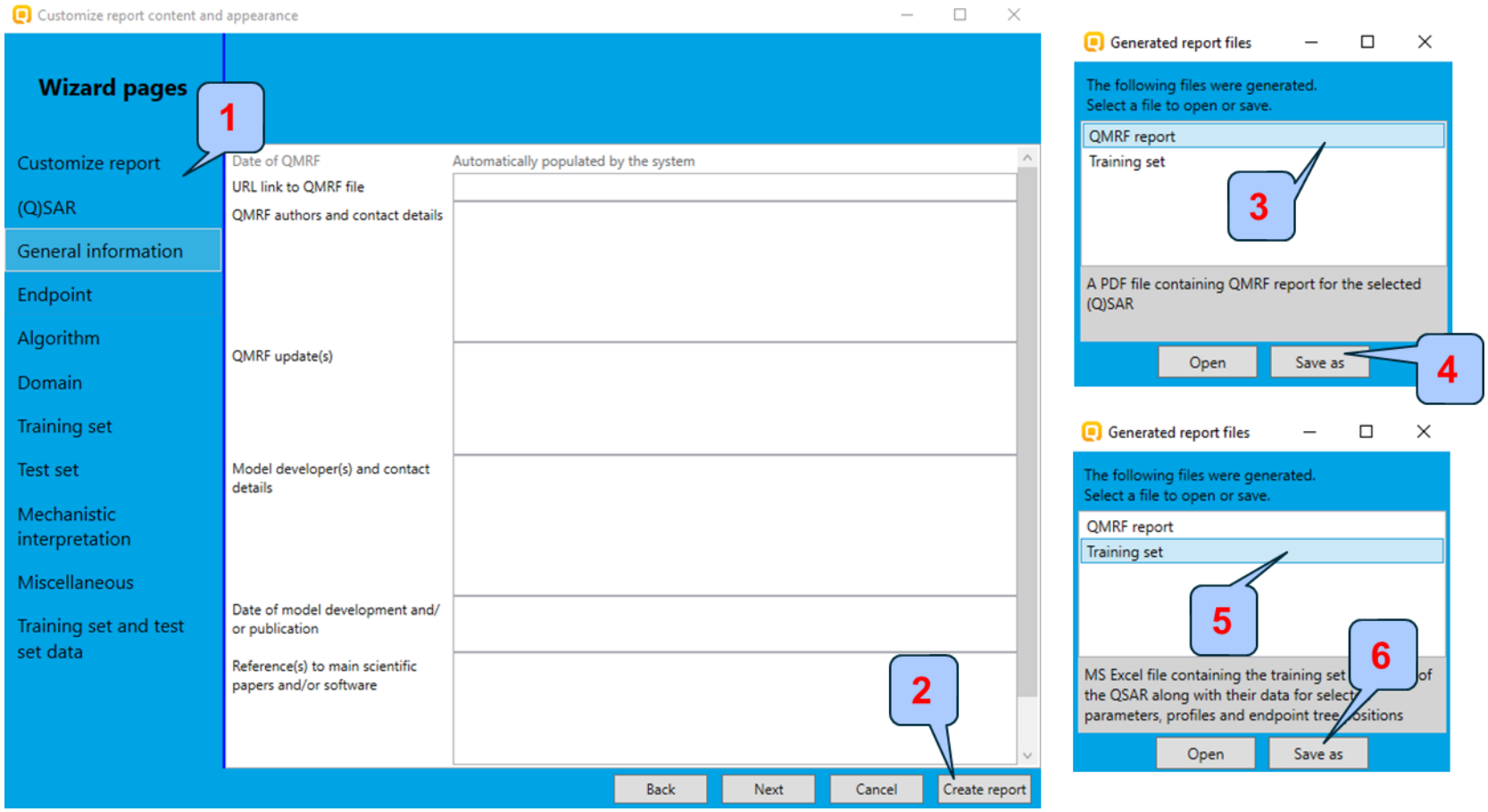
Outlook

- Background
- Objectives
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - Data
  - Category definition
  - Data gap filling
  - QSAR model
  - Export QSAR prediction
  - **Report**

Report

1. **Go** to “Report”; 2. **Select** QMRF; 3. **Mark** the user-defined QSAR model; 4. **Click** OK;

Report



1. **Navigate** through the Wizard to customize the report; 2. **Select** Create report; 3. **Choose** QMRF report to create a PDF format of the report; 4. **Click** Save as; 5. Choose Training set in order to create a MS Excel file (training set of the QSAR along with their data); 6. **Click** Save as;

Report

IGC50_Tetrahymena_Furfural 1 / 4

IGC50_Tetrahymena_Furfural

QMRF Report

A (Q)SAR model

1. (Q)SAR identifier

1.1. (Q)SAR identifier (title):
IGC50_Tetrahymena_Furfural (v.1.0)

1.2. Other related models:
Not available

1.3. Software coding the model:
QSAR Toolbox 4.1

2. General information

2.1. Date of QMRF:
Not available

2.2. QMRF author(s) and contact details:
Not available

2.3. QMRF update(s):
Not available

2.4. Model developer(s) and contact details:
Not available

2.5. Date of model development and/or publication:
Not available

2.6. Reference(s) to main scientific papers and/or software package:
Not available

2.7. Availability of information about the model:
Not available

2.8. Availability of another QMRF for exactly the same model:
Not available

3. Defining the endpoint (OECD Principle 1)

3.1. Species:
Tetrahymena pyriformis

Training set

Training set.xlsx - Microsoft Excel

			Training set #1			Training set #2			Training set #3			Training set #4		
Substance identity														
Structure			Invalid structure			Invalid structure			Invalid structure			Invalid structure		
CAS number			97-96-1			112-44-7			112-31-2			613-45-6		
Chemical name			Ethylbutanal			Undecanal			Decanal			2,4-DIMETHOXYBENZALDEHYD		
Other identifier														
SMILES			CCC(CC)C=O			CCCCCCCCCCC=O			CCCCCCCCCC=O			COc1ccc(C=O)c(OC)c1		
Parameters		**unit**												
Profilers														
Training set data and user gathered data														
Training set data														
environment	*endpoint*		*value*	*unit*	*species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value*	*unit*	*species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value*	*unit*	*species, duration, test type, type of method, assay, strain, test guideline, year, reference*	*value*	*unit*	*species, duration, test type, type of method, assay, strain, test guideline, year, reference*
Aquatic Toxicity	IGC50		112	mg/L	Tetrahymena pyriformis	3.9	mg/L	Tetrahymena pyriformis	8.2	mg/L	Tetrahymena pyriformis	191	mg/L	Tetrahymena pyriformis

Congratulations

- You have used the Toolbox to build a user-defined QSAR model.
- You now know another useful tool in the Toolbox.
- Continue to practice with this and other tools. Soon you will be comfortable dealing with many situations where the Toolbox is useful.